


RESEARCH

Open Access



Experimental study of underwater operation scene with target perception framework

Jue Gao^{1*} , Wei Ding¹ and Haiping Yang¹

*Correspondence:
gao_jue@hotmail.com

¹ School of Electrical Engineering and Automation, Changshu Institute of Technology, Changshu 215500, Jiangsu, People's Republic of China

Abstract

This paper presents a target perception framework aimed at enhancing diver safety and facilitating underwater operations by extracting critical information from underwater scenes. The framework employs a layered processing approach, which encompasses water column imaging, constant false alarm rate detection, and local feature analysis. To simulate the diver's underwater environment, we conducted experiments with three distinct fields of view: fixed down-looking, fixed front-looking, and mobile side-looking perspectives. Our experimental findings demonstrate the framework's ability to accurately differentiate between false targets, stationary targets, and moving targets within the underwater scenes, as well as to capture the motion trajectories of dynamic targets. Furthermore, the application of 3D reconstruction techniques to underwater scene data enables the generation of approximate stereoscopic representations of divers and bubble groups.

Keywords: Target perception, Underwater operation, Acoustic image, Sonar

1 Introduction

Equipment developed using acoustic technology has become the 'eyes', 'ears', and 'mouths' for humans to carry out underwater activities [1–3] such as resource surveying, environmental monitoring, meteorological observation, and ocean mapping. In recent years, there has been a rapid development of unmanned ships or underwater robots as carriers, equipped with acoustic equipment, and the use of acoustic technology to engage in various underwater operations [4, 5]. Nikolovska et al. [6] used multi-beam water body imaging technology to image bubble groups in the Black Sea, to accurately locate the actual gas leakage area. Weber et al. [7] examined the near-surface layer during a storm in order to obtain observational verification for theories on bubble clustering. Masett et al. [8] reported that divers investigated coastal waters that might pose a danger to seafarers and drew a map of sunken ships.

When considering the application of acoustic technology in underwater operations, it becomes imperative to furnish support for the following tasks:

- Scanning of extensive and intricate underwater environments: swiftly surveying vast and intricate underwater landscapes.

- Identification of Regions of Interest (ROI): discerning and isolating areas of significance within these underwater environments.
- Verification of Targets: distinguishing between genuine and false targets.
- Assessment of Target Motion: determining whether the target is stationary or in motion. In particular, the dynamic target is usually the diver, whose movement trajectory needs to be tracked to ensure safety [9].

The challenges encountered in underwater acoustic imaging within operational areas encompass low resolution, diminished SNR, non-uniform sound propagation, and pronounced sidelobe interference [3, 10]. When interpreting targets during underwater operations, the ROI typically comprises a collection of weak targets that span the entire operational region. These targets are characterized by their small reflection cross section and low echo intensity, rendering them challenging to discern from acoustic images [11, 12]. Furthermore, most of the current acoustic technology research is carried out according to a certain application branch [13–15], such as the detection of a specific underwater target, the tracking of fish or divers, and the classification of marine topography and geomorphology. Acoustic surveillance of underwater operating scenarios means integrating rapid automatic detection, continuous tracking of multiple targets, target recognition and target classification into one system, and it is necessary to develop a common framework that integrates multiple technologies.

This paper conducts experimental research on acoustic surveillance of underwater operating scenes, focusing on identifying dynamic targets associated with divers to assist the operation and ensure operational safety. The diver's operation scene in the actual acoustic monitoring area is simulated, and the experiments are under three fields including the fixed view of the down-looking, fixed view of the front-looking, and mobile view of the side-looking. It is dedicated to exploring information including the diver, bubbles, and static targets quickly, accurately, and stably in a certain underwater operation area. A target perception framework for underwater operation is proposed, which regards the acoustic imaging system as a sensory organ, and comprehensively integrates the detection, recognition, classification, and tracking technology from previous studies [16–18].

The remainder of this paper is organized as follows. Section 2 introduces the underwater target perception framework and its key technologies. Section 2 describes the simulation experiment of the underwater operation scene and gives the test results and analysis. Finally, the conclusion is given in Sect. 4.

2 Methods

By integrating a variety of acoustic techniques, a target perception framework for the underwater operating scene is designed, as shown in Fig. 1. The framework uses three layers of processing mechanism to explore specific target information: the first layer carries out acoustic imaging for underwater scenes, and the system maps underwater scenes within a certain field of view into acoustic images; the Second layer detects the potential target region in the acoustic image and removes the significant non-target region in the entire acoustic image to obtain the ROI; the Third layer further discriminates the ROI and identifies real targets from potential targets mixed with artificial clutter and natural clutter false alarms. Three layers of processing including acoustic

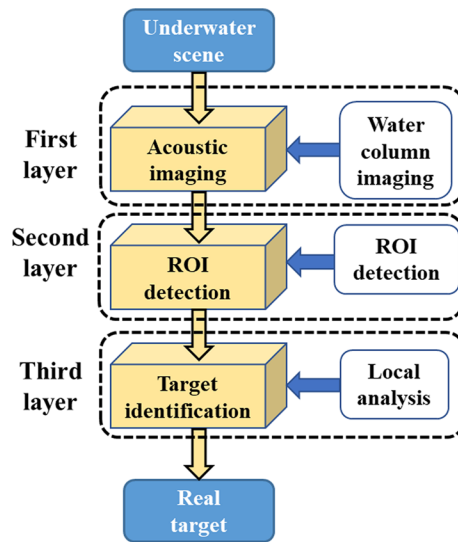


Fig. 1 Underwater target perception framework

Table 1 Parameter description

Parameter	Description
P_{fa}	False alarm probability
T	Threshold
C	Shape parameter
B	Scale parameter
n_r	Reference cells length
n_g	Guard cells length
A	Accumulated matrix
$X(m,n)$	m row and n column pixels of the acoustic image
$R_{(1,2,3)}$	Remaining subsets
I_i	The i th frame in the acoustic image sequence
D_i	Feature set of the i th frame
M	Feature set template
F_j	The j th feature in the feature set
K	Continuous unmatched frame threshold

imaging, ROI detection, and target identification can be regarded as progressive relationships, and at the same time, they are an organic cooperative whole. The key technologies for this framework are derived from previous research, and Table 1 lists the description of the main parameters.

