2024 33rd IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 26-30, 2024. Pasadena, California, USA.

2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN) | 979-8-3503-7502-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/RO-MAN60168.2024.10731209

UnRealTHASC – A Cyber-Physical XR Testbed for Underwater Real-Time Human Autonomous Systems Collaboration

Sushrut Surve[1], Jia Guo[1], Jovan C. Menezes[1], Connor Tate[2], Yiting Jin[1], Justin Walker[1], Silvia Ferrari[1]

Abstract— **Research on underwater human-robot teaming is particularly promising because of complementary sensory skills that may overcome unique challenges associated with the combination of limited connectivity and low visibility. Nevertheless, testing human-robot collaboration under water, especially in complex, real-world scenarios, poses severe safety, cost, and time constraints that significantly hinder academic research in this space. This paper presents a novel cyber-physical extended-reality (XR) testbed, referred to as Underwater Real-Time Human Autonomous Systems Collaboration (UnRealTHASC), designed to enable human-robot interactions in simulated photorealistic environments with mixed real and virtual wearables, and advanced autonomous underwater sensors in-the-loop. Novel sensor interfaces are designed to integrate real and virtual sensors for measuring physiological and cognitive human states underpinning decision-making abilities. Physics-based human and robot motion models are developed along with new sensor simulations in order to capture the couplings between underwater behaviors and perception based on measurements from optical and sonar sensors. Real-time data acquisition pipelines are created to access and share data from both real and virtual sensors and robots, such that new methods for online planning and collaboration may be tested via human-in-the-loop demonstrations.**

I. Introduction

Scuba diving presents many technical challenges that are potentially life threatening even for highly experienced divers involved in exploration and navigation. Hence, when complex tasks, physiological stresses (e.g. nitrogen narcosis), limited visibility, or confined environments are introduced they notably elevate cognitive load and operational risk often resulting in incomplete or failed missions. Cognitive, physiological, and environmental stressors all adversely impact diver's perception and cognitive state potentially resulting in anxiety, panic, or even seizures as known from the well-documented history of diving accidents [1], [2], [3]. To help overcome these challenges, scientific and military dive operations are incorporating unmanned underwater vehicles (UUVs) with increased frequency, underscoring the necessity for collaborative efforts between humans and machines to ensure safe and efficient undersea operations [4]. The primary challenges in underwater human-robot teaming include (i) the complexity of inter-agent communication, (ii) diminished perception due to environmental factors, and (iii) the risks and cost associated with *in situ* testing of collaboration approaches.

Recent studies in underwater human-robot interaction (HRI) often rely on task-centric design strategies for interaction, which however effective in simpler communication contexts of social robotics and terrestrial operations, prove insufficient for the underwater domain, where communication and perception constraints drastically affect teaming performance. Progress with this approach is further fettered by the cost to test and iterate in this domain. This highlights the need for a nuanced approach that addresses the specific challenges of underwater environments [5], [6]. Communication and perception are vital for team dynamics, used for sharing task-specific information and coordinating efforts to prevent conflicts. Yet, the unique conditions of underwater operations, including the covert nature of missions, optical properties of water, and hardware limitations, necessitate a shift towards autonomy. Acoustic sensors have proved to be a viable solution to enhance perception in underwater environments. The deficiencies in communication need to be mitigated by explicit and implicit non-verbal cues, such as body gestures [6], to aid the human diver effectively. However, these cues alone do not fully reveal the reason behind the diver's behavior, which might be driven by cognitive or environmental stressors, indicating the necessity for deeper contextual understanding. Recent advancements in physiological sensors offer promising insights into divers' physiological and cognitive states[7].

To address these challenges, this paper presents a human-centered HRI approach, integrated into an XR testbed for Underwater Real-Time Human Autonomous Systems Collaboration, *UnRealTHASC*. This work builds on the state-of-the-art XR testbed RealTHASC [8] to incorporate human physiological sensors and simulated sonar sensors in underwater human-robot collaboration, enabling robots to provide more effective and context-aware assistance, thereby fostering safer and more efficient underwater human-robot teaming. A novel method to render acoustic measurements in real time is developed. This paper presents a novel integration of real and virtual sensors enabling realistic human and robot movement and a human-centric teaming approach inside 3D underwater virtual environments.

II. System architecture

The UnRealTHASC facility, as illustrated in Fig. 1, combines physical and virtual workspaces so that human divers can interact seamlessly with programmable virtual robots defined in the virtual workspace. Let the physical laboratory space where the human diver exists be denoted by $\mathcal{W} \subset \mathbb{R}^3$. The virtual workspace where the virtual robots operate is

[1] Sushrut Surve, Jia Guo, Jovan. C. Menezes, Yiting Jin, Justin Walker and Silvia Ferrari are with Laboratory of Intelligent Systems and Control(LISC), Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14850, USA.

[2] Connor Tate is with the Human Healthspan Resilience and Performance team, Florida Institute for Human Machine Cognition, Pensacola, USA

