IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 June 2023, accepted 29 June 2023, date of publication 12 July 2023, date of current version 11 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3294710

RESEARCH ARTICLE

Activity Segmentation and Fish Tracking From Sonar Videos by Combining Artifacts Filtering and a Kalman Approach

JULIAN WINKLER, SABAH BADRI-HOEHER, AND FATNA BARKOUCH
Faculty of Computer Science and Electrical Engineering, Kiel University of Applied Sciences, 24149 Kiel, Germany

Corresponding author: Julian Winkler (julian.winkler@fh-kiel.de)

This work was supported by the German Federal Ministry of Food and Agriculture (BMEL) Based on a Decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the Innovation Support Program under Grant 2819111618.

ABSTRACT Ecosystems are highly dynamic systems that are constantly changing under the influence of a variety of external factors. This is especially true for marine ecosystems, which are under multiple stresses. The cumulative effects of overexploitation, on the one hand, and the simultaneous manifestation of anthropogenic climate change, on the other, mean that fish stocks are the most endangered components of marine ecosystems. To minimize these vulnerabilities to marine ecosystems and ensure natural and sustainable resource use, monitoring systems must be placed in oceans and seas. Examples of the development of these monitoring systems are provided by the Underwater Fish Observatory (UFO) and UFOTriNet, two projects being conducted by several researchers from marine biology, engineering, and industry in Germany between 2014 and 2016 and between 2019 and 2023, respectively. The systems collect abiotic as well as camera and sonar data to count and analyze fish populations over the seasons. This work proposes a method for robust fish counting using sonar data, supplemented by camera data. To successfully accomplish this task, activity segmentation and object tracking are important steps. Background subtraction is often used as a pre-processing step for stationary sonars. Our proposed method improves this step by bandpass filtering considering the motion of all actors in the sonar data. For the segmentation step, our method uses a simple Gaussian distribution model with positional covariances which are computed directly from the intensity image. The tracking step is performed using a classical Kalman filter which estimates the velocity and position of each object in Cartesian coordinates. Sonar detections in close range of the observation area are compared with camera detections for validation. In addition, automated parameter optimization is used to maximize the correlation with the camera detections. Furthermore, the proposed method is applied to the Caltech fish counting dataset and compared with a deep learning method based on YOLOv5. While YOLO is still superior in detection and counting metrics, the multi object tracking accuracy is somewhat higher with our method.

INDEX TERMS Acoustical imaging, activity segmentation, computer vision, fish tracking, imaging sonar, underwater image processing.

I. INTRODUCTION

For effective implementation of sustainability in fisheries, it is essential that adequate, accurate, and continuously reliable data on the diversity of fish species and their abundance, on the functioning of the ecosystems concerned, and on fishing activities are available over time. However, this is not currently the case, as the provision of high quality data requires the availability of appropriate methods, tools, and procedures for adequate appropriate monitoring, control, and advisory services. The Underwater Fish Observatory (UFO) [1] and the UFOTriNet [2] are projects that aim to

The associate editor coordinating the review of this manuscript and approving it for publication was Xuebo Zhang.

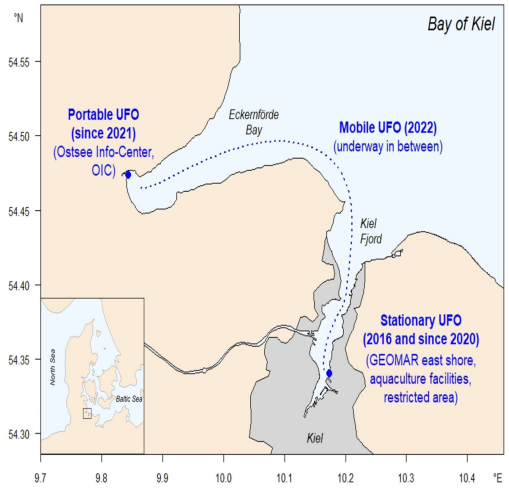


FIGURE 1. Locations of the deployment of the UFOTriNet system.

automate many tasks that can be very time consuming and costly, such as stock assessment and biomass estimation. These tasks have traditionally been performed by marine biologists using nets and vessels, as well as human underwater observation and photography [3]. These methods are considered invasive and in most cases provide results with significant inaccuracies. In addition, the amount of data collected is limited in both time and space and is not nearly sufficient to describe the observed environment [4]. An automated system for data collection offers a better alternative, and the data obtained can be extremely useful to successfully solve tasks such as fish detection, fish segmentation, fish tracking, and also fish classification with very low environmental impact.

UFO is based on a stationary lander equipped with numerous sensors to collect various information, deployed underwater and continuously capturing sonar and optical video throughout the year with very low environmental impact. UFOTriNet is based on the establishment of a trilateral network for automatic, continuous and non-invasive monitoring of fish stocks. The goal is to detect, classify and track fish in three modes: stationary UFO, portable UFO and mobile UFO (see Figure 1).

