Diver-Robot Communication Glove Using Sensor-Based Gesture Recognition

Christopher R. Walker, Đula Nađ, Derek W. Orbaugh Antillon, Igor Kvasić, Samuel Rosset, Nikola Mišković, *Senior Member, IEEE*, and Iain A. Anderson

Abstract*—Despite the popularity of the buddy system, where a pair of divers look out for each other, scuba diving still poses significant risks with 15–30 deaths per 100 000 divers per year. In the 2015 edition of "*Diving Medicine for Scuba Divers*," it is reported that 86% of diving fatalities involve a lone diver, yet divers still elect to dive alone for a variety of reasons including buddy unavailability. A solution to this may lie in a dedicated autonomous dive-buddy robot, which could also operate as an assistant. The challenge is how would the diver interact and communicate with the robot in underwater conditions often involving low visibility. In this article, we detail and demonstrate a technological solution that involves the direct motion capture of fingers and hand using a smart dive glove with integrated stretch sensors alongside an inertial measurement unit. The hardware and gesture recognition protocol were initially evaluated in a lab-based study where 10 users participated in a virtual reality environment setup and interacted with a virtual robot using gestures. In total, 140 out of 175 gestures were correctly recognized, resulting in diver-robot commands. A full sea demonstration of the underwater motion-capture system followed with the underwater dive glove used to send gesture-based acoustic commands to a dedicated dive buddy, a modified autonomous underwater vehicle. The diver was able to command the robot to move in six directions with a total delay of 6 s or less between the beginning of the gesture and the initiation of the action by the robot.

***Index Terms*—Autonomous vehicle navigation, gesture, posture, and facial expressions, sensor-based control, underwater (UW) communication.**

I. Introduction

FROM the earliest days of diving, before the invention of the demand regulator by Cousteau, Dumas, and Gagnan in 1942, divers have largely relied on hand gestures for diver-to-diver communication. Of over 240 recorded gestures [1], about two dozen are in common use today by divers for conveying distress, suggested swimming direction, comfort/discomfort, and threats. This is despite the availability of voice "through water communications systems" (TWCS) for which there are two forms: 1) human voice transmission directly from a submerged transducer at a strength for which the dive team can hear it directly through their ears [2], or 2) diver in-helmet microphone that converts the message to a higher frequency for direct transmission through the water to the ship or dive buddy where it is converted back to the normal voice frequency range [2], [3]. Drawbacks to TWCS include system price and reduction of stealth. For the latter, we refer to environmental noise contamination that will inhibit the unobtrusive study of timid marine life. In-helmet TWCS also requires unimpeded movement of the lips in a gas environment, thus requiring expensive full-face-mask systems. Therefore, the use of hand gestures for communication, taught to every novice diver, remains central to diver-to-diver communication in the inherently dangerous underwater (UW) environment.

To mitigate danger, the diving fraternity has developed the buddy system, where a pair of divers look out for each other, monitoring their partner's safety, air supply, and decompression time. Each diver's self-contained underwater breathing apparatus (SCUBA) system also provides an immediate back-up for the other. Gesture-based communication plays an important role in the buddy system—a large fraction of the common gestures used by divers involve safety, air supply, and navigation. The critical importance of the buddy system was highlighted in a study of the causes for 947 SCUBA diving fatalities over 11 years (1992–2003) [4]. The three highest triggers of a situation resulting in death were associated with insufficient gas (41%), entrapment (20%), and equipment problems (15%). They also reported that "57% of decedents who began with an assigned buddy were separated prior to death." It is our own experience that buddy separation can occur for a number of reasons that include poor visibility, currents, and distractions associated with the mission. It is reasonable to assume that the three high-incidence causes listed above could have been mitigated or overcome with an effective buddy in close proximity and in good communication.

Due to buddy unavailability, some divers elect to dive alone. And many UW photographers, with a buddy nearby that is distracted by a photographic subject, are effectively buddyless. A solution to this problem could lie in a dedicated dive-buddy robot. Such a robot could also operate as an assistant that can

Manuscript received 11 July 2022; revised 17 December 2022; accepted 3 April 2023. This work was supported by the United States Office for Naval Research under Grant N000141812011. *(Corresponding author: Christopher R. Walker.)*

Associate Editor: J. Potter.

Christopher R. Walker, Derek W. Orbaugh Antillon, Samuel Rosset, and Iain A. Anderson are with the Biomimetics Laboratory, Auckland Bioengineering Institute, University of Auckland, Auckland 1010, New Zealand (e-mail: cwal192@aucklanduni.ac.nz; dorb476@aucklanduni.ac.nz; s.rosset@auckland.ac.nz; i.anderson@auckland.ac.nz).

Đula Nađ, Igor Kvasić, and Nikola Mišković are with the Laboratory for Underwater Systems and Technologies, Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia (e-mail: dula.nad@fer.hr; igor.kvasic@fer.hr; nikola.miskovic@fer.hr).

Digital Object Identifier 10.1109/JOE.2023.3265634

read and respond to diver hand gestures. Gesture recognition has been proven to work for above-the-surface human–robot communication [5].

A UW buddy robot was described by Georgiades et al. [6]: AQUA, capable of paddle swimming and walking that could be deployed autonomously direct from a beach to a UW inspection site and back. AQUA was based on the design of a hexapod walking vehicle [7] but where the rotating feet also serve as fins for pushing water backward. The AQUA robot was guided by servos, following visual patterns. The on-board camera system could also relay gestures from the diver to the surface for interpretation and issuance of commands back down to the robot.

