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 vehicle (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...
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 $\phi_{\text{max}}$ and the total width of $N$ beam slices in the azimuth direction is $\theta_{\text{max}}$. Field of view (FoV) is controlled by range $r_{\text{min}}$ and $r_{\text{max}}$ as shown in Fig. 1b.

When considering a point $(c^x, c^y, c^z)$ in camera Cartesian coordinate system as $(\mathfrak{R}, \theta, \phi)$. The conversion of the two coordinate systems is

$$
\begin{bmatrix}
c^x \\
c^y \\
c^z
\end{bmatrix} =
\begin{bmatrix}
\mathfrak{R} \cos \phi \cos \theta \\
\mathfrak{R} \cos \phi \sin \theta \\
\mathfrak{R} \sin \phi
\end{bmatrix},
$$

(1)

where the inverse transformation is

$$
\begin{bmatrix}
\mathfrak{R} \\
\theta \\
\phi
\end{bmatrix} =
\begin{bmatrix}
\sqrt{c^2 x^2 + c^2 y^2 + c^2 z^2} \\
\tan^{-1} \left( \frac{c^2 y}{c^2 x} \right) \\
\tan^{-1} \left( \frac{c^2 z}{\sqrt{c^2 x + c^2 y}} \right)
\end{bmatrix}.
$$

(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 $\mathfrak{R} \times \theta$ matrix and after processing, a fan-shape image is presented by mapping $(\mathfrak{R}, \theta)$ to $(l_x, l_y)$ as

$$
\begin{bmatrix}
l_x \\
l_y
\end{bmatrix} =
\begin{bmatrix}
\mathfrak{R} \sin \theta \\
\mathfrak{R} \cos \theta
\end{bmatrix}.
$$

(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 Vilarneau (2014) and effective region in Cho et al. (2018). If $r_{\text{min}}$ and $r_{\text{max}}$ 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

$$
L_u \triangleq S_u \cap P,
$$

(4)

$$
L_l \triangleq S_l \cap P.
$$

(5)
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 light 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 \((x, y, z, \phi, \varphi_y, \varphi_z)\) where \((\varphi_y, \varphi_y, \varphi_z)\) 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 \(\phi = \frac{\theta_i}{2}\) is denoted as \(r_i^u\) and the ray with elevation angle \(\phi = -\frac{\theta_{max}}{2}\) is denoted as \(r_i^l\). The direction vectors \(c_{v_i}^u\) and \(c_{v_i}^l\) of rays \(r_i^u\) and \(r_i^l\) in camera coordinate are

\[
\begin{align*}
\begin{bmatrix} c_{v_i}^u \end{bmatrix} & = \begin{bmatrix} \cos \frac{\theta_{max}}{2} \cos \theta_i \cos \frac{\theta_{max}}{2} \sin \theta_i \sin \frac{\theta_{max}}{2} \end{bmatrix}^T, \\
\begin{bmatrix} c_{v_i}^l \end{bmatrix} & = \begin{bmatrix} \cos -\frac{\theta_{max}}{2} \cos \theta_i \cos -\frac{\theta_{max}}{2} \sin \theta_i \sin -\frac{\theta_{max}}{2} \end{bmatrix}^T.
\end{align*}
\]

Then, they are transferred to world coordinate by

\[
\begin{bmatrix} w_{v_1} \\ w_{v_2} \end{bmatrix} = R \begin{bmatrix} c_{v_1} \\ c_{v_2} \end{bmatrix},
\]

where \(w_{v}\) is direction vector in world coordinate and \(c_{v}\) is the direction vector in camera coordinate and rotation matrix \(R\) is

