Research Article

Saliency-Based Diver Target Detection and Localization Method

Jianjiang Zhu,1 Siquan Yu,2 Lei Gao,2 Zhi Han,2 and Yandong Tang2

1School of Electrical & Automatic Engineering, Changshu Institute of Technology, Changshu 215500, China
2State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

Correspondence should be addressed to Zhi Han; hanzhi@sia.cn

Received 19 August 2019; Revised 10 January 2020; Accepted 30 January 2020; Published 24 February 2020

Academic Editor: Gaetano Giunta

Copyright © 2020 Jianjiang Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Diver target automatic detection is indispensable for underwater defense systems, such as the unmanned harbor surveillance system. It is a very challenging task due to various poses and intensity features of diver target. In addition, the background noise in sonar images is complex, which also makes the task more difficult. In this paper, we propose a diver detection method based on saliency detection for sonar images. On the basis of studying the characteristics of diver sonar images, we first decompose the original sonar image and perform median filtering on it, which can significantly improve the quality of the sonar image saliency map. We employ saliency detection technique based on frequency analysis to segment the acoustic highlight region from its surroundings. This segmentation region roughly locates the diver target and generates a region of interest (ROI). We then extract the acoustic shadow region in ROI, which contributes to furtherly improve the localization accuracy. Finally, we merge the segmented highlight region and the extracted acoustic shadow region and compute the minimum outer rectangle of the merged region. Experimental results validate that the proposed method can well detect and locate the diver target, and it can also satisfy the demands of real-time application, and there is almost no false alarm in this method.

1. Introduction

With the characteristics of small size, high maneuverability, and difficult to detect, diver attack is a common means used by terrorist organization to attack navy vessels base, offshore platform, civil harbor, and so on [1, 2]. Therefore, many countries and organizations have made great efforts for diver detection. Compared with the visible light camera, the imaging sonar has the advantages of reliable imaging in low visibility and turbid underwater environment and is widely used for underwater target detection, recognition, and tracking [3–6]. Dual-frequency identification sonar (DIDSON) can obtain near-video-quality dynamic images and narrow the gap between the existing imaging sonar and the optical systems [7].

Traditionally, the analysis of sonar images is carried out by skilled human operators, which is a boring and heavy task for humans. Thus, automatic detection for sonar targets is very necessary for underwater defense systems, such as the unmanned harbor surveillance system. As shown in Figure 1, because of complex underwater environment and multipath effect of underwater acoustic propagation, sonar images have the characteristics of low SNR, low contrast, and dynamic background. In addition, unlike rigid targets, the pose and movement direction of the diver target are variable, which results in various shapes and intensity features of the diver target in sonar image sequences. Therefore, it is a very challenging task to detect a diver in a sonar image with low SNR, especially how to achieve low false alarm rate and high detection accuracy. The sonar images of divers with various shapes and intensity features are shown in Figure 1.

In the past 20 years, various methods of moving object detection have been reported, such as difference-based [8, 9], model-based [10, 11], template-based [12, 13], and neural network-based [14, 15] methods. Difference-based methods, such as frame difference and background difference, are widely used for moving target detection, but they are greatly affected by the dynamic background and the velocity of target to be detected. Model-based methods are computationally intensive, such as Markov random field (MRF), which cannot meet the requirement of real-time applications. Template matching method has a good effect on rigid
Object detection, such as mine and UUV, but it is difficult to adapt to diver detection as diver is a deformable object and the pose and motion direction of a diver are various in sonar image sequences. Neural network-based methods have achieved excellent performance in object detection [16], such as fast RCNN, YOLO, SSD, and so on. However, because of the high cost of collecting sonar images of divers, the positive training samples are very limited. The dataset is too small to adequately train a deep network.