2.1 Water column imaging technology

Water Column Imaging (WCI) records the scattered signals emitted by acoustic pulses to the water column and was first used in marine fisheries to study fish habits. With the rapid development of multi-beam sonar systems and corresponding imaging technology, WCI has been widely used in many fields [19, 20], including seabed gas

leakage, suspended sediment, archaeological oceanography, seaweed ecosystem, and aquaculture. The backscattered signal can be described by the sonar equation:

$$EL = SL - 2TL + TS \quad (1)$$

where EL is the echo signal level, SL is the acoustic source level, TL is the propagation loss that describes the attenuation of echo strength, and TS is the scattering strength of the target.

The receiving system undergoes a sequence of operations, including time-varying gain adjustment, A/D conversion, and beamforming, to transform the received backscattering signals into directional beam signals. These beam signals are subsequently mapped to the appropriate acoustic image resolution unit, with their placement determined by factors such as acoustic wave propagation distance and the opening angle between the beams. This paper improves the quality of acoustic imaging through the utilization of virtual beam interpolation and dynamic brightness allocation. Additionally, we employ an algorithm based on background estimation to suppress sidelobe interference, especially in high SNR case.

2.2 Constant false alarm rate detection technology

The Constant False Alarm Rate (CFAR) algorithm is widely used in adaptive target detection [21]. Under the condition of a given false alarm probability P_{fa} , the threshold T is automatically set according to the statistical distribution of the local background, and the pixel to be detected is compared with the adaptive threshold to determine whether it belongs to the target or the clutter. By traversing the whole image with the reference window, all pixels in the image can be detected automatically.

To achieve accurate ROI detection, the previously proposed Subset Censored-Constant False Alarm Rate (SC-CFAR) detection method is adopted. This strategy involves partitioning the reference window into multiple subsets, where the subset with the highest echo intensity has the greatest potential for containing interference targets. Consequently, this specific subset is excluded, and the remaining subsets are employed for background parameter estimation, thus mitigating the risk of spurious local detection threshold increases. The overview of the 2D Subset Censored CFAR method is exhibited in Fig. 2, which mainly consists of five steps:

-
- (1) calculate the shape parameter C , select the length of reference cells nr and the guard cells ng , and set the false alarm probability P_{fa} ;
 - (2) compute the accumulated matrix \mathbf{A} ;
 - (3) perform an initial search of potential targets with the threshold T_p ;
 - (4) estimate the scale parameter \mathbf{B} using the fast algorithm based on the integral image;
 - (5) compare the test cell with the local threshold T_{mnr} , then label the $x(m,n)$ as an object or clutter.
-

Taking an example of SC-CFAR with a sliding window of four subsets, the expression of the local detection threshold is:

$$T = \left(\left(P_{fa}^{-\left(\frac{4}{3nr}\right)} - 1 \right) (R_1 + R_2 + R_3) \right)^{\frac{1}{B}} \quad (2)$$

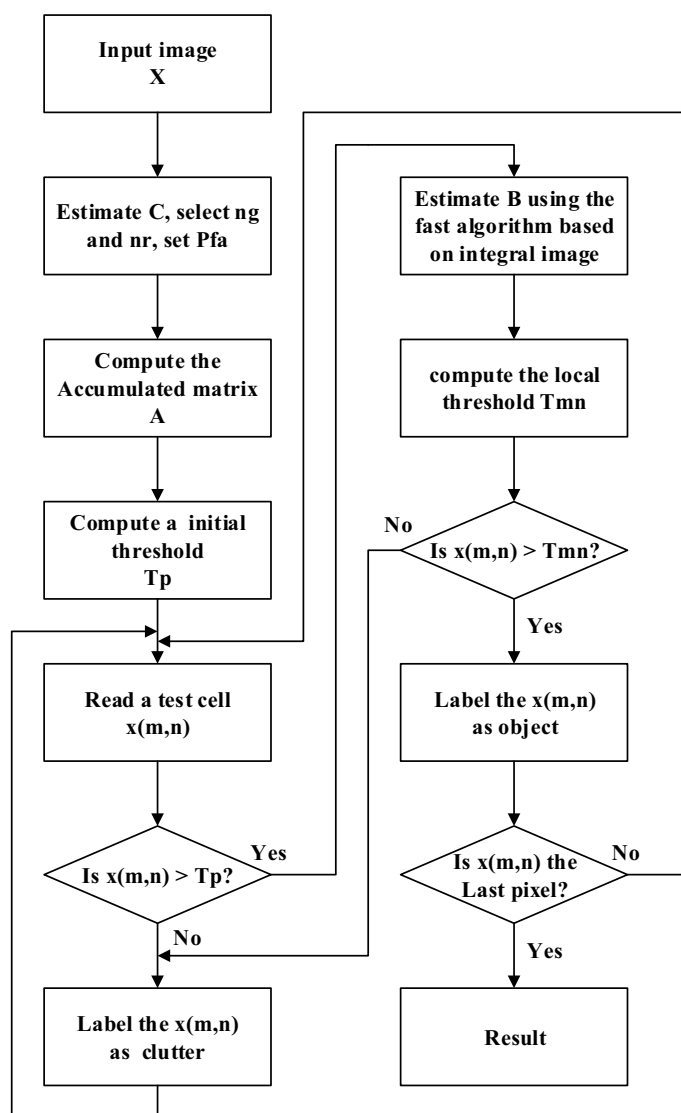


Fig. 2 Overview of the 2D subset censored CFAR

where R_1 , R_2 , and R_3 are the remaining subsets used for background parameter estimation after the largest subset is censored.

2.3 Local analysis techniques

A local feature is an image pattern that differs from the nearest neighbor, implying a change in image properties [22]. The local analysis regards each detected ROI as one or multiple potential targets. Initially, it seeks out local invariant features suitable for representing these potential targets, and subsequently identifies the potential targets through the tracking of these local invariant features.

Local feature extraction includes two parts: keypoint detection and feature description. Based on previous studies, this paper adopts Hessian keypoints to represent potential targets in acoustic images, and the matching of SURF descriptors measures the

relevance of potential targets between frames. For the acoustic image $I(x,y)$, the Hessian matrix of any point (x,y) is defined as:

$$H(x, y, \sigma) = \begin{bmatrix} \frac{\partial^2}{\partial x^2} G(\sigma) * I(x, y) & \frac{\partial}{\partial x} \frac{\partial}{\partial y} G(\sigma) * I(x, y) \\ \frac{\partial}{\partial x} \frac{\partial}{\partial y} G(\sigma) * I(x, y) & \frac{\partial^2}{\partial y^2} G(\sigma) * I(x, y) \end{bmatrix} \quad (3)$$

where σ is the scale space factor and $G(\sigma)$ is the Gaussian kernel function.