The collected data will be used to develop and train algorithms for fish detection [1] and fish species classification [5], [6], [7]. The lander has a stereo system with two cameras that capture optical images and a sonar that generates acoustic images. An acoustic image is a gray-scale image that does not clearly reveal the shape of the fish or its structure, but only a high intensity that may even be associated with an object in the water column. In many cases, random noise generated by the sensor or the environment results in a similar sudden increase in intensity, but the behavior of the noise pixels in successive images is different from the actual fish pixels, making it somewhat easier for the viewer to distinguish them. Despite the shortcomings of these images, they are still very important when it comes to detecting and tracking movements. They provide the ability to detect any movement over a very large area, especially if it is not just a single fish, but a large school of fish moving within the sonar image area. One of the main goals of the UFOTriNet project is to develop a hybrid method that combines both optical and acoustic imagery to create a more robust system for fish detection, classification, and tracking that will provide marine biologists with a highly reliable automated assistant in their research. Our main focus is on acoustic imaging to facilitate the labeling and indexing of activities within the recordings. Our work is divided into the following points:

- Filtering of background and unwanted moving particles
- Segmentation of moving objects using a Gaussian model
- Fish tracking using Kalman Filter

This paper is organized as follows: Section II provides an overview of scientific works related to fish detection, segmentation and tracking based on sonar data. Section III briefly describes the lander in the UFO and its sensor content. The proposed method for segmentation and tracking is described in Section IV. Different results and measurement evaluations are included in Section V. Section VI summarizes the whole work and gives an outlook on the future focus and perspectives.

II. RELATED WORK

Activity segmentation and fish tracking based on sonar images are essential tasks in various research issues in the underwater field. In the realm of unsupervised learning, several traditional methods, such as fuzzy-based segmentation [8], mixture-model MRF [9], and active contour (AC) techniques [10], have been extensively employed for the segmentation task. However, in the context of supervised learning, challenges have arisen in segmenting fish in sonar data, which has led researchers to prioritize fish detection over the segmentation task.

Despite the complexity, unsupervised learning for segmentation, where fishes are unlabeled, remains feasible. For instance, in [11], the local spatial mixture (LSM) segmentation method is designed to estimate pixel labels in sonar images by considering the potential spatial correlation between neighboring pixels. By incorporating an intermediate step (I-step) between the expectation (E-step) and maximization (M-step) steps in the expectation-maximization algorithm, the LSM method enhances the algorithm's ability to leverage spatial information for more accurate segmentation. Additionally, to address the challenge of intensity inhomogeneity in underwater environments, the method employs a new initialization algorithm, which automatically sets appropriate thresholds to ensure robustness and consistent performance across various underwater conditions.

Moreover, researchers have explored automated fish segmentation in videos captured by DIDSON (Dual-Frequency IDentification SONar) in [12]. This approach involves two main parts: a fixed process that includes data extraction,

pre-processing for geometric reconstruction, frame smoothing, merging into a continuous video stream, and background removal to enhance the quality and clarity of the raw DIDSON images. The second part utilizes an iterative process with optical flow analysis to track fish motion within the video, followed by custom criteria for evaluating and determining the most suitable foreground mask. To optimize the parameter set for calculating the optical flow field and assisting in fish target mask extraction, a genetic algorithm is utilized.

In recent years, Convolutional Neural Networks (CNNs) have emerged as a popular approach for fish detection in sonar analysis. These networks effectively capture spatial patterns and correlations within sonar data, enabling accurate fish detection. For instance, in [13], a deep learning model using CNN was developed to automatically detect adult American eels from sonar data, achieving high accuracy (>98%) for image-based classification, surpassing human experts. Similarly, in [14], a DCNN method was presented to enhance weakly illuminated underwater pictures. This approach efficiently addresses the issue and adapts the DCNN architecture for underwater detection and classification.

In addition to CNNs, researchers have explored the utilization of pre-trained models such as Faster R-CNN (Region Convolutional Neural Network) [15] and YOLO (You Only Look Once) [16], [17] for fish detection in sonar images. These pre-trained models leverage transfer learning by leveraging knowledge gained from large-scale visual datasets and then fine-tuning them specifically on sonar data.

While Kalman Filters are typically used in navigation problems [18], they are also well suited for multiple object tracking in videos [19]. Different variants of Kalman Filters have also previously been applied to vehicle tracking using RADAR and SONAR systems. In [20] an Extended Kalman Filter is used to track a target ship and in [21] an Unscented Kalman Filter is used for submarine tracking.

Regarding fish tracking, in [14] the Kalman Filter (KF) method is proposed for object tracking in combination with a CNN model. Similarly, in [16], the SORT (Simple Online and Realtime Tracking) algorithm is combined with YOLOv5 for real-time and online tracking. Another approach, presented in [2], utilizes a local optical flow application in which corners and key points are extracted from the estimated fish contours. Instead of the traditional local forward optical flow application, a forward and backward tracking of the corners is performed, followed by an evaluation phase for the contours' movements, trying to assign the correct track to the extracted contours.

In summary, the literature review highlights the significance of activity segmentation and fish tracking in underwater environments. It discusses various traditional and deep learning methods employed for segmentation and fish detection, as well as tracking algorithms utilized for monitoring fish motion. While existing approaches have made significant contributions, the focus has often been on detection rather than accurate segmentation. Therefore, there is a clear need to introduce a new method that overcomes these challenges and improves overall fish segmentation and tracking performance. In response to this gap, our proposed method combines traditional techniques and competes with advanced methods.

