Detection and localization of an underwater docking station in acoustic images using machine learning and generalized fuzzy hough transform

Divas Karimanzira and Helge Renkewitz

Department of Surface and Maritime systems, Fraunhofer IOSB-AST, Ilmenau, Am
Vogelherd 50, 98689, Ilmenau, Germany
`divas.karimanzira@iosb-ast.fraunhofer.de`
<http://www.iosb-ast.fraunhofer.de>

Abstract. Long underwater operations with autonomous battery charging and data transmission require an Autonomous Underwater Vehicle (AUV) with docking capability, which in turn presume the detection and localization of the docking station. Object detection and localization in sonar images is a very difficult task due to acoustic image problems such as, non-homogeneous resolution, non-uniform intensity, speckle noise, acoustic shadowing, acoustic reverberation and multipath problems. As for detection methods which are invariant to rotations, scale and shifts, the Generalized Fuzzy Hough Transform (GFHT) has proven to be a very powerful tool for arbitrary template detection in a noisy, blurred or even a distorted image, but it is associated with a practical drawback in computation time due to sliding window approach, especially if rotation and scaling invariance is taken into account. In this paper we use the fact that the docking station is made out of aluminum profiles which can easily be isolated using segmentation and classified by a Support Vector Machine (SVM) to enable selective search for the GFHT. After identification of the profile locations, GFHT is applied selectively at these locations for template matching producing the heading and position of the docking station. Further, this paper describes in detail the experiments that validate the methodology.

Keywords: Object detection and localization · Classifier · Generalized fuzzy Hough Transform · Selective search.

1 Introduction

Extended underwater applications such as the inspection and maintenance of underwater structures require an autonomous underwater vehicle(AUV) with autonomous docking capability for battery charging and data transmission. Underwater docking is a complex process and is composed of several sub-tasks, i.e., detection and localization, homing, physically attach to recharge AUV batteries and to establish a communication link, to wait in a low power state for a new mission, and to undock. In this paper we will focus on the *detection and localization* part of the docking process. It is assumed that the AUV is equipped with a

forward looking imaging sonar (FLS) as the perception system. Although sonars are not limited by turbidity, their data have some characteristics that make it difficult to process and extract valuable information. These characteristics are given in [1], and include non-homogeneous resolution, non-uniform intensity, speckle noise, acoustic shadowing and reverberation and multipath problem. In addition, the data are often only cross sections of the objects. In literature, some works which propose strategies to identify objects in acoustic images can be found, e.g. [1, 2]. Santos et al., developed a special system which uses acoustic images acquired by FLS to create a semantic map of a scene [1].

Acoustic images require preprocessing. Therefore, many works have been carried out for filtering and enhancement of acoustic images. Very important is the insonification described in [3, 4] where a sonar insonification pattern (SIP) obtained from averaging a large number of acoustic images taken from the same position is applied to each acoustic image reducing the effects of the non-uniform insonification and the overlapping problem of acoustic beams. Other artifacts such as speckle noise can thereafter be removed by Lee filtering [7]. Image enhancement intensifies the features of images. In [3], a method specifically developed for enhancing underwater images known as mixed Contrast Limited Adaptive Histogram Equalization (CLAHE) was discussed. Its results show less mean square error and high peak signal to noise ratio (PSNR) than other methods. Another technique for enhancement of acoustic images is using Dynamic Stochastic Resonance (DSR) technique. It has been used for enhancement of dark low contrast images in [5].

In image classification tasks, having proposal of the locations of potential objects tends to increase classification accuracy. Modern methods are based on a way to selectively search images for potential object locations by using a variety of segmentation techniques to come up with potential object hypotheses. For example, thresholding technique is used to group pixels in which input gray scale image is converted into binary image based on some threshold value [6]. Machado et al. proposed a method specifically for acoustic images [4], where the regions of interest are extracted by a linear search finding pixels with intensities higher than a certain value.

As for the detection methods, which are invariant to rotations, shifts and scale changes of objects, the Generalized Hough Transform (GHT), a geometric hashing, and variations on these methods have been proposed so far. However, the major weakness of GHT is that the scale and rotation of the object is handled in a brute-force approach which requires a four-dimensional parameter space and high computational cost.

In this paper, a method is provided which detect and recognize a docking station in a scene. The captured acoustic image is segmented, and the shape of each segment is described geometrically. Each shape is then classified into two main classes (aluprofile, and obstacles) using the well-known Support Vector Machine (SVM) algorithm. After identification of the aluprofile locations, GFHT is applied at these locations for template matching producing the heading and position of the docking station.