979-8-3503-7502-2/24/$31.00 ©2024 IEEE

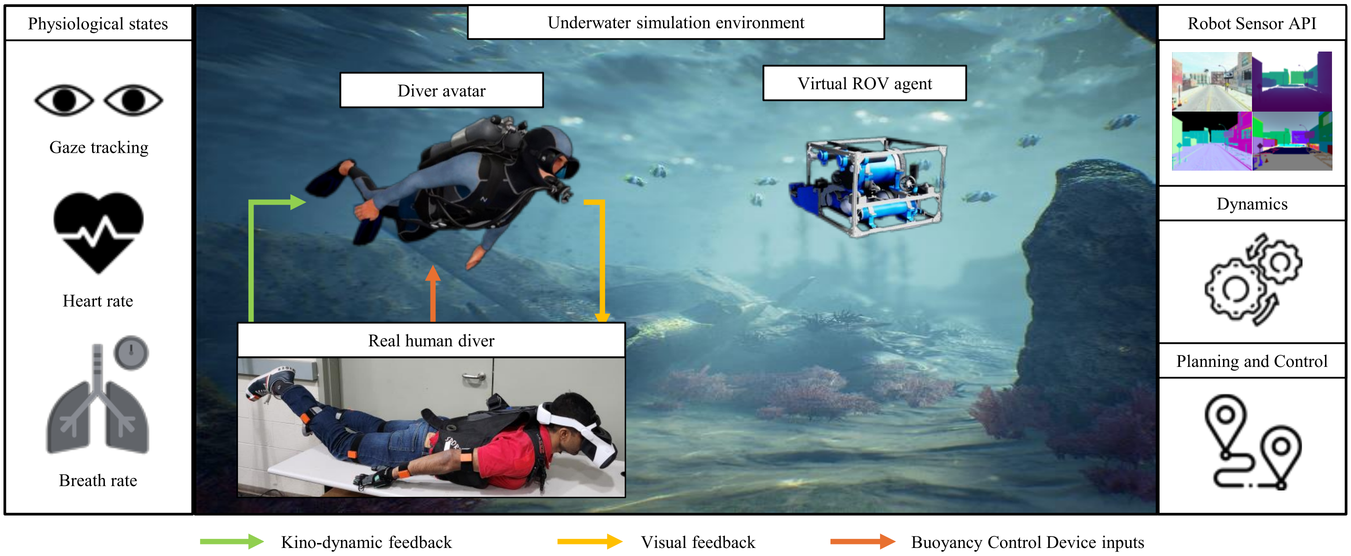


Fig. 1. A real human diver interacting with a virtual remotely operated vehicle (ROV) agent in the UE™ simulation environment, through a diver avatar. The real human diver controls the movement of the diver avatar and receives visual feedback from the environment through the VR headset. The real human diver is also equipped with physiological sensors monitoring the heart rate, breath rate, and gaze of the human. The virtual UUV agent is equipped with various acoustic and optical sensors, dynamics, and planning and control for maneuvering in the environment

created in Unreal Engine™ (UE™) and is denoted by $\mathcal{U} \subset \mathbb{R}^3$. The inertial reference frames embedded in physical workspace $\mathcal{W}$ and virtual workspace $\mathcal{U}$ are denoted by $\mathcal{F}_{\mathcal{W}}$ and $\mathcal{F}_{\mathcal{U}}$, respectively. Virtual worlds from the UE™ Marketplace are adapted to create photorealistic underwater environments to simulate realistic scenarios experienced by human divers while performing tasks underwater. The virtual environment $\mathcal{U}$ facilitates interaction between human divers and robots.

Overall, three types of agents are involved in UnRealTHASC, namely real human divers, diver avatars, and virtual robots. All the UnRealTHASC agents are characterized with a motion model and a sensor model. *Real human divers* exist in the physical laboratory $\mathcal{W}$ but sense the virtual world $\mathcal{U}$ through virtual reality (VR) headsets. *Diver avatars* are the projections of human divers into the virtual environment $\mathcal{U}$, which serve as the surrogate of the physical divers for interacting with the virtual environment including the virtual robots. The motion model of the diver avatar is governed by the motions of the human diver in $\mathcal{W}$ as described in Section.III-A.2. The human diver and associated avatar share the same field of view (FOV) as a result of the VR headset integration. The details for specifying the diver avatar are described in Section III. Virtual robots are solely defined in $\mathcal{U}$ and are programmable in terms of the motion model, sensor model, and decision-making strategies. This paper focuses on high-fidelity simulation of robot sensors for both environment attributes and physiological states of human divers. See Section IV for details. With the current system architecture, it is also straightforward to define virtual humans in UnRealTHASC besides the three aforementioned types of agents.

III. Human XR Integration

This section details the three aspects considered when integrating human divers into UnRealTHASC: swimming motion of human, physiological states, and diving apparatus. In the simulation, the state of human diver is represented by a lumped vector $\mathbf{x} = [\ \mathbf{j}^T, \mathbf{p}^T, \mathbf{l}^T\ ]^T$. Here the notation $\mathbf{j}$ denotes the collection of joint angles that characterize the limb motion of the human diver. The symbol $\mathbf{p}$ denotes the position and orientation (expressed in the quaternion format) of the diver with respect to $\mathcal{F}_{\mathcal{W}}$. The vector $\mathbf{l}$ denotes the physiological states.

A. Human Motion

The diver's swimming motion is decomposed into *limb motion* defined as the motion of limbs relative to a neutral body pose, and *body motion* which is defined as the motion of the entire body relative to the ground. Defining the motion of a human avatar requires specifying both limb and body motion. With motion capture devices, it is feasible to measure joint angles $\mathbf{j}$ in real time that define the body pose and transmit $\mathbf{j}$ to the diver avatar. However, it is prohibitively difficult to measure positions and orientations $\mathbf{p}$ that characterize the absolute body motion in water. In UnRealTHASC, a simplified yet carefully calibrated hydrodynamic model is integrated to generate the swimming motion of the human avatar based on the sampled limb motion. This approach is inspired by the work of Clarke and Gutman on the XR-based skydiving simulator [9], [10]. Additionally, divers wear a buoyancy control device (BCD) to regulate the depth at which these motions are conducted. A combination of all these factors enables the human diver to maneuver freely in the 3D underwater virtual environment $\mathcal{U}$.

1) Real-time Motion Capture of Limb Motion: By modeling the human body as a multi-body system of 16 segments (pelvis, abdomen, thorax, head, upper arms, forearms, hands, upper legs, lower legs, feet) linked by 15 joints, the limb motion of a human is fully characterized by the joint angles, which we denote by $\mathbf{j}$. UnRealTHASC leverages the IMU-based motion capture suit Xsens Awinda to record the limb motion and consequently, reconstruct the 3D body pose. The real human diver straps 17 IMU sensors to different parts of the body as shown in Fig. 2. As the experiment commences, real-time IMU data is streamed to estimate $\mathbf{j}$ and reconstruct limb motions for the diver avatar. These limb motions are also used as input of the hydrodynamic model to generate the body motion, as described in Section.III-A.2. Additionally, UnRealTHASC integrates the MANUS gloves for high-fidelity finger tracking to support the simulation of underwater tasks that involve hand manipulation [11] as well as gesture-based communication [12]. This integration provides complete control of all the joints on the diver avatar.

