Rotation Estimation of Acoustic Camera Based on Illuminated Area in Acoustic Image

Yusheng Wang [*] **Yonghoon Ji** [**] **Hanwool Woo** [***]
Yusuke Tamura [****] **Hiroshi Tsuchiya** [†] **Atsushi Yamashita** [*]
Hajime Asama [*]

[*] *Department of Precision Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan (e-mail: [wang,tamura,yamashita,asama]@robot.t.u-tokyo.ac.jp).*
[**] *Department of Precision Mechanics, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan (e-mail: ji@mech.chuo-u.ac.jp).*
[***] *Department of Intelligent Mechatronics, Akita Prefectural University, Yurihonjo City, Akita 015-0055, Japan (e-mail: woo@akita-pu.ac.jp).*
[****] *Institute of Engineering Innovation, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan (e-mail: tamura@robot.t.u-tokyo.ac.jp).*
[†] *Research Institute, Wakachiku Construction Co., Ltd, 31 Minamisode, Sodegaura, Chiba 299-0268, Japan (e-mail: hiroshi.tsuchiya@wakachiku.co.jp).*

Abstract: In this paper, the concept of illuminated area, an important characteristic in acoustic images, is formalized, which can be applied to tasks such as the 3D mapping of underwater environment. Unmanned exploration using underwater robots is gaining attention among the scientific community. A way to sense the underwater environment is to employ the acoustic camera, a next generation forward looking sonar with high resolution even in turbid water. It is more flexible than common underwater sonars; however, studies on acoustic cameras are still at the early stage. Acoustic cameras have a fixed vertical beam width able to generate a limited bright area on acoustic images which is named illuminated area. In this paper, we propose the concept of illuminated area and the method to detect the illuminated area under a flatness assumption of the ground. Then, it is shown how the knowledge of the illuminated area can be fused with depth information to estimate the roll and pitch angles of the acoustic camera. The estimated quantities can be employed to carry out a 3D mapping process. Experiment shows the validity and effectiveness of the proposed method.

Keywords: Underwater robotics, underwater sensing, illuminated area, acoustic camera, pose estimation, 3D mapping

1. INTRODUCTION

In the last few years, unmanned underwater exploration is attracting the interests of scientists and engineers. Underwater robots such as remotely operated vehicle (ROV) and autonomous underwater vechicle (AUV) have become popular to accomplish various tasks. There are several kinds of sensors mounted on the underwater robot for underwater inspection such as optical cameras, laser sensors and sonars. Usually, sonars are considered to be powerful in underwater environments as they are not influenced by illumination conditions and have a larger range with respect to the other sensors. An acoustic camera is a forward looking imaging sonar which can generate images similar to optical images based on acoustic lens. It has relatively small volume making it flexible to be mounted on the underwater robot. Considering its flexibility, pose estimation of the acoustic camera is important. Accurate pose estimates are necessary to perform tasks such as the measurement of underwater targets of the 3D modeling of underwater environments.

Estimating the rotation of the camera accurately and precisely is a complex task. One of the reason is that high precision underwater inertial measurement unit (IMU) sensors usually have a very high cost (Yang and Huang (2017)). Certain types of the acoustic cameras, e.g. dual frequency identification sonar (DIDSON) and adaptive resolution image sonar (ARIS) have built-in magnetic

⋆ This work was supported by JSPS KAKENHI Grant Number 19K21547.

compasses (Belcher et al. (2002)). These compasses can measure the roll, pitch and yaw angles of acoustic camera. However, the measurement is not reliable enough. Input control of rotator mounted on the sensor such as pan-tilt module can be used to acquire rotation information, however, they are not always available depending on situations. A possible method is to acquire rotation information directly from acoustic images, similar to visual-inertia odometry, 2D features on successive images can be taken into consideration to estimate the rotation angles. However, a problem exists that 2D features like SIFT performs poorly on acoustic images, meanwhile, finding corresponding points on each image is also a difficult task due to different geometry theory to optical camera (Negahdaripour et al. (2011)). An idea to acquire feature points robustly on acoustic image is to design an underwater landmark which is easier to be recognized by acoustic camera. In Lee et al. (2017), a circle-type artificial landmark is proposed with a probabilistic method to recognize the marker. However, further test would be needed to evaluate the performance of pose estimation using this kind of marker.

