

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>., provided by the author. The material includes database from sea experiments used in the analysis section. Contact roee.d@univ.haifa.ac.il for further questions about this work

An Active Acoustic Track-Before-Detect Approach for Finding Underwater Mobile Targets

Roe Diamant, Dror Kipnis, Eyal Bigal, Aviad Scheinin, Dan Tchernov and Adi Pinchasi

Abstract—We consider the challenge of tracking and estimating the size of a single submerged target in a high reverberant underwater environment using a single active acoustic transceiver. This problem is common for a multitude of applications, ranging from the security and safety needs of tracking submerged vehicles and scuba divers, to environmental research and management implications such as the monitoring of pelagic fauna. Considering that the target can be either slow (e.g., a scuba diver) or fast-moving (e.g., a shark), we avoid continuous signalling, and rely on the emission of wideband pulses whose reflection pattern are evaluated and reshaped in a time-distance matrix. As opposed to common approaches that track targets through template matching or by using tracking filters, we avoid making difficult assumptions about the target’s reflection patterns or motion type, and instead perform probabilistic tracking using a constraint Viterbi algorithm, whereby detection is determined based on maximum likelihood criterion. In this process, we use the expectation-maximization (EM) approach to manage stationary reflections through distribution analysis, which otherwise may be misidentified as targets. Based on the tracked path, we then evaluate the target’s size. To test our approach, we performed extensive simulations as well as eight sea experiments in different environmental settings to track both a scuba diver and a sandbar shark (*Carcharhinus plumbeus*). The simulation results show a tracking performance that is close to the Cramér-Rao lower bound, and the experiment results show a good trade-off between detection rate and false alarm rate for a low signal-to-clutter ratio of 5 [dB], and average tracking error of 1.5 [m] and 6.5 [m] in the detections of a scuba diver and sandbar shark, respectively. For reproducibility, we share our sea experiment data.

Index Terms—Underwater acoustic detection, Track-before-detect, Tracking moving targets, Detection in a reverberant environment, Shark detection, Scuba diver detection

I. INTRODUCTION

We consider the challenge of automatic tracking of a single submerged target that is slowly or fast moving in a high reverberant environment (e.g., a shallow water reef). For ease of deployment, we consider mono-static detection using a single transceiver, and focus on active detection. This capability is in high demand for security and environmental research implications. For security needs, a terrorist attack on a marine facility is considered a game changer in terms of foreign policy [1], [2], [3]. The considered scenario involves

scuba divers approaching a marine facility such as a gas rig, a drilling platform or a harbour exit, and either directly attacking it or deploying submerged mines to disable maritime traffic. The scuba divers may arrive by boats, small submerged vehicles or swimming. To secure coastal facilities, the timely detection of approaching scuba divers or deployed mines is required. In addition, it is important to evaluate the location of the suspected object so that efficient counter-measures can be employed.

With regard to environmental research and management, a reliable scheme of submerged target tracking may be useful for the detection and size estimation of pelagic species. Abundance indices of highly migratory fauna often rely on fishery-dependent data, which are known to be biased towards more vulnerable species as well as certain types of habitats or gear [4]. In addition, fishery records hinge upon an incentive to report and are thus complicated by the typical omission of illegal, unreported and unregulated (IUU) fishing as well as discarded or recreational catch [5]. To deal with these types of biases, it is necessary to develop indiscriminative and reliable, fishery-independent methods. To this end, acoustic systems have been increasingly used for the detection of marine fauna during ecological surveys [6].

A few methods showed capability to passively detect scuba divers in ranges of tens of meters [7] or at low signal-to-noise ratio [8]. Yet, these methods rely on an array of receivers to achieve directivity, and detect specific high frequency sounds produced by the diver’s breathing system, which do not apply for the detection of marine fauna. For these reasons, in this work we use active acoustics. Since the subject target may be close to the sea surface or seabed, it might be difficult to distinguish from clutter or unwanted objects such as rocks, anchors, chains, etc. Considering this challenge, one option is to use template matching, where the reflected signal is compared with a fingerprinting sample using a matched filter [9]. However, template matching requires prior knowledge about either the environment or the target, which are both considered unknown in the herein provided case. Detection of a moving target is also viewed as a tracking problem and solutions are traditionally obtained through the use of tracking filters; based on an assumed motion model, the filter is set to track multiple possible paths [10], [11]. A different approach is to track the distance-velocity matrix where, for each signal, non-zero velocity components are searched by estimating Doppler shift content in the reverberant signal [12].

While some solutions have been designed for tracking either fast or slow targets, none have been proposed for the joint problem. This problem is especially important since, in the

R. Diamant (corresponding author, email: roeed@univ.haifa.ac.il), Dror Kipnis, and A. Pinchasi are affiliated with the Dept. of Marine Technologies; E. Bigal, Aviad Scheinin and Dan Tchernov are affiliated with the Morris Kahn Marine Research Station, Department of Marine Biology. This work was funded by the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 773753 (Symbiosis), and by the NATO Science for Peace and Security Programme under grant G5293.

Parts of this work have been presented at the 176th meeting of the Acoustical Society of America, Victoria, Canada.

case of underwater target detection, both types of targets may exist simultaneously. In particular, it may be challenging to distinguish between slow mobile target and stationary reflectors such as rocks or anchors. Considering this challenge, in this paper we propose a track-before-detect approach to find, track, and estimate the size of a single submerged target in the presence of strong clutter and stationary reflectors. For this purpose, we use a combination of a constraint expectation-maximization (EM) algorithm and a constraint Viterbi algorithm that works on a time-distance (TD) matrix of received reflections. The EM identifies stationary targets and discards them, while the Viterbi algorithm tracks the mobile target's path. Detection is then determined by maximum likelihood. Once a valid path is identified, we evaluate the target's size based on a statistical analysis of widths obtained from all of the target's reflections. Since, realistically, the deploying platform may be small, we consider using only a single acoustic transceiver. While we assume a mixture of distributions to identify stationary targets, no motion model is imposed on the mobile target. Thus, our method can handle any motion pattern.

Our contribution is threefold:

- 1) A new blind track-before-detect approach that manages both slow and fast-moving targets using only a single transceiver;
- 2) A statistical approach to identify stationary targets from a TD representation of the reflection pattern;
- 3) A compact method to evaluate target size.

We have examined the performance of our approach by means of numerical simulations as well as eight sea experiments in different environmental settings. Four of the experiments involved the tracking of scuba divers, and the other four involved the tracking of a shark, captured for ecological research purposes as part of a permitted monitoring programme. The simulation results showed tracking performance close to the derived Cramér-Rao lower bound (CRLB). Moreover, compared to ground truth, the results of all experiments showed good performance in terms of tracking, detection, and size estimation in the presence of high clutter, noise transients and a low signal-to-clutter-ratio (SCR).

The remainder of this paper is organized as follows. In Section II, we describe the state-of-the-art in acoustic target tracking. Our system's model and assumptions are described in Section III. In Section IV, we describe our tracking and target's size estimation methodology. Performance is analyzed in Section V for both the simulations (Section V-A and sea experiments (Section V-B). Conclusions are drawn in Section VI. A list of the major notations is provided in Table I.

II. RELATED WORK

Two main approaches exist for the detection of targets by using acoustic emissions. The first is the transmission of pulses with injected guard intervals to suppress clutter. The target is then found by tracking its possible path after a given number of acoustic signals are transmitted [13]. The second approach is the continuous active SONAR (CAS), which employs narrowband transmissions in multiple sub-carriers to

detect Doppler components [14]. While the latter approach yields near real-time detection, it relies on the existence of a strong Doppler shift, and may thus not fit our considered case of slowly moving targets.

The main challenge in active underwater acoustic target tracking is strong clutter. To form the TD matrix with a reasonable SCR, the common approach is to accumulate matched filter (MF) responses of the received reflections [15]. Alternatively, based on the expected distribution of noise, locally optimum filters can replace the MF [16], to manage detection in the presence of arbitrary noise and interference [17]. In the presence of transient noise (produced by e.g., snapping shrimps), the MF may be preceded by noise cancellation, usually using median filters [18] or wavelet de-noising [19]. In all of the above schemes, a detection threshold is set by the known or estimated characteristics of the ambient noise, e.g., its variance. However, at sea, noise often varies rapidly over time, and estimating noise levels and their statistics may be challenging.

Once the TD matrix has been formed, a tracking procedure can be employed. Solutions include blind signal detection and tracking in non-Gaussian noise environments [20]; supervised data association and classification of signal reflections [21]; and employing tracking filters, which is considered the most common approach. A powerful approach for dealing with non-Gaussian clutter can be found in [9], where the power spectral density of the clutter is modeled as a mixture of Rayleigh distributions, enabling derivation of detection scheme optimal for the clutter. Alternatively, to manage non-Gaussian noise, various noise density functions has been used [17] with an application for magnetic resonance imaging (MRI) curve evaluation [22]. To deal with non-linearity of measurements, tracking can be achieved using variants of the extended Kalman filter [10], that track radial range and velocity measurements [23], [10]. In [24], solutions to variant tracking applications are compared.

In cases of low SCR, detection is often preceded by a track-before-detect approach [25]. This approach avoids false detections which may exist when applying a detection threshold determined by a mismatched clutter model or by signal's fluctuations, and instead performs tracking by following possible target reflection patterns in the TD image. Common track-before-detect schemes include maximum-likelihood probabilistic data association (ML-PDA), particle filtering (PF), dynamic programming tracking, and probabilistic multi-hypothesis tracking (PMHT). The first approach applies a sequential likelihood ratio test to the sequence of measured track's parameters [26], and has the ability to track very low observable targets. Applications of ML-PDA for real bistatic and multistatic SONAR data are demonstrated in [27] and [28], respectively. The second and third approaches are based on a numerical representation of the probability density function (PDF) of the observations, while the fourth scheme uses a parametric representation of the target's PDF to dramatically reduce the computational cost (e.g., [29]). PMHT was first presented in [29], to perform tracking over successive time-frames of the SONAR image by separating the superposition of target and clutter components using EM estimation. An extension of

TABLE I. List of major notations

Notation	Explanation
N^{sig}	Number of emitted signals
K	Number of samples corresponding to the reverberation delay spread
τ_1	Lower bound of the number of samples allowed to represent reflection from a target
τ_2	Upper bound of the number of samples corresponding to the drifting distance of the deploying platform
\mathbf{M} and d	The $[N^{\text{sig}} \times K]$ time-distance matrix and its elements d
w , $\hat{w}(i)$, and \hat{w}	Actual size of the target, its estimate from the i th emission, and its final estimation (in samples)
φ	Threshold for detection decision
δ_m and ω_m	Prior and distribution parameters of the m th class
λ_i and Γ_i	The i th column of \mathbf{M} and its classifier
\mathbf{t} and $\hat{\mathbf{t}}$	The time-distance track of the target and its estimate
ρ_w and ρ_t	Size error and tracking error measures
T^{measure}	Measurement time

this approach is given in [30] to combine the histogram of the intensity data. Performance of these algorithms for various applications including SONAR were studied in [25], [24], and demonstration over a towed array SONAR is presented in [31]. To improve performance in cases of low SCR, [32] offered to use a Poisson mixture model that handles fluctuation in the TD matrix. A comparison between track-before-detect approaches can be found in [33], [25]. The former concludes that PMHT has an advantage over ML-PDA in scenarios where there are two close targets with similar motion dynamics. The work in [25] compared between dynamic programming, ML-PDA, PF, and histogram probabilistic multi-hypothesis tracker (H-PMHT), and concluded that the latter method is best in terms of computational cost and position accuracy. However, H-PMHT fails to track fast-moving targets. An analytic *trackability* framework which aim to determine the clutter/interfering conditions in which tracking is feasible was developed by Schoenecker et al. in [34], [35], [36] for various types of clutter. Extension to the case of two close-by targets is given in [37].