As shown in Figure 1, one sonar image with a diver object is usually composed of acoustic shadow region, acoustic highlight region, and background region. The intensity level of the acoustic shadow region is close to that of dark background, and the intensity level of the highlight region is close to that of bright background. For such reason, it is very difficult to segment the target by using traditional threshold methods. On the other hand, it is notable that the shape of the acoustic shadow region is similar to that of the target and always located above the highlight region. In addition, the highlight region is more salient than the background region in a sonar image. Based on the psychological studies, human perception system is more sensitive to the salient objects. The saliency detection technique has been widely used in image segmentation, object recognition, and image retrieval [17–19].

In this paper, inspired by the advantages of the saliency detection technology, we propose a diver target detection and localization method based on saliency. Specifically, as shown in Figure 2, after decomposing and filtering the input

Figure 1: Sonar images of divers with various shapes and intensity features.
sonar image first, we then conduct saliency detection on the image by amplitude spectrum filtering, which can help segment the acoustic highlight region from its surroundings, and roughly locate the diver target. According to the spatial relations between the highlight region and the shadow region, we generate a region of interest (ROI) which contains only the acoustic shadow region and part of background. It can reduce the interference of dark background efficiently. Then, we extract the acoustic shadow region by integrating median filter, contrast enhancement, edged detection, and morphological operator in ROI. All operations contribute to further improve the localization accuracy of the diver target. Finally, we merge the segmented acoustic highlight region and the extracted shadow region and compute the minimum outer rectangle of the merged region. Experimental results validate that the proposed method can well detect and locate the diver target with diverse postures and motion directions in sonar images.

This work aims at recognizing objects from sonar images for real application; thus, robustness and running time are the two most considered properties of our method. The biggest contribution of our work is that after studying the characteristics of the target in a sonar image, we proposed this pipeline for the first time to solve diver target detection and localization problem, and it showed outstanding robustness and efficiency, which is qualified for real-time and real-scene application. The technical contributions of our algorithm are three folds: (1) based on studying the characteristics of targets in the sonar image, the preprocessing method of the original sonar image is designed, which effectively improves the performance of the saliency map; (2) we adopt saliency detection techniques based on frequency analysis to make the gray scale of the highlight region far higher than that of other regions in the saliency map; (3) the proposed method integrates saliency detection, threshold segmentation, morphological operation, and intersection calculation, which can effectively locate the diver target and meet the demands of real-time application, and there is almost no false alarm.

### 2. Proposed Method

The proposed method is described in detail as follows.

#### 2.1. Saliency-Based Highlight Region Extraction and ROI Generation

In general, salient regions in an image are distinctive and nonrepeated regions, and nonsalient regions are homogeneous and repeated regions. It is a simple and efficient approach to detect salient objects in image frequency domain. In frequency domain, the Fourier spectrum of an image consists of an amplitude spectrum and phasespectrum. In the amplitude spectrum, distinct spikes at low-frequency band correspond to nonsalient regions, and its amplitude is much higher than that of high-frequency band. Therefore, a proper low-pass Gaussian filter is adopted to perform the
convolution with the image amplitude spectrum, which can smooth spikes, suppress nonsaliency regions, and highlight the salient regions. The saliency map is obtained by inverse Fourier transform using the smoothed amplitude spectrum and the original phase spectrum. Inspired by literature [20–22], the saliency map \( S(x, y) \) is defined as

\[
S(x, y) = g \ast \left[ F^{-1} [ A_s(u, v) \cdot (\cos(P_s(u, v)) + i \cdot \sin(P_s(u, v)))] \right]^2,
\]

where \( g \) is a Gaussian filter to smooth the saliency map for better visual effects, \( F^{-1} \) denotes inverse Fourier transform, and \( i \) is the imaginary function.

\( A_s(u, v) \) denotes the smoothed amplitude spectrum of an input image \( I_m(x, y) \) and is defined as

\[
A_s(u, v) = \exp \left( h \ast \log \left[ F(I_m(x, y)) \right] \right),
\]

where \( h \) is a low-pass Gaussian filter used to smooth spikes in the amplitude spectrum, the logarithm function is used to enhance the amplitude spectrum, \( F \) denotes Fourier transform, and \( \exp \) is an exponential function.