In this study, instead of using the feature points, a concept named illuminated area is proposed. The acoustic camera has a fixed vertical beam width which will be explained in the next section. The acoustic camera has the similar characteristics as a torch in the underwater environment. The area that is lighted up on the acoustic image is defined as the illuminated area. With a flatness assumption of the ground, which is reasonable in most artificial environments, this area can be automatically detected. From the illuminated area, roll and pitch angles of the acoustic camera can be automatically estimated. In our previous study, a 3D mapping method is proposed and a local map was generated through the roll rotation of the acoustic camera (Wang et al. (2018)). The rotation information is from the built-in compass which is found not reliable. In this paper, attempts are made to generate 3D map by using more accurate roll and pitch angles estimated from the illuminated area.

2. PRINCIPLE OF ACOUSTIC CAMERA

An acoustic camera uses acoustic lens to form realistic acoustic images with high resolution instead of using delay lines or digital forming techniques as conventional sonar. It is a multiple beam forward looking sonar that ensonifies a fan-shape wave in the forward direction as shown in Fig. 1a. The fan-shape wave consists of N beam slices in the azimuth direction which can be considered as 2D beams. Only range information is acquirable that information in the elevation direction is lost. The average beam width in the elevation direction for beam slices is ϕ_{max} and the total width of N beam slices in the azimuth direction is θ_{max}. Field of view (FoV) is controlled by range r_{min} and r_{max} as shown in Fig. 1b.

When considering a point $({}^cx, {}^cy, {}^cz)$ in camera Cartesian coordinate system, it can also be represented in spherical coordinate system as $(\Re, \theta, \phi)$. The conversion of the two coordinate systems is

$$\begin{bmatrix} {}^cx \\ {}^cy \\ {}^cz \end{bmatrix} = \begin{bmatrix} \Re\cos\phi\cos\theta \\ \Re\cos\phi\sin\theta \\ \Re\sin\phi \end{bmatrix}, \quad (1)$$

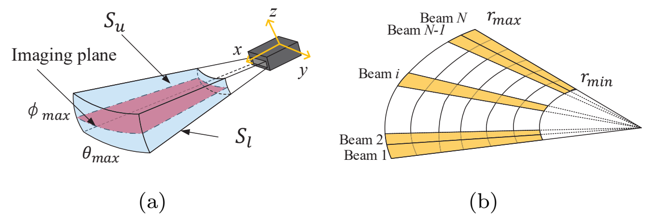


Fig. 1. Acoustic projection model:(a) geometrical model and imaging plane of the acoustic image and (b) beam slices on acoustic image.

where the inverse transformation is

$$\begin{bmatrix} \Re \\ \theta \\ \phi \end{bmatrix} = \begin{bmatrix} \sqrt{{}^cx^2 + {}^cy^2 + {}^cz^2} \\ \tan^{-1}\left({}^cy^2/{}^cx^2\right) \\ \tan^{-1}\left({}^cz/\sqrt{{}^cx + {}^cy}\right) \end{bmatrix}. \quad (2)$$

A common model of the acoustic camera is based on a non-orthogonal projection that all the points are projected to the plane where $\phi = 0$. Raw data of the acoustic image is a $\Re \times \theta$ matrix and after processing, a fan-shape image is presented by mapping $(\Re, \theta)$ to (I_x, I_y) as

$$\begin{bmatrix} I_x \\ I_y \end{bmatrix} = \begin{bmatrix} \Re\sin\theta \\ \Re\cos\theta \end{bmatrix}. \quad (3)$$

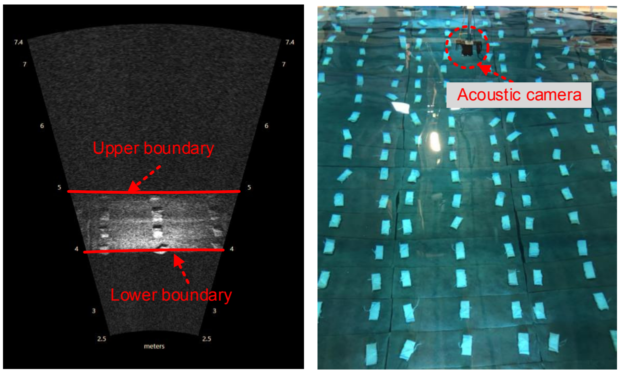
3. ILLUMINATED AREA

As we mentioned above, the acoustic camera can be considered as a torch in the underwater environment. Transducers in the acoustic camera send out sound wave to “light up” the underwater environment and receive the reflected signals. Usually, a bright area can be found on the acoustic images which is considered as “illuminated area”. Illuminated area is one of the important concept in acoustic images which have not been studied in most of the previous studies.