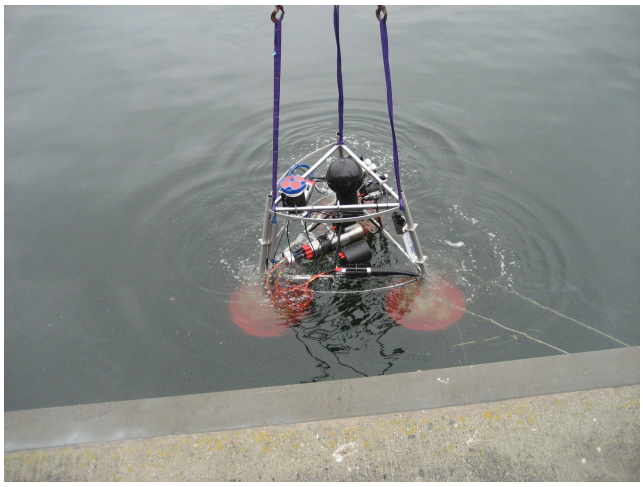


FIGURE 2. Stationary UFO lander.

III. SYSTEM OVERVIEW AND DATA DESCRIPTION

The collection of data underwater traditionally requires intensive and very sensitive use of vessels, divers or fishing nets. However, the amount of data collected and its quality cannot be considered accurate because these methods are invasive and do not capture the life form of fish underwater. In addition, the temporal and spatial coverage of these methods is very limited and the type of data that can be obtained from them is also limited. To address almost all of these shortcomings, the use of a stationary lander equipped with multiple sensors that can cover the desired location 24/7 is required. In addition to this, mobile systems can be used to optimize coverage in space. This solution guarantees the preservation of the underwater environment and collects all the necessary information, from optical and sonar images to other data such as salinity, pressure, oxygen content and temperature. This is exactly the goal of the UFOTrinet project. In this work, we will limit ourselves to the use of the stationary UFO data since suitable data with the mobile UFO is not yet available at the current time.

A. GENERAL SYSTEM DESCRIPTION

In the stationary UFO, the lander in Figure 2 contains several sensors for different tasks. The two most important for fish counting are the stereo camera system, which is used to detect, track, and classify fish species at close range, and an imaging multibeam sonar, which plays a central role in this work and is the only sensor that provides information at long range because the visual range of the optical camera is limited. This system was also used in [1]. The infrastructure

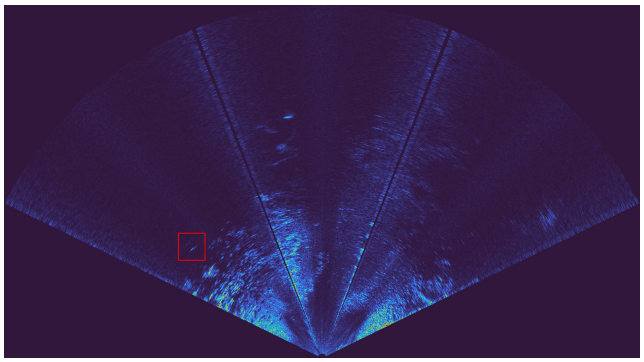


FIGURE 3. Unprocessed sonar images with swarm of fish in highlighted region.

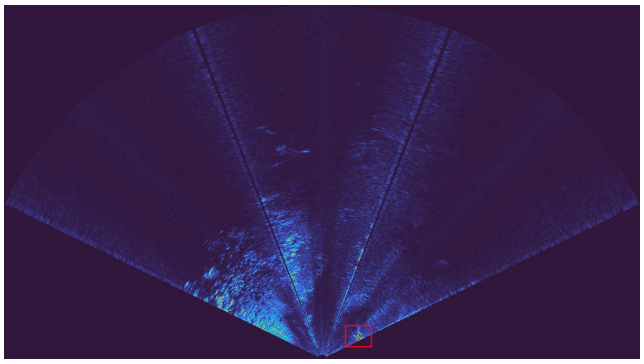


FIGURE 4. Unprocessed sonar images with disturbances in highlighted region.

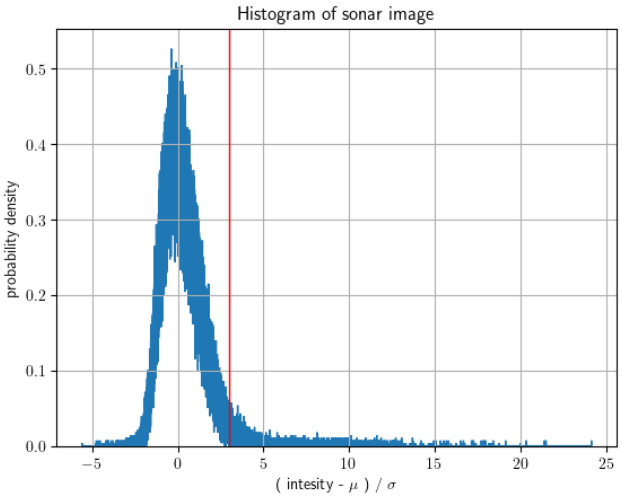


FIGURE 5. Histogram of normalized intensities on sonar image without fish activities.