A diver that is in the water alongside a robot such as AQUA would be more situationally aware than a surface operator. Thus direct diver-robot communication is more desirable. To facilitate this, Sattar et al. [8] developed a visual system for AQUA using barcode-like fiducial markers. However, preproduced markers have limited potential for conveying anything other than a small subset of possible commands. To take advantage of the large capacity for gesture-based information transfer, a diver-robot language has been developed under the auspices of the European Union FP7 funded "Cognitive Autonomous Diving Buddy" [9], [10]. Recent work has focused on extracting gesture data from camera data [11], [12], [13], [14]. The potential for diver-robot gesture-based communication with some robots to diver feedback was demonstrated by DeMarco et al. [15] who used a robot that was carrying a camera to stream gesture and other cues to the surface. Acknowledgment was communicated with the diver through dimming of the robot's lights. Again, high cognitive decisions were made on the surface.

A clear requirement for the success of vision approaches for gesture-based communication includes the need for the diver to face the camera and be in close proximity with clear uneclipsed vision of hand signals. This can be near impossible due to blurring and scattering effects as well as low contrast in UW imaging [16]. Further requirements include the capability of the system to accurately capture hand motion, discriminate between general hand movement that can involve swimming, grasping or holding objects and discrete gestures for communication, and the ability to process gestures into commands and transfer command data in a timely manner.

Wearable inertial measurement units (IMU) or strain-bending sensors can provide a means for gesture-based human–robot communication [17]. This will avoid some of the pitfalls of direct line-of-sight vision-based technologies. For direct measurement of finger motion, stretchy capacitive dielectric elastomer (DE) sensors provide a proven transduction mechanism for human motion capture. These are silicone rubber based stretchable capacitors that are linear with uniaxial stretch [18]. They have previously been demonstrated for human motion capture in dry environments [19], [20] and can work effectively as motion capture sensors UW [21], [22]. Other gloves have implemented DE sensors to capture hand motion in the past. Böse et al. [23] fabricated a glove for operational control not only using DE strain sensors but DE pressure and contact sensors on the fingertips as well. A later glove from the same group expanded on this to also include a sliding DE sensor design for analogue intuitive control [24]. Another group developed a novel DE force sensor for use in a glove prototype [25] for improving the desired motion of workers in an industrial context. There is also an opportunity to combine IMU with stretch sensor technologies.

Any method to characterize gestures from finger-hand motion will require the identification of the start and end points of a meaningful motion pattern and the segmentation of a gesture from this, against the variation in shape and duration, even from the same gesturer [26]. Information will be sent from the glove preferably to an untethered robot. For this, we can use acoustic signals [27]. A big advantage to acoustics is that there is no need for direct line-of-sight communication, and no need to face the robot directly and it is almost impervious to water visibility. However, the bandwidth of acoustic signal transmission is severely limited: typically Kilobits per second compared with optical at Gigabits per second [27]. The range of optical systems, however, can be limited to a few meters in coastal waters, making it largely unsuitable for buddy purposes [28]. The potential amount of information required to recognize a single gesture from a five-sensored glove could substantially swamp an acoustic system. It is therefore sensible to process raw sensor signals close to the source and keep acoustic transmission down to critical messages. This method also reduces unnecessary acoustic noise in the environment, as transmission will only occur when the user wishes to send a message.

Here, we present a smart glove for diver-robot communication that overcomes the limitations of acoustic signal transfer and that fuses two of the motion-capture technologies: direct strain sensing with IMU. The glove is smart in the sense that the stretch and inertial data are locally interpreted by dedicated electronics from known gestures that are turned into specific commands telemetered directly to the robot buddy using an acoustic modem integrated with part of the glove electronics. Motion is effectively measured using DE sensors and an IMU. To enable the conversion of a gesture stream to a command, we have produced a gesture library and have developed an algorithm for the recognition of discrete gestures and the conversion of a gesture stream to a command that is transferred by an acoustic signal to the robot, thus minimizing the bandwidth usage of the acoustic channel.

The article describes two activities for the development and testing of the glove system: 1) an initial lab-based study where a virtual reality (VR) glove was fabricated for use in a dry environment to develop the gesture library, validate the recognition protocol, and establish a baseline for accuracy and recognition delay; and 2) a full sea demonstration of the system in the water with a UW dive glove with capability for acoustic command transmission to a dedicated dive buddy, a modified autonomous underwater vehicle (AUV). This first example of a DE-based motion capture/acoustic telemetry approach overcomes limitations of turbidity and the need for direct facing or for close proximity to the UW robot.

II. Methods

A. General Description

The basic concept of gesture-based communication described in this article comprises a sensor-integrated glove recognizing