2) Hydrodynamic Model of Body Motion: For swimming motion, the connection between limb motion and body motion is the hydrodynamic forces and moments, which are generated from through-water limb motion and serve as the driving power of body motion relative to the ground. In UnRealTHASC, a hydrodynamic model is incorporated which adopts empirical formulas to calculate the hydrodynamic forces and moments $\mathbf{f}_{h,k}$ exerted to each body segment k for $k = 1, \cdots, 16$, based on the real-time body pose $\mathbf{j}$. Then a Newton-Euler model about the body motion $\mathbf{p}$ is propagated to generate the body motion of the human avatar. The equations of motion takes the form,

$$M\ddot{\mathbf{p}} + D\dot{\mathbf{p}} + K\mathbf{p} = \mathbf{f}_g + \mathbf{f}_b + \sum_{k=1}^{16} \mathbf{f}_{h,k}(\mathbf{j}, \mathbf{p}, \mathbf{c}),$$

where M is the mass and inertia matrix, D is the generalized damping matrix, K is the generalized stiffness matrix, $\mathbf{f}_h$ denotes the calculated hydrodynamic forces and moments, $\mathbf{c}$ denotes the model parameters, $\mathbf{f}_g$ denotes the gravity, and $\mathbf{f}_b$ denotes the water buoyancy. Since the human diver is lying flat, a simplified model is used in this paper which considers the motion of lower limbs and the rotation of the spine to define the in-plane motions of the diver avatar. The derivation and experiment validation of the hydrodynamic model are omitted in this article due to the page limit.

3) Buoyancy Control Device: The Buoyancy Control Device (BCD) is an inflatable jacket-like wearable device that is connected to the gas cylinders. Regulating the air filled in the BCD regulates the buoyancy force acting on the diver allowing the diver to change diving depth at will. The amount of air in the BCD is adjusted using the hose. The hose has two buttons, one each to gradually increase and decrease the amount of air in the BCD. Additionally, a dump valve is also installed on the BCD to quickly release the air and descend faster. A commercially available BCD was modified as shown in Fig. 2 to support digital buttons. Two of these digital buttons are placed on the inflate and deflate buttons on the hose and the third one is placed near the dump valve of the BCD. These buttons are connected to a Rasberry Pi 4 microcontroller installed inside the BCD which harbors the hose and dump valve. The microcontroller communicates wirelessly with the virtual environment over a User Datagram Protocol (UDP) port. Thus, every time any of these buttons are pressed, the depth of the diver avatar changes based on the functionality assigned to that button.

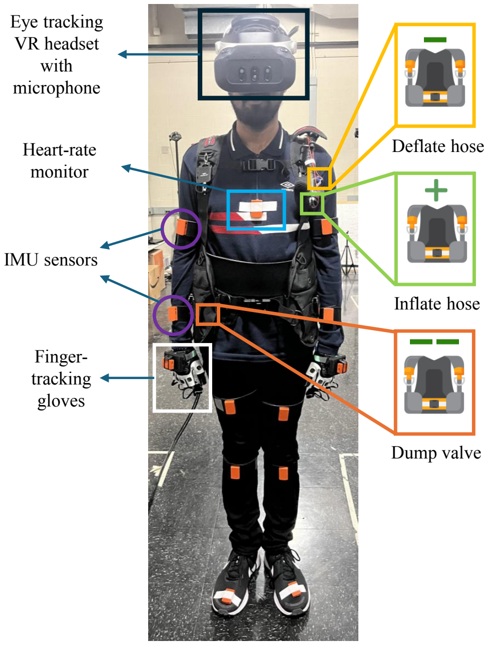


Fig. 2. Real human diver equipped with the physiological sensors, modified BCD, and IMU-based body tracking system

B. Physiological Sensors

1) Eye Tracking: Gaze tracking is the process of estimating the direction of gaze relative to the head of a person. This direction is defined by a 3D vector called the gaze vector, which is estimated by recording the images of the eyes with high-infrared speed cameras. Gaze tracking has been used in HRI applications to communicate with the robot agents non-verbally and predict human intent [13]. In the human-computer interaction (HCI) community, eye tracking is also extensively used for the analysis of human attention and focus [14]. Since the nature of underwater communication is highly non-verbal, real-time gaze tracking serves as an important tool for estimating the human cognitive states, predicting future states $\mathbf{x}$, and latent goals. This is implemented in the UnRealTHASC using the Varjo XR-4 headset. The XR-4 headset uses the OpenXR plugin to stream information about the gaze vector. This gaze vector is then projected on the 2D image frames observed by the real human to obtain gaze tracks in the image frame. The inferred gaze tracks can also be used to infer saccades and fixations.

2) Heart Rate Monitoring: Heart rate monitoring is conducted to analyze the effect of a stressor on the human diver. Polar H10 chest trap is used to measure the heart rate and associated metrics in real time. This chest strap uses ECG technology to monitor heart rate. The data recorded is transmitted over Bluetooth to a Python-based interface which estimates various metrics such as heart rate variability.

3) Breath Monitoring: Breath monitoring plays a pivotal role in predicting stress, fatigue, and the onset of negative physiological conditions such as hypercapnia [15] in technical diving. This facility uses breath sounds to investigate the breathing characteristics of the human diver during experiments in real time. Breath sounds are measured using the microphone installed in the VR headset. These sounds are communicated to a Python script that records 10 seconds of audio from the microphone. A power spectral density analysis of the breath sounds, as described in [16], is conducted to estimate the number of breaths per minute (BPM). The spectral analysis and data recording are conducted in two parallel threads in the CPU, to avoid missing out on audio signals while calculating the BPM estimate for the previous 10 seconds of audio.