The probabilistic multi hypothesis tracking uses an assumed measurement model, which may be mismatched with actual data. Instead, the PF tracking approach tries to learn this probabilistic model. The scheme involves estimating the PDF of the target's reflections by statistically sampling the discrete grid of the target's state-space (e.g., the target's bearing and range). For a multi-target scenario, a demonstration of this concept on active and passive SONAR is presented in [38] and [11], respectively. Another tracking approach that avoids imposing a measurement model is dynamic programming, which applies the Bayesian tracking framework. Assuming the target-related reflections follow a Markov process, tracking is performed following a grid search. Application for target detection in low SCR using an active SONAR is reported in [13]. While complexity is polynomial, the search is performed over the entire TD matrix, and thus computation costs is high.

The above approaches propose solutions applied for target detection using active acoustics. We observe three main gaps that limit the practicality of these methods: 1) Mismatches between the assumed clutter model and the distribution of reflections from detected targets. This mostly applies to filtering approaches, which either assume or estimate the PDF for clutter's and target's reflections, and thus are limited to cases of static PDFs which, in practice, may change due to

the time-varying sea conditions. 2) Hard assumptions about the motion pattern of the target. This is specifically correct for the ML-PDA, filtering, and the model-based dynamic programming tracking approaches, which rely on a simple dynamic model, e.g., fixed velocity, to represent and track the target. 3) Robustness to various target types. For example, due to its assumed Markov model, dynamic programming tracking do not manage well fluctuations in the strength of the target's reflections, and may loose the track completely in case of reflections that sometimes disappear in clutter. Differently, PMHT performs well when tracking slow moving targets but have difficulties in tracking fast targets [25]. Moreover, to the best of our knowledge, the available literature do not manage the case of reflections from stationary targets such as rocks, ship hulls, concrete poles, etc., that may be misidentified as targets. In the following, we describe our track-before-detect method that takes special care of these four factors, namely, the mismatches of clutter model, assumption on the target motion, robustness to various targets, and managing stationary targets. We explore the performance of our method in both simulations and in sea experiments with real targets. To the best of our knowledge, ours is the first work to report results on real submerged targets. We thus share our raw measurements for further exploration in this field.

III. SYSTEM SETUP AND ASSUMPTIONS

Our aim is to design a low-complexity, reliable and autonomous tracking system that is robust in the marine environment. Our setup includes a single acoustic transceiver deployed from an anchored or slowly drifting platform such as a buoy or boat. The transceiver emits a sequence of N^{sig} acoustic signals of duration T^{sig} [s], and records the reflections with an assumed negligible transmission/reception switch delay. The cycle lasts for $N^{\text{sig}} \cdot T^{\text{sig}}$ [s], after which target tracking is performed. The signals are emitted with guard intervals that correspond to the projector's detection range.

During the observation time, $N^{\text{sig}} \cdot T^{\text{sig}}$, the TD matrix may or may not include target-related reflections. We consider a moving object that can be slow or fast-moving (here, we demonstrate this for a scuba diver and a shark, respectively). The target moves along a smooth but unknown path, and its position relative to the transceiver is unknown. The target's shape is also unknown, and hence the reflection pattern cannot

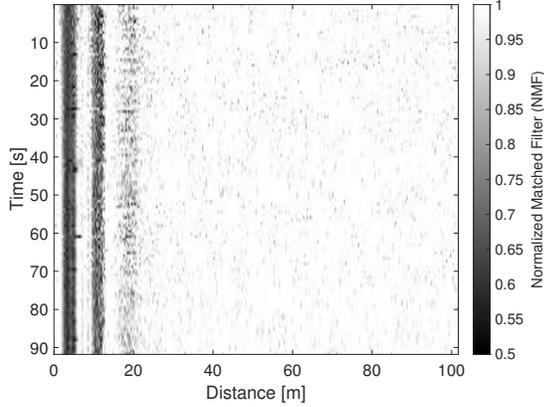


Fig. 1: Example of a TD matrix M containing reflections from stationary targets. Ship hull at distances 4 [m] and 10 [m], and an anchor at distance 20 [m]. Normalized matched filter outputs below 0.5 are neglected to emphasize reflections from stationary targets.

be compared with a template reflection. No prior knowledge is assumed regarding the sea bathymetry and bathymetry¹.

Our solution fits the case of a single mobile target, which can be either slow or fast-moving. In our solution, we assume a rough upper bound for the target's maximum speed, V_{\max} . We also assume a lower bound τ_1 [samples], such that the minimum target's size is $\tau_1 c / F_s$, where F_s [Hz] is the sampling frequency. This lower bound is used to smooth the TD matrix and thus mitigate clutter which, in turn, is modelled to be i.i.d. random in time and space. For a relatively low carrier frequency of 12 kHz, our results showed tolerance to the choice of the upper bound of the target size. We also assume an upper bound for the drifting distance of the deploying platform with a corresponding number of samples, τ_2 [samples], around a reference location during the measurement period $N^{\text{sig}} \cdot T^{\text{sig}}$. This bound is required to compensate for the movement of the platform, which can be obtained from its GPS log or from the expected magnitude of the water current. We make the assumption that τ_2 is greater than τ_1 , and in the following τ_2 as an upper bound for the target's size.

IV. DETECTING A SUBMERGED MOBILE TARGET

In the absence of prior knowledge about the target's texture, the system cannot assume that the target possesses distinctive characteristics. Yet, the fact that the desired target is mobile offers some opportunities for detection by utilizing the stability of its reflections. In particular, our approach is based on analyzing the TD reflection matrix, created by the stacking of match filter outputs for consecutive reflections. An example of such a matrix is shown in Fig. 1. The matrix is processed to distinguish between static reflectors like rocks or anchors (type I); reflections from the sea boundaries or volume scatterers (type II); and reflections from other mobile targets (type III). The key idea behind our tracking method is that, since

¹In our results, we use the sound speed, c [m/s], but only for visual inspection and not for tracking or size estimation.

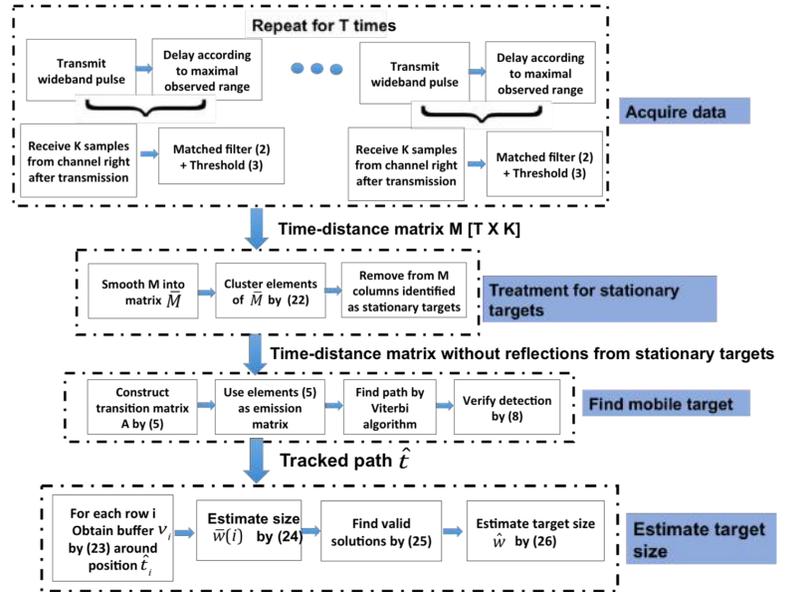


Fig. 2: Illustration of our track-before-detect procedure.

reflections from targets are stable, in the TD domain, only reflections of type III are both stable and non-stationary in distance. That is, in the TD domain, a valid target will form a continuous, but curved, line. The width of this line corresponds to the size of the reflecting object.

Referring to the illustration in Fig. 2, we start by creating the TD matrix from a sequence of emitted wideband signals. Next, we filter the matrix to remove reflections from stationary targets by analyzing the statistical distribution of each reflection instance. The clustering is performed by estimating the posterior of each reflection and jointly clustering the columns of the TD matrix. Next, we execute the Viterbi algorithm [39] over the filtering TD matrix to track remaining reflections which have a stable path. Since we already eliminated reflections from stationary targets, the established path is considered to represent a mobile target. Detection is then made according to maximum-likelihood criteria over the subject path. Finally, the size of the mobile target is estimated according to the width of valid reflections along the tracked path. Note that our method avoids making assumptions on the motion of the target, nor does it classify it as slow or fast-moving. Rather, the estimated path may belong to either target type.

A. Path Tracking

As a first step, we create the TD matrix, M , whose i th row is comprised of the reflection pattern for the signal transmitted at time index i . Let s be a baseband transmitted signal, $h(i)$, be a reverberant channel impulse response, and $r(i)$ be a received reverberant signal. Assuming s is a wideband signal and neglecting the Doppler shift effect as well as ambient noise, the i th row of M , i.e., the convolution between s and $h(i)$, is [40]

$$s * r(i) \approx s * s * h(i) \approx \mathcal{I} * h(i) = h(i), \quad (1)$$

where \mathcal{I} is the Kronecker delta function. Considering (1), we use a MF to evaluate the impulse response for the reverberant

signal that is reflected back from the channel.

To mitigate clutter, we threshold the MF output so that element $M(i, j)$ is set to zero if $M(i, j) < \text{Th}$. While most approaches use a parametric detector, i.e., one that requires parameter estimation for the noise (and sometimes also signal) distribution, in our case, this approach may lead to a high false negative rate due to the time-varying reverberant sea environment. Hence, to set Th , we use a non-parametric approach. That is, avoiding the need to estimate the noise distribution, we use the normalized matched filter

$$\text{NMF}(i) = \frac{\mathbf{s}\mathbf{r}(i)^T}{\sqrt{\mathbf{s}\mathbf{s}^T \cdot \mathbf{r}(i)\mathbf{r}(i)^T}}, \quad (2)$$

and set Th by [41]

$$P_{\text{fa}} = 1 - B\left(\text{Th}, \frac{1}{2}, \frac{N-1}{2}\right), \quad (3)$$

where P_{fa} is the desired false alarm rate, N is the time-bandwidth product of \mathbf{s} , and

$$B(a, b, z) = \int_0^a n^{b-1}(1-n)^{z-1}dn$$

is the (tabulated) regularized incomplete beta function. That is, given a desired false alarm probability and N characteristic, we calculate Th numerically by (3). The NMF (2) for the i th transmitted signal operates on a reflected buffer \mathbf{r} of duration K , where K is set by the assumed detection range, and both \mathbf{r} and \mathbf{s} are baseband shifted and downsampled. Repeating this operation for N^{sig} signals, we obtain a $[N^{\text{sig}} \times K]$ TD matrix \mathbf{M} .

Observing \mathbf{M} as a timely sequence of N^{sig} reflection observations, we detect the mobile submerged target via a track-before-detect approach. Specifically, we aim to find a sequence of consecutive reflections in the distance domain of high likelihood to originate from a target. Here, we do not assume knowledge about the target's motion pattern, but rather that it moves in an organized fashion with a bounded speed, V_{max} . Formally, for a target observed in time instance i at distance k and in time instance $i+1$ at distance j , i.e., by $M(i, k)$ and $M(i+1, j)$, respectively, we assume

$$|j - k| \leq V^{\text{max}} \cdot \Delta_T, \quad (4)$$

where Δ_T is the time elapsed between two consecutive signal transmissions or rows of \mathbf{M} . Hence, a valid target path should reflect a state change by no more than $V^{\text{max}} \cdot \Delta_T$ across the field of observations $0, \dots, K-1$.

