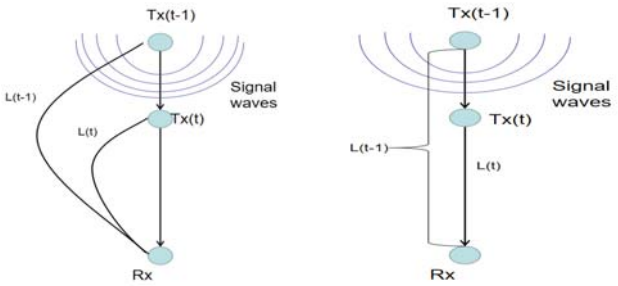
Underwater Motion and Activity Recognition using Acoustic Wireless Networks

Haochen Hu, Zhi Sun, and Lu Su
University at Buffalo, State University of New York, NY, USA, 14260
E-mail:{haochenh, zhisun, lusu}@buffalo.edu

***Abstract*— Underwater motion recognition using acoustic wireless networks has a promisingly potential to be applied to the diver activity monitoring and aquatic animal recognition without the burden of expensive underwater cameras which have been used by the image-based underwater classification techniques. However, accurately extracting features that are independent of the complicated underwater environments such as inhomogeneous deep seawater is a serious challenge for underwater motion recognition. Velocities of target body (VTB) during the motion are excellent environment independent features for WiFi-based recognition techniques in the indoor environments, however, VTB features are hard to be extracted accurately in the underwater environments. The inaccurate VTB estimation is caused by the fact that the signal propagates along with a curve instead of a straight line as the signal propagates in the air. In this paper, we propose an underwater motion recognition mechanism in the inhomogeneous deep seawater using acoustic wireless networks. To accurately extract velocities of target body features, we first derive Doppler Frequency Shift (DFS) coefficients that can be utilized for VTB estimation when signals propagate deviously. Secondly, we propose a dynamic self-refining (DSR) optimization algorithm with acoustic wireless networks that consist of multiple transmitter-receiver links to estimate the VTB. Those VTB features can be utilized to train the convolutional neural networks (CNN). Through the simulation, estimated VTB features are evaluated and the testing recognition results validate that our proposed underwater motion recognition mechanism is able to achieve high classification accuracy.**

I. INTRODUCTION

Acoustic wireless networks have a great potential to perform passive diver activity recognition and aquatic animal classification such as regalecus glesne and jellyfish in the deep seawater environment. Although there are several image-based underwater aquatic animals classification techniques such as [1][2], underwater cameras are expensive and are susceptible to underwater conditions such as light, transparency and depth. Poor seawater transparency and night time certainly deteriorate the quality of photos and limit the performance of the underwater camera. On the contrary, acoustic wireless sensing networks are more robust than underwater cameras in austere situations such as the dark deep seawater since the wireless signal won't be affected by the transparency of the underwater environments. Underwater acoustic wireless networks are able to perform 24 hours monitoring and avoid the influences of bad weather. Furthermore, with the communication devices that have been installed on the autonomous underwater vehicle (AUV), there is no extra cost of a hardware implementation for acoustic wireless sensing networks. In addition, the capability of wireless sensing networks has been shown in WiFi-based wireless sensing techniques. For example, wireless sensing techniques are able to perform passive motion recognition from small movements such as finger keystroke [3] to relatively drastic activities such as walking and falling activity recognition [4] in the indoor environments. Furthermore, indoor environment independent human activity recognition is developed in [5] by minimizing the influences of environments on the input features. However, those WiFi-based wireless sensing techniques cannot be directly utilized for underwater motion recognition.



(a) Doppler effect in the inhomogeneous underwater environment

(b) Doppler effect in the air

Fig. 1: Doppler effect when transmitter moves down to the receiver at a constant speed

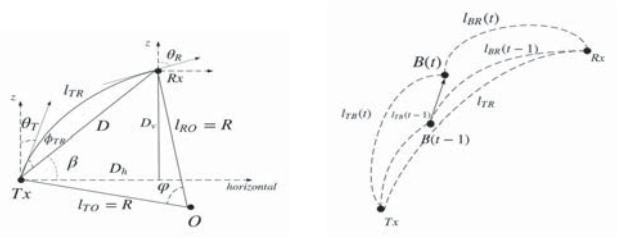
Due to the complicated influences of underwater environments, accurately estimating the environment independent features is still the main challenge for underwater target motion classification using acoustic wireless networks. Velocities of the target body (VTB) that are estimated from Doppler frequency shift values by DFS coefficients can be considered as environment independent features. However, VTB estimation in the inhomogeneous deep seawater faces four problems. Firstly, when the signal propagates through the inhomogeneous deep ocean water, the signal propagation path $L(t)$ is not a straight line as shown in Fig. 1(b) but an arc as shown in Fig. 1(a) from the transmitter to the receiver. That phenomenon is caused by the fact that acoustic signal propagation speed in the underwater environment becomes smaller with deeper depth before reaching the minimum sound speed depth. As shown in Fig.1, when the transmitter moves downwards at a constant speed, Doppler effect in the inhomogeneous underwater environment has an increasing DFS while Doppler effect in the air has a constant DFS. Thus, DFS coefficients in inhomogeneous deep seawater are different from the DFS coefficients in the air. Secondly, ambiguity exists in the VTB estimations. Doppler effects depend on both signal propagation speed and path length changing speed. In inhomogeneous underwater environments, if the signal propagation path length changing speed is equal to the acoustic signal velocity changing rate, there is no Doppler effect for one transmitter receiver pair (TRP) when the target

978-1-7281-5089-5/20/$31.00 ©2020 IEEE

moves. Different motion velocities may lead to the same DFS value for one TRP. Thirdly, VTB features extractions still need a suitable optimization algorithm to reduce the outliers. Fourth, if multiple transmitter receiver pairs are utilized to solve the ambiguity in VTB estimation, different receivers need to be synchronized. Time and frequency synchronization among all the receivers are important for the training data structure.

