A Method for Automatic Detection of Underwater Objects using Forward-looking Imaging Sonar

Jeonghwe Gu, Juhyun Pyo, Hangil Joe, Byeongjin Kim, Juhwan Kim, Hyeonwoo Cho and Son-Cheol Yu
Pohang University of Science and Technology (POSTECH), Pohang, Korea
Email: bygu111@postech.ac.kr

***Abstract*—A method for automatic detection or recognition using high-frequency forward-looking imaging sonar is proposed. The method includes segmentation of an underwater object's echo shape or its acoustic shadow shape. For accuracy of segmentation, some constraints were assumed on the underwater ground and the sonar's pose. The separated shape was compared to simulated reference shapes to know its orientation or identity by the shape matching method that use shape context concept. An experiment conducted in a wave tank validated the method qualitatively. The method can be applied to an AUV's object-searching applications.**

I. Introduction

There are a variety of sonar systems for ships, buoys, and underwater vehicles for various purposes such as navigation, obstacle avoidance, and environmental observation. The imaging sonar is one of such systems and is designed to provide visual scenes of underwater environment. The imaging sonar operated with over 1.0 MHz generates images of which resolution is high enough for human to identify objects in the images. Also, other imaging sonars having low frequency such as side-scan sonars are used to cover large areas of the sea floor. Many research have studied methods or algorithms dealing with those sonar data or images. M. Mignotte et al. studied the segmentation of sonar images using Markov random field model [1]. In this study, segmentation of sonar images follows some repeating steps considering neighbors' state for classification. This method is particularly useful for stationary sonar images covering large area. Multiple objects tracking in sonar images obtained by the sector scanning sonar was studied by David M. Lane et al. for the purpose of obstacle avoidance, visual servoing and underwater surveillance [2]. For the feature matching, S. Daniel et al. devised an algorithm where sonar images obtained by the side-scan sonar are matched [3]. These methods are usually not general but highly dependent on the types of imaging sonars. This is because the operational mechanism of imaging sonars is different and there is no standard in sonar data formats. Therefore it is normal to develop methods or algorithms specific to each imaging sonar.

In this paper, we devised a method that can be used for automatic detection or recognition using an AUV and imaging sonar together. This method is specialized to a high-frequency forward-looking imaging sonar. Firstly, we will define the problem situation with assumed constraints and then we propose a method to solve the problem. Lastly, the method will be verified by an experiment.

II. Problem statement

A. Motivation and Objective

This research is motivated from an observation that we human can identify underwater objects by inferring their shapes appearing in the sonar images. The shapes mean either the objects' echo signals that have specific forms in the images or the shapes of their acoustic shadows. Fig. 1a shows examples of the sonar images taken by the dual-frequency identification sonar (DIDSON) (Fig. 1b) and we can infer the underwater objects or structures by their outlines or shapes in the images.

Therefore the shapes appearing in the sonar images are important factors when determining the identity of underwater objects. Except those shapes, there are little information available in the sonar images; the images are gray-scaled and have blurred edge lines. Furthermore the image variation are large as the sonar pose changes. These characteristics of the sonar images made the development of advanced image process algorithms such as recognition and tracking difficult. Our objective is to develop an segmentation algorithm that separates the object-related shapes from the sonar images, and then analyze and utilize the shape information for various purposes. However, we saw the general solution for segmentaion as impossible one, so inevitably we assumed some conditions or constraints to the problem.

B. Constraints

Some constraints on underwater environment being investigated and the sonars pose are assumed to mitigate the difficulty of the problem. This is mainly for easy segmentation of the target shapes. The desirable segmentation implies the precise separation of the shapes from the background. Typical one threshold value can not solve this problem. The constraints on the underwater ground and the fixed sonar pose can restrict the form of sonar images. This will be discussed again in the next section (segmentation). The first constraint is that the underwater ground should be generally flat and homogeneous in texture, that is, in the sonar images, the region corresponding to the ground echo is almost invariable. The second constraint is that the altitude and view angle of the sonar should be fixed. (Fig. 1c) The proper values for the altitude and view angle will depend on the size of target objects. This second constraint confines the form of the sonar images by fixing the size and position of the ground echo within the images. In the next section, we will explain the segmentaion and shape matching method in a stepwise manner. before that, we briefly introduce the imaging sonar, DIDSON, used in this research.