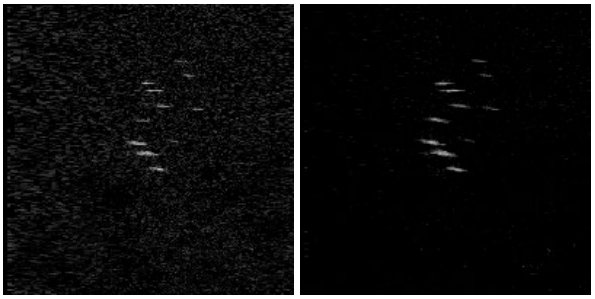


FIGURE 6. Pre-processing using background subtraction (left) and band-pass filter (right).

for power supply and data transmission is realized by cable connections to a computer cabin of a nearby fish farm.

B. DATA DESCRIPTION AND CHALLENGES

The used sonar has a horizontal field of view of 130°, a vertical beam width of 20° and is configured for a maximum range of 50 m [1]. The raw unprocessed sonar images look as shown in Figure 3. It can be seen that the fishes in the highlighted region look the same as most static objects in the image. Therefore, it is not feasible to detect them in single images. Even as humans, it is often only possible to tell them as fish by observing their motion in a time series of sonar images. Also, the used sonar had some technical defects, which caused very high intensity disturbances as shown in Figure 4. These disturbances often only affect few or single frames.

The activity detection method previously used in the UFO project was intended to detect fish activities by normalizing the intensity values of all pixels by previously calculated per pixel mean and variance over time. Then everything exceeding the 3σ border would be treated as activity. While this method works fine on some recordings, it has some problems when the water current is stronger and there are some particles, sediment or plants in the water. A histogram of the normalized intensities on a single sonar image without fish activities can be seen in Figure 5. In this sonar recording, around 2% of pixels are always above the 3σ border even in frames without fish activities. This leads to a large offset error in the activity count.

IV. FISH TRACKING METHOD

Object tracking can be split into three consecutive tasks, preprocessing, segmentation and actual tracking.

A. PRE-PROCESSING

The pre-processing step is responsible for removing the background of the sonar image caused by the seabed and static objects on it. This is typically done by calculating the average intensity of each pixel over a temporal cutout of the sonar recording and subtracting these average values from each frame [16]. As an alternative, we propose applying a finite impulse response (FIR) band-pass filter on the intensity signal of each pixel. In addition to the background, this also removes the high frequency noise resulting from the sonar beamforming. Our sonar videos have been recorded with five pings per second. The used FIR filter has 127 taps and is designed using least square optimisation. The pass band is set from 0.03 Hz to 0.3 Hz. As shown in Figure 6 the background

subtracted image includes more noise than the band-passed version.

B. SEGMENTATION

The segmentation process is typically implemented by grouping pixels with similar color or gray values together [22]. While this works well on camera images, it has some problems with sonar recordings, because the intensity of a single object is often not constant on all pixels of the object. Especially small objects like fish often have an intensity maximum in the center and fade out towards the edge. Instead, a new algorithm is proposed, which first sorts the pixels starting from the highest intensity and then labels them one by one. Each time a new pixel is labeled, the positional covariance of all pixels belonging to the labeled object is calculated. The covariance of the tracked objects is used to calculate the likelihood of the next pixels of belonging to each object. The object with the highest probability is selected to label the new pixel. If none of the tracked objects is likely enough to belong to the pixel, a new tracked object is created with the new pixel. The complete segmentation algorithm can be seen in Algorithm 1. Here, tracks is a list of objects with the fields $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and img where img is the collection of pixels already assigned to this object. pixels is a list of objects with the fields x, y and intensity and pdf() is the probability density function of a multivariate Gaussian distribution described by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. When the intensity of a pixel at position $\begin{bmatrix} x & y \end{bmatrix}^\top$ is given by the function $I(x, y)$, the positional center of mass $\boldsymbol{\mu}$ can be calculated as shown in (1) and the related positional covariance matrix $\boldsymbol{\Sigma}$ as in (2).

$$\boldsymbol{\mu} = \frac{\sum_x \sum_y \begin{bmatrix} x \\ y \end{bmatrix} \cdot I(x, y)}{\sum_x \sum_y I(x, y)} \tag{1}$$

$$\boldsymbol{\Sigma} = \frac{\sum_x \sum_y \left(\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{\mu} \right) \cdot \left(\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{\mu} \right)^\top \cdot I(x, y)}{\sum_x \sum_y I(x, y)} \tag{2}$$

Together, the properties $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ describe the position, size, proportion and orientation of each object being tracked. Larger objects with more complicated shapes could probably not be described with just these two metrics, but as shown in Figure 7, it accurately describes the shape of the fishes visible in the sonar image. The ellipses in Figure 7 visualize the 3σ border of the probability distribution of each pixel being part of the tracked objects.