To solve the problems of underwater motion recognition, in this paper, we propose an underwater target motion recognition mechanism using acoustic wireless networks, which is able to estimate the velocities of target body components as features by dynamic self-refining optimization algorithm and underwater DFS coefficients. First, to solve the ambiguity of VTB estimation in underwater environments, multiple autonomous underwater vehicles are used as transmitter and receivers to form an acoustic wireless network in deep seawater. Signal path length changing speed cannot be the same for all TRPs with different positions of receivers. Second, time and frequency synchronization among all the receivers are achieved by M^{2}I based underwater synchronization strategy to reduce the synchronization errors. Third, to derive DFS coefficients, we calculate the curve signal propagation path length for one TRP in the inhomogeneous deep seawater environment. When deriving underwater DFS coefficients, we take the changing sound speed and the curve signal propagation path into consideration. Then, we show that the doppler frequency shift is actually a linear combination of velocities that multiply with DFS coefficients. Those coefficients are functions of the positions of the target body when the positions of transmitter and receivers are known. Fourth, to reduce the outliers in the VTB estimation with multiple TRPs, we formulate a dynamic self refining optimization algorithm to fully utilize the entire motion procedure instead of estimating all the VTB at one time. Taking advantage of dynamic optimization is able to iteratively refine VTB based on the relations among velocities estimated from different time points, which generates fewer outliers in estimation results and emphasizes the changing pattern of VTB. Due to the doppler coefficients initialization and rough position estimation of the target body, estimated VTB cannot be directly utilized to identify the target motions. We feed the extracted VTB into a classic CNN as the training data. CNN has the capability of learning the pattern of VTB to perform motion identification. Through the simulation, we validate that the VTB features that are recovered from received data have fewer outliers and prove that our proposed underwater motion recognition mechanism can achieve high classification accuracy independent of the target location.

The rest of this paper is organized as follows: in Section II, we derive the length and time of flight of curve signal propagation path in isogradient acoustic signal speed deep seawater. Then we derive DFS coefficients that can be utilized for VTB estimation. In Section III, we first utilize a M^{2}I synchronization method to adjust the time clock and center frequency for different receivers. Second, we propose a dynamic self-refining VTB optimization framework to extract velocity features from the Doppler frequency shift with fewer outliers. In Section IV, we use MATLAB to simulate the received signal in the isogradient sound speed underwater environment. Then, VTB is estimated according to our proposed DFS coefficients and optimization algorithm. Finally, classification results are evaluated. Section V concludes this paper.



(a) Signal propagation between two nodes

(b) Path length changes when target moves

Fig. 2: Signal propagation path length in isogradient acoustic signal speed sea water

II. DFS COEFFICIENT DERIVATION IN ISOGRADIENT ACOUSTIC SIGNAL VELOCITY UNDERWATER ENVIRONMENT

In order to derive the DFS coefficients that can be utilized for VTB estimation, we first derive signal path length expression for one TRP by the ray-tracing method.

Although underwater environments are always assumed to be homogeneous in some underwater channel modeling researches [6][7], the real propagation model of acoustic signal in the inhomogeneous ocean environments is different from the acoustic signal propagation in the air. The underwater acoustic signal speed c is affected by many factors such as salinity, seawater density and temperature [8]. The acoustic signals propagate along with a curve instead of a straight line such as they propagate in the air.

In this paper, we assume that the underwater acoustic signal propagates in the isogradient sound speed seawater which is similar to the real ocean condition. The velocity of the acoustic signal is only influenced by the decreasing gradient g and acoustic signal speed at the surface of the seawater v_{suf}, which can be summarized as a function of depth $c(z) = g \cdot z + v_{suf}$. In [9], the propagation of underwater acoustic signal between two nodes can be considered as an arc. According to the Snell's law, the sound propagation meets the following equation:

$$\kappa = \frac{d\theta}{dl} = \frac{sin\theta \cdot dc}{c \cdot dz} = \frac{sin\theta \cdot g}{c} \tag{1}$$

where κ is the curvature of the arc. θ is the angle between the tangent line at any points on this arc and the vertical z axis. l denotes the length of this arc. Accordingly, the radius R of this arc can be calculated by the following equation:

$$R = \kappa^{-1} = \left|\frac{c(z_T)}{sin\theta_T \cdot g}\right| = \left|\frac{c(z_R)}{sin\theta_R \cdot g}\right| \tag{2}$$

where θ_T, θ_R are the angles between z - axis and tangent lines at transmitter and receiver respectively as shown in Fig.2 (a). $c(z_T), c(z_R)$ are the acoustic signal speed at depth of transmitter and receiver.