\[
R = \begin{bmatrix} c_{\varphi_y} c_{\varphi_z} & c_{\varphi_z} s_{\varphi_y} s_{\varphi_z} - c_{\varphi_z} s_{\varphi_z} s_{\varphi_y} + c_{\varphi_y} c_{\varphi_y} s_{\varphi_z} & c_{\varphi_y} s_{\varphi_y} s_{\varphi_z} + c_{\varphi_y} s_{\varphi_z} s_{\varphi_y} \\ -s_{\varphi_y} c_{\varphi_z} & c_{\varphi_z} s_{\varphi_y} s_{\varphi_z} + c_{\varphi_z} s_{\varphi_z} s_{\varphi_y} - c_{\varphi_y} c_{\varphi_y} s_{\varphi_z} & c_{\varphi_y} s_{\varphi_y} s_{\varphi_z} + c_{\varphi_y} s_{\varphi_z} s_{\varphi_y} \\ s_{\varphi_y} s_{\varphi_z} & s_{\varphi_z} s_{\varphi_y} s_{\varphi_z} - s_{\varphi_z} s_{\varphi_z} s_{\varphi_y} & c_{\varphi_y} c_{\varphi_y} s_{\varphi_z} + c_{\varphi_y} s_{\varphi_z} s_{\varphi_y} \end{bmatrix}.
\]

(9)

The measurement points can be denoted as \((\mathbf{R}_i, \theta_i)\). Note that \(\theta_i\) 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.
Finding upper boundary
Finding lower boundary

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 $\varphi_x$ 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 $\phi_{max}$

The optimization process is based on the following equation.

$$\varphi_x, \varphi_y, \hat{\phi}_{max} = \arg\min_{\varphi_x, \varphi_y, \hat{\phi}_{max}} \frac{1}{N_u} \sum_{i = u_x}^{u_c} ||\mathfrak{R}_i - \hat{\mathfrak{R}}_i||_2 + \frac{1}{N_l} \sum_{i = l_x}^{l_c} ||\mathfrak{R}_i - \hat{\mathfrak{R}}_i||_2.$$  \hspace{1cm} (15)

where $\mathfrak{R}$ is the ideal value from ray tracing model and $\hat{\mathfrak{R}}$ is the measured value from real image. $[u_x, u_c]$ is the interval that $N_u$ beam slices within this interval are taken into consideration. Similarly, $[l_x, l_c]$ 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 $\varphi_x, \varphi_y$ 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 $\phi_{max}$ to the boundaries is getting smaller. If roll angle is up to 90 degree, even if the $\phi_{max}$ 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 $\phi_{max}$. For the sake of simplicity, in the following part of this research, $z$ and $\phi_{max}$ 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, $\phi_{max} = \hat{\phi}_{max}$.

4.1 Roll and pitch angle estimation

The optimization method to acquire $\varphi_x, \varphi_y$ is similar to the optimization method above. Since $\phi_{max}$ is considered to be a known value, the cost function is as follow.

$$\varphi_x, \varphi_y = \arg\min_{\varphi_x, \varphi_y} \frac{1}{N_u} \sum_{i = u_x}^{u_c} ||\mathfrak{R}_i - \hat{\mathfrak{R}}_i||_2 + \frac{1}{N_l} \sum_{i = l_x}^{l_c} ||\mathfrak{R}_i - \hat{\mathfrak{R}}_i||_2.$$  \hspace{1cm} (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

$$z = h_1 + D \sin \alpha + 0.5L \cos \alpha.$$  \hspace{1cm} (17)

where $\alpha$ 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 $\phi_{max}$, 13 images captured from the same pose where the input control of roll angle is zero are used to estimate the $\varphi_x$, $\varphi_y$ and $\phi_{max}$ value. The average values and standard deviations are $\bar{\varphi}_x$, $\bar{\varphi}_y$ and $\phi_{max}$ are $-1.36 \pm 0.27$ degree, $55.4 \pm 0.05$ degree and $16.24 \pm 0.03$ degree. It can be seen that $\phi_{max}$ is a little bit larger than 14 degree.

$\phi_{max}$ 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 $\varphi_x$ and $\varphi_y$. 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.

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