2 Methodology

The task is to find the location of the docking station in an acoustic image which is rather noisy by optical imaging standards. Therefore, the proposed method illustrated in Fig. 1 has six steps which include data collection, image filtering and enhancement, segmentation, segment description, classification and localization. In the first step, four different filters, each of which aims to revise a special defect

Fig. 1. Proposed methodology pipeline

are applied to the acoustic images. In the second step an automatic segmentation process of the images based on intensity peak analysis is conducted. In the third step the segments are converted to Gaussians that are easily described by shape descriptors such as the width, height, inertia ratio, area, hull area, convexity and pixel intensities information. The shape descriptors form the feature vectors which are applied in the pre-last step where a Support Vector Machine is trained to recognize the aluminum profiles of the docking station and finally GFHT [8] is applied to localize the complete docking station. In the following sections the methods applied in all steps will be discussed.

Step 1: Image denoising and enhancement . The acoustic image runs through a filtering pipeline to mitigate sonar defects starting from the non-uniform intensity problem through speckle removal and finally image enhancement using dynamic stochastic resonance. The first processing stage is about blurring the homogeneous regions keeping edges unharmed. Therefore, in this step we apply an image correction process to mitigate the non-uniform intensity problem and speckle noise. Typically this problem can be reduced by a mechanism to compensate the signal loss according to the distance traveled. However, the intensity variations can also have other causes, e.g., by changing the sonar tilt angle. As in [3, 4], we first compute the sonar insonification pattern by averaging a significant number of images captured by the sonar at the same spot. The insonification pattern is applied to each acoustic image reducing the defects. Now, having a pattern-free image, the remaining speckle noise and acoustic reverberation and multipath problem can be reduced in the next steps. Several filters for eliminating speckle based upon diverse mathematical models of the phenomenon exist. Speckle elimination in wavelet domain is very popular, but has some drawbacks,

e.g., the selection of appropriate threshold is very difficult. Another method, the adaptive speckle filtering include the Lee filtering technique [7] which is based on minimum mean square error with preserving edges and the Lee filter has a special property that it converts the multiplicative model into an additive one, thereby reducing the problem of dealing with speckle noise to a known tractable case. Principally, the Lee filter works the same as the Kalman filter. During speckle elimination, the value of pixel in filtered image is determined by the gain factor ($k_{(i,j)}$). It is assumed that the noise in image is unity mean multiplicative noise. If the captured noisy image is z , true image is x and noise is n , then the noisy image model can be expressed as

$$Z_{(i,j)} = x_{(i,j)} \cdot n_{(i,j)} \quad (1)$$

where the $\bar{x}$ is local mean, the speckle free pixel value is calculated by

$$\hat{x}_{(i,j)} = \bar{x} + k_{(i,j)}(z_{(i,j)} - \bar{x}). \quad (2)$$

The Lee filter tries to minimize MSE between $x_{(i,j)}$ and $\hat{x}_{(i,j)} * k_{(i,j)}$ and the gain factor $k_{(i,j)}$ is calculated by Eq. 3.

$$k_{(i,j)} = Var(x) / (\bar{x}^2 \sigma_n^2 + Var(x)) \quad (3)$$

where, $Var(x)$ is the local variance. The coefficient of variation, σ_n gives the knowledge of ratio of standard deviation to mean i.e. $\sigma_z/\bar{Z}$ over homogeneous areas of noisy image.

Another phenomenon in acoustic images is the acoustic reverberation and multipath problem which generate effects such as ghost objects. In this work, the received signal is analyzed and the homomorphic deconvolution method is applied as a means of combating the multipath problems. Followingly, the image captured by the FLS $g(x, y)$ is decomposed into the reflectance function $r(x, y)$ and the illumination intensity $i(x, y)$ using $g(x, y) = i(x, y) \cdot r(x, y)$. Using the log of the image, the components of illumination and reflectance can be separated. The log-image is Fourier transformed and High-pass filtered using the H modified Gaussian filter to Eq. 4.

$$G(w_x, w_y) = H(w_x, w_y) \cdot I(w_x, w_y) + H(w_x, w_y) \cdot R(w_x, w_y) \quad (4)$$

Inverse Fourier transform is applied to return into the spatial domain to get the filtered image.

All the filters applied up to now blurs the acoustic image. Therefore, it is important to apply some mechanism for image enhancement to strengthen some features. In this paper we tested two methods DSR and CLAHE. The CLAHE method has been described fully in [3] and therefore we refer to this article for further details. These methods have proven to be suitable for enhancing both the grayscale and colored images. The principle of the DSR is described in [5] that if optimum noise is added with the weak input signal it boosts the signal considerably and gives better signal to noise ratio (SNR).