As shown in Fig.2 (a), φ is the central angle corresponding to arc l_{TR}. Thus, the central angle φ can be described by using cosine theorem:

$$\varphi = arccos(\frac{2R^2 - D^2}{2R^2}) \tag{3}$$

where $D = \sqrt{(x_T - x_R)^2 + (y_T - y_R)^2 + (z_T - z_R)^2}$ is the dis-

tance between transmitter and receiver. φ is also the angle between line RO and line TO and can be described by the slope of TO k_{TO} and slope of RO k_{RO} :

$$\varphi = arctan(\frac{k_{TO} - k_{RO}}{1 + k_{RO}k_{TO}}) \tag{4}$$

where the slope k_{TO}, k_{RO} are orthogonal to the slope of tangent lines of arc l_{TR}. Receiver and transmitter are located at different depth and horizontal positions. So the slope k_{TO}, k_{RO} can be described as $k_{TO} = -tan\theta_T, k_{RO} = -tan\theta_R$. By solving the following equation:

$$\begin{gathered} arccos(\frac{2R^2 - D^2}{2R^2}) - arctan(\frac{k_{TO} - k_{RO}}{1 + k_{RO}k_{TO}}) = 0 \\ \frac{2(\frac{c(Z_T)}{sin\theta_T \cdot g})^2 - D^2}{2(\frac{c(Z_T)}{sin\theta_T \cdot g})^2} - \frac{1}{\sqrt{1 + (\frac{tan\theta_R - tan\theta_T}{1 + tan\theta_T \cdot tan\theta_R})^2}} = 0 \end{gathered} \tag{5}$$

We firstly take cos to all the terms in the equation and then substitute $\theta_R = arcsin\frac{sin\theta_T c_{z_R}}{c_{z_T}}$ into equation (5). Equation solving procedures are omitted due to page limitation. Then the angle θ_T can be obtained as:

$$\begin{gathered} \theta_T = \frac{\pi}{2} - arctan(\frac{g\sqrt{(x_T - x_R)^2 + (y_T - y_R)^2}}{2v_{suf} + gz_T + gz_R}) - \\ arctan(\frac{z_R - z_T}{\sqrt{(x_T - x_R)^2 + (y_T - y_R)^2}}) \end{gathered} \tag{6}$$

where x_T, y_T, z_T are the coordinates of transmitter. x_R, y_R, z_R denote the coordinates of receiver. As shown in the Fig.2 (a), the chord angle ϕ_{TR} is the angle between arc l_{TR} and the straight line D. Angle $\beta = arctan\frac{z_R - z_T}{\sqrt{(x_T - x_R)^2 + (y_T - y_R)^2}}$ and angle $\theta_T = 0.5\pi - \phi_{TR} - \beta$. Since the ϕ_{TR} is the chord angle, the central angle φ can be calculated as following equation:

$$\varphi = 2\phi_{TR} = 2arctan(\frac{g\sqrt{(x_T - x_R)^2 + (y_T - y_R)^2}}{2v_{suf} + gz_T + gz_R}) \tag{7}$$

Then, the total underwater transmission length for acoustic signal between transmitter and receiver can be calculated as:

$$l_{TR} = R \cdot \varphi = \frac{c(z_T)}{sin\theta_T \cdot g} \cdot 2arctan(\frac{g\sqrt{(x_T - x_R)^2 + (y_T - y_R)^2}}{2v_{suf} + gz_T + gz_R}) \tag{8}$$

The traveling time of signal that propagates along with arc l_{TR} cannot be calculated by dividing length by sound speed, since the acoustic signal speed varies with different depth. Traveling time can be calculated as an integration:

$$\tau = \left|\int \frac{dl}{c(z)}\right| = \left|\int_{\theta_T}^{\theta_R} \frac{d\theta}{sin\theta \cdot g}\right| = \frac{1}{g}\left|ln(\frac{sec\theta_R + tan\theta_R}{sec\theta_T + tan\theta_T})\right| \tag{9}$$

The doppler effect is caused by the signal propagation path length changing speed and signal propagation speed according to [10]. According to the classic doppler effect model, doppler effect can be described as $f_{DFS} = f_0\frac{\Delta v}{c}$ when the velocity difference between transmitter and receiver is much smaller than signal propagation speed, where f_0 is the original frequency. This Δv can be regarded as the signal propagation path length changing speed due to the assumption that the source is moving towards or away from a receiver with constant speed. Since the underwater transmitter and receiver are the AUVs and could be considered as roughly static, the only thing that changes the signal path length is the motion of target as shown in Fig.2 (b) where target body moves from location $B(t-1)$ to the location $B(t)$. Transmitted signals encounter the target body and then are reflected towards the receivers. Underwater doppler effect is significantly different from the doppler effect in the air as shown in Fig .1 due to two reasons. First, as we explained above, the signal propagation path is not a straight line but an arc. Second, in the air, the signal velocity can be considered as a constant. However, in the isogradient acoustic speed seawater, the acoustic signal velocity changes with different depths. Thus, as shown in Fig. 1, when the transmitter moves vertically downward to the receiver in the air at a constant speed, the doppler frequency is a constant as shown in Fig. 1(b). The doppler frequency in the isogradient deep sea water is increasing even with a constant transmitter moving speed as shown in Fig. 1(a). The reason is that the acoustic signal speed becomes smaller when transmitter moves downward in the isogradient sound speed seawater. It is noteworthy that only the sound velocity at the location where motions happen affects the doppler effect. When acoustic signals that are actually mechanical wave propagate through other space in the seawater which are not involved in the target motions, the frequency $f = \frac{c}{\lambda}$ will not change.