3.1 Definition of illuminated area and its property

The illuminated area is an area in the acoustic image which is brighter than the other part during sensing of the ground due to the fixed vertical beam width of the acoustic camera. This concept is similar to the effective area mentioned by Vilarnau (2014) and effective region in Cho et al. (2018). If r_{min} and r_{max} are set properly, this area can be recognized clearly as two line-like boundaries under the assumption that the ground is flat. These boundaries are caused by the intersection of upper surface S_u and lower surface S_l of the ensonified sound wave with the ground, which are defined as upper line L_u and lower line L_l in this paper. Then, upper line L_u and lower line L_l are projected into camera imaging plane, which generate the upper boundary B_u and lower boundary B_u in acoustic image as shown in Fig. 2a while sensing the environment in Fig. 2b. When assuming that the ground is a plane, the boundaries in acoustic image have some specific features that can be applied in various under water tasks. This study is under the assumption that the ground is a plane. Denote the plane base as P, L_u and L_l can be expressed as

$$L_u \triangleq S_u \cap P, \quad (4)$$
$$L_l \triangleq S_l \cap P. \quad (5)$$



(a) (b)

Fig. 2. Acoustic image and sensed environment: (a) the upper boundary and the lower boundary on acoustic image where roll angle of camera is close zero. Due to the limited beam width, a lighted up area can be seen from the image which is the illuminated area. (b) is the environment where (a) is captured.

Let's define the six degrees of freedom (DoF) pose of the acoustic camera as $(x, y, z, \varphi_x, \varphi_y, \varphi_z)$ where $(\varphi_x, \varphi_y, \varphi_z)$ refers to roll, pitch and yaw. A ray tracing method similar to Kwak et al. (2015) is used to calculate the boundaries B_u and B_l on acoustic images. Considering the i-th beam slice, the ray with elevation angle $\phi = \frac{\phi_{max}}{2}$ is denoted as r_u^i and the ray with elevation angle $\phi = -\frac{\phi_{max}}{2}$ is denoted as r_l^i. The direction vectors ${}^c v_u^i$ and ${}^c v_l^i$ of rays r_u^i and r_b^i in camera coordinate are

$$ {}^c v_u^i \triangleq \left[\cos\frac{\phi_{max}}{2}\cos\theta_i \;\; \cos\frac{\phi_{max}}{2}\sin\theta_i \;\; \sin\frac{\phi_{max}}{2} \right]^T, \quad (6) $$

$$ {}^c v_l^i \triangleq \left[\cos\frac{-\phi_{max}}{2}\cos\theta_i \;\; \cos\frac{-\phi_{max}}{2}\sin\theta_i \;\; \sin\frac{-\phi_{max}}{2} \right]^T. \quad (7) $$

Then, they are transferred to world coordinate by

$$ \begin{bmatrix} {}^w v_1 \\ {}^w v_2 \\ {}^w v_3 \end{bmatrix} = R \begin{bmatrix} {}^c v_1 \\ {}^c v_2 \\ {}^c v_3 \end{bmatrix}, \quad (8) $$

where ${}^w v$ is direction vector in world coordinate and ${}^c v$ is the direction vector in camera coordinate and rotation matrix R is

$$ R = \begin{bmatrix} c_{\varphi_y} c_{\varphi_z} & c_{\varphi_z} s_{\varphi_y} s_{\varphi_x} - c_{\varphi_x} s_{\varphi_z} & s_{\varphi_x} s_{\varphi_z} + c_{\varphi_x} c_{\varphi_z} s_{\varphi_y} \\ c_{\varphi_y} s_{\varphi_z} & c_{\varphi_x} c_{\varphi_z} + s_{\varphi_y} s_{\varphi_x} s_{\varphi_z} & c_{\varphi_x} s_{\varphi_y} s_{\varphi_z} - c_{\varphi_z} s_{\varphi_x} \\ -s_{\varphi_y} & c_{\varphi_y} s_{\varphi_x} & c_{\varphi_y} c_{\varphi_x} \end{bmatrix}. \quad (9) $$

c, s refers to cosine and sine. Points (X, Y, Z) on the ray can be represented as

$$ \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} x +^w v_1 t \\ y +^w v_2 t \\ z +^w v_3 t \end{bmatrix}, \quad (10) $$

where $t \triangleq \Re$.