Step 2: Sonar image segmentation, feature extraction and annotation . As objects suspended in water reflect acoustic waves more than the water environment, they are characterized by high-intensity regions on the images. Therefore, the approach for segmentation is to distinguish and separate the objects from the background [1, 4]. Due to this fact, an approach based on the acoustic image formation to detect peaks of intensity is adopted as in Santos et al. [1]. Briefly, a sonar image is composed of beams B and bins. Therefore, every acoustic beam of the acoustic image is analyzed individually for every bin. The average intensity $\bar{I}(b, B)$ is calculated for each bin b of a given beam B by Eq. 5.

$$\bar{I}(b, B) = 1/(win_{sz}) \sum_{(i=b-win_{sz})}^b I(i, B) \quad (5)$$

where win_{sz} is the window size, in number of bins, admitted in the averaging; b and i are bin identifiers; $I(i, B)$ is the intensity of i^{th} -bin of B^{th} beam. The intensity $I_{peak}(b, B)$ is an offset of $\bar{I}(b, B)$ as shown in Eq. 6

$$I_{peak}(b, B) = \bar{I}(b, B) + h_{peak} \quad (6)$$

where h_{peak} determines the minimum height of a peak of intensity. A sequence of bins with an intensity $I(b, B)$ greater than $I_{peak}(b, B)$ are considered part of a peak and are not considered on the $\bar{I}(b, B)$ computation.

Along the beam B , the bin b_{peak} with the greater intensity $I(b_{peak}, B)$ is adopted to build the segmentation parameters. Fig. 2b shows in red values of $\bar{I}(b, B)$, in blue values of $I(b, B)$ and in green values of $I_{peak}(b, B)$ of all bins of a single beam B . The peaks detected b_{peak} are represented by colored circles. After the detection of all peaks, a neighboring search for connected pixels is performed for each peak. The 8-way neighborhood criterion is adopted by the BFS algorithm. All the connected pixels are visited if $I(i, j) > \bar{I}(b_{peak}, B)$.

Fig. 2. a) Peak analysis of a single beam and (b) beam intensity profile

To be able to apply SVM, features need to be extracted from the segments. So after segmentation, a Gaussian probabilistic function is applied and shape descriptors for each segment are calculated (see Fig. 3a). Using the Singular

Value Decomposition (SVD), the eigenvalues and eigenvectors of the covariance matrix are computed, from which the largest eigenvalue and the second largest eigenvalue are used to define the width and the height, respectively. In addition, other shape descriptors can be calculated starting with the segments area which is computed using the Green's theorem, the convex hull area and the perimeter, the inertia ratio, the convexity, the mean and the standard deviation of the acoustic intensity of each segment. Almost all data are geometrical information, however the mean and the standard deviation of the intensities represents the acoustic data. After the Gaussians are automatically calculated, a manual segment annotation process is conducted (Fig. 3b).

Fig. 3. a) Acoustic image gaussians and b) annotated acoustic image

Step 3: Segment classification using supervised learning . After the description and the annotation of the segments, they are now ready for classification. Well-known supervised classifiers such as Support Vector Machine, Random Trees and K-Nearest Neighbors can be used for this purpose. The OpenCV implementation of the supervised classifiers were used on this work. Four classes of objects available in our dataset (aluprofile, obstacles(1-3)) were adopted for learning.

Step 4: Generalized Fuzzy Hough Transform . After finding all the possible positions of the aluprofiles, the next step is to use GFHT for localizing the docking station in the acoustic images using its template. For detailed description of GFHT we refer to Suetake et al. [8]. In GFHT, the fuzzy concept is introduced to the voting process. Therefore we consider the area C_k containing the feature points, with radius R_c (pixel) from the point (x_c, y_c) in question. In order to consider the effect of the neighboring feature points on (x_c, y_c) in the area C_k , membership value is given to each point in the area C_k , according to the distance from (x_c, y_c) . The following is the vote value in the voting process of the GFHT: $\text{Vote value} = \sum \text{membership value of the feature point in } C_k$. This means that the effects of the feature points around (x_c, y_c) are counted up in voting.

3 Experimental results

The experimental results were obtained using real acoustic images collected from a test basin in which the docking station was installed. A video was taken using

a 2D Forward looking sonar mounted on an AUV. From the video, 16-bits gray scale images with a resolution of 1429x768 were generated. The AUV dives to the level of the underwater docking station starting from different position of the test basin, left, right, center and moves towards the docking station while recording data.