In this paper, we consider the receivers and transmitters as static in a short time when they collect data. In addition, we also assume that there is no surface reflection and bottom reflection in the deep seawater. Then, the doppler frequency caused by the motions of the target can be calculated as a function of t:

$$f_d(t) = f \cdot \frac{v_{path}}{c(z)} = f \cdot \frac{dL(t)}{dt \cdot c(t)} \tag{10}$$

where $L(t) = l_{TB} + l_{BR}$ is the total path length which combines the path length from transmitter to target body and path length from target body to the receiver. According to the equation (8), the total path length can be calculated as:

$$\begin{aligned} L(t) &= l_{TB}(t) + l_{BR}(t) \\ &= \frac{c(z_B)}{sin\theta_{BT} \cdot g} \cdot 2arctan(\frac{g\sqrt{(x_B + v_x t - x_R)^2 + (y_B + v_y t - y_R)^2}}{2v_{suf} + gz_B + gv_z t + gz_R}) + \\ &\quad \frac{c(z_B)}{sin\theta_{BR} \cdot g} \cdot 2arctan(\frac{g\sqrt{(x_B + v_x t - x_T)^2 + (y_B + v_y t - y_T)^2}}{2v_{suf} + gz_B + gv_z t + gz_T} \end{aligned} \tag{11}$$

where θ_{BT} denotes the angle between z axis and the tangent line of arriving signal from transmitter to target body at the location of target body. θ_{BR} denotes the angle between z-axis and the tangent line of reflected signal from target body to receivers respectively at the location of target body. v_x, v_y, v_z are the velocity of target motion along x-axis, y-axis and z-axis. The expressions of $sin\theta_{BR}, sin\theta_{BT}$ are really complicated due to the θ_T expression. Let's denote a sub function $u(t)$:

$$\begin{aligned} u_{BR}(t) = &\frac{g[(x_B + v_x t - x_R)^2 + (y_B + v_y t - y_R)^2]}{(2v_{suf} + 2gz_R) \cdot \sqrt{(x_B + v_x t - x_R)^2 + (y_B + v_y t - y_R)^2}} + \\ &\frac{[2v_{suf} + g(z_B + v_z t + z_R)] \cdot (z_B + v_z t - z_R)}{(2v_{suf} + 2gz_R) \cdot \sqrt{(x_B + v_x t - x_R)^2 + (y_B + v_y t - y_R)^2}} \end{aligned} \tag{12}$$

Then the term $\frac{c(z_B)}{sin\theta_{BR} \cdot g}$ can be further derived as:

$$\frac{c(z_B)}{sin\theta_{BR}\cdot g} = \frac{[g(z_B + v_z t) + v_{suf}]\cdot\sqrt{1+u_{BR}^2}}{g} \tag{13}$$

$\frac{c(z_B)}{sin\theta_{BT}\cdot g}$ can be derived with a similar procedures. Calculation of $\frac{dL(t)}{dt}$ are omitted due to the page limitation, doppler frequency effect can be represented as a linear summation of v_x, v_y, v_z and the DFS coefficients a_x, a_y, a_z:

$$f_d = \frac{f}{c(t)}\frac{dL(t)}{dt} = \frac{a_x v_x + a_y v_y + a_z v_z}{\lambda(t)} \tag{14}$$

where $\lambda(t) = \frac{f}{c(t)}$ denotes the acoustic signal wavelength at the location where target motions happen. In this paper, locations of transmitter and receivers are known. x-axis DFS coefficient a_x can be represented as following equation:

$$a_x = \frac{x_B(t) - x_R}{\sqrt{(x_B(t)-x_R)^2 + (y_B(t)-y_R)^2}} + \frac{x_B(t) - x_T}{\sqrt{(x_B(t)-x_T)^2 + (y_B(t)-y_T)^2}} \tag{15}$$

where $x_B(t), y_B(t), z_B(t)$ denote the coordinates of target body as functions of time t. a_x is a function of the location of target body $a_x = f_1(x_B(t), y_B(t))$ when receiver and transmitter are static and their positions are known. The y-axis DFS coefficient a_y has a similar expression:

$$a_y = \frac{y_B(t) - y_R}{\sqrt{(x_B(t)-x_R)^2 + (y_B(t)-y_R)^2}} + \frac{y_B(t) - y_T}{\sqrt{(x_B(t)-x_T)^2 + (y_B(t)-y_T)^2}} \tag{16}$$

a_y is also a function of the location of target body $a_y = f_2(x_B(t), y_B(t))$ when target is moving. a_z has a more complex expression which can be represented as following equation:

$$a_z = \frac{\frac{g[(x_B(t)-x_R)^2+(y_B(t)-y_R)^2]+[2v_{suf}+g(z_B(t)+z_R)]\cdot(z_B(t)-z_R)}{(2v_{suf}+2gz_R)\cdot\sqrt{(x_B(t)-x_R)^2+(y_B(t)-y_R)^2}}}{\sqrt{1+(\frac{g[(x_B(t)-x_R)^2+(y_B(t)-y_R)^2]+[2v_{suf}+g(z_B(t)+z_R)]\cdot(z_B(t)-z_R)}{(2v_{suf}+2gz_R)\cdot\sqrt{(x_B(t)-x_R)^2+(y_B(t)-y_R)^2}})^2}} + \frac{\frac{g[(x_B(t)-x_T)^2+(y_B(t)-y_T)^2]+[2v_{suf}+g(z_B(t)+z_T)]\cdot(z_B(t)-z_T)}{(2v_{suf}+2gz_T)\cdot\sqrt{(x_B(t)-x_T)^2+(y_B(t)-y_T)^2}}}{\sqrt{1+(\frac{g[(x_B(t)-x_T)^2+(y_B(t)-y_T)^2]+[2v_{suf}+g(z_B(t)+z_T)]\cdot(z_B(t)-z_T)}{(2v_{suf}+2gz_T)\cdot\sqrt{(x_B(t)-x_T)^2+(y_B(t)-y_T)^2}})^2}} \tag{17}$$

a_z is a function of the location of target body $a_z = f_3(x_B(t), y_B(t), z_B(t))$. As shown in Fig.2 (b), the received signals can be divided into line-of-sight l_{TR} and reflection signal from target body for one transmitter and receiver link.