\( P_s(u, v) \) denotes the phase spectrum of an input image \( I_m(x, y) \) and is defined as

\[
P_s(u, v) = \angle \left[ F(I_m(x, y)) \right].
\]

For the diver sonar image, more than one saliency region appears in the saliency map. So, we must identify the one corresponding to the acoustic highlight region. The highlight region is located below the acoustic shadow region, and its gray scale is higher than that of other regions. We first segment the saliency map according to gray threshold and obtain the candidate region sets \( R_c \):

\[
R_c = \{ (x, y) \mid (x, y) \in S(x, y), \ \text{th}_{\min} \leq g(x, y) \leq \text{th}_{\max} \},
\]

where \( g(x, y) \) is the gray value of a point in the saliency map and \( \text{th}_{\min} \) and \( \text{th}_{\max} \) are the minimum and maximum threshold, respectively.

Then, we extract the region set \( R_{ca} \) from \( R_c \) according to the area parameter:

\[
R_{ca} = \{ R \mid R \in (R_c \cap R_s), \ \text{area}_R \geq T_A \},
\]

where \( \text{area}_R \) is the area of each region in the intersection of \( R_c \) and \( R_s \), \( T_A \) is a threshold, and \( R_s \) is the visible sector region in the sonar image after performing morphological erosion operation, as shown in Figure 3(b).

At last, we calculate the central coordinates of each region in \( R_{ca} \) and determine the region with the largest row coordinates in \( R_{ca} \) as the acoustic highlight region \( R_h \):

\[
R_h = \{ R \mid R \in R_{ca}, \ \text{row}_R = \max(\text{row}[R]) \},
\]

where \( \text{row}[R] \) is an array of central row coordinates of each region in \( R_{ca} \) and \( \text{row}_R \) is the central row coordinates of a region in \( R_{ca} \).

As shown in Figure 3(a), more than one saliency region appears in the saliency map, such as the highlight region, acoustic shadow region, and interference region. The gray scale of highlight regions is far higher than that of other regions in the saliency map. Thus, to segment the highlight region, we can use \( \text{th}_{\min} \) and \( \text{th}_{\max} \) to segment the saliency map and obtain the candidate regions \( R_c \). As shown in Figure 3(b), to improve the robustness of the proposed method, we set a wider threshold range. We redefine \( \text{th}_{\min} \) as 110 and \( \text{th}_{\max} \) as 255 to ensure that highlight regions in various situations can be all extracted. Although it may simultaneously cause the issue that more interference regions are also included, such as water depth signs at the boundary of the visible sector region in a sonar image, we can use further processing to remove them. To eliminate the interference region, we segment the visible sector region in the sonar image and perform morphological erosion operation to get a reduced sector region \( R_s \). Then, we calculate the intersection of \( R_c \) and \( R_s \), which makes the area of interference regions far smaller than that of the highlight region. Thus, we can use the area parameter to eliminate the interference regions, and we redefine \( T_A \) as 600, which can adapt to different input images. As shown in Figures 3(c)–3(e), after extracting the highlight region by using the row coordinate parameter, we can generate a rectangle region above it as ROI, which mainly contains the acoustic shadow region and bright background.
2.2. Image Preprocessing for Improving Saliency Map.

Because of the diver sonar image with low contrast and complex background, the saliency map calculated directly from the original image by equations (1)–(3) is not good enough. We decompose the original sonar image into three channels and perform median filter on the channel $I_d(x, y)$ with the highest contrast. We take the preprocessed image $I_m(x, y)$ as the input of equations (1)–(3):

$$I_m(x, y) = \text{median}(I_d(x, y)).$$

Figure 4 shows the comparison results using different preprocessing methods for improving saliency map, such as gray image transformation, image decomposing, and median filtering. The first row shows the results using different preprocessing methods. The second row shows the corresponding saliency map. Obviously, the saliency map of column (c) has not only high contrast but also less interference, which is the best.