Before detecting the target, we run an initial step to reduce the clutter in \mathbf{M} . For this purpose, we observe that a valid target-related reflection is one that corresponds to a non-unity target size. That is, we expect valid reflections to have some *width* of unknown size w , such that for a target identified at the i th row of \mathbf{M} , there exist elements $\{M(i, j), \dots, M(i, j+w-1)\}$ that correspond to a target centred around column $j+w/2$. We also note that, while samples associated with reflections from the target are dependent or even non-stochastic, clutter-based reflection samples are i.i.d. Hence, the *width* of the reflection allows us to differentiate between valid targets'

reflections and clutter reflections. Still, in the case of a small target or in the case of emitted acoustic signals of low carrier frequency, the unknown value w may be small. Thus, in an attempt to avoid making hard decisions, we smooth \mathbf{M} . To this end, we use a median filter of length τ_1 .

To find the target's path, we use the Viterbi algorithm for the Hidden Markov Model [42]. In the context of a Hidden Markov Model, the Viterbi algorithm finds the most probable set of states within a sequence of observations. Reflecting this idea to our case, we identify the columns of the TD matrix (i.e., the distance samples) as states, and the rows of the TD matrix (i.e., the time samples) as observations. Then, using the Viterbi algorithm, we find the set of time-distance entries in the TD matrix that most probably fits a mobile target. To that end, we set the values of the TD matrix's indexes (i.e., the normalized matched filter outputs) as emissions, and use the pre-knowledge about the target's maximum velocity for the transition matrix between the different states. The former is set since the output of the normalized matched filter can serve as a reliability measure, while the latter indicates how far a target can shift its location between two states in a given time. We chose the Viterbi algorithm because it efficiently optimizes a sequence of emission probabilities, while allowing the use of preliminary knowledge by formulating transmission constraints between the observed states. Moreover, since the Viterbi algorithm is structured as a private case of the forward-backward algorithm [39], it does not require initial conditions, and is thus suitable for a track-before-detect approach. As mentioned above, to set the transition probability from one state to another, we limit transition jumps by ensuring that transition probability from one state to another is zero unless the maximal assumed velocity is met. This way, we make sure the solution of the Viterbi algorithm follows a smooth line and avoids non-physical rapid changes. Formally, we set a $K \times K$ transition matrix, \mathbf{A} , by

$$A(j, k) = \begin{cases} 0 & \text{if } |j - k| > V_{\text{max}} \cdot \Delta_T \\ \frac{1}{V_{\text{max}} \cdot \Delta_T} & \text{otherwise} \end{cases}, \quad (5)$$

such that, without pre-knowledge of the mobility model of the target, uniform distribution is considered for allowed transitions. The result is an estimated path $\hat{\mathbf{t}}$ that meets the criterion in (4), whose element $\hat{t}(i)$ corresponds to the distance of the target at time instance i . Using the Viterbi algorithm, path $\hat{\mathbf{t}}$ is chosen as the path that maximizes the set of normalized MF coefficients - a set of N^{sig} responses. The complexity is $\mathcal{O}(N^{\text{sig}} \cdot K^2)$ [42]. The resulting path, $\hat{\mathbf{t}}$, can lock on a true target path that lasts for the whole observation time, $N^{\text{sig}} \cdot T^{\text{sig}}$, or for part of the TD matrix \mathbf{M} . This is particularly important since submerged targets may appear or disappear in the clutter, depending on their position along the water column.

B. Detection Decision

Once the target path has been estimated, we make a detection decision via a maximum likelihood criterion for binary hypothesis, with state '1' being a target path, and state '0' being a clutter path (non-target path). Given the accumulated probability obtained by operating the Viterbi algorithm along

the chosen path, $P(\hat{\mathbf{t}}|\mathcal{H}^1)$, and along other arbitrary paths, $P(\hat{\mathbf{t}}_b|\mathcal{H}^0)$, $\hat{\mathbf{t}}_b \neq \hat{\mathbf{t}}$, we obtain the optimal hypothesis testing strategy by computing the log-likelihood ratio (LLR) [43],

$$\text{LLR}_b = \beta^1 - \beta_b^0, \quad (6)$$

where $b = 1, 2, \dots, B$ is a running index with B being a user's choice,

$$\beta^1 = \log P(\hat{\mathbf{t}}|\mathcal{H}^1), \quad (7a)$$

$$\beta_b^0 = \log P(\hat{\mathbf{t}}_b|\mathcal{H}^0). \quad (7b)$$

Assuming a single target, we make a decision of the existence of a target by comparing the likelihood ratio (6) to a threshold. As we show further on, the threshold is chosen as a trade-off between the detection and false alarm rates for a given SCR level. Formally, the decision procedure for the hypothesis testing problem becomes

$$\hat{\mathcal{H}} = \begin{cases} \mathcal{H}^1 & \forall b: \text{LLR}_b > \varphi, \\ \mathcal{H}^0 & \exists b: \text{LLR}_b \leq \varphi, \end{cases} \quad (8)$$

where $-\infty < \varphi < 0$ is a threshold that trades off between false alarms and detection probabilities. To determine φ , we evaluate LLR_b on-the-fly by taking paths b for which $P(\hat{\mathbf{t}}_b|\mathcal{H}^1)$ is not maximal. That is because, assuming a single mobile target, it will be detected by the most probable Viterbi path, whereas least probable paths will mark a path consisting of clutter reflections. Since reflections from a target are expected to occupy several columns in the TD matrix, for the log-likelihood test we choose paths b that are positioned at least τ_2 samples from the maximal path $\hat{\mathbf{t}}$.

C. Ignoring Stationary Targets

One particular drawback of the Viterbi algorithm is that it locks onto the best path in terms of maximum emission probability; however, in contrast to a tracking filter, it does not follow a specific motion model. The solution is therefore vulnerable to strong reflections from stationary targets. These can be reflections from objects like anchors, chains, pipes or rocks, but also reflections from the deploying vessel itself. With respect to reflections from the mobile target, all the above are characterized by steady reflections, that is, for a stationary reflector at distance j ,

$$M(i_1, j) \approx M(i_2, j \pm \epsilon), \quad (9)$$

where i_1 and i_2 are two, not necessarily consecutive, observation time indexes, and ϵ is a bounded distance set by the transceiver's self-motion (e.g., its drifting motion). Note that we allow also the target to sometimes stay in equal range to the transceiver. This may occur when the target moves in a circle around the transceiver or when it stops to e.g., change course. However, the number of these cases is assumed small relative to N^{sig} . The result of reflections from a stationary target is a seemingly straight line in the TD image representation of M . An example of such an image is shown in Fig. 1. The image was formed by stacking reflection patterns for a transmission of a 10 kHz bandwidth linear frequency modulation signal deployed at a 2 m depth from a vessel anchored in a water

depth of roughly 20 m. We observe two straight lines at distances of roughly 4 m and 20 m, which correspond to the direct and second reflections from the deploying vessel, respectively. Note that the observed reflection lines are only semi-straight lines. This is due to the boat's drifting motion around its anchor.

Since reflections from a stationary target are often strong, the Viterbi algorithm may lock on these reflections and consider it a valid target. Our solution to this problem is based on observation (9). This observation states that values in the columns of M related to stationary reflections have lower variance and higher mean average compared to clutter- or target-related columns. We therefore consider the problem of detecting stationary targets as a clustering problem, for which all entries in M are classified with two labels: either 'belong to a stationary target' or 'belong to clutter or a moving target'. Utilizing the statistical relations between the two clusters, we solve the problem using the EM algorithm [43].

1) Formalizing the EM Solution:

Recall that, due to the self-motion of the deploying vessel, delays (or distances) of reflections from a stationary target will likely slowly vary over time. Hence, instead of straight lines in M , reflections from stationary targets would more likely form curved lines. If the precise self-motion of the deploying vessel is known, e.g., by using a differential GPS, the matrix M can be corrected to compensate for this drift. Otherwise, as an initial step in detecting reflections from stationary targets, we form a modified matrix \bar{M} whose (i, j) entry is the accumulation

$$\bar{M}(i, j) = M(i, j) + \dots + M(i, \min[j + \tau_2, K]), \quad (10)$$

where τ_2 corresponds to the maximum drift of the deploying vessel. This way, each column of \bar{M} takes into account all measurements in a window of length τ_2 .

Our goal is to cluster each element in \bar{M} . Let $m = 1$ indicate a stationary target, and $m = 2$ indicates a clutter or a moving target. Also, let $\bar{M}_{i,j}$ represent the (i, j) entry of the TD matrix, and set ω_m as the distribution parameters of the m th class. Assuming the elements in \bar{M} are i.i.d., we employ a mixture distribution model,

$$p(\bar{M}|\Omega) = \sum_{m=1}^2 \prod_{i,j} \delta_m p(\bar{M}_{i,j}|\omega_m), \quad (11)$$

where δ_m is the prior for class $m \in \{1, 2\}$ such that

$$\delta_1 + \delta_2 = 1, \quad (12)$$

with $\Omega = \{\delta_1, \omega_1, \delta_2, \omega_2\}$ as the model's parameter set, and $p(\bar{M}_{i,j}|\omega_m)$ is the distribution of the single reflection element, which we model as the Gaussian

$$p(\bar{M}_{i,j}|\omega_m) = \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{|\bar{M}_{i,j} - \mu_m|^2}{2\sigma_m^2}}. \quad (13)$$

To cluster each element $\bar{M}_{i,j}$, we utilize four observations:

$$\mu_1 > \mu_2 \quad (14a)$$

$$\sigma_1 < \sigma_2 \quad (14b)$$

$$\delta_1 < \delta_2 \quad (14c)$$

Elements from the same column should be jointly clustered. $(14d)$

Constraint (14a) results because the reflection from a stationary object is expected to be more powerful than the reflection from clutter or a moving target. Constraint (14b) results because reflections from a stationary target are expected to remain stable. Constraint (14c) results because, across the K spread of the reflections, we expect to find more clutter than stationary targets. Finally, constraint (14d) results when a column is associated either with a stationary target or with clutter/a moving target.

To derive the EM algorithm, we formalize the log-likelihood function of the elements in \bar{M} . Let $\lambda_i = \{\bar{M}(1, i), \dots, \bar{M}(N^{\text{sig}}, i)\}$, $i = 1, \dots, K$ denote the i th column of \bar{M} . In addition, let Γ_i be the classifier for all the elements in λ_i such that

$$\Gamma_i = \begin{cases} 1 & \text{if column } \lambda_i \text{ is classed as a stationary target} \\ 2 & \text{otherwise} \end{cases}, \quad (15)$$

and let $\Gamma = \{\Gamma_1, \dots, \Gamma_K\}$. Based on (11), to label Γ_i , we estimate the posterior,

$$P(\Gamma_i = m | \lambda_i, \Omega) = \frac{\delta_m p(\lambda_i | \omega_m)}{p(\lambda_i | \Omega)} = \frac{\delta_m \prod_{d \in \lambda_i} p(d | \omega_m)}{\sum_{n=1}^2 \delta_n \prod_{d \in \lambda_i} p(d | \omega_n)}. \quad (16)$$

Note that (16) is somewhat different than the posterior for the traditional Gaussian mixture model (e.g., in [43]). This is because it encompasses the prior information, stating that elements from the same column should be classed together. Assuming a preliminary parameter set Ω^x , the posterior expectation of the complete-data log-likelihood function is

$$\begin{aligned} \mathcal{L}(\Omega | \Omega^x) &= E[\log P(\bar{M}, \Gamma | \Omega) | \bar{M}, \Omega^x] = \\ &= \sum_{m=1}^2 \sum_{i=1}^K P(\Gamma_i = m | \lambda_i, \Omega^x) \cdot \\ &= \left[\log \delta_m + \sum_{d \in \lambda_i} \log p(d | \omega_m) \right]. \end{aligned} \quad (17)$$

We find Ω by maximizing the log-likelihood function (17).