$$\begin{aligned} h(t) &= e^{-j\psi_j t}[\alpha_{l_{TR}} e^{-j2\pi f \tau_{l_{TR}}} + \alpha_{l_{TBR}}(t) e^{-j2\pi f \frac{V_{path}(t)}{c(t)}\tau_{l_{TBR}}(t)}] \\ &= e^{-j\psi_j t}[h_{static} + \alpha_{l_{TBR}}(t) e^{-j2\pi f \frac{V_{path}(t)}{c(t)}\tau_{l_{TBR}}(t)}] \end{aligned} \tag{18}$$

where ψ_j denotes the phase offset of jth TRP. Phase offsets are caused by the carrier frequency unsynchronization and time clock unsynchronizaiton. Phase offset changes with different transmitter receiver link. $\alpha_{l_{TR}}$ denotes the attenuation coefficient for LoS path. Since the transmitting AUV and receiving AUV are regarded as static, only the path l_{TBR} has the doppler frequency shift due to the target motion. In addition, time delay $\tau_{l_{TRB}}$ and attenuation coefficient $\alpha_{l_{TBR}}$ would also change with the motions of target due to the changes of signal propagation path length. The unknown phase offset can be mitigated according to [10] by calculating the conjugate multiplication of channel state information. In addition, the doppler frequency shift could also be extracted from frequency domain according to [10].

III. ESTIMATION OF VELOCITIES OF TARGET BODY

A. M^2I Assisted Underwater Synchronization

As we mentioned in Section I, we utilize multiple receivers and one transmitter to solve the ambiguity of VTB estimation. Time and frequency synchronization should be performed among different receivers.

Compared with indoor wireless sensing networks, the underwater wireless networks encounter a more serious unsynchronization problem due to the long distance of signal propagation and the slower velocity of the acoustic signal. Time clock unsychronization would lead to the problem that the extracted DFS values form each receiver are not describing the motions in the same time window. Frequency unsynchronization would lead to inaccurate DFS values. In this paper, we utilize the metamaterial inspired (M^2I) assisted synchronization [11] to synchronize the frequency and time clock for AUVs. M^2I channel have a significantly faster transmission speed that leads to a tiny time delay.

To briefly explain, a predetermined master AUV acts as the reference. The first step of synchronization is that the master AUV broadcasts beacon signal to all other slave AUVs using M^2I signals. This broadcast contains a beacon signal at a known frequency and a synchronization preamble, which is used for carrier frequency synchronization for each slave AUV. According to the [11], the first step frequency synchronization can be shown as:

$$\Delta f_{c,j} = \frac{\hat{f}_{s,1,j} - f_{s,j}}{k} = (\bar{b_j} - 1) f_{c,1} + \frac{\epsilon_{s,j}}{k} \tag{19}$$

where $\bar{b}_j$ is the average clock drift of jth receiving AUV to the master AUV. k is the frequency multiplier. $\hat{f}_{s,1,j}$ is the estimated frequency of beacon signal with local oscillator of jth slave AUV. $f_{s,j}$ is the operating frequency of jth slave AUV and $f_{c,1}$ is the oscillator frequency of master node. $\epsilon_{s,j}$ is the frequency estimation error. The time synchronization error for jth receiver can be calculated as:

$$\epsilon_{j,t} = \epsilon_{j,f}[\frac{P_{1,j}}{B} + \frac{d_{1,j}}{v} + \sum_{i=j+1}^{J}(\frac{P_{1,i} + P_{i,1}}{B} + 2\frac{d_{1,i}}{v})] \tag{20}$$

where $P_{1,j}$ and $P_{i,1}$ are the total packet length delivered from the master node to the slave node and packet length from the slave node to the master node respectively. B is the bandwidth, $d_{1,j}$ is the distance between the master node and the slave node j, v is the propagation speed of the signals used for synchronization. $\epsilon_{j,f} = \frac{\epsilon_{s,j}}{k}$ is the frequency estimation error.

According to the above two equations, the time synchronization error and frequency synchronization error of M^2I mainly depend on the bandwidth B, signal propagation speed v and frequency multiplier k. M^2I has a obviously larger bandwidth, faster transmission speed and bigger frequency multiplier k in underwater environments compared with acoustic signal based

synchronization method. Thus, M^{2}I assisted synchronization for underwater wireless sensing networks has a smaller both time synchronization error and frequency synchronization error.

B. Dynamic Self-Refining Optimization for VTB Estimation

After synchronizing different receivers and extracting doppler frequency shift values from the received data, the most important step is to estimate the velocities of the target body. In a certain type of motion of the target body, different parts of the target body actually move at different velocities, which causes different doppler frequency values in the frequency domain. The amplitude of doppler frequency indicates the reflection attenuation from those parts to the receivers. As we mentioned, there is an ambiguity between doppler frequency and the velocities of the target body. For example, as shown in equation (10), the doppler frequency shift is equal to the signal propagation path length changing speed over the acoustic signal speed. It is easy to find a situation that the path length becomes longer at a direction where the acoustic signal speed decreases. This situation may introduce an unchanged doppler frequency shift although the target is moving. Furthermore, different path length changing speeds may lead to the same doppler frequency shift due to the fact that the underwater acoustic signal speed changes with different depths.

To solve the ambiguity between VTB and DFS value, we take the advantages of multiple J receivers which are actually AUVs to form a deep water wireless sensing network. With the receivers distributed at different positions especially at a different depth, the path length of signal propagation has different changing speeds with the same velocities of target motion. To briefly explain, different positions of receivers introduce the different values of θ_{BR}. According to the expression of a_y, a_x, different horizontal positions of receivers lead to the different values of a_x, a_y, Furthermore, with the same horizontal positions, DFS values still can be distinguished by receivers at different depth. With receivers at different depth z, the same z-axis velocity v_z can cause different values of DFS according to the expression of a_z.