C. Dive Computer and VR Interface

In addition to the visual feedback provided by the VR headset, an interface is designed to provide feedback from the BCD and a simulated dive computer worn by the diver avatar. The BCD control input provided by the human diver in the $\mathcal{W}$ is displayed on the VR headset. Dive computers are used by professional divers to track their dives. A decompression model running on these devices uses built-in depth sensors and timers to track the amount of dissolved nitrogen in the body. As a result, these devices calculate the remaining dive time and thus form an important part of the decision-making process while indirectly providing information about the diver's physiological state. The dive computer interface designed for UnRealTHASC simulations integrates the DecoTengu library to run a decompression model with inputs from the virtual environment. The dive computer also acts as an important source of information for the robot agent collaborating with the human. For example, the robot can warn or intelligently disobey the human if a certain task cannot be done because the remaining dive time prohibits its completion. Thus, Python-based socket communication ports are created to stream this information, whenever required, from the virtual environments to the planners running on the robots. The VR diver user interface with the dive computer and BCD feedback is shown in Fig. 3.

IV. Robot XR integration

Since the robot in UnRealTHASC is virtually defined, fidelity of sensor simulation largely determines the fidelity of the downstream planning and control simulations. In this section, we investigate XR simulation of various underwater sensors including imaging sonar and profiling sonar. Then we develop a simple pipeline for simulating robot planning and control. In this section, we denote the motion state of a robot by $\mathbf{s} \in \mathbb{R}^{12}$ which includes its 6D pose and the first-order derivatives of the pose. The geometry of the robot is denoted by $\mathcal{A} \in \mathcal{U}$, and the robot fixed frame is denoted by $\mathcal{F}_{\mathcal{A}}$.

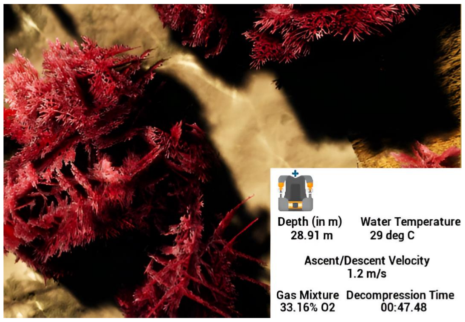

Depth (in m)
28.91 m
Water Temperature
29 deg C
Ascent/Descent Velocity
1.2 m/s
Gas Mixture
33.16% O2
Decompression Time
00:47.48

Fig. 3. Dive computer and BCD control inputs displayed on the VR interface for the real human diver

A. Robot Sensors

Robot and other autonomous agents inside the virtual environment interact with diver avatars and perceive the environment using both exteroceptive and propioceptive sensors. UnRealTHASC builds on the sensor suite developed in RealTHASC [8], which integrated sensor APIs from UnrealCV [17], traditional computer vision (CV) algorithms, and recent advances in CV to provide a wide array of optical sensors in VR. These sensors included RGB cameras, online panoptic segmentation, surface normal maps, depth cameras, and optical flow maps. An RGB-image-based side-scan sonar measurement model developed in [18] is also implemented in RealTHASC. However, underwater environments pose unique challenges for robot perception, navigation, and communication which must rely on a combination of optical and acoustic sensors [19]. HoloOcean [20] is one approach to simulating sonar sensors in UE™ by leveraging the Octree representation. This paper presents an alternative pipeline to augment several existing UE™-based 3D computer vision pipelines such as UnrealCV [17] and AirSim [21] to render sonar images in simulation environments. This section describes a novel approach to simulating imaging sonar or forward-looking sonar, profiling sonar and echo sounder measurements, using depth maps and surface normal maps.

Game engines and recent computer vision algorithms provide depth maps as well as surface normals in real time. Thus it is assumed that the depth map, $\mathbf{D}$, and surface normal map, $\mathbf{N}$, can be accessed depending on the desired sonar sensor frequency required. The algorithm for this approach is summarized in Algorithm. 1 . Firstly, raytracing [22] is used to project each pixel in $\mathbf{D}$ to reconstruct the target geometry, $\mathcal{T} \subset \mathcal{U}$, being observed as a 3D point cloud. Note that this 3D point cloud is defined in the camera or robot reference frame $\mathcal{F}_{\mathcal{A}}$. Now, this point cloud is transformed into spherical coordinates and pruned according to the sensor FOV geometry, denoted by $\mathcal{S} \subset \mathbb{R}^3$, which defines the limits on the range, azimuthal angle, and elevation angle. This pruned 3D point cloud is an estimate, $\hat{\mathcal{T}}$, of the 3D target geometry lying inside the FOV of the sonar sensor $\mathcal{S}$. For every 3D point, $\mathbf{t} \in \hat{\mathcal{T}}$, the surface normal map $\mathbf{N}$ defines a vector $\mathbf{n} \in \mathbb{R}^3$ as a vector containing 8-bit RGB values $[r, g, b]$, where the unnormalized surface normal vector is estimated by using

$$\mathbf{n} = 2 \ \left[\left(\frac{r}{255}\right)^{\gamma} - 1, \left(\frac{g}{255}\right)^{\gamma} - 1, \left(\frac{b}{255}\right)^{\gamma} - 1\right]^T \quad (1)$$