If the depth of acoustic camera z is known and the base plane is represented as $z = c$, $\Re$ can be calculated as

$$ \Re = \frac{c - z}{{}^w v_3}. \quad (11) $$

For the i-th beam, $\Re_u^i$ and $\Re_l^i$ can be acquired from equation (11) and θ_i can be calculated by

$$ \theta^i = -\frac{\theta_{max}}{2} + \frac{\theta_{max} i}{N}. \quad (12) $$

Finally, B_u and B_l can be represented as

$$ B_u = \{(\Re_u^i, \theta^i) | i = 0, 1, \ldots, N\}, \quad (13) $$

$$ B_l = \{(\Re_l^i, \theta^i) | i = 0, 1, \ldots, N\}. \quad (14) $$

Theoretically, B_u and B_l are curves instead of straight lines based on the camera model using in this paper. Another important characteristic of the boundaries is that they are independent to x, y, φ_z since ${}^w v_3$ in equation (11) is independent to x, y, φ_z. In other words, the boundaries will be influenced by z, φ_x, φ_y and ϕ_{max} which will be discussed later. It is worth mentioning that either upper boundary or lower boundary might not be seen from the image because of the limited effective range of sensor. In this paper, we only consider the situation that upper and lower boundaries appear on the image simultaneously.

3.2 Detection of illuminated area

A method of automatic detection of the illuminated area is proposed in this research. This method works when there is no large objects on the base plane. A pipline of pre-processing of acoustic images is shown in Fig. 3. The raw image which is the $\Re - \theta$ image is used as the input image. As shown in Fig. 3a, the vertical direction is the beam number and the parallel direction is the samples, which corresponds to θ and $\Re$ respectively. First, a median filter is used to smooth the image as shown in Fig. 3b. Second, the image is binarized using a fixed value of threshold. This binarization process may influence the position of boundaries. With different threshold, the lower boundary almost keeps the same, however, the upper boundary will be different. Here, a threshold which does not filter the pixels on the boundary is chosen, though much noise is left on the image as shown in Fig. 3c. Third, a small object removal process is implemented to filter the small, discrete pixels on the image by removing all connected components that have fewer pixels which is also known as area opening. After that the discrete components are totally removed as shown in Fig. 3d. At last, detection of boundaries is carried out as shown in Fig. 3e.

(1) *Initialize index $i = 0$*
(2) *For the i-th row of image matrix, start from the first column.*
(3) *Finding the column that the intensity of pixels changes from 0 to 1 and return the index of the column*
(4) *Start from the last column*
(5) *Repeat step (3)*
(6) *Increment index $i = i + 1$, stop if i exceeds the last row*

The measurement points can be denoted as $(\hat{\Re}^i, \theta^i)$. Note that θ^i is the same with ideal condition since it is also calculated from equation (12).

The intensity on the side of the image is low due to absorption, which leads to the difficulty of detecting full upper boundary. Figure 3e shows this phenomenon that for upper boundary, the detected result consists of the red line and the green line. In this research, green line detections are simply discarded by ignoring the beam slices on the side.