Algorithm 1 Dynamic Self-Refining VTB Optimization

Input: Doppler frequency shift profile $D = \{D_1, D_2, ..., D_{N_t}\}$
Output: $V = \{V_1, V_2, ..., V_{N_t}\}$ velocity profile of human body components

1: initialize the linear coefficients profile $A_1^1, ..., A_1^J$
2: **for** i = 1 to N_t **do**
3: perform $min_V \sum_{n=1}^{i} \sum_{j=1}^{J} \|A_n^j V_n I/\lambda(z_n^k) - D_n^j\|_F^2$
4: **for** m = 2 to i+1 **do**
5: **for** j = 1 to J **do**
6: **for** k = 1 to K **do**
7: $A_m^{jk} = [f_1^j(x_m^k, y_m^k), f_2^j(x_m^k, y_m^k), f_3^j(x_m^k, y_m^k, z_m^k)]$
8: $[x_m^k, y_m^k, z_m^k] = [x_{m-1}^k + \Delta t \cdot v_x^{mk}, y_{m-1}^k + \Delta t \cdot v_y^{mk}, z_{m-1}^k + \Delta t \cdot v_z^{mk}]$ where $\Delta t = \frac{T}{N_t - 1}$
9: **end for**
10: **end for**
11: **end for**
12: **end for**

The underwater wireless networks contain one transmitter and multiple receivers. Before performing target motion identification, AUVs first localize each other and the target body. AUVs are assumed to know position of themselves and keep static. We consider K as the total number of parts of target body which can caused the doppler frequency shift. This K is also the total number of doppler frequency bins that can be extracted from frequency domain for one TRP at one time point. The number of resolution in frequency domain can be calculated by $\frac{B}{K+1}$.

Localization of target body could only give us the rough estimations of positions of different parts. We cannot know which part of human body corresponds to which doppler frequency shift. Thus, we propose a dynamic self-refining VTB estimation optimization algorithm. In the proposed optimization algorithm, we utilize an insight of target motion. To illustrate, the doppler coefficients a_x, a_y, a_z are the functions of target body position when the transmitter and receivers are considered as static. The velocity $V_i(k) = [v_x^k \quad v_y^k \quad v_z^k]$ at time i can be estimated by following equation:

$$min_V \sum_{n=1}^{i} \sum_{j=1}^{J} \|A_n^j V_n I/\lambda(z_n^k) - D_n^j\|_F^2 \tag{21}$$

where I is the identity matrix. V_i contains the K kinds of estimated velocities.

$$V_i = \begin{vmatrix} v_x^1 & ... & v_x^K \\ v_y^1 & ... & v_y^K \\ v_z^1 & ... & v_z^K \end{vmatrix} \tag{22}$$

And A_i^j contains doppler coefficients:

$$A_i^j = \begin{vmatrix} a_x^1 & a_y^1 & a_z^1 \\ ... & ... & ... \\ a_x^K & a_y^K & a_z^K \end{vmatrix} \tag{23}$$

where D_i^j is a KxK diagonal matrix contains K DFS values for jth TRP at time i. $A_i^{jk} = [a_x^k, a_y^k, a_z^k]$ are the DFS coefficients for jth TRP at time i.

If the time interval between two consecutive received signals is small enough, then we can reasonably assume that all parts of the target body will keep the same direction and the same velocity V_i^j. Therefore, the DFS coefficient A_{i+1} of next time point can be obtained as shown in algorithm 1 step 7 and step 8 where f_1^j, f_2^j, f_3^j denote the DFS coefficients functions that are described in equations (15)(16)(17) for jth TRP. By adding the newly obtained DFS coefficients A_{i+1} into optimization framework, all the velocities from i = 1th to i + 1th are re-estimated. Then, the newly obtained are used v_i to update A_i. Utilizing newly obtained v_{i+1} to predict A_{i+1} and estimate the velocity of k part of target body again. With more DFS coefficients at different time points added to the optimization problem, the estimated accuracy would be more accurate. In addition, the wavelength $\lambda(z_i^k)$ could also be refined with the newly obtained positions of the target body. With more accurate estimated velocities, the DFS coefficients A will be more valid. Iteratively performing proposed velocity optimization lets the velocities V be more accurate. Furthermore, the velocity changing pattern becomes more smooth with fewer outliers. All the procedures are described in Algorithm 1. Although the time complexity of the proposed algorithm 1 is $O(n^4)$, the total calculation time is affordable. The number of transmitter receiver links and the number of VTB features are often limited.

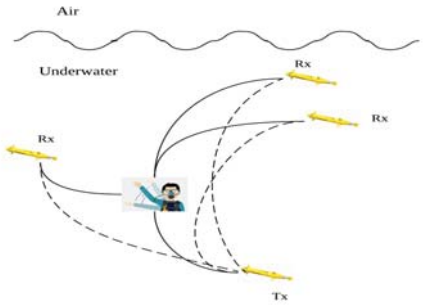


Fig. 3: Underwater diver activity recognition using acoustic wireless networks

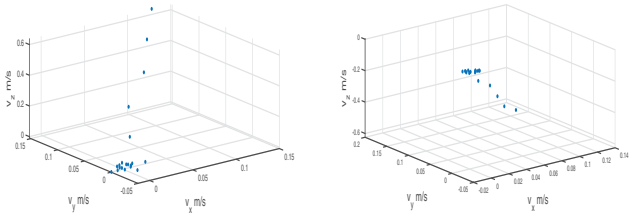
IV. SIMULATIONS AND PERFORMANCE EVALUATION

Before embarking the simulation results, we first illustrate the underwater environments simulation setup using MATLAB.