To generate SURF feature vectors, the square area is divided into 4×4 sub-areas along the main direction with key points as the center, and the sub-block vector is obtained by calculating the sub-window response values using the wavelet template:

$$V = \left[\sum dx, \sum |dx|, \sum dy, \sum |dy| \right] \quad (4)$$

Since there is a 4×4 sub-block, each sub-block has 4 vectors, forming 64-dimensional SURF feature vectors.

Following the extraction of local ROI features, the strategy of tracking before detection (TBD) is adopted to simultaneously track multiple local features. The potential target can be classified as a static target, dynamic target, or false target based on criteria such as mean offset, start and end offset, and the continuity of its characteristic trajectory. The overview of feature tracking is shown in Fig. 3, which comprises five main stages:

-
- (1) Input the first frame I_1 , obtain the feature set D_1 , and save it as a template M ;
 - (2) Read the subsequent frame I_i and acquire the feature set D_i , match the extracted feature F_j from M ;
 - (3) Make the matching feature F_j as potential targets, and update the corresponding feature in F_j from M ;
 - (4) Remove the mismatching feature F_j of consecutive k frames from M (Considering that acoustic imaging is susceptible to environmental interference resulting in insufficient stability, k is rounded to 10% of the total number of frames);
 - (5) After traversing the entire image sequence, determine whether the remaining feature F_j represents the real target, then obtain the feature trajectory.
-

3 Experiment results

In the experiment, the underwater operation scene of the diver is simulated, and the sonar system is installed on a boat or a robot for target monitoring and positioning to assist the diver and ensure safety. The experiments are designed under three fields of view: fixed view of the down-looking, fixed view of the front-looking, and mobile view of the side-looking, and the target information in underwater operation scenes is explored by the proposed target perception framework. The relevant system parameters are as follows: The operating frequency is 200 kHz, the pulse width of the transmitting CW pulse is 0.1 ms, the sampling frequency is 88 kHz, the ping rate is 0.25 frame per second, and the receiving array has 100 elements, the number of beams is 512, and the beam coverage is $160^\circ \times 1^\circ$ in the horizontal direction.

3.1 Fixed view of the down-looking

In the fixed downward view, the simulated underwater operational scenario involves the sonar system, placed on the underwater vessel or Autonomous Underwater Vehicle (AUV), monitoring the activities of a diver engaged in various tasks. The layout of

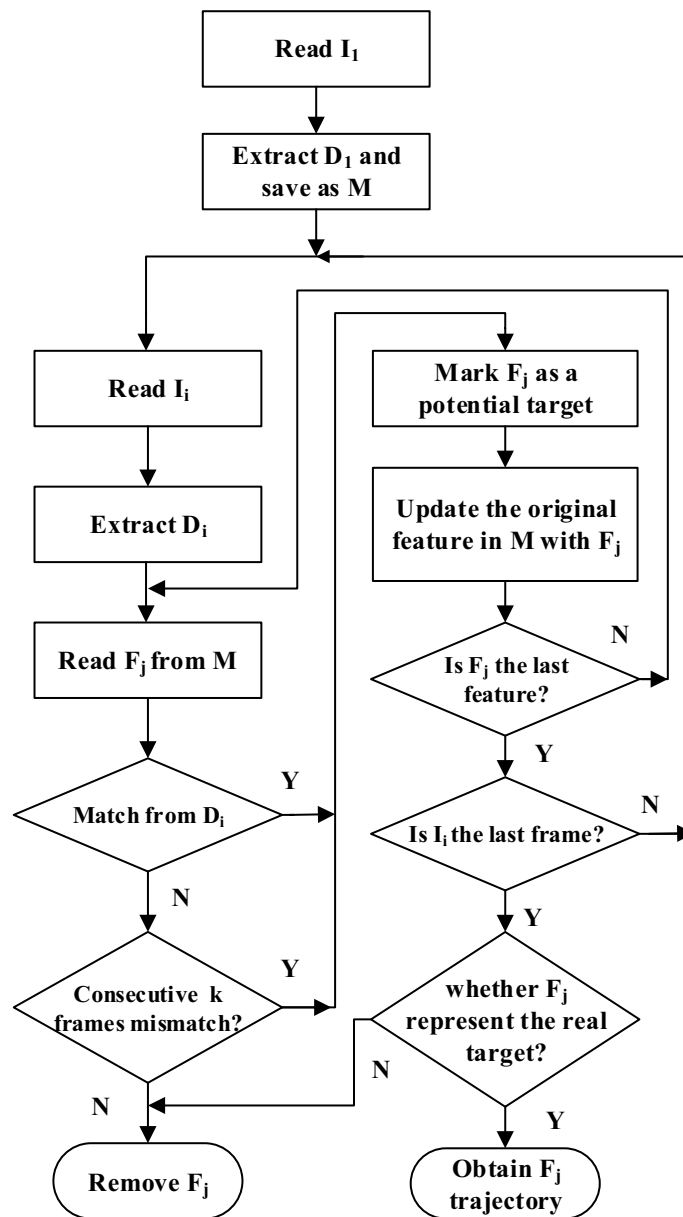


Fig. 3 Overview of the feature tracking

the sonar system is shown in Fig. 4, the receiving and transmitting arrays are placed along the x -axis and the y -axis to form a T-shaped plane perpendicular to the z -axis, and the beam sector is perpendicular to the water surface. Taking the center position of the sonar system as the origin of the motion coordinates, the diver's motion along the z -axis is specified as the vertical direction, the x -axis as the horizontal direction, and the y -axis as the trajectory direction. The diver moves in the horizontal direction, first from $x = -6$ m to $x = 6$ m (part-I), and then back to the starting point (part II).

In the first layer, water column imaging processing of the first layer is used to generate a sequence of 49 acoustic images. The acoustic image has a size of 1341×881 pixels, reflecting an underwater scene of $13.3 \times 8.9\text{m}^2$ with a resolution grid of $0.01 \times 0.01\text{m}^2$.