978-0-933957-43-5 ©2015 MTS

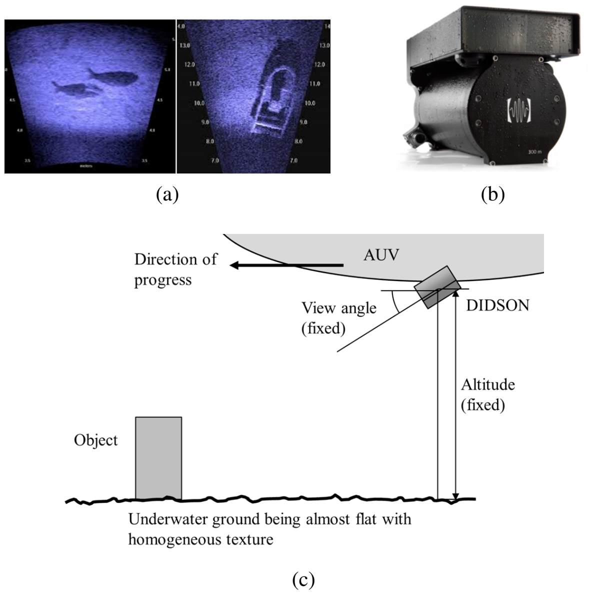


Fig. 1. (a) Examples of sonar images by the DIDSON (b) DIDSON, the imaging sonar studied in this paper (c) Problem constraints on the sea ground and the sonar pose

C. Imaging Sonar

The DIDSON is a forward-looking imaging sonar operated with two different frequency (1.1 MHz/1.8 MHz). In 1.8 MHz mode, the sonar transmit 96 acoustic beams in the forward direction with a field of view of 29 $^\circ$ and the observation range is up to 15 m [4]. Thanks to its relative high frequency, the raw image generated has 96 $\times$ 512 pixel resolution that is enough to identify underwater objects. The sonar can be mount on an AUV and used for obstacle avoidance. The method proposed in the paper is designed based on the DIDSON.

III. Proposed Method

A. Overview

Overally, the proposed method can be divided into two steps. The first step is segmentation where the object-related shapes (object echo and acoustic shadow) are separated from the background. The second step is an analysis of the extracted shapes. The analysis will be done by comparing the shapes with well-known reference shapes and investigating the similarity between them. The reference shapes can be generated by simulating the sonar. As the comparison method, the shape matching method using shape context is chosen [5]. The segmentation and shape matching method will be discussed in detail throughout the next paragraphs.

B. Segmentation

In the segmentation step, the ground echo should be removed in the images because it is a unnecessary and subordinate information in the analysis of the object-related shapes. Fig. 2 shows a sonar image sequence where an object is observed when the sonar is in motion. A single threshold value is not enough to solve the segmentation problem, so we need additional assumptions. As stated earlier, we fixed the altitude and view angle of the sonar. This constraint makes

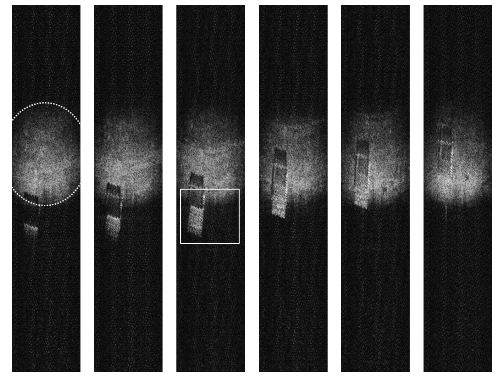


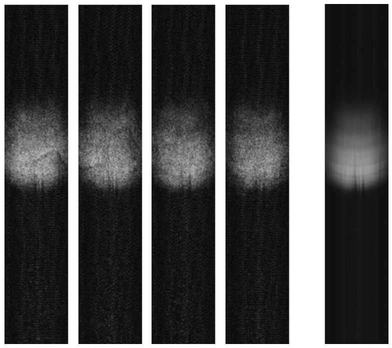
Fig. 2. Sonar image sequence obtained by the sonar showing an object appearance. (white ellipse: the size and position of the background echo, white rectangular: the ROI region where the object echo is well observed.)