2) Estimating the Prior:

Considering constraint (12), we use a Lagrange multiplier γ , and obtain δ_m by solving

$$0 = \frac{\partial}{\partial \delta_m} \left[\sum_{m=1}^2 \sum_{i=1}^K \log(\delta_m) P(\Gamma_i = m | \lambda_i, \Omega^x) + \gamma \left(\sum_{m=1}^2 \delta_m^x \right) \right] = \sum_{i=1}^K \frac{P(\delta_i = m | \lambda_i, \Omega^x)}{\delta_m} + \gamma. \quad (18)$$

Since $\sum_{i=1}^K P(\Gamma_i = m | \lambda_i, \Omega^x) + \gamma \delta_m = 0$, it follows that $\gamma = -K$, and therefore,

$$\delta_m = \frac{1}{K} \sum_{i=1}^K P(\Gamma_i = m | \lambda_i, \Omega^x). \quad (19)$$

3) Estimating the distribution parameters:

To estimate ω_m , we consider only the relevant parts of the first term of (17),

$$\begin{aligned} \ddagger(\mu_m, \sigma_m) &= \sum_{i=1}^K P(\Gamma_i = m | \lambda_i, \Omega^x) \cdot \\ &= \sum_{d \in \lambda_i} \left(-\log(2\sigma_m) - \frac{|d - \mu_m|^2}{2\sigma_m^2} \right). \end{aligned} \quad (20)$$

Then, based on the constraints in (14), ω_m is found by solving the following optimization problem:

$$\omega_m = \underset{\mu_n, \sigma_n}{\text{argmin}} - \sum_{n=1}^2 \ddagger(\mu_n, \sigma_n) \quad (21a)$$

$$\text{s.t.} : \mu_1 > \mu_2, \quad (21b)$$

$$\sigma_1 < \sigma_2, \quad (21c)$$

$$\delta_1 < \delta_2, \quad (21d)$$

which is a convex optimization problem solved through interior point methods (cf. [44]).

4) EM Operation:

The EM algorithm is used prior to the Viterbi algorithm in Section IV-A, and its purpose is only to identify stationary targets. The EM is used directly on the TD matrix. To execute the EM algorithm, we first form matrix \bar{M} in (10) and then estimate Ω iteratively. Specifically, based on Ω^{x-1} from iteration $x-1$, we estimate the new δ_m^x and ω_m^x through (19) and (21), respectively. The iterations stop once parameters converge or after a predefined number of iterations, X . Following convergence, we calculate the posterior (16) and cluster each column of the TD matrix by the class that maximized the posterior, $P(\Gamma_i = m | \lambda_i, \Omega^X)$. Formally, the decision is made by

$$\hat{\Gamma}_i = \begin{cases} 1 & \text{if } P(\Gamma_i = 1 | \lambda_i, \Omega^X) > P(\Gamma_i = 2 | \lambda_i, \Omega^X) \\ 2 & \text{otherwise} \end{cases}. \quad (22)$$

Since the complexity of the interior point optimization is polynomial and that of the EM algorithm is $\mathcal{O}(N^{\text{sig}} \cdot K)$, the overall complexity of our method for finding stationary targets is $\mathcal{O}((N^{\text{sig}} \cdot K + K^2) \cdot X)$ (cf. [43]).

Finally, to prevent the Viterbi algorithm from locking onto strong stationary reflections, we must remove those L columns of \bar{M} clustered as 'stationary targets', i.e., for which $\hat{\Gamma}_i = 1$. Still, due to the constraint in (5) that allows transition only between a maximal number of columns, simply deleting these columns would change the TD matrix and may affect the tracking solution. Hence, as an alternative, we keep the same matrix size by replacing columns identified as stationary targets with replicas of L randomly chosen columns identified as 'non-stationary targets'. The result is a filtered matrix \bar{M} . Note that by replacing columns we may distort the path of a mobile

target crossing a column identified to include reflections from a stationary target. Yet, since the target is assumed moving, this intersection should occur only in a small number of rows of matrix M . Then, since the Viterbi algorithm uses a sequence of observations rather than single ones, the effect is limited.

To initialize Ω^0 , we cluster the elements of \bar{M} using the K-means clustering algorithm [43]. Since K-means does not support constraints that will ensure clustering elements of the same column together, we perform clustering per column of \bar{M} using the following heuristics. We start by separately clustering the elements of each column λ_i , $i = 1, \dots, K$ into two possible classes. Then, we make the column labelling decision based on the majority vote, and statistically evaluate the mean and variance of the inspected elements. Following this operation, we average the measured mean and variance of columns equally labelled, and consider these average values for Ω^0 . Here, we order the classes based on relation (14b). We note that in the cases where no column is labelled as a stationary target, the fallback operation is to simply cluster all of the elements of \bar{M} . We justify this decision by assuming that in any deployment of an omni-directional transducer, reflections from some stationary targets will always exist.

D. Estimating the Target's Size

Once the target's path, \hat{t} , has been determined, we evaluate its size. Besides the benefit of exploring the target's size, e.g., for ecological purposes, the size of the target can help validate detection. First, the number of samples for a valid target's size should be greater than τ_1 , and second, following the target's motion, the estimated size vector for all time instances should be smooth.

Since the acoustic emissions to and from the target traverse through the fading two-way channel transmitter-to-target and target-to-receiver, and since the target's reflection surface may not have a constant acoustic target strength, reflections related to the target may not be smooth. An example recorded in a sea experiment for reflections related to a scuba diver is shown in Fig. 3. The estimated position of the target appears as a circle. Around the estimated position, strong fading is observed. Moreover, due to the strong clutter, the structure of the target's reflection does not differ much from that of clutter-related reflections. The task of defining the reflection's width is therefore a challenge.

Without pre-knowledge about the target's reflection pattern, we estimate its size statistically. Let v_i be a vector of samples corresponding to the reflected signal at time instance i such that

$$v_i = [M(i, \hat{t}(i) - \tau_2), \dots, M(i, \hat{t}(i) + \tau_2)] . \quad (23)$$

That is, v_i encompasses the reflection samples around the estimated target position, and recall we use τ_2 as a rough upper bound for the target's size. An example of v_i is given in Fig. 3. The example shows a section of the normalized MF output (blue line) that was identified as a reflection from a target. The black circle marks the position found by the Viterbi algorithm, and the black 'X's mark the positions used for size estimation. The example shows that the task of defining the target's size

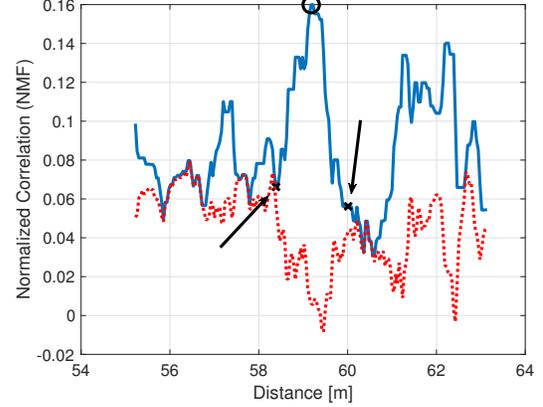


Fig. 3: A reflection associated with a scuba diver. The curve marked in red is the reflection's envelope. The black circle represents the estimated location. The black 'X's represent the positions chosen for the size estimation.

is challenging. Due to the non-homogeneous reflection plane and the fading underwater acoustic channel, the reflection pattern from the target is not homogeneous, and it is hard to distinguish the clutter reflections from the target and determine the latter's width. Still, we consider the target's reflection as a low frequency component that modulates higher frequency components of multipath reflections, fading, etc. Our solution is then obtained by calculating the lower envelope (cf. [45]) of v_i , denoted by \tilde{v}_i . We calculate the latter by the magnitude of the analytic signal, which we obtain by the Hilbert transform. The result is a smoothed version of v_i , which ideally crosses the original signal v_i at the edges of the reflection peak. The estimate of the target's size, \hat{w}_i , can then be obtained by the difference between the zero-crossing points of $v_i - \tilde{v}_i$ around the target's position $\hat{t}(i)$, while tolerance is defined by the user. Formally,

$$\bar{w}(i) = l_1 - l_2 , \quad (24a)$$

$$l_1 = \underset{n}{\operatorname{argmin}} |v_i(n) - \tilde{v}_i(n)|, \quad n > \hat{t}(i) , \quad (24b)$$

$$l_2 = \underset{n}{\operatorname{argmin}} |v_i(n) - \tilde{v}_i(n)|, \quad n < \hat{t}(i) , \quad (24c)$$

where (24b) and (29) hold element-wise absolute operations for a given index n . The result is a vector, \bar{w} , of size estimations, $\bar{w}(i)$. This somewhat heuristic approach may fail if the clutter is stronger than the target-related reflection, but works well even for a low SCR level of 2 dB. Moreover, it has the advantage of finding the target's size with no prior information, and is thus robust to the target types as well as different sea conditions.

We now continue to fuse all of the found target's sizes, \bar{w} , to estimate a scalar size for the target, \hat{w} . The intuition is to set the size estimation, \hat{w} , as the average over \bar{w} . However, since the target moves during the data acquisition and, as a fish would often do, may change its orientation relative to the transceiver, e.g., from end-fire to broad-side, an average would not capture the true size of the target. For this reason, we estimate \hat{w} as a maximum over the set of size estimations. As opposed to the minimum value which may

capture the target when it faced the receiver by its end-fire, only the maximum value may hold realistic evaluation about the size of the target. In other words, by taking the maximum value we better approach the true size estimation, which corresponds to the case when the target faces the transceiver in its broad-side. Further, due to our rather heuristic approach of size estimation, some elements of \bar{w} are non-valid and may form outliers. Hence, before estimating \hat{w} , we identify time instances, $\hat{i}_1, \hat{i}_2, \dots$, during which tracking is assumed valid, and take a sample sequence $\bar{w}(\hat{i}_1), \bar{w}(\hat{i}_2), \dots$ with the corresponding indexes. Considering (8), a time instance i is determined valid if

$$P(\hat{t}(i)|\mathcal{H}^1) > \varphi. \quad (25)$$

The target's size is then obtained by

$$\hat{w} = \max(\bar{w}(\hat{i}_1), \bar{w}(\hat{i}_2), \dots), \quad (26)$$

to yield a single scalar value.

E. Algorithm Summary & Discussion

Algorithm 1 Pseudo-Code for our tracking algorithm.

- 1: Perform NMF (2), and stack NMF responses to form \mathbf{M}
 - 2: Filter clutter using the threshold in (3)
 - 3: Smooth \mathbf{M} via median filter of window size τ_1
 - 4: Initialize Ω^0 using K-means
 - 5: **for** $x := 1$ to X (or convergence) **do**
 - 6: Calculate the log-likelihood (17)
 - 7: Estimate δ_m (19)
 - 8: Estimate ω_m (21)
 - 9: Estimate the posterior (16)
 - 10: **end for**
 - 11: Estimate the posterior $P(\Gamma_i = m | \lambda_i, \Omega^X)$
 - 12: **if** $P(\Gamma_i = 1 | \lambda_i, \Omega^X) > P(\Gamma_i = 2 | \lambda_i, \Omega^X)$ **then**
 - 13: $\Gamma_i = 1$
 - 14: **else**
 - 15: $\Gamma_i = 2$
 - 16: **end if**
 - 17: Remove all columns for which $\Gamma_i = 1$ to construct the filtered matrix $\tilde{\mathbf{M}}$
 - 18: Contract transition matrix \mathbf{A} based on (5)
 - 19: Run Viterbi algorithm on the filtered matrix $\tilde{\mathbf{M}}$ and obtain track $\hat{\mathbf{t}}$
 - 20: Validate detection by (8)
 - 21: **for** $i := 1$ to N^{sig} **do**
 - 22: Compare \mathbf{v}_i from (23) with $\tilde{\mathbf{v}}_i$
 - 23: Estimate target's time varying width \bar{w} by (24)
 - 24: **end for**
 - 25: Filter vector \bar{w} to obtain the target's size estimate \hat{w} by (26)
-

We now summarize the operation of our detection algorithm. Referring to the pseudo-code in Algorithm 1, we start by forming the TD matrix \mathbf{M} (lines 1-2). Next, we smooth the signal (line 3), and initialize Ω^0 (line 4). The EM operation is performed iteratively (lines 5-11) by calculating the log-likelihood (line 6), and deriving the prior (line 7), distribution

parameters (line 8) and posterior (line 11). The clustering of stationary targets is then performed (lines 12-16), followed by the construction of the filtered matrix $\tilde{\mathbf{M}}$ (line 17). Then, the Viterbi algorithm is activated to find the target's path (lines 18-19), and a detection decision is obtained (line 25). Finally, the reflection size, \hat{w} , is estimated by comparing \mathbf{v}_i with its lower envelope signal, $\tilde{\mathbf{v}}_i$ (lines 21-24) and filtering the resulting vector of peak widths \bar{w} (line 25).