2.3. Edge Detection of Acoustic Shadow Region.

It is difficult to identify whether it is a diver only by the shape of the highlight region. To further improve the detection accuracy of the diver target, we extract the edges of acoustic shadow regions in ROI. As shown in Figure 5(a), we first perform median filtering and contrast stretching on the inverse image of $I_m(x, y)$ to obtain the preprocessed image $I_{pre}(x, y)$:

$$I_{pre}(x, y) = a \cdot \text{median}([\text{inverse}(I_m(x, y))] + b.$$  

Then, we extract the candidate edges $R_{ec}$ of the acoustic shadow region in ROI by "canny" operator, as shown in Figure 5(b):

$$R_{ec} = \{\text{canny}(I_{pre}(x, y)) \cap \text{ROI}\}.$$  

At last, we eliminate the interference region by the area parameter and "opening" operator to obtain the edges $R_e$ of the acoustic shadow region, as shown in Figure 5(c):
\( R \in \mathbb{R} \), area \( R \geq \text{TB} \), (10)

\( R_c = (R_e \ominus B) \oplus B \), (11)

where \( \text{TB} \) is an area threshold, \( \ominus \) is the erosion operator, \( \oplus \) is the dilation operator, and \( B \) is a structure element.

The proposed method is summarized in Algorithm 1.

**Algorithm 1:** Saliency-based diver detection and localization method.

| Type | Reference | Description | Sample size |
|------|-----------|-------------|-------------|
| I    | Figures 1(a) and 1(b) | Diver with different posture in the image is relatively clear. | 54          |
| II   | Figures 1(c) and 1(d) | Diver with different motion direction in the image is not clear enough. | 33          |
| III  | Figure 1(e) | There is only highlight region, and the highlight region is blurred. | 23          |
| IV   | Figure 1(f) | There is only background, no target. | 28          |

### 3. Experimental Results and Discussion

In this section, to evaluate the performance and robustness of the proposed method, we conduct experiments of diver detection with various postures and motion directions in sonar images. The images used for the experiments are collected by imaging sonar in real underwater environment, which can be divided into four categories. The reference, sample size, and image description are shown in Table 1. The experiments are implemented on an Intel Core i5-4210U, 1.7 GHz laptop.

#### 3.1. Experiment of Comparing Detection and Localization Accuracy

To evaluate the detection and localization accuracy of the proposed method, we compare our approach with some classical moving object detection methods, such as background difference based on Kalman filtering (BDKF) and frame difference (FD). We have also compared our method with random forest (RF). Representative results are shown in Figure 6. To compare the localization accuracy, we have labeled the target in different input sonar images with bounding box. The ground truth is indicated by yellow bounding box, and the detected region is indicated by red bounding box. We have calculated intersection over union (IoU) of the ground truth and detected region by using the following equation:

\[
\text{IoU} = \frac{|A \cap B|}{|A \cup B|},
\]

![Figure 5: Edge extraction processes of the acoustic shadow region in ROI. (a) The preprocessed image. (b) The candidate edges. (c) The edges of the acoustic shadow region. (d) The merged region. (e) The minimum outer rectangle of the merged region.](image-url)
Figure 6: Comparison results of different methods.
where $A$ is the area of ground truth and $B$ is the area of the detected region. The boxplot of several methods is shown in Figure 7. The mean value of the IoU index of the proposed method is 0.73, which is much bigger than its counterparts. The highest localization accuracy is reached by the proposed method.

As shown in Figure 6, the FD method can detect most of the diver targets in sonar images of type I and II. However, its localization accuracy is very low, especially for diver sonar images of type II. On the other hand, the FD method detects some false targets, especially for type III and IV. The false alarm rate of this method is very high. All the results demonstrate that the FD method is too sensitive to the dynamic background.