C. TRACKING

The proposed segmentation algorithm is relatively slow when starting with an empty list of tracked objects because the positional covariances of the objects have to be recalculated each time a pixel is assigned to an object. A python implementation takes 1.2 seconds to segment 26 objects on a single sonar image. The performance can be significantly improved by using the results from the previous frame as a starting point. This is where the actual tracking comes into play. To reliably predict the position of objects on a video, the velocity of each object has to be estimated first. The most common methods for velocity estimation in videos include optical flow [2] and Kalman filters [23]. While optical flow seems to work well on camera videos, it has some problems with sonar recordings, because recognisable details of the tracked objects are often hardly visible. Kalman filter based tracking methods estimate the velocity by the positions given by the segmentation of the previous frames. In this work, a four state Kalman filter instance is created for each tracked object. The states estimation vector $\hat{\mathbf{x}}$ consists of the positions x and y and the associated velocities as shown in (3). The state transition $\mathbf{F}$ and observation $\mathbf{H}$ are defined in (4) and (5)

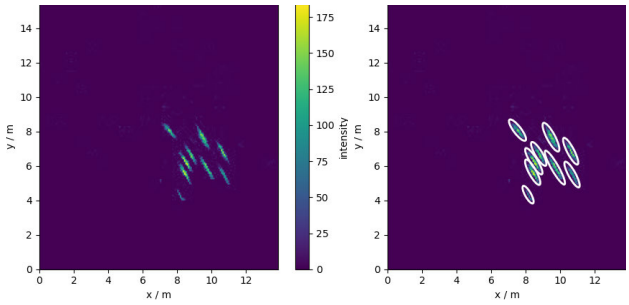


FIGURE 7. Preprocessed sonar image (left), segmented sonar image (right).

Algorithm 1 Segmentation Algorithm Using Positional Covariance

```
tracks ← {}                                          // empty list
pixels ← {{x, y, intensity}}                         // list of pixels
sort pixels by intensity
for pixel in pixels do
    // find maximum probability density
    max_pdf ← 0
    for track_ in tracks do
        pdf ←pdf(track_.μ, track_.Σ, pixel.x, pixel.y)
        if pdf > max_pdf then
            max_pdf ← pdf
            track ← track_
        end if
    end for
    if max_pdf < threshold then
        // add new object to list of tracks
        track ← new Track()
        tracks +← track
    end if
    // assign pixel to object and update μ and Σ
    track.img +← pixel
    track.μ ← center_of_mass(track.img)
    track.Σ ← covariance_of_mass(track.img)
end for
```

with Δt being the time difference between two consecutive frames.

$$\hat{\mathbf{x}} = \begin{bmatrix} x & y & \dot{x} & \dot{y} \end{bmatrix}^\top \tag{3}$$

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{5}$$

The tracking and segmentation steps are performed alternately. The prediction of the Kalman filter is used as starting point for $\boldsymbol{\mu}$ in the next segmentation iteration. The resulting $\boldsymbol{\mu}$ is then inserted as correction into the Kalman filter, before predicting the next frame. The previous iterations $\boldsymbol{\Sigma}$ is used as starting point for the next segmentation without further prediction. The segmentation step can now assign all pixels inside the 3σ borders of the objects at the positions predicted by the Kalman filter at once and only the remaining pixels exceeding a threshold have to be processed one by one. In this configuration, the algorithm runs at a speed of around 40 frames per second, depending slightly on the number of objects being tracked. This should also allow it to be used in real-time applications.

Beside fish, other moving objects like particles in the water are tracked as well. The metrics estimated by the filter can be used to sort out these unwanted objects. In addition to size, speed and proportion, the tracker outputs include timestamps and positions of start and end of each track. For example, particles are often smaller in size, more round in their proportions and move slower than most fishes. As an example, a selection of tracks to be classified as fish could be used as defined in Table 1. The exact conditions can be automatically optimized using annotations if available. Since good annotations are not available for the collected sonar recordings, the optimization is done by maximizing the correlation with detections by the cameras of the UFO-lander. Fishes more than 3 m away from the camera are ignored during the optimization process, as these can probably not be detected by the camera. Also, the times with bad visual conditions such as at night are ignored. The limited visual range of the camera leads only 10% of tracked objects taken into account. The limitation to time spans with good enough visual conditions leads to only 16% of the sonar recording time being taken into account. In total this gives around 1.6% of all available sonar data being used in the correlation, resulting in 1635948 tracked objects.

TABLE 1. Minimal bounds for tracked objects to be classified as fish.

size	>	0.001 m²
average speed	>	0.1 m/s
proportion (length/width)	>	2
time from track start to end	>	5 s
distance from track start to end	>	0.5 m

For the optimization, the features of the tracks are normalized to have a mean of 0 and a standard deviation of 1. Then the SciPy implementation of a Nelder-Mead optimizer was used to optimize the upper and lower bounds for each feature for maximal correlation with the camera detections. Optimizing to maximum correlation with the camera detection class fish_clupeidae leads to the bounds given in Table 2.