Assuming isotropic sound emission by the acoustic sensor [23], the intensity contribution of every point $\mathbf{t} \in \hat{\mathcal{T}}$ with an associated surface normal vector $\mathbf{n}$ is expressed as,

$$I(\mathbf{t}) = \cos(\alpha) = \frac{\langle \mathbf{n}, \mathbf{t} \rangle}{|\mathbf{n}| \ |\mathbf{t}|} \quad (2)$$

where α defines the angle of incidence of the acoustic ray with respect to the surface normal, intrinsically representing how incoming sound wave is reflected by the surface. All the points in the sensor FOV that do not belong to the target geometry, denoted by $\mathcal{S} \backslash \hat{\mathcal{T}}$), do not contribute to the intensities observed in the measurement. For numerical computations, the sensor FOV is discretized in the spherical coordinate frame according to pre-specified discretizations n_r, n_θ and n_ϕ along the radial, azimuthal, and elevation direction respectively. The location $[r, \theta, \phi]$ of every point $\mathbf{t} \in \hat{\mathcal{T}}$ in this discretized sensor FOV is estimated, and stored in a matrix I with dimensions (n_r, n_θ, n_ϕ) such that

$$I(r, \theta, \phi) = \begin{cases} \cos(\alpha) & \mathbf{t} \in \hat{\mathcal{T}} \\ 0 & \mathbf{t} \in \mathcal{S} \backslash \hat{\mathcal{T}}. \end{cases} \quad (3)$$

Based on the sensor FOV geometry $\mathcal{S}$ and their corresponding measurement model, operations are performed on these intensity contributions I to render the measurement.

1) Imaging sonar: Imaging or forward-looking sonar is a high-frequency acoustic imaging sensor used for tasks such as path planning, localization, and mapping. Measurements from imaging sonar are visualized as polar plots which resolve the range and azimuthal angle of the reflected acoustic waves but do not preserve the elevation angle. Thus imaging sonar measurements suffer from an elevation ambiguity issue. The intensity measurement at a given $[r, \theta]$ is represented using the model,

$$\mathbf{z}(r, \theta) = \int_{\phi_{min}}^{\phi_{max}} I(r, \theta, \phi) d\phi \quad (4)$$

where ϕ_{min} and ϕ_{max} are the minimum and maximum elevation angles respectively, as defined by the imaging sonar FOV geometry. This integration is approximated numerically by adding the intensity contributions for a given r and θ along the elevation axis and dividing by the total number of bins along the azimuth axis to normalize the measurement value.

$$\hat{\mathbf{z}}(r, \theta) = \frac{1}{n_\phi} \Sigma_{n_\phi} I(r, \theta, \phi) \quad (5)$$

2) Profiling sonar: Profiling sonar are acoustic sensors mounted on larger vessels such as ships facing downwards towards the ocean floor. These sensors generate large-scale bathymetric maps of the ocean floor. Similar to the imaging sonar, profiling sonar measurements are also be represented in a polar plot. The FOV geometry is exactly similar to that of imaging sonars however profiling sonar operates at smaller elevation angles and larger azimuthal angles as compared to imaging sonar. Thus, the sensor measurement model is described in (4) has been used to render the profiling sonar measurements.

3) Echo sounder: Echo sounders emit acoustic pulses and measure the echo return time to estimate the ocean floor depth. The FOV geometry for these sensors is characterized by a 3D cone with a specified semi-vertical angle ψ. The intensity contributions of all the points lying inside the FOV of the sensor are integrated along the elevation and azimuthal direction. Thus, the final rendered measurement is the range value measured along the radial direction of the sensor which is mathematically represented as

$$\mathbf{z}(r) = \int_{\theta_{min}}^{\theta_{max}} \int_{\phi_{min}}^{\phi_{max}} I(r, \theta, \phi) d\phi d\theta. \quad (6)$$

The numerical approximation of the measurement in (6) is estimated by taking an average of the intensity contributions along both the azimuth and elevation.

$$\hat{\mathbf{z}}(r) = \frac{1}{n_\theta n_\phi} \Sigma_{n_\theta} \Sigma_{n_\phi} I(r, \theta, \phi) \quad (7)$$

The FOV geometries and corresponding measurements for the three types of sonar sensors are summarized in Fig. 4. Since the operations described in Algorithm.1 are conducted on depth and normal maps generated by the environment with ray-casting, explicit checking for occlusions as well as shadows is not required. Unlike [20], direct ray tracing without marching along each ray is possible and hence computationally more efficient than searching in an Octree grid to find intersections. Another important advantage over existing sonar simulators is that the presented algorithm can be augmented to include the target semantic information to generate semantic segmentation masks for imaging sonar measurements useful for underwater target classification and reconstruction [24]. Owing to the pixel-based rendering model, this novel approach can be applied to any set of depth, normal, and semantic segmentation images to render sonar measurements. Thus, several existing pixel-based datasets can be augmented to synthesize sonar measurements to train neural networks for sonar sensors, without conducting resource-intensive data-collection experiments for real and synthetic sonar sensor data.

B. Robot Planning and Control

To integrate robot dynamics and environmental physics, UE™Physics Engine PhysX is leveraged for physical simulation computations. Each UUV is modeled as a rigid body in the virtual world with the thrusters as the control inputs. The location of each thruster is defined with respect to the robot body frame $\mathcal{F}_\mathcal{A}$ to resolve torques on the body due to the thrust forces. For instance, in the case of an underwater remotely operated vehicle (ROV), the positions of the four horizontal thrusters and four vertical thrusters are specified in the robot model. The drag force and buoyancy force acting on

```
Algorithm 1: Render acoustic measurement from point cloud $\hat{\mathcal{T}}$ and surface normal map $\mathbf{N}$
Input: $\hat{\mathcal{T}}$, N
Output: $z$
$I = \mathbf{0}_{n_r \times n_\theta \times n_\phi}$
for $i = 0;\ i < m;\ i = i + 1$ do
    $r_i, \theta_i, \phi_i = Cartesian2Spherical(\mathbf{p}_i) \quad \mathbf{p}_i \in \hat{\mathcal{T}}$
    if $r_{min} < r_i < r_{max}$ and $\theta_{min} < \theta_i < \theta_{max}$ and $\phi_{min} < \phi_i < \phi_{max}$ then
        $r_i^d, \theta_i^d, \phi_i^d = DiscretizedSpherical(r_i, \theta_i, \phi_i, n_r, n_\theta, n_\phi)$
        $I(r_i^d, \theta_i^d, \phi_i^d) = d_i$
```