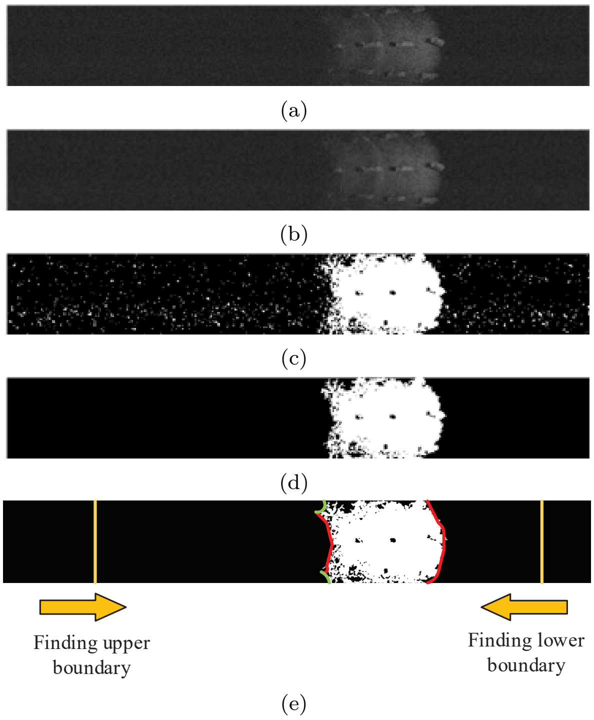


Fig. 3. Preprocessing and detection: (a) is the raw image, (b) is the smoothed image, (c) is the image after binarization, (d) is the image after removing small noise, and (e) shows how to detect the boundaries.

3.3 Estimation of vertical beam width

In this study, ARIS EXPLORER 3000 (Sound Metrics) is used with an average vertical beam width of 14 degree in the specification. However, after experiment, we recognized that this angle is slightly larger than 14 degree. It may be caused by non-calibration of acoustic camera or a complex phenomenon caused by ultrasound. To make this problem simpler, a simple camera model is to set all the vertical beam width as the same for each beam slice, which is a fixed value estimated from the following algorithm.

(1) *Preparing n images captured from the same camera pose where roll angle φ_x is close to zero*
(2) *Set the known value of z*
(3) *Initialize index $j = 1$*
(4) *Detect the boundaries B_u and B_l on the j-th image*
(5) *Carry out optimization*
(6) *Increment index $j = j + 1$, execute (4), (5)*
(7) *Stop until $j > n$, calculate the average ϕ_{max}*

The optimization process is based on the following equation.

$$\tilde{\varphi_x}, \tilde{\varphi_y}, \tilde{\phi}_{max} = \underset{\varphi_x, \varphi_y, \phi_{max}}{\operatorname{argmin}} \frac{1}{N_u} \sum_{i=u_s}^{u_e} ||\mathfrak{R}_u^i - \hat{\mathfrak{R}}_u^i||_2 + \frac{1}{N_l} \sum_{i=l_s}^{l_e} ||\mathfrak{R}_l^i - \hat{\mathfrak{R}}_l^i||_2, \tag{15}$$

where $\mathfrak{R}$ is the ideal value from ray tracing model and $\hat{\mathfrak{R}}$ is the measured value from real image. $[u_s, u_e]$ is the interval that N_u beam slices within this interval are taken into consideration. Similarly, $[l_s, l_e]$ is the interval for lower boundary and N_l is the number of the beam slices. Optimization is implemented using Levenberg-Marquardt (LM) algorithm. Initial inputs are using the values of φ_x, φ_y from the built-in compass with $\phi_{max} = 14$ degree. Images for optimization are taken from camera pose that roll angle is close to zero. The reason is that along with roll angle getting larger, the influence from ϕ_{max} to the boundaries is getting smaller. If roll angle is up to 90 degree, even if the ϕ_{max} is changed, it is hard to detect from the image.

Estimation of vertical beam width may be influenced by binarization process. In this study, threshold is manually tested when the upper boundary can be seen clearly. This value is kept during the whole process.

4. ROLL AND PITCH ANGLE ESTIMATION FOR 3D MAPPING

As we mentioned above, the boundaries are influenced by z, r, p and ϕ_{max}. For the sake of simplicity, in the following part of this research, z and ϕ_{max} are considered to be known values and the aim is to estimate r, p from one single image. In practice, z is much easier to be measured than the other values, which can be acquired using depth sensor, increasing p to make the sound wave vertical to the ground or even using a ruler. Here, $\phi_{max} = \tilde{\phi}_{max}$.