TABLE 2. Automatically optimized minimal and maximal bounds to select tracks.

	min	max
size	0.0321 m²	0.354 m²
average speed	0 m/s	2.76 m/s
proportion (length/width)	1	3.83
time from track start to end	0 s	20.8 s
distance from track start to end	0 m	26.5 m

V. EXPERIMENTAL RESULTS

A. EVALUATION USING CALTECH FISH COUNTING DATASET

To compare the detection accuracy with previous methods, the proposed algorithm is applied to the Caltech fish counting dataset [16]. For comparison, HOTA [24], CLEAR MOTA [25], IDF1 [26], nMAE [16] metrics are computed for the results of the new tracker and compared to the results of the baseline and baseline++ methods from [16]. Both methods are based on YOLOv5. The baseline++ method includes additional preprocessing with background subtraction and difference between consecutive frames. The HOTA and MOTA metrics describe object tracking accuracy. The IDF1 metric describes the ID matching accuracy and nMAE defines the normalized mean absolute error when counting fish crossing a virtual line. The evaluation is done on the elwha river recordings. The preprocessing step has been extended to downsample the resolution from 0.012 m to 0.024 m. The evaluation script has been modified to ignore the annotations of the first and last 64 frames of each series because the bandpass filter needs this time to initialize. The results are shown in Table 3. It can be seen that the new tracker performs on par with the baseline method according to the HOTA and IDF1 metrics. In the (Multiple Object Tracking Accuracy) MOTA metric, it performs better than the baseline++ method, but in the normalized Mean Absolute Error (nMAE) metric it is worse than both baseline and baseline++ results. The reason why the new method scores better in the MOTA metric is, that this metric strongly penalizes false positives. As it can be seen in Table 4, the new method has around four times fewer false positives than baseline++, but also four times fewer true positives.

TABLE 3. Evaluation results on Caltech fish counting dataset.

	HOTA	MOTA	IDF1	nMAE
baseline	22.135	-327.94	20.371	30.667
baseline++	39.647	-40.633	49.402	21.667
UFO tracker	23.442	-13.451	22.524	40.667

B. EVALUATION BASED ON THE CAMERA RESULTS

The stationary UFO lander collected continuous sonar and camera recordings from March 2021 to May 2022.

TABLE 4. True positive (TP), false negative (FN) and false positives (FP) on Caltech fish counting dataset according to CLEAR (CLR) and Identity (ID) benchmarks.

	CLR TP	CLR FN	CLR FP	ID TP	ID FN	ID FP
baseline	21002	12134	129483	18703	14433	131782
baseline++	25642	7494	39043	24163	8973	40522
UFO tracker	6448	26688	10771	5671	27465	11548

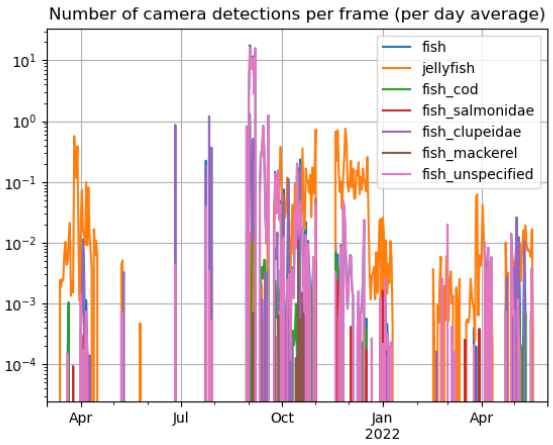


FIGURE 8. Camera detection per fish class from March 2021 to May 2022.

The camera recordings have been processed as described in [27]. To verify that the fish detections from the sonar data are correct, they are correlated with the camera detections.

A correlation of the whole time span between camera and sonar detections gave only a correlation value of 0.0104. This poor correlation is probably caused by the fact that the number of camera detections changes significantly over the course of the year as shown in Figure 8. It can be seen that the camera has 10 times more fish detections in September than in any other month. This causes all other months to become meaningless in the correlation. Therefore, the correlation is calculated for each month separately. The camera classifies detected objects into 15 different classes of fish and jellyfish. The tracker results with the selection conditions from Table 1 from each month are correlated with the different classes of camera detection. The correlation results are shown in Figure 9. It can be seen that there is some correlation with unspecified fishes in April and May. The tracker results correlate with clupeidae in August, with codfishes in October and with salmonidae in January.

The same monthly correlation is performed again while applying the selection conditions given in Table 2. The results are shown in Figure 10. As can be seen, the results in April, June, July and August start to correlate with the camera detections of type clupeidae. It is noticeable that the correlation values never reach more than 0.5. This is caused by the fact, that the field of view of sonar and camera in the water column are not hundred percent overlapped. The sonar has a larger horizontal field of view with 130° instead of 80°, while the camera has a higher vertical field of view with 64° instead of 20°, additionally, the sonar is blind in the first 50 cm.

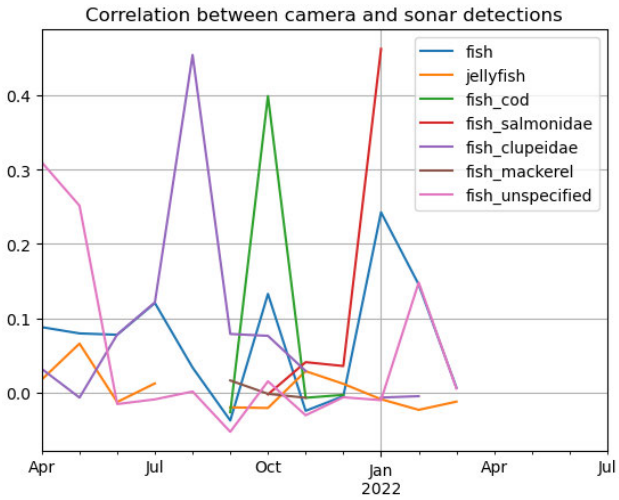


FIGURE 9. Correlation of tracker results with different classes of camera detections.