V. PERFORMANCE EVALUATION

We now explore the performance of our track-before-detect approach in both numerical simulations and in multiple sea experiments. Simulations results are compared to the Cramér-Rao lower bound, while performance in sea experiments is compared to the ground truth of the target's location. We also provide results for a baseline algorithm that performs tracking using an unconstrained Viterbi algorithm. That is, a tracking solution that does not consider the limitations in (5) but allows the path to freely propagate between states. For both the simulation and sea experiments, we use $F_s = 10$ [kHz] (after baseband conversion) and $c = 1530$ [m/s].

A. Simulations

1) Simulation Model:

Since our detection method focuses on the analysis of the TD matrix \mathbf{M} , in our simulations we avoid preparing a full simulation scenario with simulated acoustic signals. Instead, relying on our eight sea experiments to analyze performance in realistic environments, we simulate \mathbf{M} directly. We start with a $[N^{\text{sig}} \times K]$ matrix $\mathbf{M}^{\text{clutter}}$ composed of i.i.d. zero mean Gaussian distributed samples of variance $\sigma_{\text{clutter}}^2$.

We consider reflection of time-varying energy, e^{target} , from a single target. We model the target to move with fixed velocity and an i.i.d. white Gaussian acceleration (but, recall, both the structure and parameters of the motion are considered unknown to the tracking algorithm), such that

$$\alpha(i) = [t(i), v(i), v(i-1)] , \quad (27a)$$

$$\alpha(i) = \mathbf{B}\alpha(i-1) + \mathbf{N}\mathbf{n}^\alpha(i) . \quad (27b)$$

where $t(0)$ is an initial location uniformly randomized to be $1, \dots, K-1$, $v(i)$ is the target's speed at time instance i uniformly randomized between $[-v^{\text{Max}}, v^{\text{Max}}]$, $\mathbf{n}^\alpha(i) = [n^{t(i)}, n^{v(i)}, n^{v(i-1)}]$ is a zero-mean Gaussian vector with covariance matrix $\mathcal{R}^{\text{model}}$, and the advance and noise matrices are

$$\mathbf{B} = \begin{bmatrix} 1 & \Delta_T & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} \frac{\Delta_T^2}{2} & 0 & 0 \\ \Delta_T & 0 & 0 \\ \Delta_T & 0 & 0 \end{bmatrix},$$

respectively, and Δ_T is the time resolution between each signal emission. The target's size, w , is set uniformly at random between 0.1 [m] and $\tau_1 c / F_s$, and the target is injected into $\mathbf{M}^{\text{clutter}}$ as a Gaussian window of size w , centred around locations $\hat{\mathbf{t}}$. This window simulates the appearance of a target's reflection at the output of the MF. To simulate stationary targets, we consider a uniformly random number $F \in [1, 5]$ of stationary targets, each of uniformly randomly

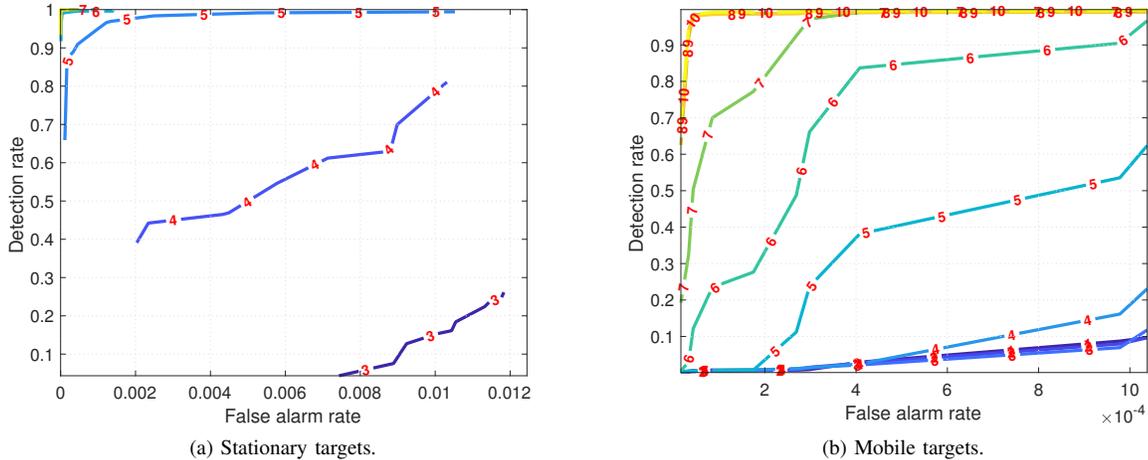


Fig. 4: ROC curves. Numbers on the contour lines indicate the SCR level in [dB]. Different pairs of false alarm-detection rate are obtained by changing the threshold in (8).

set size $a^{\text{stationary}}$ between 0.1 [m] and $\tau_1 c / F_s$ and time-varying energy $e^{\text{stationary}}$, at uniformly distributed distances $t_f^{\text{fix}}(0) \in \{1, \dots, K\}$, $f = 1, \dots, F$. To simulate a drifting motion, we use

$$t_f^{\text{fix}}(i) = t_f^{\text{fix}}(i-1) + n^{\text{fix}}, \quad (28a)$$

$$\text{s.t.} : |t_f^{\text{fix}}(0) - t_f^{\text{fix}}(i)| < \tau_2, \quad (28b)$$

i.e., the maximum drifting motion assumed is maintained, where n^{fix} is a zero mean i.i.d. Gaussian noise of variance σ_{fix}^2 .

The reflected time-varying energies, e^{target} and $e^{\text{stationary}}$ of the target and the stationary reflectors, respectively, are determined by a desired simulated SCR. We consider two definitions for the SCR: one for the mobile target, ρ^{target} , and one for the stationary reflectors, $\rho^{\text{stationary}}$. These are calculated by

$$\rho^{\text{target}} = E \left[\frac{K e^{\text{target}}(i)}{\sum_{k=1}^K M^{\text{clutter}}(i, k)} \right], \quad (29a)$$

$$\rho^{\text{stationary}} = E \left[\frac{K e^{\text{stationary}}(i)}{\sum_{k=1}^K M^{\text{clutter}}(i, k)} \right]. \quad (29b)$$

A Monte-Carlo set of 10,000 simulations is considered. In each simulation, we fix the SCR values, ρ^{target} and $\rho^{\text{stationary}}$, the clutter level $\sigma_{\text{clutter}}^2$, the width w , σ_{fix} , and the maximal absolute speed v^{max} , and randomize $t(0)$, $v(i)$, $i = 1, \dots, N^{\text{sig}}$, \mathbf{n}^α , F , $a^{\text{stationary}}$, and n^{fix} . In all cases, we considered $N^{\text{sig}} = 100$ signal emissions, which related to a measurement time of $T^{\text{measure}} = 50$ [s]. For both stationary targets/reflectors and mobile targets, results are measured in terms of the detection rate and false detection, P_d and P_{Fa} , respectively. We also analyze results in terms of the width estimation error rate, ρ_w ,

and in terms of the track estimation error rate, ρ_t , defined as

$$\rho_w = \frac{|\hat{w} - w|}{w}, \quad (30a)$$

$$\rho_t = \frac{\sum_n (\hat{t}(n) - t(n))}{\sum_n t(n)}, \quad (30b)$$

where $\hat{t}(n)$ and $t(n)$ are the n th element of $\hat{\mathbf{t}}$ and \mathbf{t} , respectively, and ρ_t is a time rate related to the measurement time, T^{measure} .

While many track-before-detect benchmarks exist, an initial study showed that none of these can manage the strong stationary targets. Instead, we compare our simulation results with the CRLB. The derivations for the CRLB are given in Appendix A.

2) Simulation Analysis:

We start by analyzing the capability of our approach to detect stationary targets. Fig. 4 shows the receiver operating characteristic (ROC) for the detection of stationary targets (Fig. 4a), and for the detection of mobile targets (Fig. 4b). The figures show the probability of detection (y-axis) vs. the probability of false alarm for different SCR levels, $\rho^{\text{stationary}}$ and ρ^{target} from (29a) for stationary targets and from (29b) for mobile targets, respectively. The latter being represented by the different contour lines. Each contour line is obtained by changing the set detection threshold, φ from (8) to obtain different detection performances. Then, threshold φ is set as a tradeoff between the detection and false alarm performance from the ROC for a certain SCR level. To form Fig. 4a, we consider detection as the identification of a stationary target in-between the

$$[t^{\text{fix}}(0) - a^{\text{stationary}}, t^{\text{fix}}(0) + a^{\text{stationary}}]$$

range, and a false detection as identification of such a target outside this range. We observe a good trade-off between P_d and P_{Fa} starting from $\rho^{\text{stationary}} = 5$ [dB]. Above $\rho^{\text{stationary}} >$

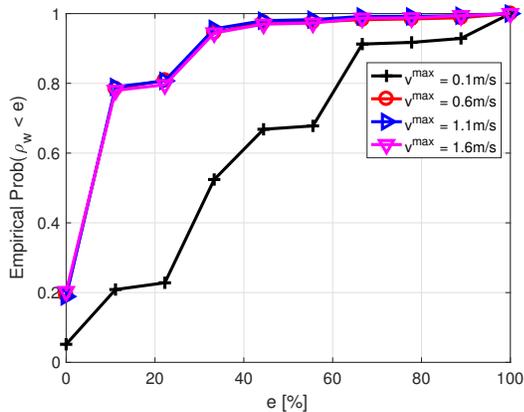


Fig. 5: Empirical cumulative distribution function (CDF) of ρ_w from (30a). y-axis shows the probability of ρ_w being smaller than e , and x-axis shows e [%]. Results obtained for $\rho^{\text{target}} = 8$ [dB].

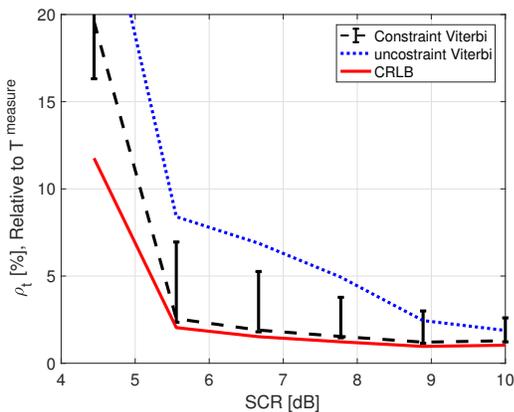


Fig. 6: Tracking error rate ρ_t as a function of the SCR. $T^{\text{measure}} = 50$ [s]. $v^{\text{max}} = 0.6$ [m/s].

7 [dB], results show almost false-less detection with minimal false alarms.

ROC results for the mobile target are shown in Fig. 4b. Here, a false alarm is determined when no target exists, but the detection rule in (8) determined \mathcal{H}^1 . The ROC is determined by changing the threshold level φ . A good trade-off is obtained for low SCR of $\rho^{\text{target}} = 6$ [dB], and for SCR $\rho^{\text{target}} = 8$ [dB], only a few miss detections are observed.