the echo of the ground shown in the images fixed in term of the position and size (Fig. 2) while only that of the object is changing. If we know the position and size of the ground's echo as *prior* information, we can subtract the ground echo signal from target sonar images to remove the ground shape. Other observation can help solve the segmentation problem. In sonar images, there is a specific domain where an objects echo shape is well seen or observed. This domain is normally located in the front of the ground region. Note that when the object is at the same distance from the imaging sonar (the sixth frame in Fig. 2), it is difficult to differentiate them because the scattering strengthes from the object and ground are similar. However if the object is located a little closer, that is, in the front of the ground, the object can be well seen (the third frame in Fig. 2). Based on this observation, we can set the region of interest (ROI) within the images for computational efficiency. The exact definition of ROI is not necessary, so it can be set manually.

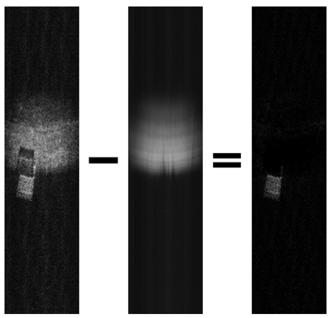
As the next step in segmentation, we need the sonar images where only the ground is shown. (Fig. 3a) We will call these images as the training images. The training images are collected to find their average and standard deviation by dealing them as 2 $\times$ 2 matrix. Then, the average image is subtracted from target sonar images to remove the ground shape (Fig. 3b) and the resulted image is classified into three levels: object, ground, and shadow. The decision in that classification will be based on the standard deviation value computed. Depending on the interested shape, either the object echo or acoustic shadow, it becomes two-level classification. Because the classification is performed for each pixel, there can be many speckles in the images. As the final step in segmentation, the simple post process is done to remove those speckles. These steps are summarized in Fig. 4.

C. Shape Matching

After segmentation, the separated shape is analyzed by a shape matching method. In the shape matching method, the shape is compared with other shapes and the similarity between them is quantified. This can be used for a given shape to find the closest one. Among many shape matching methods, we chose the shape matching method which uses the concept of shape context [5]. Shape context can be considered as a point feature where it describes its neighbor points' distribution in



(a)



(b)

Fig. 3. (a) Training images where only the background echo exists. (the rightmost is the average image) (b) Segmentation by subtracting the background from the target sonar image

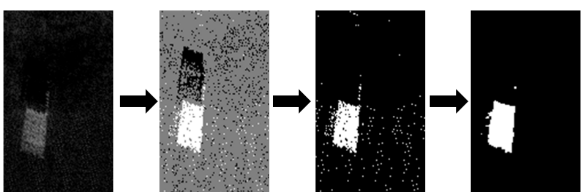


Fig. 4. Segmentation procedure (background subtraction, three-level classification, two-level classification, and speckle reduction from left to right)

terms of distance and direction. The neighbor points' relative distribution is specified by two-dimensional histogram and the histogram is defined as the shape context. We briefly introduce the steps of shape matching method here, but for detailed description you may refer to [5]. In the first step, the edges of the shapes to be compared are found using typical edge detection algorithms. For binary images, edge detection using the mask is simple and fast [6]. In the second step, from the edges, points are sampled so that they are evenly spaced from each other. The degree of gab between points can be chosen by users and it implies the precision of the method. In other words, the dense point sampling can raise the precision of the method, but in the same time it increases computational load. In the third step, the shape context is computed for each point sampled. The computation of the shape context means the histogram making which summarizes information of the neighbor points' distribution. In the last step, the shape contexts or the histograms of two shapes are one to one matched so that the matched shape contexts have the similar tendency. In other words, every shape context from the two shapes is compared by chi-squared test and their similarity is computed as a cost, and then it finds the best one to one

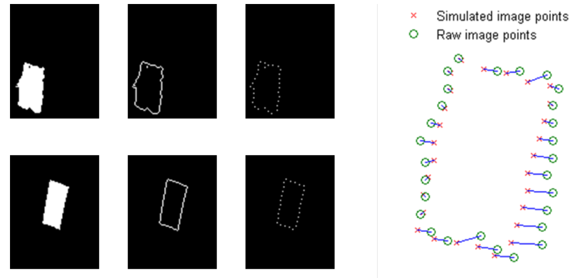


Fig. 5. Procedure of shape matching method (two shapes to be compared, edge detection, point sampling, and one to one match using the shape context from left to right)

match so that their total matching is lowest. The final matching cost can be considered as the degree of similarity between the shapes. Fig. 5 shows the shape matching procedures stated.