The situation for underwater target motion identification is considered as the deep inhomogeneous seawater. According to [8], the deep seawater environment can be modeled as the isogradient acoustic signal speed environment where the sound speed decrease with depth increases at a constant gradient. We assume the sound speed at the surface of seawater v_{suf} = 1470m/s and the sound speed decreasing gradient g = 0.03. The total number of receivers is 10, that is, the total number of TRP J = 10. We applied wavelet transformation to convert the received signal into the frequency domain. Compared with short time fourier transformation (STFT), discrete wavelet transformation (DWT) is able to achieve high resolution in both time and frequency domain. The k = 20 means that the target body may have 20 different velocities during the motion. The time period for a certain type of motion is defined as 5 seconds. The time resolution of DWT is defined as 0.1 second and the VTB can be assumed as constant during this small period. All the receivers and the transmitter are located from 91 meters to 181 meters in depth.

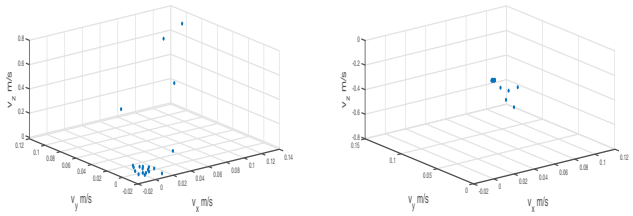
Before collecting data, all the AUVs perform time and frequency synchronization. Frequency synchronization mitigates the carrier frequency offset. Time synchronization is more important for AUVs. In the underwater environment, the sound speed is lower than the speed of EM signal in the air. The acoustic signal propagation path length is also longer than the signal path length in the indoor environment. Those two facts lead to a larger time delay of received signals. The training data can be collected immediately when the target motion starts in the indoor environment. However, AUVs need to calculate the corresponding time delays to choose the beginning time and the ending time for training data. Local time clocks need to be synchronized for different AUVs to ensure they utilize the same length of received data that only describes the target motion. Otherwise, the length of training data may be different and the training data may contain the time period without motion.

To summarize the procedures of simulation, first, we simulate the received signals according to our proposed channel model and doppler frequency model in Section II. The second step is to select the time period of the received signal when the motion happened. Third, DWT is performed to extract the doppler frequency shift from received data. In the fourth step,



(a) lifting arm upward (b) putting arm downward

Fig. 4: VTB estimation results using dynamic self-refining optimization and underwater DFS coefficients



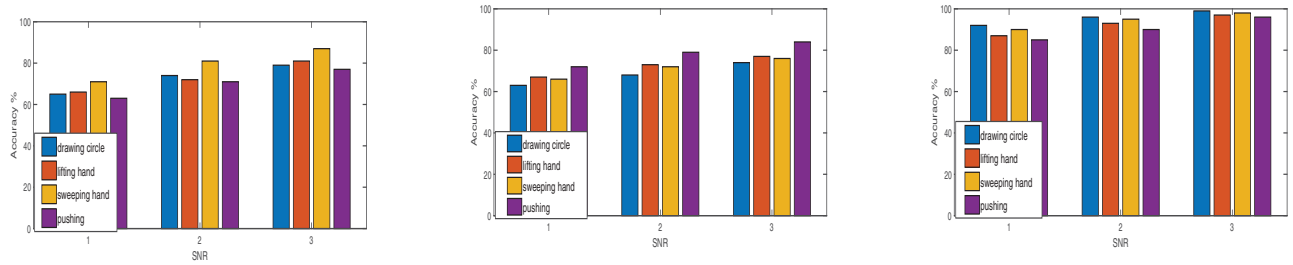
(a) lifting arm upward (b) putting arm downward

Fig. 5: VTB estimation using one time estimation and DFS coefficients in the air

the doppler frequency shift is fed into the dynamic self-refining VTB optimization framework to generate the VTB. Finally, we train the CNN with VTB and test our proposed motion recognition mechanism. It is noteworthy that the same motion happens at different locations to generate VTB features. The testing data is the VTB that is generated by target motion in different locations from the training locations.

The whole underwater wireless sensing scenario can be described in Fig. 3. As shown in Fig.4 and Fig.5, VTB estimations at a time point are performed with different types of optimization algorithms and different DFS coefficients. The target motion is defined as the human lifting and laying arms in the underwater environments as shown in Fig.3. Since most parts of the human body keep static, most points of VTB are around zero velocity as shown in Fig.4 and Fig. 5. When lifting arm upwards and laying arm downwards, different parts of the arm have different velocities. Compared with Fig.4, the VTB estimation results in Fig.5 have more outliers and larger values of v_z due to the one-time optimization and wrong DFS coefficients. Our proposed DSR optimization framework considers the relations between adjacent VTB values and iteratively self refines both DFS coefficients and VTB values, which make the VTB estimation smoother. With fewer variations in VTB values, CNN is easier to find the pattern hidden in VTB along with the time to achieve high recognition accuracy.