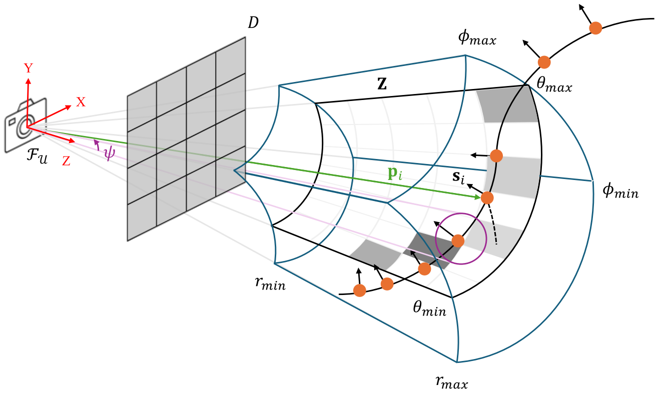


Fig. 4. Acoustic sonar measurements rendered from surface normal maps S and depth maps D. The field-of-view geometry of an imaging sonar and profiling sonar is characterized by a 3D sector (blue) while an echo sounder FOV (pink) is characterized by a 3D cone with vertical angle ψ

the robot agent are modeled by specifying linear and damping constraints, water density, and considering the physical attributes of the agent such as the mass, volume, center of mass, and center of buoyancy of the rigid body defining the ROV. Consequently, these forces and torques are provided to the PhysX engine that simulates the motion of the robot in the virtual world. The robot agents employ a Proportional-Integral-Derivative (PID) control with a waypoint-following policy to maneuver in the virtual environment. Given a current state of the robot $\mathbf{s}$ and a desired state of the robot $\mathbf{s}^*$, the control input $\mathbf{u}$ is defined as:

$$\mathbf{u} = K_P\ (\mathbf{s}^* - \mathbf{s}) + \int K_I\ (\mathbf{s}^* - \mathbf{s})\ dt - K_D\ \dot{\mathbf{s}} \qquad (8)$$

where, K_P, K_I and K_D are the proportional, integral, and derivative gains respectively. These control inputs are clipped by control bounds to avoid unrealistic motion trajectories. Finally, the force and torque values to be applied by PhysX are calculated by considering the control commands as the linear and angular accelerations while taking into account the physical attributes of the robot.

V. Communication

Although UnRealTHASC aims at facilitating real-time interaction between the human diver and simulated robot agent, it can also be leveraged for offline tasks such as learning interaction models using training data from the environment. An overview of the communication framework that supports these interactions is shown in Fig. 5. The virtual environment is hosted by an Alienware Aurora R13 system and handles the graphics rendering requirements of the environment. The human diver in the laboratory is equipped with eye tracking VR headset which, over a Local Area Network (LAN) connection, communicates the gaze vector to the virtual environment while acting as a source of visual feedback for the diver. The IMU modules attached to the human diver transmit joint angles and angular velocities at 120 Hz over Wi-Fi to the human motion model which calculates the body position and velocities of the avatar in the virtual environment. These body positions and velocities are streamed over a UDP connection to the virtual environment. All physiological sensors wirelessly transmit data to a Python script which transmits these values to the virtual environment. The buttons provided on the BCD are also connected directly to a Raspberry Pi 4 microcontroller which publishes the boolean arrays, interpreted as presses, over WiFi. The virtual environment consists of the virtual robot agent equipped with the dynamics, control strategies, and planners for waypoint following. This robot agent is controlled by a Python script which acts as a remote or local client to the virtual environment based on the type of application setting that is being simulated. This client can, synchronously or asynchronously, access all the sensors deployed in UnRealTHASC and can add controllable and simulated delays in sensor data transmission which is often the case in real-world experiments.

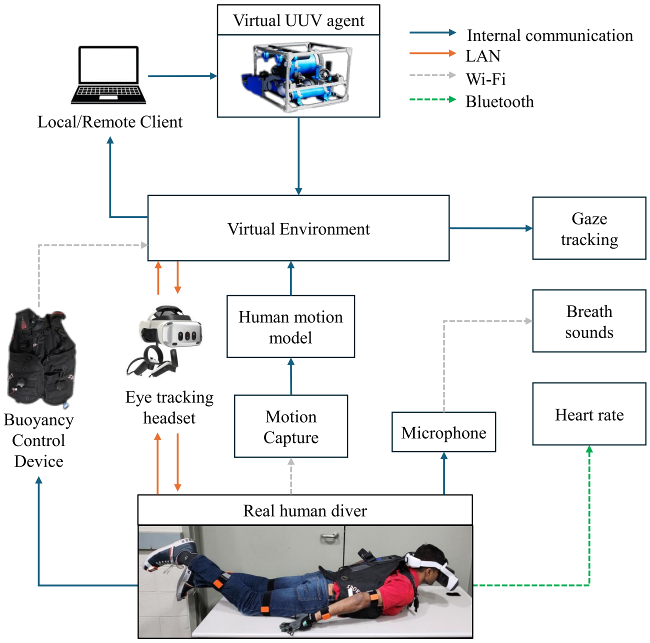


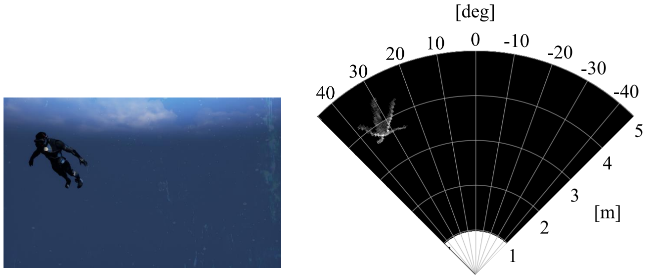
Fig. 5. Real-time communication between various real and virtual agents as well as sensors operating in UnRealTHASC.

VI. Experiments

This section presents two experiments designed to showcase some functionalities of the real and virtual sensors in the UnRealTHASC facility and demonstrate how data from these sensors can be used to simulate underwater human-robot teaming scenarios. Specifically, the first experiment showcases the novel sonar rendering approach and the second experiment demonstrates how physiological sensors can be used as an alternate medium of implicit human-robot communication.