The shape matching method needs two shapes for comparison. In other words, we should prepare some reference shapes to be compared with the extracted shape. For the reference shapes, we perform a simulation where artificial sonar images that have compliance to the sonar are generated. As the Input parameters, we give the object information as 3D graphics and the sonar's observation conditions such as altitude and view angle. As a result, we obtain artificial but ideal sonar images under identical observation condition as in real experiments. The precision of the simulated images in terms of size and shape are verified to be reasonable. The reference shapes can prepared by extracting the corresponding shape from simulated images. The principle of the simulation is described in [7].

The comparison of the extracted shape with the reference ones provides us with useful information about an object. If we know an object in advance, we can estimate the orientation of the object, which is usually unknown. If the object is one of the objects known, we can identify the object by comparison. These need to be checked by an experiment. In the next chapter, we will introduce the experiment that is conducted to verify the algorithm.

IV. Experiment

A. Overview

An experiment was conducted to check the feasibility of the proposed method and furthermore its usefulness. The experiment was done in a wave tank (Fig. 6). The target object to be observed is a container (Fig. 7a). The four of the objects were put on the tank floor upside down with about 3 m space apart. The tank floor as the background was made from cement and almost flat throughout the tank. The sonar was attached to the mechanical device that can move along the wave tank with desired speed. The sonars altitude from the ground was set to 3 m and the view angle 45 degree. With this fixed pose, the sonar proceeded along the wave tank with the speed of 1 m/s. With this experimental setting, we obtained the sonar images of the objects lying on the tank ground.

For segmentation, About 200 frames of sonar image were used as the training images that only contain the echo of the ground. Fig represents the examples of those training images. The variance of training images was low because

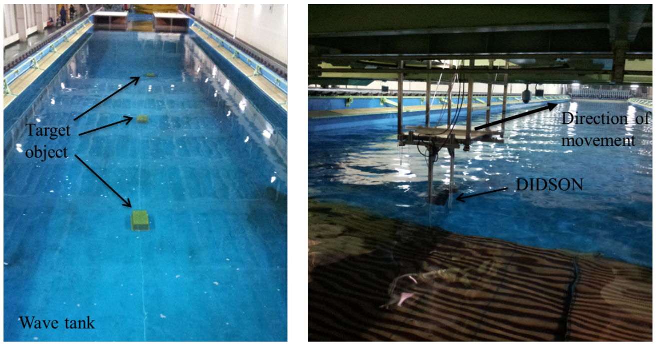
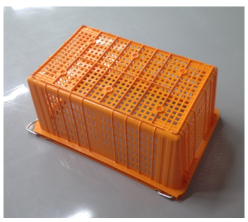
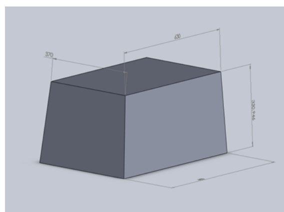


Fig. 6. Pictures of experimental setting showing the wave tank having the dimension of 50×4×4 m (L×W×D), arrangement of target objects and the sonar.



(a)



(b)

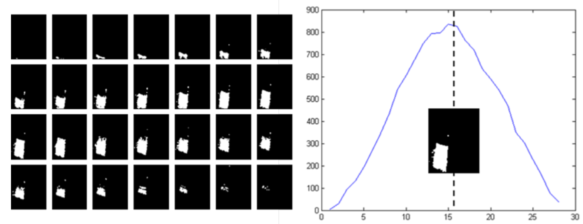
Fig. 7. (a) Picture of the target object having the dimension of 0.63×0.37×0.33 m (L×W×H) (b) Approximated 3D model of the object for the simulation

the ground was a uniform structure (flat and clear). In the shape matching step, the identical measurement condition was used to generate reference shape in the simulation. The object was approximated as the 3D model, but the dimension was set identically with real ones. (Fig. 7b) As a result of the simulation, we obtained the reference shapes with different orientations (label 1 through 12 in Fig. 9). Also we get some other objects simulated images for shape comparison (label 13 through 16 in Fig. 9)