4.1 Roll and pitch angle estimation

The optimization method to acquire φ_x, φ_y is similar to the optimization method above. Since ϕ_{max} is considered to be a known value, the cost function is as follow.

$$\tilde{\varphi_x}, \tilde{\varphi_y} = \underset{\varphi_x, \varphi_y}{\operatorname{argmin}} \frac{1}{N_u} \sum_{i=u_s}^{u_e} ||\mathfrak{R}_u^i - \hat{\mathfrak{R}}_u^i||_2 + \frac{1}{N_l} \sum_{i=l_s}^{l_e} ||\mathfrak{R}_l^i - \hat{\mathfrak{R}}_l^i||_2. \tag{16}$$

LM algorithm is used to get the estimated r, p value. From the result, it seems that this optimization converges fast and does not rely on the initial input. By carrying out this process, roll and pitch angles of acoustic camera can be estimated.

4.2 3D mapping

3D mapping method is based on a 3D occupancy mapping framework (Wang et al. (2018)). The acoustic images are binarized and considered as range image. Each pixel is classified into free, occupied and unknown area. Point cloud is generated based on an idea similar to back projection. Then, the point cloud is inputted into a 3D occupancy framework using OctoMap (Hornung et al. (2013)). By using multiple images captured from different positions to update OctoMap, a 3D model of underwater target can be generated. Usually pure roll rotation is effective to generate 3D map. In previous study, compass data are used for 3D mapping; however, it is not reliable. In this study, we utilize the estimated roll and pitch angles to generate a 3D map.

5. EXPERIMENT

The proposed method is evaluated by data collected in a water tank. ARIS EXPLORER 3000 was used under 3.0 MHz mode. 3.0 MHz mode has a higher resolution but the absorption rate is higher. The acoustic camera is mounted on AR2 rotator (Sound Metrics) working in roll-tilt mode, which can change roll and pitch angle and record the input control.

One important task is to measure the height of the acoustic camera. In this paper, height z is measured using a ruler as shown in Fig. 4a. The middle of transducer array is considered to be the camera center as shown in Fig. 4b so that the height is calculated as

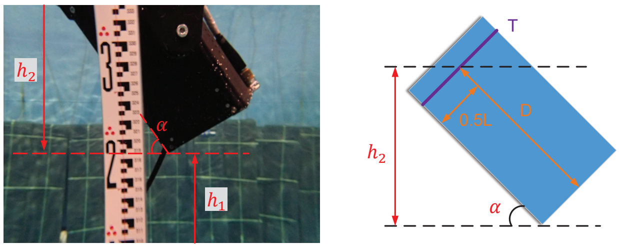
$$z = h_1 + D \sin \alpha + 0.5L \cos \alpha. \tag{17}$$

where α is read from the image as shown in Fig. 4a. As the transducer array is close to the bottom of camera, D and L are the length and the height of the acoustic camera.

In order to acquire the vertical beam width ϕ_{max}, 13 images captured from the same pose where the input control of roll angle is zero are used to estimate the φ_x, φ_y and ϕ_{max} value. The average values and standard deviations are $\bar{\varphi}_x, \bar{\varphi}_y$ and $\bar{\phi}_{max}$ are -1.36 ± 0.27 degree, 55.4 ± 0.05 degree and 16.24 ± 0.03 degree. It can be seen that $\bar{\phi}_{max}$ is a little bit larger than 14 degree.

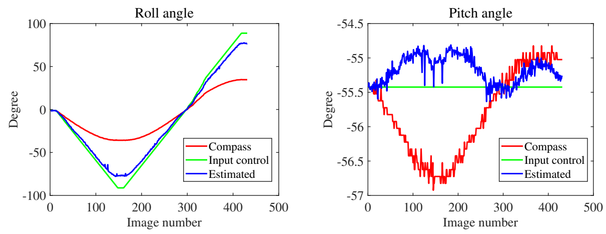
$\bar{\phi}_{max}$ is used to estimate the roll and pitch angles during one roll rotation of the camera. Pitch angle is controlled to keep 60 degree and roll angle is controlled to change from 0 to -90 to 90. Since the compass is not calibrated and the input control value is a relative pose, the roll and pitch angles are initialized with $\bar{\varphi}_x$ and $\bar{\varphi}_y$. Figure 5 shows the roll and pitch angles from compass data, control input of rotator and estimated from the image. Pitch and roll angles are estimated from one single image without considering the relationship between each frame. All the initial guesses are set to $\varphi_x = 0, \varphi_y = 0$ without any priori information. Red curve is the data from the built-in compass and the green one is from the input control of rotator. Green curve is the estimated result. Usually the input control is more reliable than the compass data, especially for roll angle. Considering roll rotation, as shown in Fig. 5a, the estimated result is close to the result from control input; however, roll angle from compass is not reliable with some significant error. For pitch angle, there are no significant difference between the curves. Comparing pitch from input control and estimated result, the maximum difference is about 0.5 degree, which is within the error in the allowed range. In addition, pitch angle from the compass is more reliable.