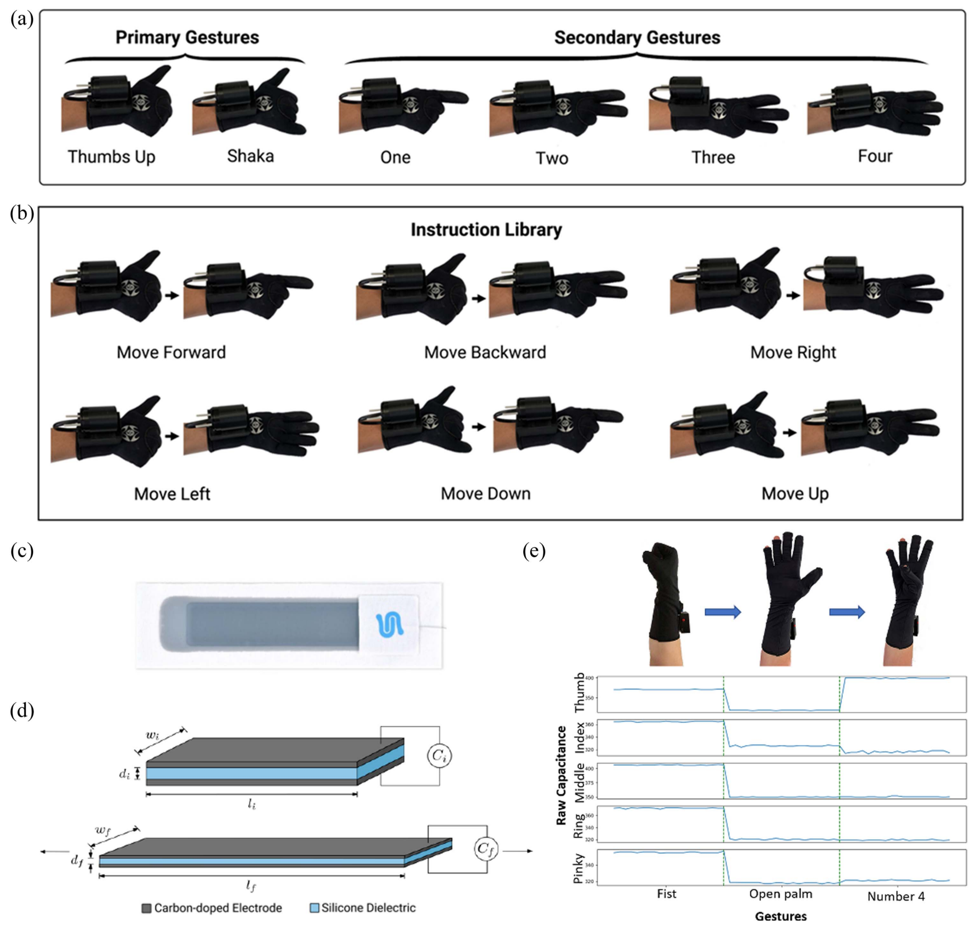


Fig. 1. (a) Individual gesture library. (b) Instruction library. (c) DE sensor. (d) Illustration showing the changes in geometry of the sensor with stretch. (e) Sensor capacitances as a user performs the "fist," "open palm," and "four" gestures.

performed gestures, interpreting these as messages with onboard electronics, and transmitting them to the robot as an instruction. The gesture library [see Fig. 1(a)] consisted of six different gestures in total, with two used as primary gestures and four as secondary gestures. To form an instruction for the robot, a combination of two gestures was required; the user first performed a primary gesture and followed this with the desired secondary gesture [see Fig. 1(b)]. Requiring two gestures per instruction was implemented to improve accuracy, reduce false positives, and allow for a greater number of instructions without increasing the number of recognized gestures.

In terms of sensing technologies, both IMU and DE sensors were used, but for different purposes. The IMU allowed the user both to initiate and terminate/cancel the gesture recognition processes of the glove by simply rotating the wrist side-to-side while performing the "Shaka" gesture. This initiation process allowed the diver to use the glove as they would a normal glove until gesture recognition was required. Additionally, if the user performed an incorrect primary gesture, they were able to cancel the recognition and start again. To capture gestures, however, finger motion of the user was tracked with a DE strain sensor sewn into each finger and thumb. These sensors are stretchable capacitors with carbon-doped silicone rubber electrode layers sandwiching compliant silicone rubber dielectrics. Sensors were bonded to a fabric backing for sewn attachment to the glove [see Fig. 1(c)]. They were purchased commercially (StretchSense, New Zealand) with an additional layer of silicone applied in-house to reduce water ingress. When stretched uniaxially [see Fig. 1(d)], the electrode area increases, and electrode separation decreases resulting in an increase in capacitance represented by the equation for a parallel plate capacitor

$$C = n\frac{\varepsilon_r \varepsilon_0 l w}{t} \tag{1}$$

where ε_r is the relative dielectric constant of the material, ε_0 is the permittivity of the free space, l, w, and t are the sensor length, width, and thickness, respectively, and n is the number of dielectric layers. By sewing the sensors into the glove, finger flexion mechanically deforms the sensor by applying uniaxial stretch. Fig. 1(e) shows the capacitance of each sensor changing as the user performs the "fist," "open palm," and "four" gestures.

B. Algorithm

Capacitances from each strain sensor were measured by a commercially available capacitance measure circuit (StretchSense, New Zealand). Each time a glove was used, by the same or a new user, a calibration process was performed to obtain custom recognition thresholds. This was a critical step

TABLE I
EACH INDIVIDUAL GESTURE PERFORMED AS WELL AS TO WHAT QUINTILES EACH DIGIT STRETCHED THE RELATIVE DE SENSOR

Gesture	***Lookup table values***	***Gesture***	***Lookup table values***
 ***Thumbs up***	*{0,3-4,3-4,3-4,3-4}*	 ***Two***	*{2-3,0,0,3-4,3-4}*
 ***One***	*{2-3,0,3-4,3-4,3-4}*	 ***Three***	*{2-3,0,0,0,3-4}*
 ***Shaka***	*{0,3-4,3-4,3-4,0}*	 ***Four***	*{3-4,0,0,0,0}*

The quintile vales from left to right run from thumb to pinky.

as previous investigations have shown up to 17% variance in capacitive signals from different users performing the same gesture [29]. The calibration process, shown in Fig. 1(e), requires the user to perform the three gestures designed to stretch and relax the sensors between their maximum and minimum capacitance values. The maximum and minimum capacitances measured during the calibration process were used in (2) to normalize live sensor readings into quintiles

$$mapped_{value} = int\left(\frac{(C_s - C_{\min}) * (4.99 - 0)}{(C_{\max} - C_{\min})}\right) \quad (2)$$

where C_s is the sensor capacitance, $mapped_{value}$ is the quintile the sensor capacitance is mapped to, and *int* is a function that transforms the input into an integer by simply removing values after the decimal point. Each individual gesture was then characterized based on which quintile range each finger position fell into (see Table I), with "0" representing the unstretched sensor (lower quintile) and "4" representing the maximum stretch relative to the users' calibration process. For example, when a "Thumbs Up" gesture was performed, the thumb-based sensor is relatively unstretched and therefore its capacitances would fall into the lowest quintile, 0. The four finger-based sensors, however, are all near their maximum stretched positions and would fall into the higher quintiles, 3 or 4. Different users perform the same gesture slightly differently; therefore, by allowing for a range of quintiles in Table I, gestures were more likely to be recognized across multiple users.