When training the CNN classifier, we define 4 different motions which are lifting hands, pushing, sweeping and drawing circles. To simulate the received signal caused by those 4 types of motions, 5 points move at different speeds represent the motion of the human arm. For example, 5 points move at different angular velocities but at the same absolute speed when the diver is drawing circles. After each time interval, we update the positions of points. Then, we simulate the received signals according to the proposed model in Section



(a) Receivers are unsynchronized

(b) Using wrong DFS coefficients and one time optimization

(c) Using underwater DFS coefficients and DSR optimization

Fig. 6: Underwater target motion recognition accuracy in different situations

II. It is noteworthy that we add addictive Gaussian noise to the received signals to simulate the real noise in the underwater environments. We generate 2100 samples for different types of motions. In addition, the samples are generated in 3 different locations. In each location, we generate 700 samples. For each sample, it contains data collected from 50 times points, 10 transmitter receiver pairs and 20 different body components. In other words, we generate a data set that has 4x2100x50x20x10 = 8.4x10e7 as our input training features.

We test our trained classifier in 3 different situations as shown in Fig. 6 to illustrate the influences of unsynchronization, wrong DFS coefficients, and our proposed underwater DFS coefficients. In Fig. 6(a), we input the testing data collected from unsynchronized receivers. All the receivers have different local time clock. Therefore, data collected from different receivers are not caused by the motions happen at the same time, which leads to the low accuracy of classification results. Although the testing accuracy increases with high signal to noise ratio (SNR), the average classification accuracy doesn't reach 90 percent.

As shown in Fig. 6(b), we use the DFS coefficients when signals propagate in the air and one-time optimization which estimates the VTB values of a period at one time. As we explained, using DFS coefficients in the air will make the velocity along z-axis v_z larger than the real values. Estimating all VTB values at one time could lead to more outliers, which makes the VTB changing pattern ambiguous. Even the training data are generated with the DFS coefficients in the air and the one-time estimation, the classification accuracy is still low when SNR value is small.

Finally, we test our proposed underwater target motion classification mechanism using derived underwater DFS coefficients a_x, a_y, a_z and proposed DSR optimization. DSR optimization produces fewer outliers in VTB values and estimates VTB values that are more similar to the real velocities by considering the relation between two sets of DFS coefficients at consecutive times for a certain TRP. As shown in Fig.6 (c), our proposed underwater recognition accuracy could achieve high accuracy even with relatively low SNR values. When SNR increases, the average classification accuracy could reach 97 percent. The classification results show that our proposed DFS coefficients in the inhomogeneous deep seawater and dynamic self-refining optimization have the capability of performing the accurate estimation of VTB features.

V. CONCLUSION

In this paper, we proposed a target motion recognition mechanism using acoustic wireless networks in inhomogeneous deep water environments. First, we derived the length of signal propagation path in the isogradient sound speed underwater environments where signals propagate along with an arc to figure out DFS coefficients. Second, with derived DFS coefficients, we developed a dynamic self-refining optimization algorithm to estimate the VTB from doppler frequency shift, which has more accurate VTB estimation results and reduces the outliers of VTB estimation. Finally, we tested our proposed target motion recognition mechanism through MATLAB simulations. VTB estimation results are evaluated to verify the advantages of our proposed DSR optimization algorithm. Classification testing results showed that the underwater target motion can be identified with high accuracy.

REFERENCES

[1] Zhao, Minghao et al. "Real-Time Underwater Image Recognition with FPGA Embedded System for Convolutional Neural Network." Sensors (Basel, Switzerland) vol. 19,2 350. 16 Jan. 2019, doi:10.3390/s19020350

[2] Xiu Li, Min Shang, Hongwei Qin, Liansheng Chen. " Fast accurate fish detection and recognition of underwater images with Fast R-CNN." OCEANS 2015 - MTS/IEEE Washington, 1–5. https://doi.org/10.23919/OCEANS.2015.7404464

[3] K. Ali et al, "Recognizing Keystrokes Using WiFi Devices," IEEE Journal on Selected Areas in Communications, vol. 35, (5), pp. 1175-1190, 2017.

[4] W. Wang et al, "Device-Free Human Activity Recognition Using Commercial WiFi Devices," IEEE Journal on Selected Areas in Communications, vol. 35, (5), pp. 1118-1131, 2017.

[5] Jiang, Wenjun et al, "Towards Environment Independent Device Free Human Activity Recognition," 289-304. 10.1145/3241539.3241548.

[6] Baktash, Ebrahim, et al. "Shallow Water Acoustic Channel Modeling Based on Analytical Second Order Statistics for Moving Transmitter/Receiver." IEEE Transactions on Signal Processing, vol. 63, no. 10, IEEE, May 2015, pp. 2533–45, doi:10.1109/TSP.2015.2411219.

[7] Naderi, Meisam, et al. "A Geometry-Based Underwater Acoustic Channel Model Allowing for Sloped Ocean Bottom Conditions." IEEE Transactions on Wireless Communications, vol. 16, no. 4, IEEE, Apr. 2017, pp. 2394–408, doi:10.1109/TWC.2017.2664829.

[8] Domingo MC. "Overview of channel models for underwater wireless communication networks." Phys Commun 2008;1(3):163–82. https://doi.org/10.1016/j.phycom.2008.09.001.

[9] Casalino, G. Caiti, A. "RT^2: A Real-Time Ray-Tracing Method for Acoustic Evaluations Among Cooperating AUVs." In Proceedings of the OCEANS 2010 IEEE—Sydney, Sydney, NSW, Australia, 24–27 May 2010; pp. 1–8.

[10] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. " Understanding and modeling of wifi signal based human activity recognition." In Proc. of ACM MobiCom 2015.

[11] S. Desai, V. Sudev, X. Tan, P. Wang, and Z. Sun, "Enabling Underwater Acoustic Cooperative MIMO Systems by Metamaterial-Enhanced Magnetic Induction", in Proc. IEEE WCNC 2019, Marrakech, Morocco, April 2019.