The BDKF method can update background dynamically by Kalman filter, which can achieve a higher detection rate than the FD method. However, except the real target, some false objects are also detected, such as rocks and bubbles. There are both the real target and false objects in the detected target regions. The target location accuracy is not high enough, and the false alarm rate is high, especially for type IV. It is prone to be disturbed by different scene switching.

The RF method needs training before target classification. We select 165 images as the training set, including 40 images of type I, type II, and type III, respectively, and 45 images of type IV, and still use the 138 sample images as the testing set. The number of classification trees is set to 100. The RF method achieves higher classification accuracy. All sample images in type I and type IV are classified correctly. Some of the sample images in type II are classified incorrectly, which are classified as type III or type IV. Most of the sample images in type III are also classified correctly, except a few which are classified as type II. Representative images wrongly classified are shown in rows 3 and 5 of Figure 6. Besides, the RF method cannot locate the target in sonar images.

The proposed method can not only correctly detect the diver target with diverse postures and motion directions in sonar images but can also precisely locate the target. It is remarkable that the proposed method has almost no false alarms and is not interfered by dynamic background and different scene switching. The comparative data of different methods are shown in Table 2.

### 3.2. Experiments on Robustness and Efficiency Evaluation of the Proposed Method

To validate the robustness of the proposed method, we conduct a series of experiments on different types of data.

We first discuss the effect of different parameter settings on the final performance. As shown in Figure 8, when $\text{th}_{\text{min}}$ is set between 110 and 120, $\text{th}_{\text{max}}$ is set to 255, which can adapt to different input images. As shown in the first row and third row, when $\text{th}_{\text{min}}$ is set below 110, the segmented interference regions of the input image are too large to be completely eliminated, which results in the decrease of localization accuracy and, more importantly, the increase of false alarm rate. As shown in the second row, when $\text{th}_{\text{max}}$ is set above 120, some blurry targets in the sonar image cannot be detected. As shown in the fifth row, when $\text{th}_{\text{min}}$ is set to 110 and $T_A$ above 700, some targets cannot be detected. As shown in the last row, when $T_B$ is set above 110, the segmented edges of the acoustic shadow region are incomplete. Parameter settings are in Table 3.

For all of the following experiments, each parameter is a fixed value shown in Table 3. The corresponding experimental results are as follows.

For sonar images of type I, there are various postures of the diver during swimming and the acoustic highlight region and shadow region of the diver are relatively salient, but submerged in bright and dark background. With the proposed method, all divers in 54 sonar images of type I can be correctly detected and achieve very high detection accuracy. Representative results are given in Figure 9. The experimental results validate the robustness of the proposed method for diver detection with various postures.

For sonar images of type II, the motion directions of the diver are diverse during swimming and the acoustic highlight region of the diver is relatively salient in the whole sonar image, but the gray gradient of the acoustic shadow region is not big enough. With the proposed method, all divers in 33 sonar images of type II can be effectively detected and achieve higher detection accuracy. As shown in Figure 10, the gray scale of the highlight region is far higher than that of other regions in the saliency map, which makes it easy to segment. Because of the low edge gradient, the edges of the acoustic shadow region extracted even in ROI are not complete enough, but the location of the acoustic shadow region is relatively accurate. The experimental results validate the robustness of the proposed method for diver detection with various motion directions.