Next, in Fig. 5, we explore the cumulative distribution function (CDF) of the size estimation error, ρ_w (30a), for different values of the maximal speed, v^{max} . We test performance for $\rho^{\text{target}} = 8$ [dB], and for targets of maximal size $w < 1$ [m]. As we change v^{max} , we observe stability in the results for $v^{\text{max}} \geq 0.6$ [m/s]. This is because, below a certain maximal speed, for some time instances, the target may be mistaken to be stationary. Still, for all values explored, we observe that in most cases, the error is less than 50%. When the target is moving at maximal speed $v^{\text{max}} \geq 0.6$ [m/s], the error in most cases is below 10%, which corresponds to an absolute size estimation error of 0.1 [m].

In our last simulation analysis, we explore the average

tracking error rate, ρ_t (30b). Results in Fig. 6 show the mean and 90% confidence interval for ρ_t as a function of the SCR ρ^{target} and for $v^{\text{max}} = 0.6$ [m/s]. Compared to the performance of the no-constraint Viterbi algorithm, we observe that the constraint solution performs much better, with a gain spanning from roughly 0.6% at the high SCR levels and more than 10% at the lower ones. This is because setting limitations over the path, the Viterbi algorithm helps mostly when the clutter is as strong as the reflection. Results are also compared with the CRLB (36). For a low SCR of 6 [dB], we observe a small error of less than 2% for the estimated track. Results are also extremely close to the CRLB. The small spread of the confidence interval shows that this result is stable throughout the simulations.

B. Sea Trials

1) Experiment Setup:

To properly evaluate the performance of our tracking algorithm in a realistic environment, we performed eight sea experiments. Seven experiments were performed in the Mediterranean Sea (MS), and one experiment was performed in the Red Sea (RD). Four experiments included tracking scuba divers (with closed and open circuit systems) and four experiments tracked a sandbar shark (*Carcharhinus plumbeus*). The details for each of the experiments are summarized in Table II, and bathymetric maps of the explored areas of all experiments but D1 (whom we couldn't obtain the data) are shown in Fig. 7. A variety of sea environments were explored. Experiments D1 and D2 were conducted in open water with a sandy bottom; Experiments D3 and D4 were conducted in an open water with a reef environment with a rocky bottom and shallow water; and Experiments S1-S4 were conducted in a sheltered harbour environment with a sandy bottom and shallow water.

To track scuba divers, we deployed a single acoustic transducer from a vessel (experiments D2 and D3) and from a buoy (experiments D1 and D4). The position of the deploying platform was recorded via a GPS receiver, while the position of the scuba diver was measured by letting the scuba divers tow a surface buoy with a GPS receiver. In all experiments D1-4, the scuba divers performed a random motion, starting from the deploying vessel. To test the system's performance for shark tracking, we caught a 1.8 m sandbar shark using regular fishing gear². See image from this experiment in Fig. 10. We then tracked the shark from three directions, while it was connected to the fishing buoy, which had a GPS receiver, and upon its release. We note that since sharks do not possess a swimming bladder, their acoustic target strength is low, and thus the SCR is low. The operations for shark tracking took place in Hadera harbour, Israel, next to a power plant facility, where the sharks gather seasonally.

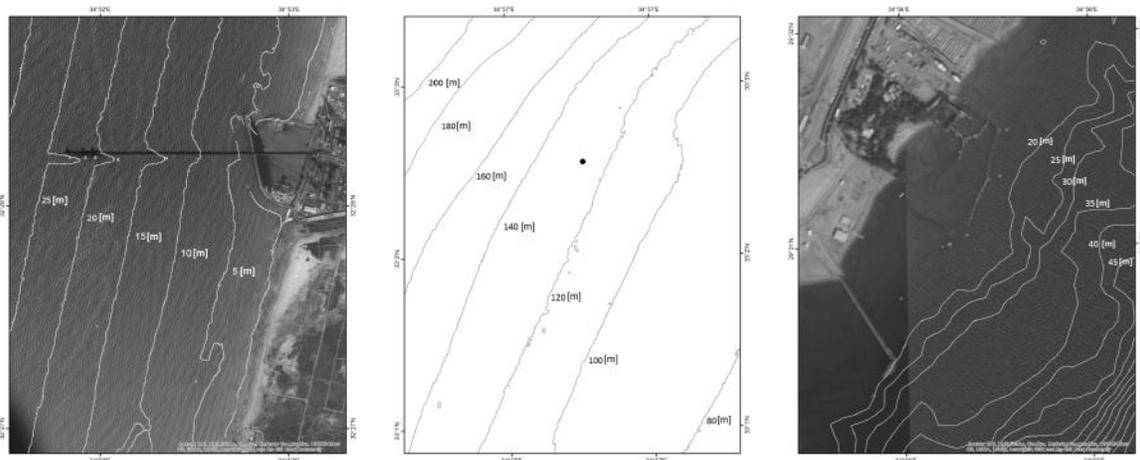
The reflection level (RL) follows the expression,

$$\text{RL} = \text{SL} - 2\text{TL} + \text{TS}, \quad (31)$$

²The fishing of sharks was performed under the approval and supervision of the Israeli Nature and Parks Authority. After tagging, all sharks were released with no harm caused.

TABLE II. Details of the sea experiments.

Exp. Tag	Location	Date	Sea State	Target Type	Water Depth	Transducer Depth	Target Depth	Target Max. Distance	Target Max. Speed
D1	MS; Achziv	Apr. 17	1	One diver	15 [m]	5 [m] (buoy)	10 [m]	180 [m]	0.5 [m/s]
D2	RS; Eilat	Mar. 18	0	Two divers (closed circuits)	18 [m]	10 [m] (buoy)	10 [m]	200 [m]	0.8 [m/s]
D3	MS; Hedera	Apr. 18	1	One diver	20 [m]	2 [m] (boat)	10 [m]	350 [m]	1 [m/s]
D4	MS; Naharia	Oct. 17	2	Two divers	120 [m]	10 [m] (boat)	10 [m]	250 [m]	0.3 [m/s]
S1	MD; Hadera	Apr. 18	1	One shark (buoy attached)	18 [m]	10 [m] (boat anchored)	10 [m]	200 [m]	2 [m/s]
S2	MD; Hadera	Apr. 18	1	One shark (buoy attached)	18 [m]	10 [m] (boat motoring)	10 [m]	200 [m]	2 [m/s]
S3	MD; Hadera	Apr. 18	1	One shark (buoy attached)	18 [m]	10 [m] (boat drifting)	10 [m]	200 [m]	2 [m/s]
S4	MD; Hadera	Apr. 18	1	One shark (free swimming)	18 [m]	10 [m] (boat drifting)	10 [m]	200 [m]	5 [m/s]



(a) Mediterranean Sea, Shallow water. Exp D3, S1-S4. (b) Mediterranean Sea, Deep water. Exp D4. (c) Red Sea, shallow water. Exp D2

Fig. 7: Bathymetry maps of the explored areas.

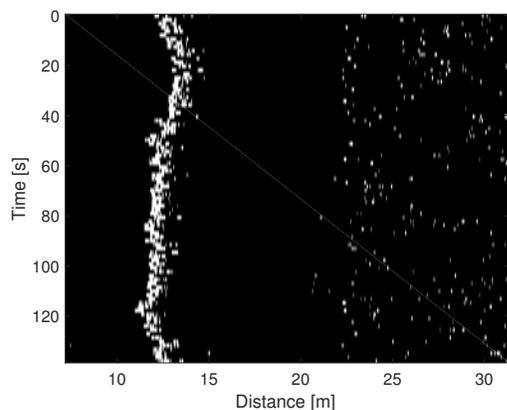


Fig. 8: Example of a detected stationary target (Exp D4). Normalized matched filter outputs below 0.5 are neglected to emphasize reflections from stationary targets.

where SL is the source level, TS is the target strength, and TL is the transmission loss, which is counted twice due to the round trip of the acoustic signal from source \rightarrow to target \rightarrow to receiver. For transmission range η [m], the TL can be simplified as

$$TL = 10k \log_{10}(\eta) + a \cdot r/1000, \quad (32)$$

where a [dB/km] is the absorption loss, and k can be bounded between a spherical propagation loss, $k = 2$, and a cylindrical one, $k = 1$. To ensure the safety of the scuba divers and of the captured shark, transmissions were made with a low SL of 165 [dB Re 1uPa@1m]. Considering a carrier frequency of 12 [kHz] that leads to $a = 0.9$ [dB/km], and given a TS of -20 [dB] [46] and a distance of 200 [m] from source to target, we expect an RL of roughly 100 [dB Re 1uPa]. In all experiments, the transmitted signals consisted of a sequence of $N^{\text{sig}} = 100$ linear chirp signals of 10 ms duration and 7 [kHz] -17 [kHz] frequency band. The signals were spaced by a guard interval of 0.7 s considering a detection range of roughly 700 [m]. The signals were transmitted from an EvoLogics S2C R 7/17

software defined modem that was used as an analog front end, and received by a Cetacean uRadar II self-recorder system. The recording system was deployed next to the projector, and both systems were omni-directional. Due to the nature of the recorder, in all experiments, analysis was made off-line.

The experiments involved great effort and investment in data collection, equipment, logistics, and the approval of the relevant authorities. For reproducibility, we share the obtained TD matrices from the eight experiments ³.

2) Experiment Results:

We start by showing an example for the detection of stationary targets. Results in Fig. 8 show a zoom-in on the TD matrix around the position of a stationary target observed in experiment D4. Two dark dotted lines mark the area identified by the algorithm to include a stationary target. We observe that in spite of reflections' power and delay fluctuations, a good match is obtained. We note that both when the scuba divers used closed breathing systems (re-breathers) and open systems, we could not detect them by their emitted noise. Moreover, comparing the spectrum of the reflection patterns for both system types, we could see no difference. We thus conclude that the effect of bubbles exhaled from the scuba divers with open circuits has no observable effect on the reflection pattern.

Results of scuba diver tracking and size estimation for experiments D1-D4 are given in Fig. 9. The upper figure shows the TD matrix; The middle figure shows the estimated track; and the bottom figure shows the vector $\hat{\mathbf{w}}$ (24) and the scalar estimate \hat{w} (26). Note that \hat{w} is obtained from the maximum of all validated size estimation results. Hence, since some outliers are detected, the obtained \hat{w} does not show the maximum over $\hat{\mathbf{w}}$. Since only a single transceiver is used, all TD matrices include a high clutter level, as well as strong observed reflections from stationary targets. In Exp D1 and Exp D2, the latter fits the location of the buoy's chain; in Exp D3, the fixed lines (i.e., reflections from stationary targets) matches the position of nearby pier; and in Exp D4, the locations match the boat's anchor. In all but for Exp D4, we observe that the SCR is very low and the scuba diver's path is hardly visible. Still, a close observation reveals accurate path findings. This conclusion is supported by comparing the estimated path to the GPS-based ground truth. The maximum and average absolute error is reported in the middle figure for each experiment.

Given the low SCR values, the results in Fig. 9 can confirm the simulation results in Fig. 6, which, for SCR=6 [dB] show an error of roughly 5%. Moreover, similar to the simulation results, comparing performance to the baseline unconstrained Viterbi algorithm, we observe that at low SCR levels setting the constraints in (5) allows proper tracking. Analyzing the size estimating results, we note that in experiments D1 and D3 the scuba divers mostly swam from the boat and were detected from the end-fire; while in experiments D2 and D4 the scuba divers changes their angle relative to the transceiver. Moreover, since D2 and D4 involved two scuba divers, the

size of the reflecting surface changes over time, relative to the distance between the scuba divers. As a result, we observe a wider spread in size estimations for D2 and D4. Nevertheless, considering an average length of a scuba diver of 1.8 [m], the results of \hat{w} show only a small error (in D4, the captured size is for the two scuba divers together). Since the scuba divers in experiments D2 and D4 swam roughly 1 or 2 [m] apart, we could not separate between them. This is also shown by the size estimation results shown in Fig. 9.