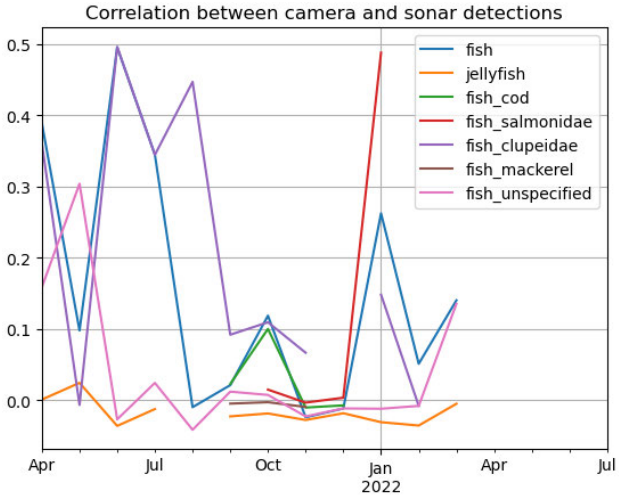


FIGURE 10. Correlation of tracker results with different classes of camera detections optimized to detect fish of type clupeidae.

VI. CONCLUSION

In this work a new method for fish segmentation and tracking was developed and successfully tested. It has been shown that the proposed tracking method could successfully track different fishes on the sonar recordings of the UFO project. The correctness of the detections has been evaluated by correlating them with the camera-based tracker. The correlation values were significantly increased. In contrast to deep learning methods, the described method does not depend on the accuracy of manual annotations and the computation time is moderate. The proposed method can be fine-tuned for different types of fish as shown by optimizing it to detect fish of type clupeidae.

As the comparison with deep learning method has shown superiority of deep learning in some metrics, it is planned to combine deep learning with our proposed method in a future research work.

AKNOWLEDGMENT

The authors gratefully acknowledge the support of Joachim Groeger, Boris Cisewski, Catriona Clemmesen-Bockelmann, Gordon Boer, Hauke Schramm, and Karin Boos for the valuable discussions and support.

REFERENCES

[1] L. M. Wolff and S. Badri-Hoeher, "Imaging sonar-based fish detection in shallow waters," in *Proc. OCEANS*, St. John's, NL, Canada, Sep. 2014, pp. 1–6.

[2] A. Bouzaouit, D. Fietz, and S. Badri-Höher, "Fish tracking based on sonar images by means of a modified optical flow," in *Proc. OCEANS*, Sep. 2021, pp. 1–7.

[3] S. Jenkins, P. Åberg, G. Cervin, R. Coleman, J. Delany, S. Hawkins, K. Hyder, A. Myers, J. Paula, A. Power, P. Range, and R. Hartnoll, "Population dynamics of the intertidal barnacle *Semibalanus balanoides* at three European locations: Spatial scales of variability," *Mar. Ecology Prog. Ser.*, vol. 217, pp. 207–217, Jul. 2001.

[4] C. Spampinato, D. Giordano, R. Di Salvo, Y.-H.-J. Chen-Burger, R. B. Fisher, and G. Nadarajan, "Automatic fish classification for underwater species behavior understanding," in *Proc. 1st ACM Int. Workshop Anal. Retr. Tracked Events Motion Imag. Streams*, Firenze, Italy, Oct. 2010, pp. 45–50.

[5] M. A. Iqbal, Z. Wang, Z. A. Ali, and S. Riaz, "Automatic fish species classification using deep convolutional neural networks," *Wireless Pers. Commun.*, vol. 116, no. 2, pp. 1043–1053, Jan. 2021.

[6] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager, and E. Harvey, "Fish species classification in unconstrained underwater environments based on deep learning," *Limnology Oceanogr., Methods*, vol. 14, no. 9, pp. 570–585, Sep. 2016.

[7] H. Qin, X. Li, J. Liang, Y. Peng, and C. Zhang, "DeepFish: Accurate underwater live fish recognition with a deep architecture," *Neurocomputing*, vol. 187, pp. 49–58, Apr. 2016.

[8] A. Abu and R. Diamant, "Enhanced fuzzy-based local information algorithm for sonar image segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 445–460, 2020.

[9] N. Sun, T. Shim, and H. Hahn, "Sonar image segmentation based on Markov gauss-Rayleigh mixture model," in *Proc. Int. Workshop Educ. Technol. Training, Int. Workshop Geosci. Remote Sens.*, Dec. 2008, pp. 704–709.

[10] G. Huo, S. X. Yang, Q. Li, and Y. Zhou, "A robust and fast method for sidescan sonar image segmentation using nonlocal despeckling and active contour model," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 855–872, Apr. 2017.

[11] A. Abu and R. Diamant, "Unsupervised local spatial mixture segmentation of underwater objects in sonar images," *IEEE J. Ocean. Eng.*, vol. 44, no. 4, pp. 1179–1197, Oct. 2019.

[12] T.-M. Perivolioti, M. Tuser, D. Terzopoulos, S. P. Sgardelis, and I. Antoniou, "Optimising the workflow for fish detection in DIDSON (Dual-frequency IDentification SONar) data with the use of optical flow and a genetic algorithm," *Water*, vol. 13, no. 9, p. 1304, May 2021.