The gesture recognition algorithm (see Fig. 2) looped continuously, reading sensor values, normalizing data using (2), and comparing the mapped quintiles with values stored shown in Table I. A gesture was recognized if the mapped values of each finger matched the corresponding gesture for 500 ms. This helped to reduce instances of false-positive recognitions and allowed the algorithm to differentiate between intentional finger gesture motion from motion associated with swimming or tasks.

C. VR Environment

A VR environment and VR glove were developed to help validate the recognition algorithm across a number of users and provide a baseline for the UW experiment to compare against in terms of accuracy and recognition delay. The lycra-based glove had integrated DE sensors on each finger with discrete electronic components housed on the back of the wrist [see Fig. 3(a)]. In terms of electronics, the glove comprised of a capacitance measure circuit for sensor measurement, a microcontroller for data processing, an IMU for movement analysis, a Bluetooth circuit for instruction transmission, a Li-Po battery, and a haptic motor for feedback.

The gesture-based communication process had the user holding a primary gesture (for at least 500 ms) until two short (300 ms) vibrations occurred. This feedback confirmed recognition of a primary gesture, indicating the user could proceed to a secondary gesture or cancel the recognition. After holding a secondary gesture (for at least 500 ms), the instruction [see Fig. 1(b)] was transmitted over Bluetooth to a nearby computer and confirmed to the user via two longer (500 ms) vibrations. Therefore, the minimum required time was 500 ms for an individual gesture to be recognized and 1600 ms (two 500-ms recognitions separated with 600 ms of feedback) for an instruction to be recognized excluding feedback given after transmission of the instruction. The minimum time required for each step in the recognition process is summarized in Fig. 3(b), which excludes the final feedback as this occurs after transmission.

A UW VR environment was developed in Unity with a CAD model [Fig. 3(c)] of the AUV and humanoid glove [Fig. 3(d)]. The VR setup included the HTC Vive (Vive, New York) headset [Fig. 3(e)] and VR glove [Fig. 3(a)], which was transmitting instructions over Bluetooth into the Unity platform. Users, while wearing the setup, were able to move the virtual glove and interact in real time with the virtual AUV [see Fig. 3(f) and (g)] in a UW scene.

The laboratory VR experiment involved 10 participants casually interacting with the virtual AUV. The users were able to instruct the AUV to move in the horizontal plane (left, right, forward, and backward). After placing the VR glove on, participants performed the calibration process [see Fig. 1(e)], customizing the recognition algorithm to each participant. With the HTC Vice headset on, each participant had access to a virtual instruction library menu and started interacting with the robot

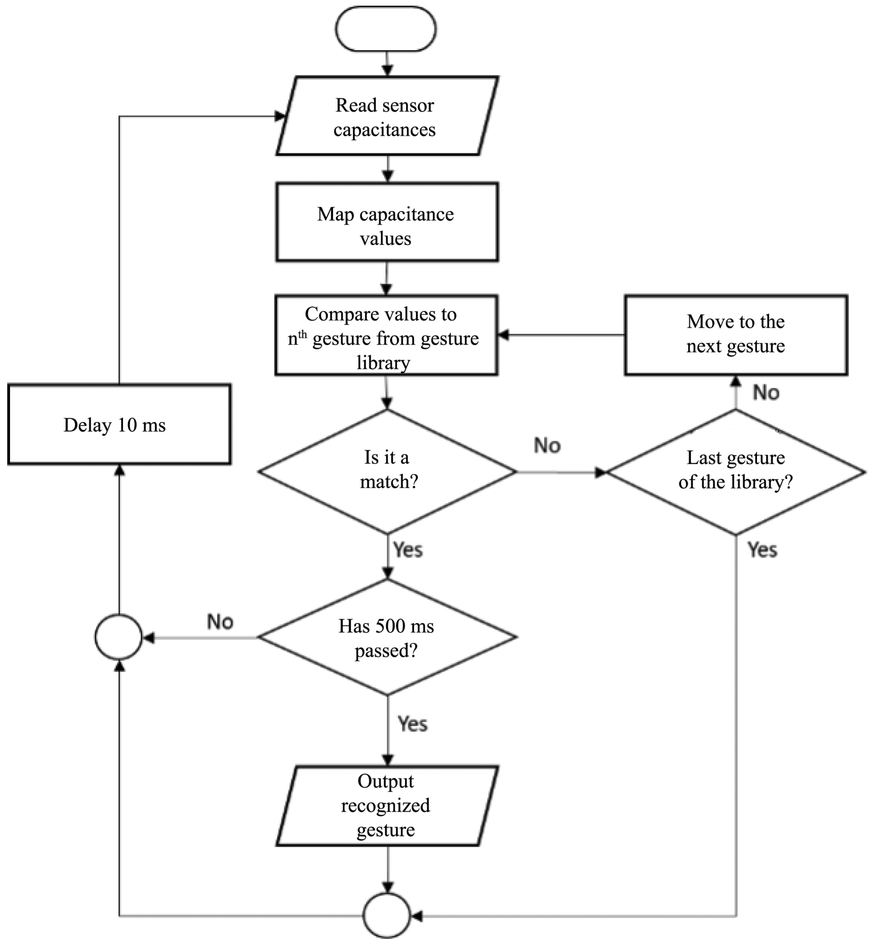


Fig. 2. Algorithm flowchart showing the data flow from sensor readings to a recognized gesture.