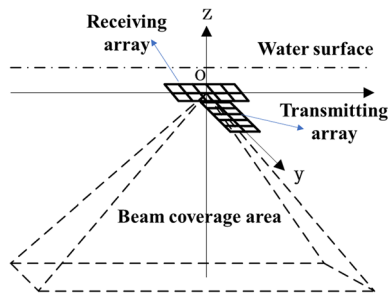


Fig. 4 Layout of the sonar system

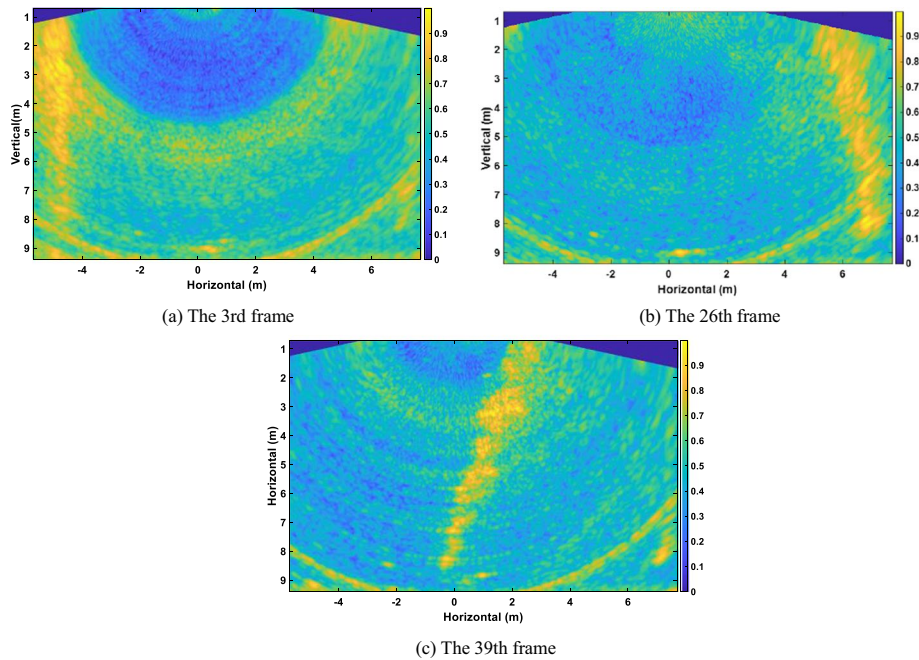


Fig. 5 Water column imaging

The 3rd frame, the 26th frame, and the 39th frame of the acoustic image sequence are shown in Fig. 5. In Fig. 5a, the diver is at $x = -4.7$ m and begins to move to the right. In Fig. 5b, the diver is at $x = 6.9$ m, near the far-right end of the field of view. In Fig. 5c, the diver is at $x = -0.4$ m, returning to the left. The diver is equipped with open breathing apparatus, which produced many bubbles. The bottom-up bubble swarm follows the frogman's moving slowly. Although a series of image processing is performed on the acoustic image, the noise and side lobes in the image still make the image resolution low and the target blurry.

In the second layer, the SC-CFAR algorithm is used to detect ROIs from the acoustic image sequence, setting $P_{fa} = 0.1$, $Nr = 5$, and $Ng = 12$. The results of the previous three frames are shown in Fig. 6a–c respectively, which displays that numerous ROIs are extracted from the underwater scene. Relative to the preceding layer, this layer's processing involves the interpretation of more explicit information within the underwater scene, specifically pertaining to the identification of potential targets through

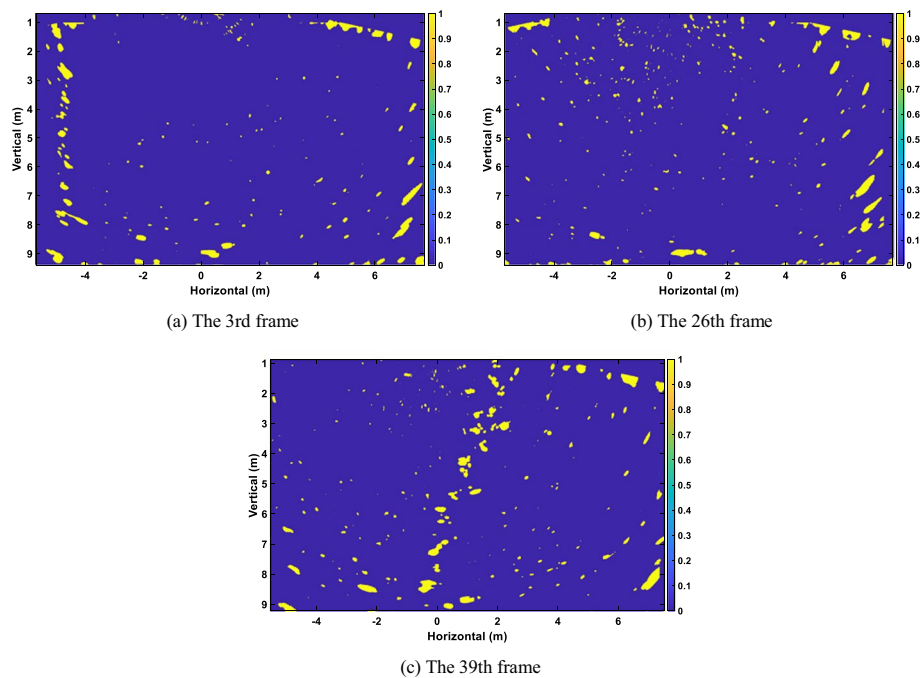


Fig. 6 ROI detection results

Table 2 Statistical information on feature tracking

Motion	Tracking frames	Tracking rate (%)	Moving distance (m)	Average offset (m)	Type
Part-I	17	65.38	0.46	0.17	Static
	16	61.54	11.74	0.92	Dynamic
	13	50.00	11.62	0.90	Dynamic
Part-II	15	71.43	9.65	0.96	Dynamic
	12	57.14	10.77	1.11	Dynamic

ROI extraction. Given the continuous fluctuations in environmental reverberation and noise, it remains challenging to entirely detect the area occupied by the underwater diver and the bubble clusters. Especially in Fig. 6b, most of the real targets are missed, due to insufficient SNR. It seems that the traditional method cannot complete the detection task well in some low SNR frames.