A. Sonar Sensor Measurements

This experiment is conducted to demonstrate how the presented sonar measurement rendering approach can be used to simulate measurements for imaging sonar, profiling sonar, and echo sounder. In the first part of this experiment, an ROV observing a diver avatar through an onboard imaging sonar sensor is considered as shown in Fig. 6. The sonar range is assumed to be between 1 meter to 5 meters while the azimuthal and elevation aperture angles are assumed to be 90°and 28°respectively. In the second part of the experiment, a profiling sonar mounted on a ship vessel is employed to create a bathymetric map of the environment. The profiling sonar measurement at a certain time-step is shown in Fig. 7. The azimuthal range is 90°while the elevation is small (0.5°) representing a thin but wide swath measuring the ocean floor. In the final experiment, an ROV equipped with an echo sounder is tasked with measuring the ocean floor depth. The ROV slowly moves down towards the floor at a speed of 5 m/min, as shown in the range-time graph shown in Fig. 8. The sensor FOV is discretized into 500 bins along each axis of the spherical coordinate frame in all the three experiments.

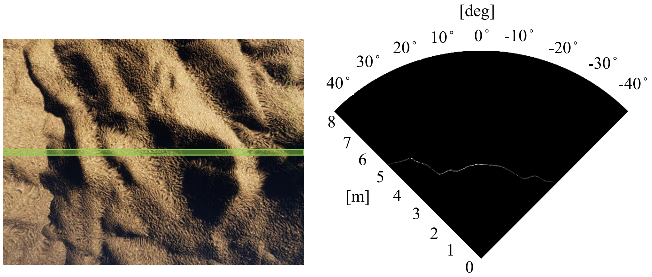


(a) (b)

Fig. 6. (a)ROV is collaborating with a diver by (b) observing the diver inside the FOV of the imaging sonar sensor mounted on it and (c) rendering an imaging sonar sensor measurement.

B. Physiological Sensor Integration with the Virtual Environment

This experiment is conducted to demonstrate the use of gaze tracking to enable human-robot collaboration in a task. A team of human diver and an ROV is tasked with searching the workspace around a sunken submarine. The human diver wearing the eye-tracking VR headset suddenly observes a shark. A Faster RCNN-based detector [25] processes a fixed-size bounding box around the gaze track to identify whether the human is observing any stressors. The detector detects that the human is observing a shark with 96% accuracy. Consequently, the category of the stressor, “shark” in this instance, and its relative position with respect to the diver is sent to the ROV. The onboard ROV planner is programmed to conduct certain maneuvers when a certain type of stressor is reported by the detector. Without explicit communication from the human diver, the ROV is autonomously informed about the shark and its relative position. Within 3 seconds, the ROV conducts a defensive maneuver during which it positions itself between the shark and the diver avatar to continuously observe the movements with the onboard forward-looking sonar of the shark giving the human diver time to escape or conduct some other task safely. This experiment is summarized in Fig. 9.



(a) (b)

Fig. 7. Profiling sonar mounted on a ship vessel (a) scanning the seafloor and (b) recording the depth

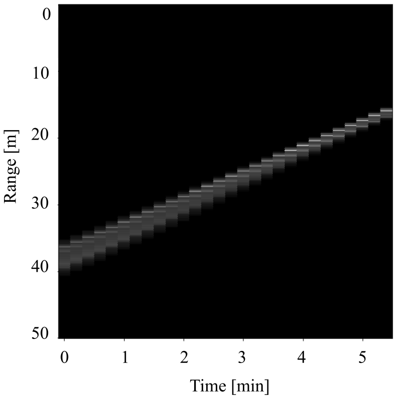


Fig. 8. Range measured by an echo sounder mounted on a UUV moving vertically down towards the seafloor at 5 m/min.

VII. Conclusion

This paper presents UnRealTHASC, an XR-testbed to study underwater HRI and develop human-centric collaboration strategies for teaming. Seamless integration of motion capture technology, embedded systems, as well as underwater dynamics modeling facilitates realistic human and robot motion generation in the virtual environment. Various physiological sensing modules are integrated to enable the robot agent to analyze human diver performance, infer

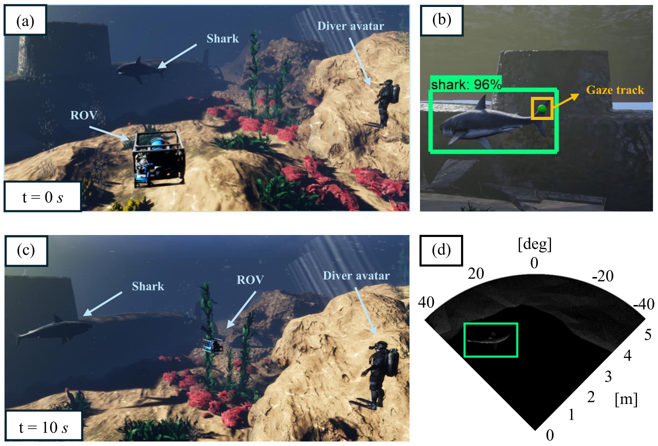


Fig. 9. (a) The diver avatar observes a shark swimming (b) A bounding-box image around the gaze track is used by the detector to identify the shark (c) ROV receives the information from the detector and positions itself in between the diver and the shark (d) The ROV continues to observe the shark with the onboard forward-looking sonar sensor

and predict diver states, and better inform robot decision-making for collaboration. Communication pipelines enable real-time data acquisition from real and virtual sensors while seamlessly interfacing with real and virtual workspaces. A novel method to render real-time acoustic measurements in simulation and post-process existing datasets is developed. Future work will include qualitative studies with novice and expert divers to evaluate the effectiveness of the testbed. This facility will be leveraged to develop and conduct studies on physiological sensor-driven robot decision-making, as well as acoustic perception-enabled HRI strategies for mapping, localization, and collaborative planning.