TABLE II
VR Glove and Experiment; Recognition and Time Delay Statistics for the 175 Gestures of the 10 Participants Broken Down by Gesture Type

Gesture name	Success/Failure	Percentage accuracy	False recognition	Average recognition time (ms)
Thumbs Up	78/20	79.6%	1	579
One	20/10	66.7%	0	780
Two	12/1	92.3%	0	1749
Three	16/4	80.0%	1	571
Four	14/0	100%	0	1008

at their own speed. Video footage and screen recordings of each participant were analyzed with 175 gestures recorded in total. Each gesture was timestamped from when the participant finished transitioning to the gesture position to when it was recognized by the system. Furthermore, each instruction was also timestamped from when the user finished transitioning to a primary gesture position, to when the secondary gesture was recognized by the system. The results of the VR experiment are shown in Tables II and III.

D. UW Environment

The UW experiment was a field trial performed to demonstrate gesture-based communication in a marine environment. The demonstration involved one participant scuba diver who was trained in the VR experiment on the gesture library. The UW glove used the same horizontal-plane movement instructions as the VR glove; however, the instructions for "up" and "down" were added specifically for this demonstration as the

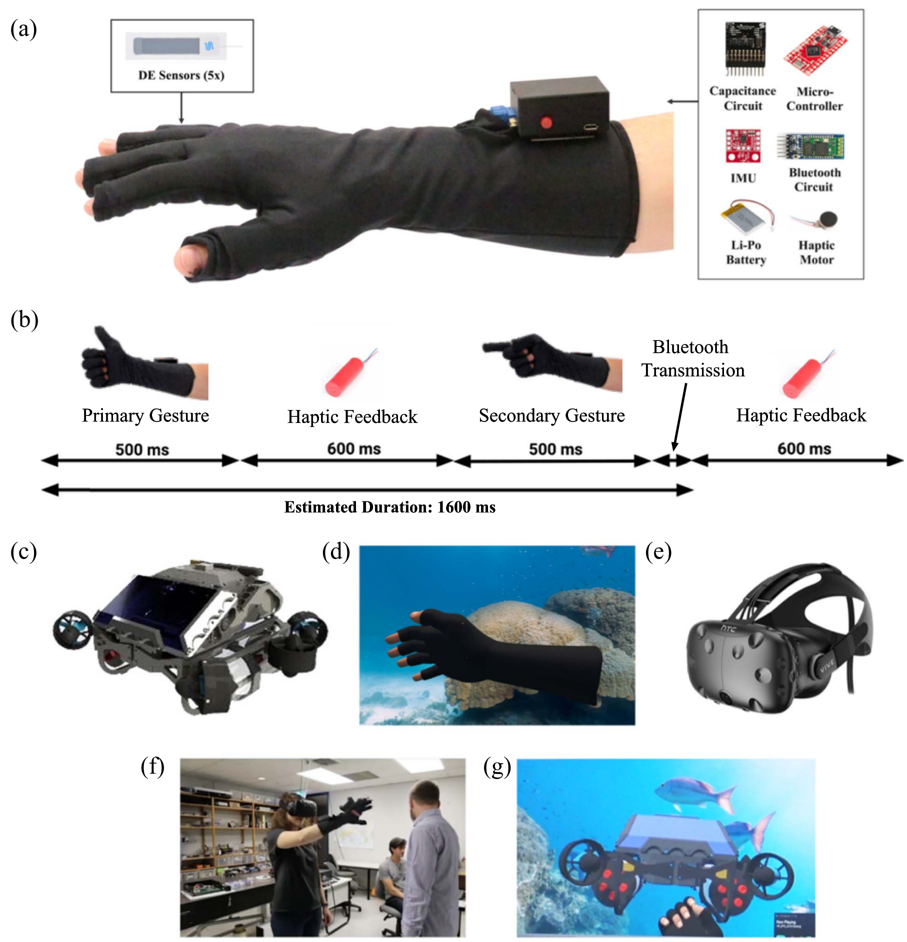


Fig. 3. (a) VR glove, DE sensors, and electronics. (b) Gesture-based communication process to instruct the virtual AUV. (c) CAD model of the AUV. (d) CAD model of the VR glove. (e) HTC Vive headset. (f) User wearing the headset and glove interacting with the AUV. (g) Scene from the UW VR unity environment.

TABLE III
COMMAND DELAYS FROM THE UW AND VR EXPERIMENTS COMPARED

Instruction name	VR delay mean (ms)	UW delay mean (ms)	Time difference (ms)
Move Left	2741	3210	469
Move Right	3552	5610	2058
Move Forward	2247	4110	1863
Move Backward	2930	4810	1880

Note that the transmission method was Bluetooth in the case of VR and acoustics for the UW glove.

capabilities of the AUV had been developed further. The cloth-backed neoprene UW glove was fabricated in a similar way to the VR glove; however, the UW glove also had inductive charging, a waterproof haptic motor, acoustic communication circuitry, a red LED for visual feedback, and was potted in UR5048 (Electrolube Ltd., U.K.) polyurethane resin [see Fig. 4(a)]. All of the electronics were potted inside a three-dimensional printed PLA enclosure mounted on the back of the glove. Only the red LED protruded externally on the left side of the housing for visual feedback.