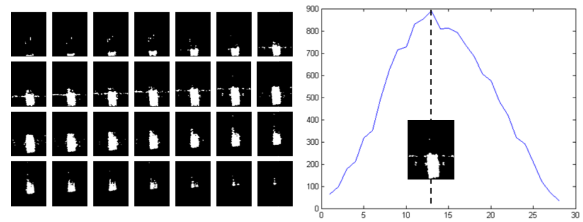
B. Results on Segmentation

We conducted the segmentation process to the four objects that appeared in the image streams. The process was done only on the ROI region, and as results the left figures in Fig. 8 represents the resulted image sequence after the segmentation. We can observe the variation of objects echo shapes from appearance to disappearance as the frame goes. In usual AUV applications using the imaging sonar, objects will be observed in this manner. Shape (a) and (d) in Fig were well separated from the background, and the shapes were distinct with few speckles. Shape (b) and (d) had side noises that deteriorated the shape form, but still the SNR of shape (c) is low. However, shape (d) had much larger noise than (c), which made the identification of the original shape difficult. These noises might come from the interference phenomenon between acoustic waves.

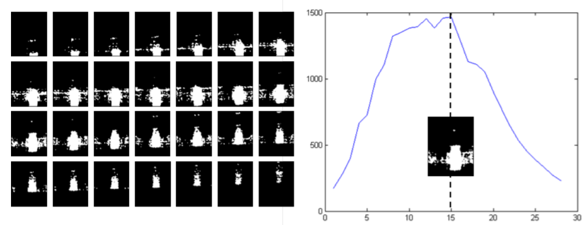
We should choose one shape among these shape sequences. Obviously we human can select the most appropriate one for shape analysis. However, we want to make the decision automatically. The key lies on an area of the shapes. The area means the number of pixels that were considered as shape class (or white pixels in the images). The variation of the area as the frame changes was drawn and the shape corresponding to the peak in the graph was well agreed to our intuitive selection (Fig. 8). Fig. 9a shows the final shapes (labeled as I to IV) selected for the next shape analysis.



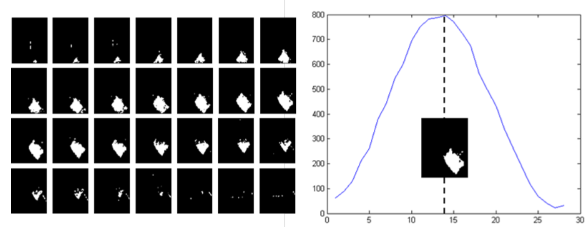
(a)



(b)



(c)



(d)

Fig. 8. (a)-(d) Segmentation results on the four objects (left: Image frames on the ROI from top-left to bottom-right in sequence, right: The shape's area variation as the frame changes

C. Results on Shape Matching

After the segmentation and selection of the appropriate shape, the selected shapes were compared to corresponding reference shapes. The reference shapes are shown in Fig. 9. Shape labels 1 through 12 in Fig. 9 represent the same object but different orientation, while 13 through 15 represent the different objects to target object. We implemented the shape matching method using shape context. The result summarized in Fig. 10. The shape similarity was defined by the average cost that is the result of chi-squared test between shape contexts. Therefore the average cost of 0 means that two shapes are identical while 1 means the maximum value the cost can

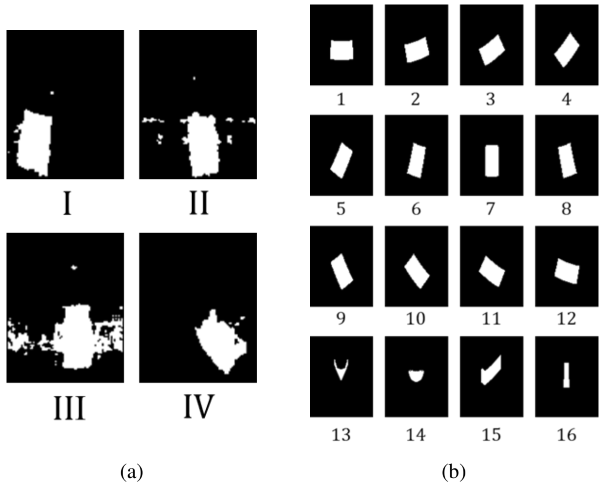


Fig. 9. (a) Extracted and selected shapes labeled as I through IV (b) Simulated reference shapes, label 1 through 12 representing different orientations and label 13 through 16 representing different objects