Results for experiments S1-S4 are shown in Fig. 11. The structure of the figure is similar to Fig. 9, except for Exp S4, where no ground truth for the shark's location was available. In all cases, no visible path is observed in the TD matrices, and the SCR was extremely low. Furthermore, results show strong stationary targets in the first 20 m range. Still, in all cases, detection was successful. The formed path shows higher errors than the scuba diver experiment. We believe this is mostly due to the long line of roughly 20 m that connected the shark to the buoy that held the GPS receiver. Hence, in contrast to the case of the scuba divers, here the location of the buoy did not accurately represent the shark's location. Yet, the low error-to-distance rate allows us to argue that the observed path is not random. We also observe that, due to the low SCR level, the performance gain compared to the baseline unconstrained Viterbi algorithm is higher. In all experiments, the estimated target's size was roughly 1.5 [m]. Comparing this length to the measured total length of the shark (1.8 [m]), and noting that, due to its low mass, the shark's tail has a low acoustic target strength, we argue for accurate size estimation.

Similar to our numerical simulations, in terms of target detection and tracking, our experimental results showed accurate tracking performance for both slow (scuba diver) and fast (shark) mobile targets in different sea environments. In terms of size estimation, since modelling the reflection width is a difficult task, the results show occasional inaccuracies sometimes twice the actual size. Still, both in the simulations and experiments, we observe that the determined size value, obtained as the maximum over the self-validated sequence of size estimates, fits the ground truth results. We therefore conclude that our tracking method is able to robustly track mobile targets in a realistic sea environment composed of stationary reflections and strong clutter.

VI. CONCLUSIONS

In this paper, we presented a low-complexity tracking algorithm for submerged mobile targets in the presence of strong clutter and stationary reflectors. The targets considered are both slow-moving ones, like scuba divers, and fast-moving such as marine predators. Our method is based on transmitting a sequence of short wideband signals and forming a time-distance (TD) matrix. To manage only mobile targets, our algorithm first detects stationary targets by clustering the TD matrix using a constraint EM procedure. Then, considering reflections as states, we track the target using a constraint Viterbi algorithm matched to some bounds (if known) about the target's motion pattern. Detection is then formed by maximum likelihood. We also presented a heuristic approach

³Data can be downloaded from https://drive.google.com/drive/folders/1WqooS8jXGuOSZcEb-59_LZRe0ACUJWjI?usp=sharing.

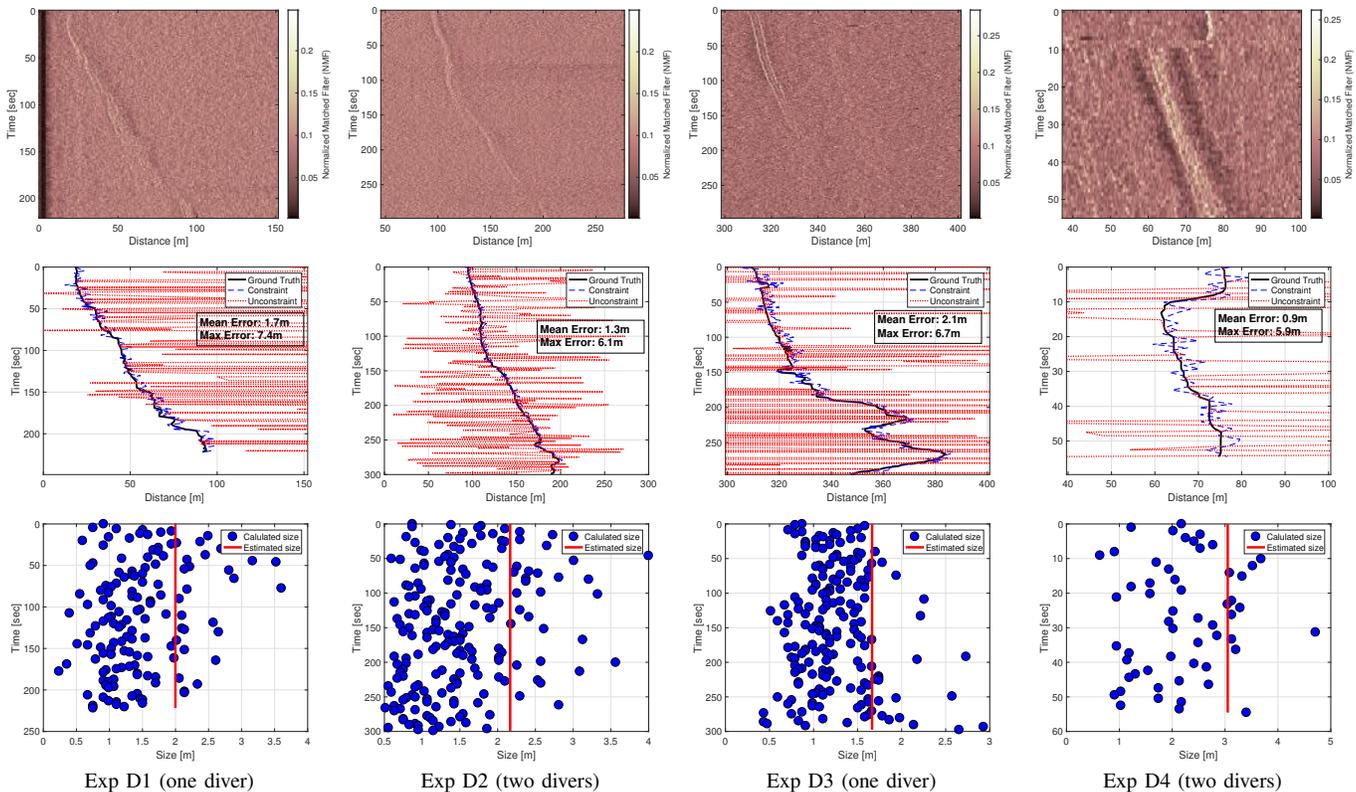


Fig. 9: Results for diver detection: TD matrix after removal of reflections from stationary targets (upper row); Obtained path (middle row); estimated width \hat{w} , fixed black line shows \hat{w} (26) (bottom row). SCR levels are 6 [dB] (D1), 7 [dB] (D2 and D3), 8 [dB] (D4). In all cases, signal-to-noise ratio was above 20 [dB]. SCR levels are listed on the caption of the subfigures. Matrix elements are shown in units of normalized matched filter (NMF).



Fig. 10: Sandbar shark prior to release and acoustic sampling.

to estimate the size of targets, once tracked, by comparing reflections with their envelope. We tested our algorithm in both extensive simulations and in eight sea experiments of different sea environmental settings and for different targets - including scuba divers with open and closed circuits rebreathers and a sandbar shark of low acoustic target strength. The results show

accurate tracking and detection, as well as fine-size estimation in the presence of a very low signal-to-clutter ratio and in the presence of multiple stationary reflectors. Further work will extend this work for detection of multiple targets.

REFERENCES

- [1] M. Greenberg, P. Chalk, H. Willis, I. Khilko, and D. Ortiz, *Maritime terrorism: Risk and liability*. Rand Corporation, 2006.
- [2] N. O. Bakir, "A brief analysis of threats and vulnerabilities in the maritime domain," in *Managing critical infrastructure risks*. Springer, 2007, pp. 17–49.
- [3] B. Wilson, "Maritime energy security," An ARW Book. IOS Press <http://www.ensecoc.org/download/126/maritimeenergysecurity.pdf>, 2012.
- [4] A. Bertrand and E. Josse, "Acoustic estimation of longline tuna abundance," *ICES Journal of Marine Science*, vol. 57, no. 4, pp. 919–926, 2000.
- [5] T. Letessier, P. Bouchet, and J. Meeuwig, "Sampling mobile oceanic fishes and sharks: implications for fisheries and conservation planning," *Biological Reviews*, vol. 92, no. 2, pp. 627–646, 2017.
- [6] M. J. Parsons, I. Parnum, K. Allen, R. McCauley, and C. Erbe, "Detection of sharks with the Gemini imaging SONAR," *Acoustics Australia*, vol. 42, no. 3, p. 0, 2014.
- [7] L. Fillinger, A. Hunter, M. Zampolli, and M. Clarijs, "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. 310–316, 2012.
- [8] V. N. Hari, M. Chitre, Y. M. Too, and V. Pallayil, "Robust passive diver detection in shallow ocean," in *OCEANS*, May 2015, pp. 1–6.
- [9] K. J. Sangston and K. R. Gerlach, "Coherent detection of radar targets in a non-Gaussian background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 2, pp. 330–340, April 1994.

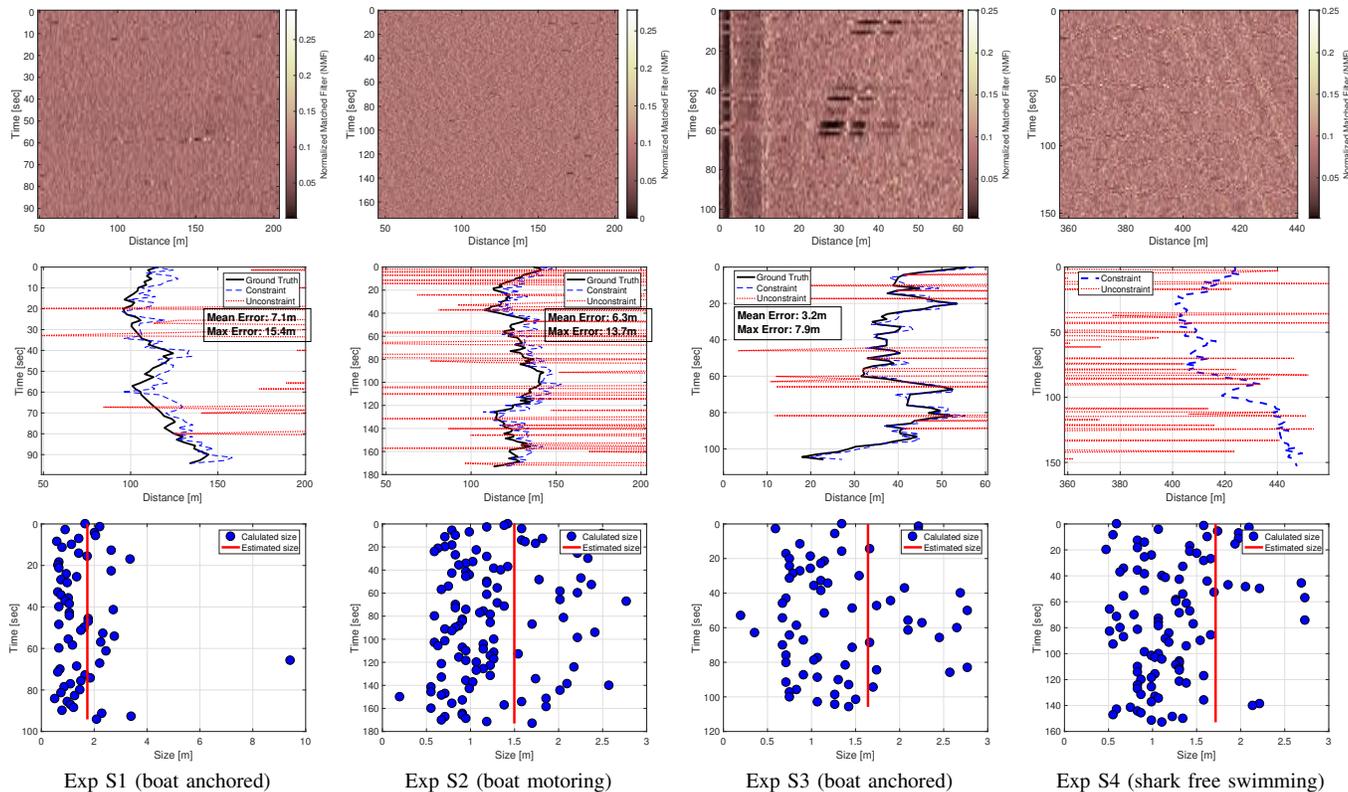


Fig. 11: Results for shark detection: TD matrix after removal of reflections from stationary targets (upper row); obtained path (middle row); estimated width \hat{w} , fixed black line shows \hat{w} (26) (bottom row). In all cases, signal-to-noise ratio was above 20 [dB], and SCR levels are 1 [dB] or below. Matrix elements are shown in units of normalized matched filter (NMF).