Rather than Bluetooth, the UW glove used a low-cost miniature acoustic modem for command transmission, which was developed and sourced from the Intelligent Sensing and Communications research group at the University of Newcastle. A pair of such acoustic modems, one incorporated in the glove electronics and one mounted on the UW vehicle enabled half-duplex communication using chirp signal modulation in the 24–28 kHz frequency range. Sound communication provided a relatively long transmission range ($\sim$2 km) but with a limited bandwidth of up to 64 B/s. This limitation restricted the ability to send raw capacitance values of each sensor to the AUV or surface and instead required all raw gesture processing to be performed by the glove's on-board potted electronics. As gesture instructions [see Fig. 1(b)] were recognized, they were sent as byte-sized instructions to the AUV. Water as a transmission medium attenuates waveforms quicker the higher the carrier frequency. All individual data messages were therefore arranged into acoustic packets preceded with a robust ping chirp waveform, which the listening modem could detect, followed by a command string that included the message type, destination address, and instruction. An example command string is "$U001011" where "$U" indicates that this is a unicast message

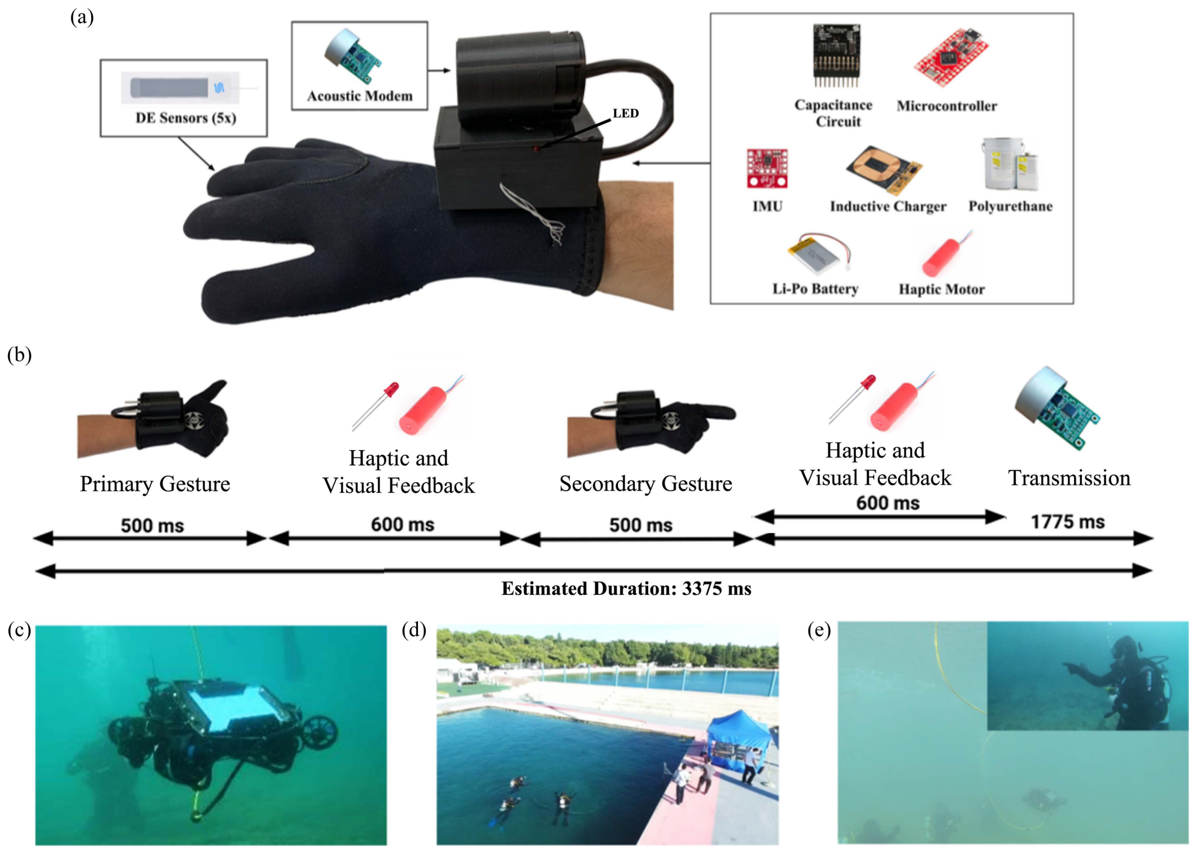


Fig. 4. (a) UW glove, DE sensors, and electronics. (b) Gesture-based recognition process to send an instruction to the AUV. (c) AUV. (d) Pool in Biograd na Moru where the demonstration was performed. (e) Diver, wearing the UW glove, performing a “one” gesture to instruct the AUV to move.

(sent to a specific receiver rather than broadcasted), “001” is the address ID of the receiver, “01” is the number of bytes of data in the command string, and, finally, “1” is the byte-sized instruction. Aside from using acoustic transmission instead of Bluetooth and incorporating visual feedback, the recognition process to send an instruction was the same as the VR glove [see Fig. 4(b)].

The AUV used as the diving buddy in the UW experiment was a prototype vehicle developed for the primary purpose of assisting and interacting with divers [see Fig. 4(c)]. The AUV was developed in the scope of ONR project Advancing Diver-Robot Interaction Capabilities described in [30]. The vehicle was based on a commercial hand-held diver navigator D2 developed by Kenautics (Kenautics Inc., USA) upgraded with a propulsion skid for autonomous navigation. Six 200-W thrusters mounted around the vehicle—four in an x-configuration and two vertical—enabled motion in five degrees of freedom. The vehicle was equipped with a wide range of navigation sensors including an inertial motion unit and attitude, heading, and roll sensor, Doppler velocity log, global navigation satellite system (GNSS), depth sensor, and imaging sonar. The vehicle also featured a low-light HD 1080p USB camera, 2 $\times$ 1500 lumen adjustable LED camera lights, 10.1-in thin-film transistor liquid-crystal display (LCD) display, and a pair of acoustic modems that could be used for communication or localization purposes. The AUV was powered by two 266.4-Wh lithium-ion batteries that provided 6–10 h autonomy. All sensor processing was done onboard using two dedicated processing boards. Weighing 20–25 kg in air with a form factor of 50 $\times$ 60 $\times$ 30 cm, the AUV was easy to deploy for a single person and relatively approachable to divers when UW.