Figure 6 shows the results of the detection of illuminated area which are shown in blue lines and the ideal border lines generated from estimated pitch and roll angles which are plotted in red lines. As shown in Fig. 6a, it is difficult to detect the whole upper boundary due to absorption; thus, the boundary in the middle part is used in this study. During roll rotation, the upper boundary and lower boundary become inclined and when roll angle is close to 90 degree, the boundaries are extremely close. Under ideal condition, if $\varphi_x = 90$ degree, the upper boundary and lower boundary will become one single line. Figure 6 shows the effectiveness of our boundary detection method and the accuracy of the estimation of roll and pitch angles. The red lines can follow the blue lines well though the blue lines are noisy and exist some outliers as shown in Fig 6c.



(a) (b)

Fig. 4. Height measurement: (a) is the image took by a diver, (b) shows how to calculate the height.



(a) (b)

Fig. 5. Estimation of roll and pitch angles: (a) roll angle from compass, input control and proposed method, and (b) pitch angle from compass, input control and proposed method.

For a further test of the estimated values, roll and pitch angles are used for 3D mapping, with an assumption that the other extrinsic parameter does not change during the roll rotation. OctoMap is used with a minimum resolution of 2 cm. It is worth to be mentioned that the bricks in the underwater environment is about 5 cm, which is difficult to be reconstructed due to the noisy data. Point cloud is generated from OctoMap by taking the centroid of each voxel and filtered by statistical outlier filter (Rusu and Cousins (2011)). Angle information from the compass, input control and the estimated result are used for 3D mapping. The result is shown in Fig. 7. Mapping result using compass data is with large error that we can barely recognize the plane as shown in Fig. 7a. Result using input control data is much better that we are able to recognize the plane and some bricks as shown in Fig. 7b. Figure 7c indicates the mapping result using the estimated angles. The generated map is very similar to the one using input control information and some bricks can be even more clear to be recognized. Under the situation that input control information is unavailable, rotation information can be directly estimated from the images based on our method. Mapping results will be compared to a true map in future.

6. CONCLUSION

In this paper, a new concept named illuminated area is proposed which is an important characteristic of the acoustic image with a method that can automatically detect this area if the ground is a plane. With known height information of acoustic camera, vertical beam width can be estimated. Roll and pitch angles can also be estimated using illuminated area information. With roll and pitch angles, 3D mapping of underwater environment can be carried out. Experiment proves the effectiveness of our proposed method.