In the third layer, the continuous identification of the actual target within the sequence is achieved through feature tracking implemented in accordance with the TBD strategy. According to Eq. (3) and Eq. (4), local features composed of Hessian keypoints and SURF descriptors are extracted in the ROI. The statistics of feature tracking are shown in Table 2. In part-I, three features are successfully tracked in 17 frames, 16 frames, and 13 frames, respectively, which can be judged according to the moving distance and the average offset, two of which are dynamic targets and the other is a static target. In part II, two features are consistently tracked over 15 frames

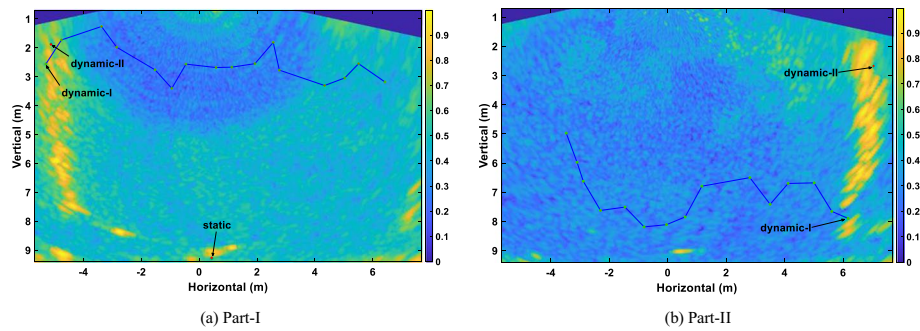


Fig. 7 Feature tracking results and trajectories

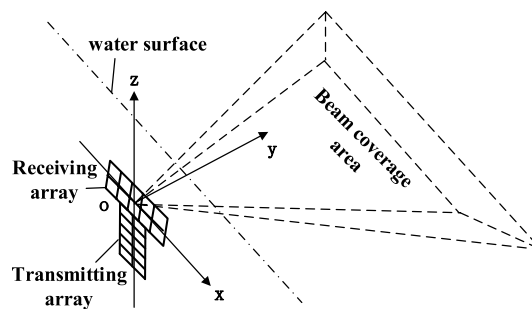


Fig. 8 Deployment of the sonar system

and 13 frames, respectively. Due to the relatively substantial displacement of these two features, it enables the determination of the associated dynamic target.

As illustrated in Fig. 7, the coordinates of these matched features are delineated on the respective acoustic image. It is evident that the dynamic target corresponds to the bubbles generated by the diver; while, the static target denotes stationary objects near the pool's bottom. By connecting the dynamic target features, an approximate trajectory of the diver can be derived.

3.2 Fixed view of the front-looking

In the fixed frontal view, the simulated underwater scenario involves a sonar system securely positioned within the operational area to monitor the long-distance movements of the diver for safety. The layout of the sonar system is shown in Fig. 8. The receiving array and the transmitting array are placed along the x -axis and z -axis, forming a T-shape perpendicular to the y -axis, and the beam sector is parallel to the water surface. The diver moves along the y -axis direction and approaches the origin of coordinates from a distance by straight, diagonal, and curve motions, respectively.

The sonar system collected the underwater scene of the diver in three motion modes. The acoustic image has a size of 401×776 pixels, reflecting an $8 \times 15.5\text{m}^2$ scene parallel to the water surface, with a resolution grid of $0.02 \times 0.02\text{m}^2$. Through processing by the first layer, Fig. 9 displays typical acoustic images of the diver captured at various positions. The strips at $x = -3\text{ m}$ and $x = 3\text{ m}$ are the left and right walls of the pool, with the diver in the far center ($x = 0\text{ m}$, $y = 13.3\text{ m}$), the left side of the pool wall ($x = -2.1\text{ m}$, $y = 13\text{ m}$), the center of the pool ($x = -0.4\text{ m}$, $y = 8.8\text{ m}$) and near the origin ($x = -0.3\text{ m}$,

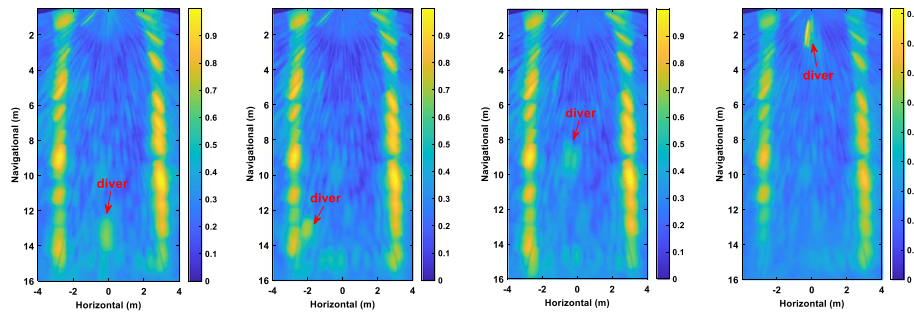


Fig. 9 Water column imaging

$y = 1.9$ m). Obviously, the SNR, shape, and size of the highlighted area change dramatically during the diver’s movement.

The second layer of processing is applied to the acoustic image sequence with parameters set at $P_{fa} = 0.1$, $Nr = 6$ and $Ng = 10$. The detection outcomes are depicted in Fig. 10, demonstrating accurate identification of the diver with minimal instances of false alarms.

The third layer of processing involves the tracking of Hessian + SURF features within the ROI region, and the corresponding statistical data are presented in Table 3. By analyzing the start and end offsets as well as the average offsets, it can be discerned that one of the tracked features, following a linear path, corresponds to the dynamic target, whereas the other three correspond to static targets. Similarly, the feature tracked with diagonal motion aligns with the dynamic target, and two features traced through curved motion correspond to the dynamic and static targets, respectively.

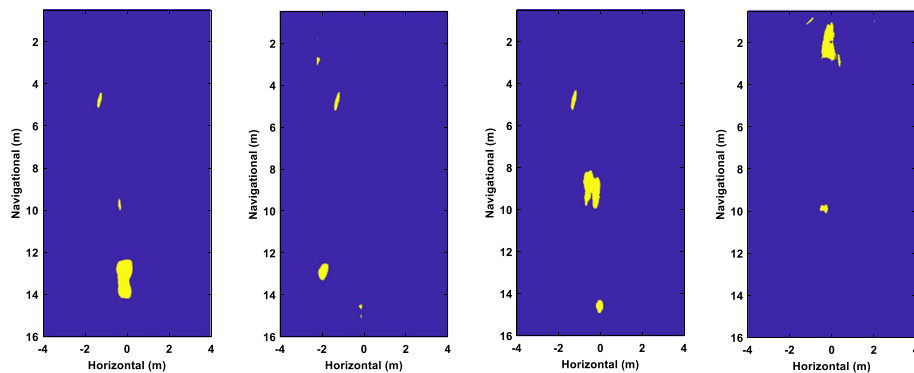


Fig. 10 ROI detection results

Table 3 Statistical information on feature tracking

Motion mode	Tracking frames	Tracking rate (%)	Moving distance (m)	Average offset (m)	Type
Straight	22	73.23	0.33	11.54	Diver
	0	0.00	0.05	0.05	False
	0	0.00	0.02	0.14	False
	0	0.00	0.05	0.08	False
Diagonal	13	61.90	0.34	7.59	Diver
Curve	20	71.43	0.41	10.70	Diver
	0	0.00	0.06	0.07	False