VIII. Acknowledgements

This research was funded by the Office of Naval Research Undersea Medicine Grant N00014-23-1-2469. The authors would like to thank Daniele Fragiacomo for his contributions on technical scuba-diving subject matter and experiments.

References

[1] Diver Alert Network. Annual diving report. `https://www.dansa.org/annual-diving-report`.

[2] Rogelio Morales, Peter Keitler, Patrick Maier, and Gudrun Klinker. An underwater augmented reality system for commercial diving operations. In *OCEANS 2009*, pages 1–8. IEEE, 2009.

[3] Paul O'Connor, Angela O'Dea, and John Melton. A methodology for identifying human error in us navy diving accidents. *Human factors*, 49(2):214–226, 2007.

[4] Oceanographic Staff. Humans to permanently live underwater from 2027. `https://oceanographicmagazine.com/news/deep-making-humans-aquatic/`.

[5] Andreas Birk. A survey of underwater human-robot interaction (u-hri). *Current Robotics Reports*, 3(4):199–211, 2022.

[6] Nikola Mišković, Marco Bibuli, Andreas Birk, Massimo Caccia, Murat Egi, Karl Grammer, Alessandro Marroni, Jeff Neasham, Antonio Pascoal, Antonio Vasilijević, et al. Caddy—cognitive autonomous diving buddy: Two years of underwater human-robot interaction. *Marine Technology Society Journal*, 50(4):54–66, 2016.

[7] Jeffrey Phillips and Connor Tate. Development and validation of an underwater occulometric assessment tool. in press.

[8] Andre Paradise, Sushrut Surve, Jovan Clive Menezes, Madhav Gupta, Vaibhav Bisht, Kyung Rak Jang, Cong Liu, Suming Qiu, Junyi Dong, Jane Shin, et al. Realthasc-a cyber-physical xr testbed for ai-supported real-time human autonomous systems collaborations. *Frontiers in Virtual Reality*, 4:1210211.

[9] Anna Clarke and Per-Olof Gutman. An automatic control system with human-in-the-loop for training skydiving maneuvers: Proof-of-concept experiment. *International Journal of Human-Computer Studies*, 170:102960, 2023.

[10] Anna Clarke and Per-Olof Gutman. A dynamic model of a skydiver with validation in wind tunnel and free fall. *IFAC Journal of Systems and Control*, 22:100207, 2022.

[11] Dhruv Jain, Misha Sra, Jingru Guo, Rodrigo Marques, Raymond Wu, Justin Chiu, and Chris Schmandt. Immersive scuba diving simulator using virtual reality. In *Proceedings of the 29th annual symposium on user interface software and technology*, pages 729–739, 2016.

[12] Arturo Gomez Chavez, Andrea Ranieri, Davide Chiarella, Enrica Zereik, Anja Babić, and Andreas Birk. Caddy underwater stereo-vision dataset for human-robot interaction (hri) in the context of diver activities. *Journal of Marine Science and Engineering*, 7(1):16, 2019.

[13] Svetlin Penkov, Alejandro Bordallo, and Subramanian Ramamoorthy. Inverse eye tracking for intention inference and symbol grounding in human-robot collaboration. In *Robotics: Science and Systems (RSS), Workshop on Planning for Human-Robot Interaction*, pages 5–7, 2016.

[14] Yuehua Wang, Shulan Lu, and Derek Harter. Multi-sensor eye-tracking systems and tools for capturing student attention and understanding engagement in learning: A review. *IEEE Sensors Journal*, 21(20):22402–22413, 2021.

[15] Jeffrey Phillips, Arash Mahyari, Ian Perera, Adrien Moucheboeuf, Madison McInnis, Anil Raj, Allison Bew, and Andrew Dorsey. Identification of hypercapnia through voice analysis and associated neurological and performance effects. Technical Report AFRL-RH-WP-TR-2021-0108, Air Force Research Laboratory, Ohio, 2021.

[16] Yunyoung Nam, Bersain A Reyes, and Ki H Chon. Estimation of respiratory rates using the built-in microphone of a smartphone or headset. *IEEE journal of biomedical and health informatics*, 20(6):1493–1501, 2015.

[17] Weichao Qiu and Alan Yuille. Unrealcv: Connecting computer vision to unreal engine. In *Proc. of the Computer Vision – ECCV 2016 Workshops*, pages 909–916. Springer International Publishing, 2016.

[18] Jane Shin, Shi Chang, Matthew J. Bays, Joshua Weaver, Thomas A. Wettergren, and Silvia Ferrari. Synthetic sonar image simulation with various seabed conditions for automatic target recognition. In *OCEANS 2022, Hampton Roads*, pages 1–8, 2022.

[19] Shahriar Negahdaripour, Hicham Sekkati, and Hamed Pirsiavash. Opti-acoustic stereo imaging: On system calibration and 3-d target reconstruction. *IEEE Transactions on Image Processing*, 18(6):1203–1214, 2009.

[20] Easton Potokar, Kalliyan Lay, Kalin Norman, Derek Benham, Tracianne B. Neilsen, Michael Kaess, and Joshua G. Mangelson. Holoocean: Realistic sonar simulation. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8450–8456, 2022.

[21] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics*, 2017.

[22] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, USA, 2 edition, 2003.

[23] Eric Westman and Michael Kaess. Wide aperture imaging sonar reconstruction using generative models. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8067–8074, 2019.

[24] Dongdong Zhao, Weihao Ge, Peng Chen, Yingtian Hu, Yuanjie Dang, Ronghua Liang, and Xinxin Guo. Feature pyramid u-net with attention for semantic segmentation of forward-looking sonar images. *Sensors*, 22(21):8468, 2022.

[25] J. Jenrette, Z. Y.-C. Liu, P. Chimote, T. Hastie, E. Fox, and F. Ferretti. Shark detection and classification with machine learning. *Ecological Informatics*, 69:101673, 2022.