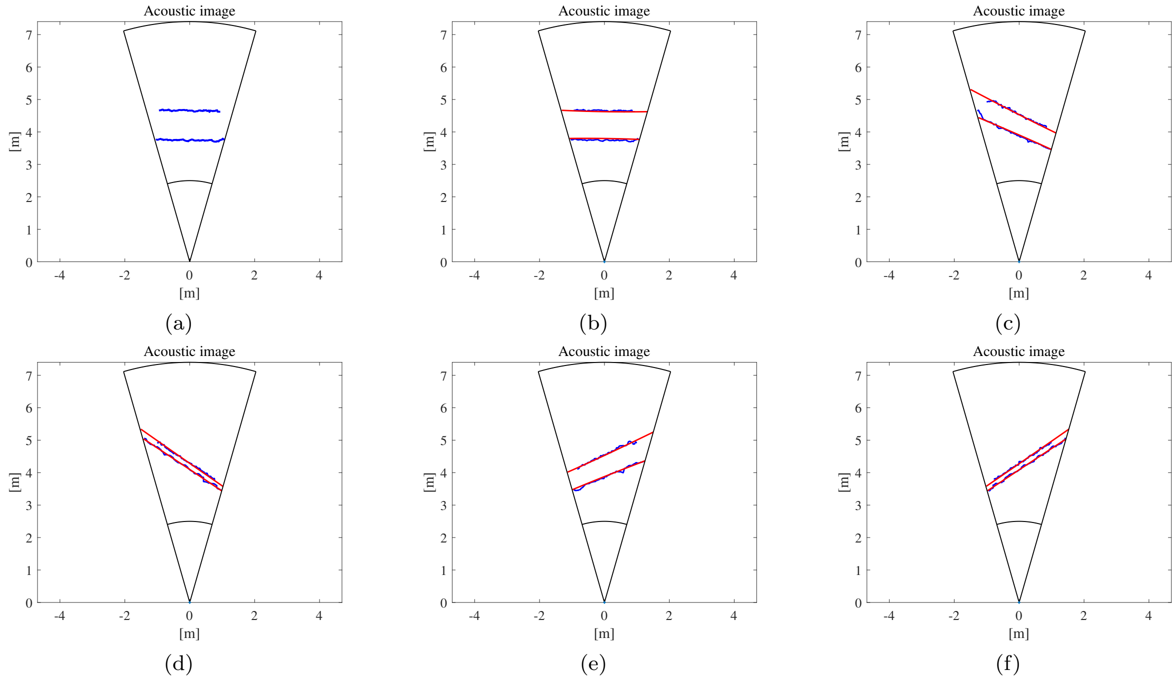


Fig. 6. Detection result and ideal boundaries from estimated angles: (a) boundaries detection when roll angle from input control is zero, (b) ideal boundaries from estimated angles with detected boundaries when roll input is zero, (c) roll input is -45 degree, (d) roll input is -90 degree, (e) roll input is 45 degree, and (f) roll input is 90 degree.

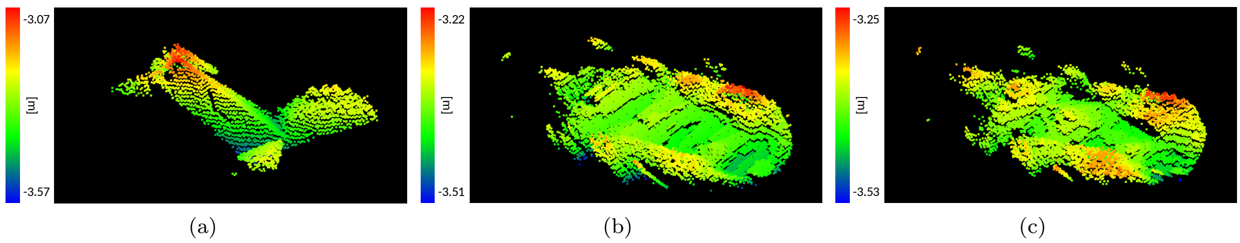


Fig. 7. 3D Mapping result:(a) from compass data, (b) from control input, and (c) from estimated angle.

Future work may include test on estimation from only one boundary, estimation of roll, pitch and depth simultaneously, 6 DoF pose estimation of acoustic camera combining illuminated area with other information, an efficient way to acquire ground truth like the true map to better evaluate our method and extracting the boundaries even if there are objects.

REFERENCES

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

Cho, H., Kim, B., and Yu, S. (2018). Auv-based underwater 3-d point cloud generation using acoustic lens-based multibeam sonar. *IEEE Journal of Oceanic Engineering*, 43(4), 856–872.

Hornung, A., Wurm, K.M., Bennewitz, M., Stachniss, C., and Burgard, W. (2013). OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*, 34(3), 189–206.

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

Lee, Y., Choi, J., Ko, N.Y., and Choi, H.T. (2017). Probability-based recognition framework for underwater landmarks using sonar images. *Sensors*, 17(9), 1953.

Negahdaripour, S., Aykin, M.D., and Sinnarajah, S. (2011). Dynamic scene analysis and mosaicing of benthic habitats by fs sonar imaging - issues and complexities. *Proceedings of OCEANS 2011 MTS/IEEE KONA*, 1–7.

Rusu, R.B. and Cousins, S. (2011). 3d is here: Point cloud library (pcl). *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA2011)*, 1–4.

Vilarnau, N.H. (2014). Forward-looking sonar mosaicing for underwater environments.

Wang, Y., Ji, Y., Woo, H., Tamura, Y., Yamashita, A., and Asama, H. (2018). 3d occupancy mapping framework based on acoustic camera in underwater environment. *IFAC-PapersOnLine*, 51(22), 324–330. 12th IFAC Symposium on Robot Control SYROCO 2018.

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