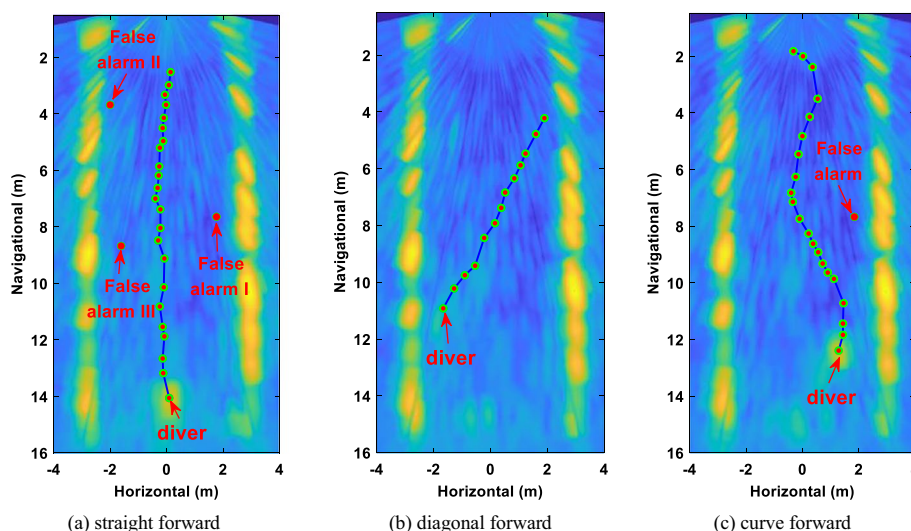


Fig. 11 Feature tracking results and trajectories

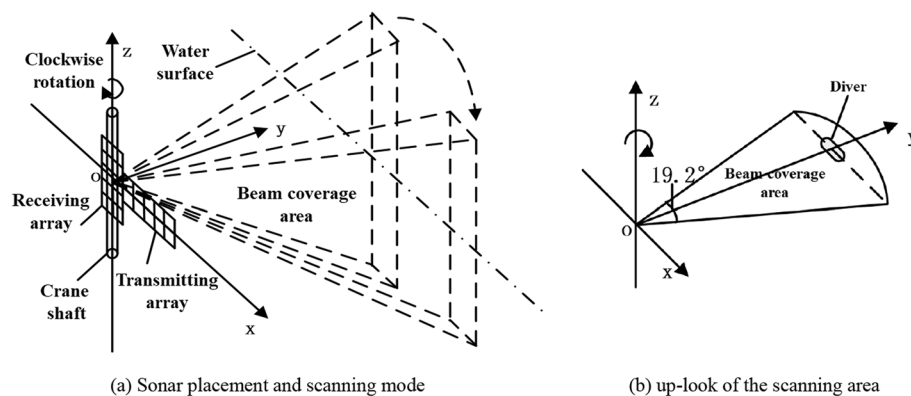


Fig. 12 Deployment of the sonar system

As shown in Fig. 11, the initial coordinates of the tracked features are annotated on the acoustic images, revealing that the dynamic target corresponds to a diver; while, the static targets lack distinct counterparts and are considered as false alarms. By linking the feature coordinates across each frame, the motion trajectory is determined to follow three distinct patterns: straight motion, diagonal motion, and an S-shaped curve.

3.3 Mobile view of the side-looking

In the mobile side-looking perspective, the simulated scene encompasses the deployment of sonar systems within harbor areas, underwater vessels, or Autonomous Underwater Vehicles (AUVs) for the purpose of monitoring underwater operations, which can involve both fixed and mobile surveillance modes. The layout of the sonar system is shown in Fig. 12a. The sonar system relates to the rotating axis of the traveling crane, the beam sector is vertically forward with the water surface, and the side scan is achieved by rotating the traveling crane clockwise at a speed of 0.2°/s. The uplook of the beam

covering plane during the scanning process is shown in Fig. 12b. The coverage range of the beam sector is $-9.6^\circ \sim 9.6^\circ$, and the angle interval between the two frames is 0.8° .

The sonar system rotates 19.2° clockwise to scan the underwater operation scene, and the acoustic image size is 601×851 pixels, reflecting a $12 \times 17\text{m}^2$ underwater scene, with a resolution grid of $0.02 \times 0.02\text{m}^2$. Figure 13 shows the acoustic images when the beam sector is rotated to 0° , -3.2° , and 3.3° . The horizontal highlight area located in the vertical direction $z = -5\text{ m}$ and $z = 5\text{ m}$ is the water surface and the bottom of the pool respectively, while the vertical highlight area located in the navigational direction $y = 16\text{ m}$ is the front pool wall. As can be seen from Fig. 13a, when the fan surface coincides with the y -axis, the highlighted area in the center of the red box ($z = 1.6\text{ m}$, $y = 8.6\text{ m}$) is the section of the diver, while the highlighted area in the red box directly above it is the section of the bubble group generated by the diver. Combined with Fig. 13b, c, when the fan surface is on both sides of the y -axis, the highlighted area of the section of the diver and the bubble group is significantly weakened, and the size changed, and the position of the center area also fluctuated slightly.