For sonar images of type III, there is only the highlight region with various poses and motion directions. Specifically, the highlight region is blurring in the whole sonar image. With the proposed method, 19 divers in 23 sonar

![Figure 7: Comparison results of localization accuracy of different methods.](image-url)

| Method    | FD | BDKF | RF | Proposed |
|-----------|----|------|----|----------|
| Detection rate (%) | 79.1 | 88.1 | 83.6 | 96.4 |
| False alarm rate (%) | 39.2 | 21.4 | 0 | 0 |

Table 2: Detection rate and false alarm rate using different methods.
images of type III can be exactly detected and 4 divers cannot be detected due to their blurring highlight region. As shown in the first 5 columns of Figure 11, the highlight region in the saliency map is very obvious. As shown in the last two columns of Figure 11, the target is blurry because it is almost

**Figure 8**: Results of the effects of different parameter settings on the final performance.

**Table 3**: Parameters setting.

| Parameter | Value |
|-----------|-------|
| \( \theta_{\text{min}} \) | 110   |
| \( \theta_{\text{max}} \) | 255   |
| \( T_A \) | 600   |
| \( T_B \) | 100   |
Figure 9: Experimental results of diver detection for type I. (a) The original images. (b) The saliency map, in which, the highlight region is detected automatically and marked in red. (c) The edges of the acoustic shadow region to be extracted in ROI, marked in red colour. (d) The final results of diver detection with diverse postures.

Figure 10: Continued.
Figure 10: Experimental results of diver detection for type II. (a) The original images. (b) The saliency map, in which, the highlight region is detected automatically and marked in red colour. (c) The edges of the acoustic shadow region to be extracted in ROI, marked in red colour. (d) The final results of diver detection with various motion directions.

Figure 11: Continued.
beyond the scope of sonar scanning. The highlight region cannot be segmented with the same parameter setting because the gray scale of the highlight region in the saliency map is lower than that of the adjacent region.

For sonar images of type IV, there is no diver target in the sonar image, which can verify whether the proposed method is disturbed by background. Using the proposed method, there is no target detected in 28 sonar images of type IV. As shown in Figure 12, there is no salient target in the saliency map. Therefore, there are almost no false alarms. The experimental results validate that the proposed method is robust for complex background.

The detection rate and mean detection time are as shown in Table 4.

As shown in Table 4, we perform the proposed method on 138 diver sonar images, the detection rate achieves 96.4%,
false alarm rate is close to 0, and the average detection time is 0.021 second per frame. Experimental results show that the proposed method achieves good performance and can meet the requirements of real-time applications. This is mainly due to the following advantages of the proposed method:

1. We first perform decomposition and median filtering on the original sonar image before conducting saliency detection based on frequency analysis, which can significantly improve the quality of the saliency map.

2. We adopt saliency detection techniques based on frequency analysis, which makes the gray scale of saliency regions far higher than that of other regions in the saliency map. We integrate threshold segmentation, morphological operation, and intersection calculation, which makes the area of interference regions far smaller than that of the highlight region. Thus, we can easily segment the highlight region from the whole sonar image, and the proposed method is robust for diver targets with diverse poses and motion directions in sonar images.

3. We generate a ROI above the highlight region and extract the edges of the acoustic shadow region in ROI. It can reduce the interference from background and improve the localization accuracy of the diver target.

### 4. Conclusions

In this paper, we propose a diver detection method based on the saliency of sonar images. On the basis of studying the characteristics of diver sonar images, we employ saliency detection techniques based on frequency analysis to segment the acoustic highlight region from its surroundings, which can adapt to diverse postures and motion directions of the diver target in sonar images. We find that decomposition and median filtering on the original sonar image before conducting saliency detection can significantly improve the effect of the saliency map. Extracting the acoustic shadow region in generated ROI can reduce the interference of background and improve the localization accuracy. The experimental results show that the method can effectively detect the diver target with diverse postures and motion directions in sonar images and can meet real-time application requirement.

### Data Availability

The image-type data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 61773367, 61821005, and 61303168) and the Youth Innovation Promotion Association CAS (no. 2016183).

### References

[1] D. Stiller, S. Benen, and P. Berkel, “An experiment for detection of underwater intruders with different kinds of sensors,” in *Future Security 2012*, pp. 327–333, Springer, Berlin, Germany, 2012.