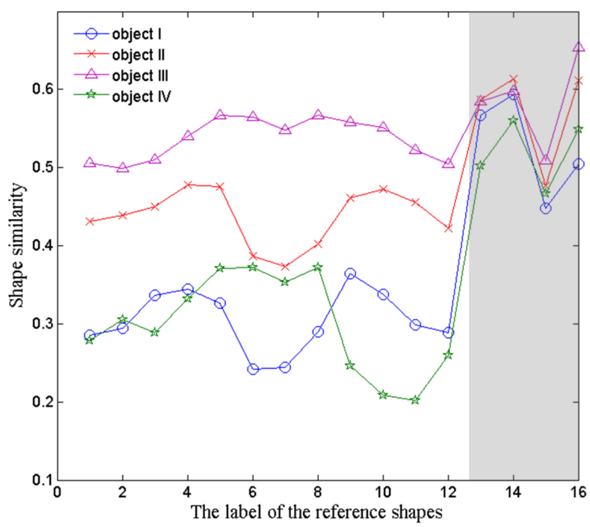


Fig. 10. The shape similarity between the extracted shapes and the reference shapes (non-shaded area: orientation comparison, shaded area: object comparison

reach. In short, the smaller the cost become, the more similar the shapes are. Lable I, II, and IV were matched to label 6, 8 and 11 respectively although label II had some noise in it. These matching results seem accurate based on our intuition. Label III had relatively high cost and was less variable compared to others. A large amount of noise would result in this tendency. The shaded area in Fig. 10 indicates the comparison with different objects. The costs were higher than the cases when compared to itself and showed uniform tendency. In conclusion, the shape matching result showed that the extracted shape can be analyzed by comparison, so its orientation and identification can be estimated.

V. Conclusion

We proposed a method for automatic detection or recognition using high-frequency forward-looking imaging sonar. The method includes segmentation of an underwater object's echo shape by removing the ground echo shape. Then, the separated shape is compared to simulated reference shapes to know its orientation or identity using the shape matching method. In the experiment, the objects echo shapes were well separated from the ground echo. Also, the shape matching method identified their orientation and identity. This method can be applied to an AUVs detection mission such as mine hunting.

Acknowledgment

This research was supported by the Ministry of Science, ICT and Future Planning, Korea, under the ICT Consilience Creative Program (IITP-2015-R0346-15-1007) supervised by the Institute for Information & communications Technology Promotion, the project titled ”Gyeongbuk Sea Grant Program”, funded by the Ministry of Oceans and Fisheries, Korea and the ONR Global grant (N62909-13-1-N290).

References

[1] M. Mignotte *et al.*, “Sonar image segmentation using an unsupervised hierachical MRF model,” *IEEE Trans. Image Process.*, vol. 9, pp. 1216-1231, Jul. 2000.

[2] D. M. Lane *et al.*, “Rubust tracking of multiple objects in sector-scan image sequences using optical flow motion estimation,” *IEEE J. Oceanic Eng.*, vol. 23, pp. 31-45, Jan. 1998.

[3] S. Daniel *et al.*,“Side-scan sonar image matching,” *IEEE J. Oceanic Eng.*, vol. 23, pp. 245-259, Jul. 1998.

[4] E. Belcher *et al.*, “Dual-frequency identification sonar (DIDSON),” in *Proc. 2002 Int. Symp. Underwater Technology*, 2002, pp. 187-192.

[5] S. Belongie *et al.*, “Shape matching and object recognition using shape contexts,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 509-522, Apr. 2002.

[6] A. M. Bahaa-Eldeen *et al.*, “Edge detection of binary images using the method of masks,” Univ. Ain Shams, Faculty of Eng. Scientific Bulletin, vol. 35, no. 3, pp. 349-355, 2000. Available: http://arxiv.org/abs/1401.5245

[7] J. Gu *et al.*, “Development of image sonar simulator for underwater object recognition,” presented at *OCEANS 2013*, San Diego, CA, 2013