[13] X. Zang, T. Yin, Z. Hou, R. P. Mueller, Z. D. Deng, and P. T. Jacobson, "Deep learning for automated detection and identification of migrating American eel *Anguilla rostrata* from imaging sonar data," *Remote Sens.*, vol. 13, no. 14, p. 2671, Jul. 2021.

[14] K. Sreekala, N. N. Raj, S. Gupta, G. Anitha, A. K. Nanda, and A. Chaturvedi, "Deep convolutional neural network with Kalman filter based objected tracking and detection in underwater communications," *Wireless Netw.*, pp. 1–18, Mar. 2023.

[15] L. Zeng, B. Sun, and D. Zhu, "Underwater target detection based on faster R-CNN and adversarial occlusion network," *Eng. Appl. Artif. Intell.*, vol. 100, Apr. 2021, Art. no. 104190.

[16] J. Kay, P. Kulits, S. Stathatos, S. Deng, E. Young, S. Beery, G. Van Horn, and P. Perona, "The Caltech fish counting dataset: A benchmark for multiple-object tracking and counting," 2022, *arXiv:2207.09295*.

[17] J. Li, L. Chen, J. Shen, X. Xiao, X. Liu, X. Sun, X. Wang, and D. Li, "Improved neural network with spatial pyramid pooling and online datasets preprocessing for underwater target detection based on side scan sonar imagery," *Remote Sens.*, vol. 15, no. 2, p. 440, Jan. 2023.

[18] N. H. Ali and G. M. Hassan, "Kalman filter tracking," *Int. J. Comput. Appl.*, vol. 89, no. 9, pp. 15–18, 2014.

[19] X. Li, K. Wang, W. Wang, and Y. Li, "A multiple object tracking method using Kalman filter," in *Proc. IEEE Int. Conf. Inf. Autom.*, Jun. 2010, pp. 1862–1866.

[20] A. S. D. Murthy, S. K. Rao, K. S. Naik, R. P. Das, K. Jahan, and K. L. Raju, "Tracking of a manoeuvering target ship using radar measurements," *Indian J. Sci. Technol.*, vol. 8, no. 28, Oct. 2015.

[21] K. L. Raju, "Passive target tracking using unscented Kalman filter based on Monte Carlo simulation," *Indian J. Sci. Technol.*, vol. 8, no. 1, pp. 1–7, Jan. 2015.

[22] Z. Chen, Y. Wang, W. Tian, J. Liu, Y. Zhou, and J. Shen, "Underwater sonar image segmentation combining pixel-level and region-level information," *Comput. Electr. Eng.*, vol. 100, May 2022, Art. no. 107853.

[23] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3464–3468.

[24] J. Luiten, A. Osep, P. Dendorfer, P. Torr, A. Geiger, L. Leal-Taixe, and B. Leibe, "HOTA: A higher order metric for evaluating multi-object tracking," *Int. J. Comput. Vis.*, vol. 129, pp. 548–578, Feb. 2021.

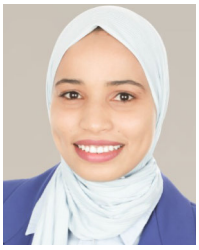
[25] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," *EURASIP J. Image Video Process.*, vol. 2008, pp. 1–10, Jan. 2008.

[26] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Computer Vision—ECCV*, G. Hua and H. Jegou, Eds. Cham, Switzerland: Springer, 2016, pp. 17–35.

[27] G. Böer, J. P. Gröger, S. Badri-Höher, B. Cisewski, H. Renkewitz, F. Mittermayer, T. Strickmann, and H. Schramm, "A deep-learning based pipeline for estimating the abundance and size of aquatic organisms in an unconstrained underwater environment from continuously captured stereo video," *Sensors*, vol. 23, no. 6, p. 3311, Mar. 2023.

JULIAN WINKLER received the master's degree in electrical engineering from the Kiel University of Applied Sciences, Kiel, where he is currently pursuing the Ph.D. degree with the Signal Processing Group. His research interests include underwater localization, navigation, and communication by means of acoustical signals.

SABAH BADRI-HOEHER received the master's degree in physics from the University of Casablanca, Casablanca, Morocco, in 1991, the Dipl.-Ing. (M.Sc.) degree in electrical engineering from the University of Paderborn, Paderborn, Germany, in 1996, and the Dr.-Ing. (Ph.D.) degree in electrical engineering from the University of Erlangen, Erlangen, Germany, in 2001. She has been several years with the Fraunhofer Institute of Integrated Circuits, Erlangen, and the Christian-Albrechts-University of Kiel. Since 2009, she has been a Full Professor with the Faculty of Computer Sciences and Electrical Engineering, Kiel University of Applied Sciences, Kiel, Germany. Her research interests include signal processing and communication techniques with a focus on underwater applications.



FATNA BARKOUCH received the master's degree in data science and big data from the Hassan II University of Casablanca, Morocco, in 2020. She is currently pursuing the Ph.D. degree with the Signal Processing Group, Kiel University of Applied Sciences. Her research interests include computer vision and image processing.