[2] L. Fillinger, A. J. Hunter, M. Zampolli, and M. C. Claris, “Passive acoustic detection of closed-circuit underwater breathing apparatus in an operational port environment,” *The Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. EL310–EL316, 2012.

[3] N. Hurtós, N. Palomeras, A. Carrera, and M. Carreras, “Autonomous detection, following and mapping of an underwater chain using sonar,” *Ocean Engineering*, vol. 130, pp. 336–350, 2017.

[4] X. Liu and Z. Xiong, “Underwater small target tracking algorithm based on diver detection sonar image sequences,” in *Proceedings of the 2012 International Conference on Industrial Control and Electronics Engineering*, pp. 727–730, Xi’an, China, August 2012.

[5] L. Zheng and K. Tian, “Detection of small objects in side scan sonar images based on POHMT and Tsallis entropy,” *Signal Processing*, vol. 142, pp. 168–177, 2018.

[6] R. Fandos, A. M. Zoubir, and K. Siantidis, “Unified design of a feature based ADAC system for mine hunting using synthetic aperture sonar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2413–2426, 2014.

[7] E. O. Belcher, W. Hanot, and J. Burch, “Dual-frequency identification sonar (DIDSON),” in *Proceedings of the 2002 International Symposium on Underwater Technology*, pp. 187–192, Tokyo, Japan, April 2002.

[8] Monnet, Mittal, Paragios, and V. Ramesh, “Background modeling and subtraction of dynamic scenes,” in *Proceedings of the 9th IEEE International Conference on Computer Vision*, pp. 1305–1312, Washington, DC, USA, October 2003.

[9] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, “Image change detection algorithms: a systematic survey,” *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, 2005.

[10] S. Wei and H. Leung, “A Markov random field approach for sidescan sonar change detection,” *IEEE Journal of Oceanic Engineering*, vol. 37, no. 4, pp. 659–669, 2012.

[11] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, “Sonar image segmentation using an unsupervised hierarchical MRF.
model,” *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1216–1231, 2000.

[12] A. Sibiryakov, ”Fast and high-performance template matching method,” in *Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1417–1424, Providence, RI, USA, June 2011.

[13] T. Dekel, S. Oron, M. Rubinstein, S. Avidan, and W. T. Freeman, ”Best-buddies similarity for robust template matching,” in *Proceedings of the 2015 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2021–2029, Boston, MA, USA, October 2015.

[14] D. P. Williams, ”Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks,” in *Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 2497–2502, Cancun, Mexico, December 2016.

[15] P. L. D. Roberts, J. S. Jaffe, and M. M. Trivedi, ”Multiview, broadband acoustic classification of marine fish: a machine learning framework and comparative analysis,” *IEEE Journal of Oceanic Engineering*, vol. 36, no. 1, pp. 90–104, 2011.

[16] Z. Zou, Z. Shi, Y. Guo, and J. Ye, ”Object detection in 20 years: a survey,” 2019, https://arxiv.org/abs/1905.05055.

[17] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, ”Unsupervised extraction of visual attention objects in color images,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141–145, 2006.

[18] U. Rutishauser, D. Walther, C. Koch, and P. Perona, ”Is bottom-up attention useful for object recognition?” in *Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 37–44, Washington, DC, USA, June-July 2004.

[19] M. M. Cheng, D. Walther, C. Koch, and P. Perona, ”Global contrast based salient region detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[20] X. Hou and L. Zhang, ”Saliency detection: a spectral residual approach,” in *Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, Minneapolis, MN, USA, June 2007.

[21] C. Guo, Q. Ma, and L. Zhang, ”Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform,” in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, Anchorage, AK, USA, June 2008.

[22] J. Li, M. D. Levine, X. An, X. Xu, and H. He, ”Visual saliency based on scale-space analysis in the frequency domain,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 996–1010, 2016.