The UW field trial was performed in a 20 $\times$ 50 m ocean pool at Biograd na Moru, on the Croatian Adriatic coast [see Fig. 4(d)]. The diver, wearing the glove at a depth of $\sim$5 m, performed each of the six gesture instructions while facing the AUV [see Fig. 4(e)]. Two additional divers accompanied the main diver for safety. Tethering the vehicle to a surface-based Wi-Fi buoy enabled real-time monitoring and data logging. The demonstration was filmed by Go Pro video cameras worn by each diver, and the AUV’s on-board video camera. Commands were time-stamped in postanalysis from when the diver finished transitioning to a primary gesture, to when the AUV started to move. Results can be seen in Fig. 5 with a comparison to the VR experiment shown in Table III.

III. Results

The percentage accuracy, number of false positives, and time-to-recognition for each of the 175 gestures are given in Table II.

In total, 140 gestures were correctly recognized with 33 gestures failing to be recognized and two gestures mischaracterized as an alternative gesture. The majority (85.6%) of the 35 failures (30 out of 35) were attributed to three of the ten participants. Contrastingly, all remaining participants only experienced two or fewer recognition failures with three of the

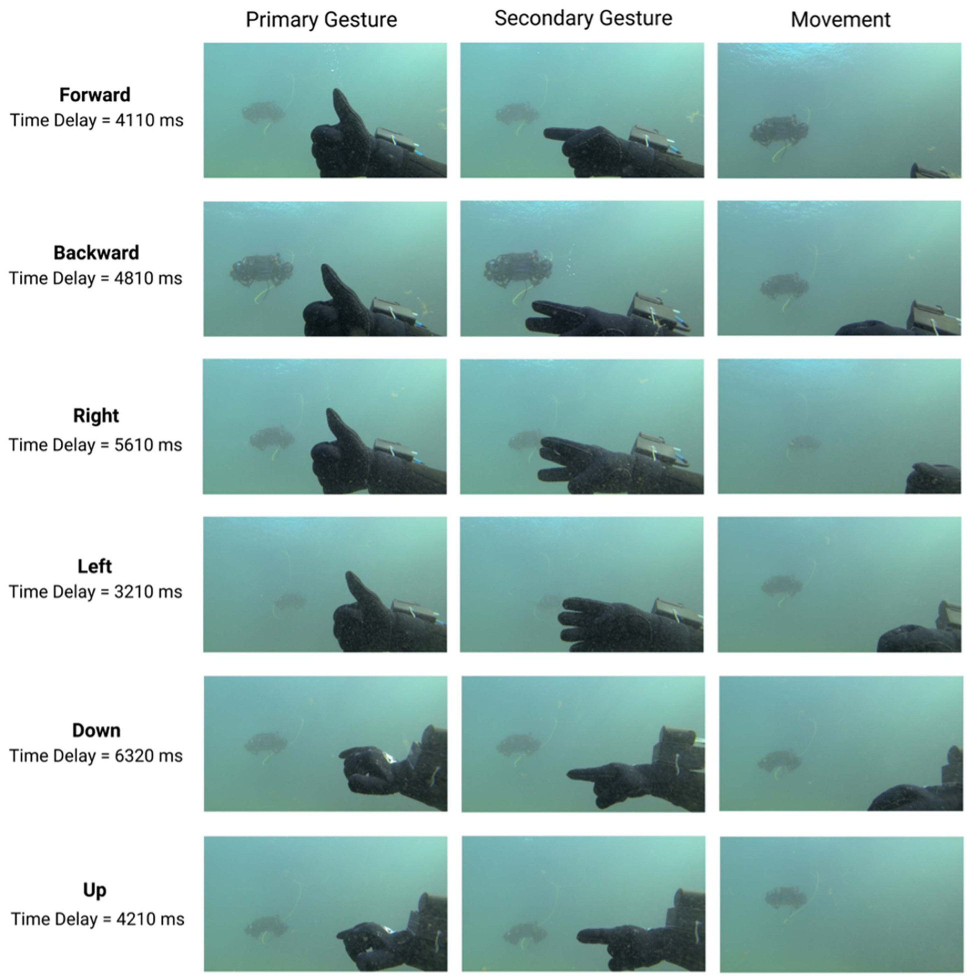


Fig. 5. Recorded footage of each gesture combination being performed to maneuver the AUV forward, backward, right, left, down, and up. In each row, the primary gesture is performed first. Once the diver receives feedback that it was recognized, they proceed to perform the secondary gesture. After recognition the second time, the relevant command is transmitted acoustically to the AUV, with the result shown in the final column.

participants having their gestures recognized 100% correctly. The gestures for "two" and "four" were correctly recognized 92.3% and 100% of the time, respectively. The gestures with the lowest recognition success were "Thumbs Up," "One," and "Three," with percentage accuracy 79.6%, 66.7%, and 80%, respectively. If the three participants with the lowest recognition success were removed from the data set, the success rate of each gesture would move above 90% for the remaining participants.

Due to its position as a primary gesture in four of the combination commands the "Thumbs Up" gesture was performed the most (98 occurrences). In terms of recognition time, the "Thumbs Up," "One," and "Three" were recognized in less than 800 ms on average. Gesture "Two" averaged a significantly longer recognition time, when compared to the other gestures. This can again be attributed to a small number of participants where the recognition took significantly longer bringing up the average. The recognition time would likely improve with the training of the participants or through the integration of machine learning algorithms trained on a large sample size.

The UW field trial showcased gesture-based communication from diver to AUV. The AUV was commanded left, right, forward, backward, up, and down solely by gesture combinations performed by the diver. An illustration of the maneuvers is shown in Fig. 5 with a video provided in the supplementary material. In each row of Fig. 5, the primary gesture is the first gesture performed as part of the combination commands from the library [see Fig. 1(b)]. Once the diver received visual and haptic feedback confirming recognition, they proceeded to perform the secondary gesture (column 2 of Fig. 5). After recognition of the second gesture, the relevant command was transmitted acoustically to the AUV, moving it approximately 1 m in the direction instructed by the diver. Time delays for each combination command are shown in the first column of Fig. 5 with captions. The delay refers to the time elapsed between initiation of the primary gesture and the AUV starting to move.