The acoustic image sequence's slice data are fused along the three dotted lines in the vertical direction, as illustrated in Fig. 13, resulting in the creation of the mosaic map presented in Fig. 14. Figure 14a shows the section splicing in the vertical direction $z = 1.6\text{ m}$. The section splicing of the diver is mainly distributed in the red box and the scattered bright spots in the lower right corner of the box are caused by the fluctuation of the position of the diver. Figure 14b shows the section splicing in the vertical direction $z = 0.8\text{ m}$. Most of the splicing of the bubble group is in the red box, which is close to the

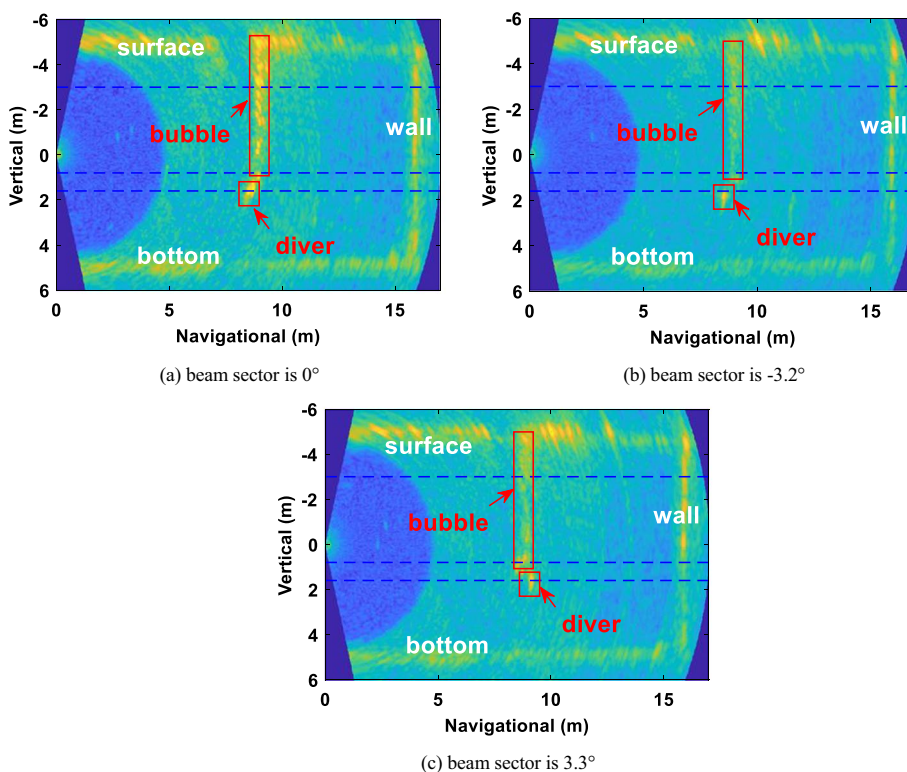


Fig. 13 Imaging with different rotation angle of beam sectors

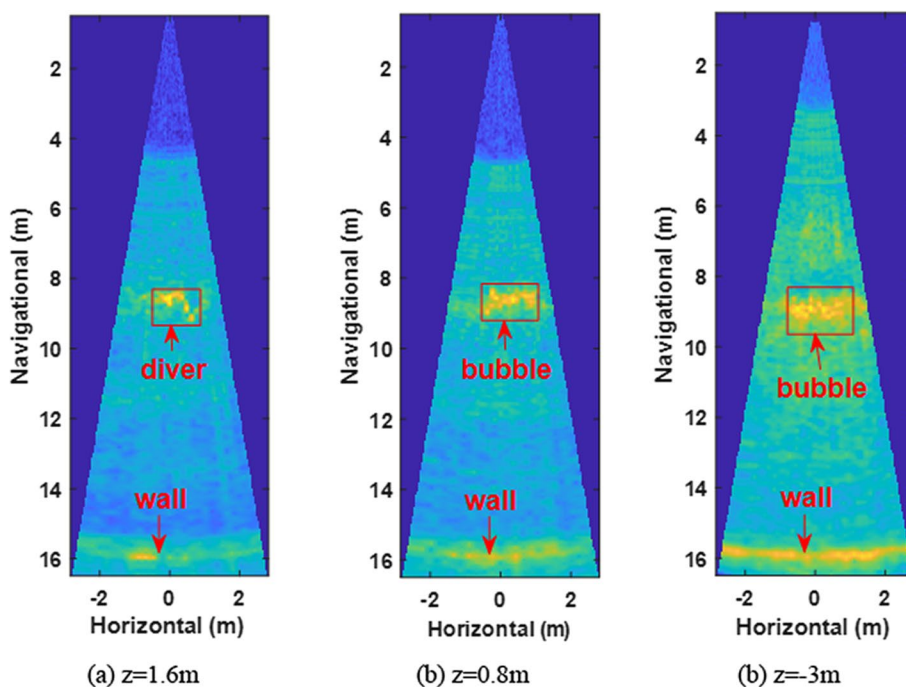


Fig. 14 Slice mosaic along different vertical directions

diver and at the bottom of the bubble group. Figure 14c shows the section splicing in the vertical direction $z = -3$ m. This splicing is far away from the diver and is the upper part of the bubble group. The bubble group gradually spreads and the section splicing area gradually expands.

3D reconstruction is carried out for the collected underwater operation scene, and multiple slices are spliced to form 3D volume data (x, y, z, v) . The x -axis represents the horizontal direction, the y -axis represents the navigational direction, the z -axis represents the vertical direction, and v represents the echo intensity of the coordinate point. A box filter with a size of $5 \times 5 \times 5$ is used to conduct 3D smoothing of the data, extract the isosurface and connect it. The effect of 3D reconstruction is shown in Fig. 15. The approximate stereoscopic profiles of the diver and the bubble group are plotted, and their distribution is consistent with the front view shown in Fig. 13 and the side view shown in Fig. 14. In summary, intensity and positional data of the beam are acquired through side view rotational scanning, and the beam's approximate contour is determined via three-dimensional reconstruction.

4 Conclusion

In order to facilitate underwater operations, we have developed an underwater target perception framework. This framework employs a layered processing approach, encompassing acoustic imaging, Region of Interest (ROI) detection, and target identification, which collectively extract information from the underwater operational environment to support tasks and ensure safety. To simulate diverse underwater working scenarios that could be encountered during underwater operations, three sets of experiments were designed, each with a different field of view: a fixed down-looking

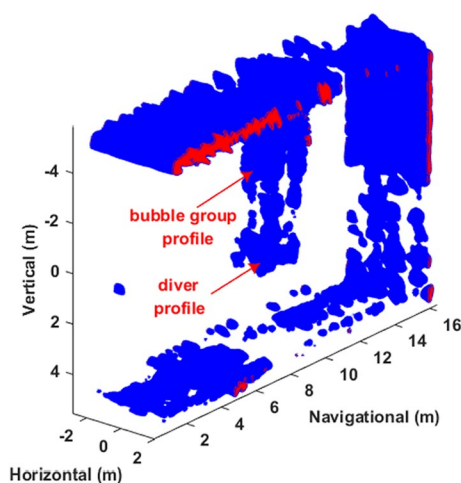


Fig. 15 3D reconstruction of the underwater scene

view, a fixed front-looking view, and a mobile side-looking view. The results of these experiments are as follows:

1. The proposed framework can autonomously capture ROIs within underwater operational scenes, and the differentiation between static and dynamic targets is based on feature offset analysis from the acoustic image sequence. Notably, it does not rely on individual frame-level target presence assessment but makes decisions based on the continuity and consistency of feature trajectories, thereby enhancing target recognition performance in underwater operational scenarios.
2. In the fixed down-looking perspective, most acoustic images exhibit a blend of the diver and bubbles, making their distinction challenging. Given the intrinsic correlation between these elements, the diver's trajectory can be indirectly inferred by tracking bubble characteristics.
3. Under the fixed front-looking view, factors such as noise, scene reverberation, beam widening, and others introduce dramatic changes in the SNR, shape, and size of the highlighted area as the diver moves closer along the navigation path. Connecting the tracked dynamic feature coordinates and overlaying them on the acoustic image reveals a motion trajectory consistent with the actual diver's movement mode.
4. In the mobile side-looking view, acoustic image sequences from different positions are fused to obtain cross sectional slices of the diver and bubble group. Through 3D reconstruction of the underwater scene data, an approximate stereoscopic contour of the diver and bubble group can be generated.

It is important to note that the experiments conducted in this study were carried out in a controlled pool environment, where water conditions were relatively calm. In real-world underwater operational scenarios, wave action can be more pronounced, resulting in lower SNR for acoustic imaging and potential deviations in the positions of static targets. In the future, the detection and tracking algorithms within the framework need further refinement based on experimental outcomes. Consideration

should also be given to the integration of cutting-edge technologies such as deep learning to enhance the framework's performance in target recognition."

Acknowledgements

The author would like to thank the reviewers, whose constructive comments have contributed to improving the quality of the paper.

Author contributions

JG conducted the study, designed experiments, data processing, and drafted the manuscript. WD participated in the experiment and the analysis of the results. HY participated in its design and coordination. All authors read and approved the final manuscript.

Funding

This work was supported in part by the China Postdoctoral Science Foundation under Grant No. 2023M732551.

Availability of data and materials

Please contact author for data requests.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no competing interests to declare.

Received: 27 March 2023 Accepted: 21 November 2023

Published online: 30 November 2023

References

1. K.G. Foote, Underwater acoustic technology: review of some recent developments, in *Proceedings of the MTS/IEEE Oceans* (Quebec, 2008), pp. 1–6
2. S.W. Cui, Y. Wang, S. Wang, Real-time perception and positioning for creature picking of an underwater vehicle. *IEEE Trans. Veh. Technol.* **69**, 3783–3792 (2020)
3. A. Trucco, M. Garofalo, S. Repetto, Processing and analysis of underwater acoustic images generated by mechanically scanned sonar systems. *IEEE Trans. Instrum. Meas.* **58**, 2061–2071 (2009)
4. H. Yang, K. Lee, Y. Choo, Underwater acoustic research trends with machine learning: passive SONAR applications. *J. Ocean Eng. Technol.* **34**, 227–236 (2020)
5. K. Sun, W. Cui, C. Chen, Review of underwater sensing technologies and applications. *Sensors* **21**, 1–28 (2021)
6. A. Nikolovska, Hydroacoustic methodology for detection, localization, and quantification of gas bubbles rising from the seafloor at gas seeps from the eastern Black Sea. *Geochem. Geophys. Geosyst.* **6**, 66 (2008)
7. T.C. Weber, Observations of clustering inside oceanic bubble clouds and the effect on short-range acoustic propagation. *J. Acoust. Soc. Am.* **124**(5), 2783 (2008)
8. G. Masetti, B. Calder, Remote identification of a shipwreck site from MBES, backscatter. *J. Environ. Manag.* **111**, 44–52 (2012)
9. G. Delyon, Clutter map detector for active diver detection sonar. *IET Radar Sonar Navig.* **14**, 177–186 (2020)
10. R. Lefort, R. Fablet, L. Berger, J. Boucher, Spatial statistics of objects in 3-d sonar images: application to fisheries acoustics. *IEEE Geosci. Remote Sens. Lett.* **9**, 56–59 (2012)
11. M. Kumar, S. Mondal, Recent developments on target tracking problems: a review. *Ocean Eng.* **236**, 66 (2021)
12. G. Neves, M. Ruiz, J. Fontinelel, Rotated object detection with forward-looking sonar in underwater applications. *Expert Syst. Appl.* **140**, 66 (2020)
13. X. Wang, Q. Li, Y. Yu, Evaluation criterion of underwater object clustering segmentation with pulse-coupled neural network. *IET Image Process.* **14**, 4076–4085 (2020)
14. A. Abu, R. Diamant, Unsupervised local spatial mixture segmentation of underwater objects in sonar images. *IEEE J. Ocean. Eng.* **44**, 1179–1197 (2019)
15. G. Mishne, R. Talmon, I. Cohen, Graph-based supervised automatic target detection. *IEEE Trans. Geosci. Remote Sens.* **53**, 2738–2754 (2015)
16. J. Gao, P.Y. Zhu, Underwater target perception in local HOS space. *Comput. Intell. Neurosci.* **2021**, 1–12 (2021)
17. J. Gao, H. Li, B. Chen, Fast two-dimensional subset censored CFAR method for multiple objects detection from acoustic image. *IET Radar Sonar Navig.* **11**, 505–512 (2017)
18. J. Gao, Y. Gu, P.Y. Zhu, Feature tracking for target identification in acoustic image sequences. *Complexity* **2021**, 1–11 (2021)
19. K. Colbo, T. Ross, C. Brown, A review of oceanographic applications of water column data from multibeam echosounders. *Estuar. Coast. Shelf Sci.* **145**, 41–56 (2014)
20. J. Clarke, Applications of multibeam water column imaging for hydrographic survey. *Hydrogr. J.* **120**(120), 66 (2006)

21. M.E. Smith, P.K. Varshney, Intelligent CFAR processor based on data variability. *IEEE Trans. Aerosp. Electron. Syst.* **36**, 837–847 (2000)
22. S. Gauglitz, T. Höllerer, M. Turk, Evaluation of interest point detectors and feature descriptors for visual tracking. *Int. J. Comput. Vis.* **94**, 335–360 (2011)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Jue Gao is an assistant professor of Electrical Engineering and Automation at the Changshu Institute of Technology in Jiangsu, PR China. I received the Ph.D. degree in Information and Communication Engineering from the Harbin Engineering University, Harbin, CHINA, in 2018. Research interests involve Underwater Object Perception Techniques based on Acoustic Image.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