- [10] J. Wang, A. von Trojan, and S. Lourey, "Active sonar target tracking for anti-submarine warfare applications," in *IEEE OCEANS*, May 2010.
- [11] T. Northardt and S. Nardone, "Track-before-detect bearings-only localization performance in complex passive SONAR scenarios: A case study," *IEEE Journal of Oceanic Engineering*, 2018.
- [12] J. Soli, G. Hickman, and J. Krolik, "Range-velocity estimation bounds for nonuniform sonar comb signals," *IEEE Transactions on Aerospace and Electronic Systems*, 2018.
- [13] M. Wei, B. Shi, C. Hao, and S. Yan, "A novel weak target detection strategy for moving active SONAR," in *IEEE OCEANS*, May 2018.
- [14] J. Bates, D. Grimmer, G. Canepa, and A. Tesei, "Towards Doppler estimation and false alarm rejection for continuous active SONAR," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. 1972–1972, 2018.
- [15] K. Sangston and K. Gerlach, "Coherent detection of radar targets in a non-Gaussian background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 2, pp. 330–340, 1994.
- [16] C. Brown and A. Zoubir, "A nonparametric approach to signal detection in impulsive interference," *IEEE transactions on signal processing*, vol. 48, no. 9, pp. 2665–2669, 2000.
- [17] M. Desai and R. Mangoubi, "Robust Gaussian and non-Gaussian matched subspace detection," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3115–3127, 2003.
- [18] D. Brownrigg, "The weighted median filter," *Communications of the ACM*, vol. 27, no. 8, pp. 807–818, 1984.
- [19] M. Jansen, *Noise reduction by wavelet thresholding*. Springer Science & Business Media, 2012, vol. 161.
- [20] J. Renard, L. Lampe, and F. Horlin, "Sequential likelihood ratio test for cognitive radios," *IEEE Transaction on Signal Processing*, vol. 64, no. 24, pp. 6627–6639, 2016.
- [21] K. Lau, M. Salibian-Barrera, and L. Lampe, "Modulation recognition in the 868 Mhz band using classification trees and random forests," *AEU-International Journal of Electronics and Communications*, vol. 70, no. 9, pp. 1321–1328, 2016.
- [22] M. Desai, R. Mangoubi, J. Shah, W. Karl, H. Pien, A. Worth, and D. Kennedy, "Functional MRI activity characterization using response time shift estimates from curve evolution," *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1402–1412, Nov 2002.
- [23] D. Lerro and Y. Bar-Shalom, "Tracking with debiased consistent converted measurements versus EKF," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 3, pp. 1015–1022, Jul 1993.
- [24] S. Davey, M. Wienenke, and H. Vu, "Histogram-PMHT unfettered," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 435–447, June 2013.
- [25] S. Davey, M. Rutten, and B. Cheung, "A comparison of detection performance for several track-before-detect algorithms," in *International Conference on Information Fusion*, June 2008.
- [26] C. Jauffret and Y. Bar-Shalom, "Track formation with bearing and frequency measurements in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 6, pp. 999–1010, Nov 1990.
- [27] P. Willett and S. Coraluppi, "Application of the MLPDA to bistatic sonar," in *IEEE Aerospace Conference*, March 2005, pp. 2063–2073.
- [28] W. Blanding, P. Willett, and S. Coraluppi, "Sequential ML for multistatic sonar tracking," in *OCEANS*, June 2007, pp. 1–6.
- [29] R. Streit and T. Luginbuhl, "Maximum likelihood method for probabilistic multi-hypothesis tracking," in *Signal and Data Processing of Small Targets*, vol. 2235. International Society for Optics and Photonics, 1994, pp. 394–406.
- [30] R. Streit, M. Graham, and M. Walsh, "Multitarget tracking of distributed targets using histogram-PMHT," *Digital Signal Processing*, vol. 12, pp. 394–404, 2002.
- [31] H. Vu, S. Davey, F. Fletcher, S. Arulampalam, R. Ellem, and C. Lim, "Track-before-detect for an active towed array SONAR," in *Proceedings of Acoustics*, November 2013.
- [32] H. Gaetjens, S. Davey, S. Arulampalam, F. Fletcher, and C. Lim, "Histogram-PMHT for fluctuating target models," *IET Radar, Sonar Navigation*, vol. 11, no. 8, pp. 1292–1301, 2017.
- [33] S. Schoenecker, P. Willett, and Y. Bar-Shalom, "ML-PDA and ML-PMHT: Comparing multistatic sonar trackers for VLO targets using a

- new multitarget implementation,” *IEEE Journal of Oceanic Engineering*, vol. 39, no. 2, pp. 303–317, April 2014.
- [34] —, “Extreme-value analysis for ML-PMHT, part 1: threshold determination,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 2500–2514, October 2014.
- [35] S. Schoenecker, T. Luginbuhl, P. Willett, and Y. Bar-Shalom, “Extreme-value analysis for mlml-pmht, part 2: target trackability,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 2515–2527, October 2014.
- [36] S. Schoenecker, P. Willett, and Y. Bar-Shalom, “The effect of K-distributed clutter on trackability,” *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 475–484, Jan 2016.
- [37] —, “Resolution limits for tracking closely-spaced targets,” *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–1, 2018.
- [38] C. Jing, Z. Lin, and J. Li, “Detection and tracking of an underwater target using the combination of a particle filter and track-before-detect,” in *IEEE OCEANS*, April 2016, pp. 1–5.
- [39] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, Massachusetts: MIT Press, 2009.
- [40] R. Diamant and L. Chorev, “Emulation system for underwater acoustic channel,” in *International Undersea Defence Technology Europe conference (UDT)*, vol. 2, 2005, pp. 1043–1046.
- [41] R. Diamant, “Closed form analysis of the normalized matched filter with a test case for detection of underwater acoustic signals,” *IEEE Access*, vol. 4, pp. 8225–8235, 2016.
- [42] B. Schuster-Böckler and A. Bateman, “An introduction to hidden Markov models,” *Current protocols in bioinformatics*, vol. 18, no. 1, pp. A–3A, 2007.
- [43] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [44] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [45] M. Caetano and X. Rodet, “Improved estimation of the amplitude envelope of time-domain signals using true envelope cepstral smoothing,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 4244–4247.
- [46] Z. Zhang, “Diver detection sonars and target strength: review and discussions,” *International Conference on Sound and Vibration (ICSV)*, 2007.
- [47] P. Ruud, “Extensions of estimation methods using the EM algorithm,” *Journal of Econometrics*, vol. 49, no. 3, pp. 305–341, 1991.
- [48] P. Robertson, E. Villebrun, and P. Hoeher, “A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain,” in *IEEE International Conference on Communications (ICC)*, vol. 2. IEEE, 1995, pp. 1009–1013.
- [49] H. Trees, *Detection, Estimation, and Modulation Theory*, 1st ed. John Wiley and Sons, 2001.
- [50] P. Tichavsky, C. Muravchik, and A. Nehorai, “Posterior Cramer-Rao bounds for discrete-time nonlinear filtering,” *IEEE Journal of Signal Processing*, vol. 46, no. 5, pp. 1386–1396, May 1998.

APPENDIX

Since our tracking algorithm is a sequence of the EM algorithm and the Viterbi algorithm, which are both unbiased estimators [47], [48], we argue that also our scheme is unbiased. Hence, we can derive the Cramér-Rao lower bound (CRLB) for mobile target tracking to bound the performance of our scheme. We consider the following measurement vector

$$y(i) = \mathbf{g}(\alpha(i)) + \mathbf{n}^{\text{measure}}(i), \quad (33)$$

where $\mathbf{g}(\cdot)$ represents any observation model one chooses for the state vector, α , and $\mathbf{n}^{\text{measure}}(i)$ is a measurement noise with a covariance matrix $\mathcal{R}^{\text{measure}}$.

Let \mathbf{S} and \mathbf{Y} be the vectors of states $\alpha(i)$ and $y(i)$, respectively. For any unbiased estimator, the CRLB gives the lower bound on the covariance (cf. [49])

$$E \left[\left(\hat{\mathbf{S}} - \mathbf{S} \right) \left(\hat{\mathbf{S}} - \mathbf{S} \right)^{\text{T}} \right] \geq \mathbf{J}^{-1}, \quad (34)$$

where

$$\mathbf{J} = E \left[-\frac{\partial^2}{\partial^2 \mathbf{S}} \log P(\mathbf{S}, \mathbf{Y}) \right] \quad (35)$$

is the inverse of the Fisher information matrix with elements $J_{a,b}$, and $P(\cdot)$ denotes the probability density function. In [50], it was shown that if only estimation $\alpha(i)$ in vector \mathbf{S} is of interest, (35) can be formulated recursively such that

$$\mathbf{J}(i) = J_{1,i} - J_{2,i}^{\text{T}} (\mathbf{J}(i-1) + J_{3,i})^{-1} J_{2,i}, \quad (36)$$

where

$$J_{1,i} = -E \left[\frac{\partial^2}{\partial^2 \alpha(i)} \log P(\alpha(i) | \alpha(i-1)) \right] - E \left[\frac{\partial^2}{\partial^2 \alpha(i)} \log P(y(i) | \alpha(i)) \right] \quad (37a)$$

$$J_{2,i} = -E \left[\frac{\partial^2}{\partial \alpha(i-1) \partial \alpha(i)} \log P(\alpha(i) | \alpha(i-1)) \right] \quad (37b)$$

$$J_{3,i} = -E \left[\frac{\partial^2}{\partial^2 \alpha(i-1)} \log P(\alpha(i) | \alpha(i-1)) \right]. \quad (37c)$$

Recall both $\mathbf{n}^{\alpha}(i)$ from (27) and $\mathbf{n}^{\text{measure}}(i)$ from (33) are modeled to be zero-mean Gaussians with corresponding covariance matrices $\mathcal{R}^{\text{model}}$ and $\mathcal{R}^{\text{measure}}$, respectively. Observe $\frac{\partial \mathbf{g}(\alpha(i))}{\partial \alpha(i-1)} = \mathbf{B}$ from (27). Thus, introducing $\mathbf{G}(i) = \frac{\partial \mathbf{g}(\alpha(i))}{\partial \alpha(i)}$, (37) becomes

$$J_{1,i} = \left(\mathbf{N} \mathcal{R}^{\text{model}} \mathbf{N}^{\text{T}} \right)^{-1} + E \left[\mathbf{G}(i)^{\text{T}} \left(\mathcal{R}^{\text{measure}} \right)^{-1} \mathbf{G}(i) \right] \quad (38a)$$

$$J_{2,i} = \mathbf{B}^{\text{T}} \left(\mathbf{N} \mathcal{R}^{\text{model}} \mathbf{N}^{\text{T}} \right)^{-1} \quad (38b)$$

$$J_{3,i} = \mathbf{B}^{\text{T}} \left(\mathbf{N} \mathcal{R}^{\text{model}} \mathbf{N}^{\text{T}} \right)^{-1} \mathbf{B}. \quad (38c)$$

In our simulations we consider a direct measurement of the speed through, e.g., Doppler shift estimation, such that $\mathbf{G}(i)$ becomes a unity.