Using the pipeline described in Fig. 1, the images went through filtering and enhancement, segmentation and finally manual annotation. The training data comprises a total of 627 segments over 33 acoustic frames manually classified into mainly two classes (aluminum profile = 330 and obstacle). The obstacle class is divided in to three categories according to their shapes to cover everything which is not an aluminum profile (Obst1=33, Obst2 = 165 and Obst3 = 99). The shape of the segment is the most distinctive feature for recognition. Therefore, the annotation for the classes was performed accordingly, the class obstacle 1 are the largest segments; the class obstacle 3 are the smallest segments and the aluminum profiles are small and the most convex segments. Overfitting is avoided by choosing the supervised classifier parameters (C , gamma for the SVM) using 5-folds cross validation. The folds are applied repeatedly and the average accuracy is used to choose the best parameters. Furthermore, normalization is required so that all the inputs are at a comparable range. For segmentation, the parameters such as the separation distance allowed between segments were defined empirically in several trials.

The best result was obtained using the SVM classifier with radial basis function kernel and $\nu = 1.442$ and $C = 11.739$. This Classifier reaches a hit rate of 98% and 93% for training and validation, respectively. In Fig. 4a, the ellipses in red are automatically detected by the segmentation algorithm, and the yellow labels have been manually defined. After running the classifier training, the labels in magenta, red (incorrect) or green (correct) appear to represent the classification assigned by the classifier. The magenta labels indicate segments without annotation to compare. The performance of the detection and localization system using GFHT is measured using the detection rate, i.e., the total number of detections compared to the actual docking station position in all images and the localization accuracy, i.e., the correct location detection of the docking station compared to its actual location in an image for all 674 images. The detection rate is quite high for all images with docking data achieved above 80% detection rate. Further, the system is able to localize the docking station on all sonar images quite accurately with an average position *error* = 3cm and average 2D orientation *error* = 3.9°.

4 Conclusions and future work

A method to automatically detect and localize a docking station in acoustic images is proposed. The acoustic image is automatically segmented and the shape of each segment is described geometrically and argued by its acoustic intensity reflected by the object. The object classification is performed by SVM classifier. The image segments are manually annotated for training. The results

Fig. 4. a) Example results of classification and b) docking station localized by GFHT

show that it is possible to identify and classify objects such as aluprofiles in real underwater environments. With the known positions of the aluprofiles, GFHT can be performed much faster for template matching of the real docking station. From the GFHT, the 2D position and orientation of the docking station is obtained and can be used by the homing algorithm. Future works will be focused on exploring and make comparisons with end-to-end deep learning.

5 Acknowledgements

The data described in the paper were collected by students based at the Fraunhofer IOSB-AST in Ilmenau. We thank these students for their hard work during data collection, especially Nguc Pham Tham.

References

1. Santos, M., Drews, P., Núñez, P., Botelho, S.: Object Recognition and Semantic Mapping for Underwater Vehicles Using Sonar Data, *Journal of Intelligent and Robotic Systems*, volume **91**(2) pp 279–289. (2018)
2. Cho, H., Pyo, J., Gu, J., Jeo, H., Yu, S.C.: Experimental results of rapid underwater object search based on forward-looking imaging sonar. In: *Underwater Technology (UT)*, 2015 IEEE, pp. 1–5 (2015).
3. Hitam, M. S., Yussof, W., Awalludin, E. A. and Bachok Z.: Mixture contrast limited adaptive histogram equalization for underwater image enhancement. *IEEE*, (2013).
4. Machado, M., Ballester, P., Zaffari, G., Drews-Jr P., and Botelho S.S.C.: topological descriptor of acoustic images for navigation and mapping in *IEEE 12th Latin American Robotics Symposium LARS*, pp. 1–6. (2015)
5. Deole M.T., Hingway S.P., Suresh S.S. Dynamic stochastic resonance for low contrast image enhancement. *IOSR JVSP*. **4** 1–5.(2014)
6. Evelin S. G., Lakshmi Y.V.S., Wiselin J.G.: MRI Brain Image Segmentation based on Thresholding, *Int. Journal of Advanced Computer Research* **3** (8). (2013)
7. Lee, J. S.: Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. PAMI; PAMI* **2**(2) 165–168. (1980)
8. Suetake, N., Uchino, E., Hirata, K.: Generalized Fuzzy Hough Transform for Detecting Arbitrary Shapes in a Vague and Noisy Image. *Soft Computing*. **10**. 1161–1168. (2006)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.