In Table III, the command delays of both the UW field trial and VR experiment are compared. Four commands overlapped between the two demonstrations: left, right, forward, and backward. The recognition algorithm and measurement electronics were the same for both experiments.

For each instruction, the gesture communication was significantly slower in the UW environment compared to the VR environment. That effect is expected due to the greater message transfer delay associated with the acoustic channel; as opposed to the air medium when using Bluetooth. The largest time difference occurred for the "Move Right" instruction, which took 2058 ms longer in the UW environment.

IV. Discussion

In this article, we have presented a glove for enhancing diver-to-robot communication with the ability to interpret a stream of on-board gestures and map them to instructions transmitted using sound to an AUV. The diver has successfully used the UW glove to command the robot using acoustic signals to move both horizontally and vertically in an ocean pool in Croatia (see Fig. 5), overcoming the limitations associated with camera-based vision and line-of-sight systems [13], [14], [15].

The VR experiment validated the multiuser reliability of the gesture recognition process with only three of the 10 participants contributing to 85.6% of the recognition failures. With some further development, the system reliability could be substantially improved. Analysis of video footage from these participants highlighted two main causes. The first related to how fast participants performed each gesture. If a gesture was not held for 500 ms, it was not recognized by the system. Furthermore, if a user transitioned slowly between gestures, the algorithm falsely detected an incorrect gesture while the user was still transitioning; their finger positions were in one quintile for over 500 ms. For both of these cases, software developments, real-time feedback, and training could help improve the reliability of the system. For instance, time-series analysis would highlight if the user was still transitioning and could prevent false-positive recognitions. Furthermore, analysis of the time series could also detect users holding gestures for less than 500 ms and send a warning to the user.

The last observed issue was glove fit. One glove was used across all 10 participants despite varying hand sizes. The recognition algorithm employed a calibration process, shown in Fig. 1(e), to customize thresholds for each user. However, if the user's hand was significantly smaller than the glove, the glove may have shifted relative to their hand after the calibration procedure was performed. Changing the position of a strain sensor results in different deformation. Ideally, each participant would have a custom-fit glove ensuring repeatable sensor readings or at minimum the option of a few general sizes. If this is not possible, providing a simple and fast way for the user to recalibrate the glove could improve recognition rates.

In terms of the time required for recognition, the algorithm needed each gesture to be held for a minimum of 500 ms. The "Two" gesture, however, had a significantly larger average time to recognition, 1749 ms. Analysis of video footage showed that users performed the same gesture differently, which was especially noticeable for the "Two" gesture. To perform this gesture, users needed to extend just the index and third finger while closing the remaining digits. However, depending on the user, the fourth finger extended to varying degrees along with the third finger. Since users were able to see this occurring visually on the virtual glove in the VR environment, they continuously adjusted their fourth finger position until the gesture was recognized. This resulted in an accuracy of 92.3% but an average time to recognition of 1749 ms. Furthermore, the discomfort of performing the gesture may have contributed to it being performed the least, 13 times across the 10 participants. Including the "Two" gesture in the calibration procedure would enable the algorithm to customize the quintiles for this gesture to each participant.

The time required to instruct the AUV can be compared in both the VR experiment and UW field trial. For both, the system was able to respond in a timely manner with a robot command generated from a gesture stream. The average time for a command to be transmitted in the UW demonstration was 4717 ms. In contrast, the average time required for a command in the VR experiment was 2867 ms. The time required for an instruction can be attributed to gesture measurement, recognition, and transmission [see Fig. 3(b)]. Gesture measurement and recognition were the same across both experiments, as the same sensors, electronics, and software were used in both. Regarding transmission, the acoustic modem introduced additional delay due to its low data throughput. Although the bandwidth utilization was minimized by encoding the gesture data to a single command byte, the data still had to be packed into an acoustic message that ensured robust transmission. The total encoded gesture message waveform amounted to 1775 ms of the acoustic message. From Table III, the time differences between the VR and UW commands are aligned well with the additional acoustic delay for the "Move Right," "Move Forward," and "Move Backward" commands. The "Move Left" command had a significantly smaller time difference; however, from video analysis, the user transitioned faster between the primary and secondary gestures for this command. The time differences would likely average higher if the gesture was performed more times and by more participants.

It is also worth noting that as the range of communication was increased, we introduced additional propagation delay of the communication channel. Sound speed through water depends on a number of factors including temperature, pressure, salinity and can vary between 1450 and 1570 m/s. Considering that the envisioned scenario as well as the UW experiment feature the AUV at the 3–5 m distance from the diver, the sound propagation time of 2–4 ms was negligible compared to the total message duration. Larger distances, however, would have produced a more significant delay.

Future improvements will include an adaptable gesture recognition algorithm trained on a larger data set to ensure accurate recognition across different users. An aim should be to reduce false positives and false negatives in the recognition algorithm as this could be detrimental during dive communication. Furthermore, while the diver had feedback acknowledging a gesture was detected by the algorithm, there was no distinction between commands. This meant the user was unsure if the correct instruction was recognized until the AUV moved in the appropriate way.

We acknowledge that diver-to-robot communication is part of a bigger challenge. Untethered autonomous robots must be able to track a diver to remain nearby for the diver to see the robot. Concept robots have relied on a vision for tracking this [31]. For instance, the Aqua robot described by Sattar et al. [32], [33], [34] was guided by "visual serving on color targets." Diver tracking and recognition could potentially be performed using imaging sonar, requiring image thresholding, background subtraction, and analysis of the clustered pixel data for diver identification [19]. Both stereo vision and multibeam sonar have been together trialed in the same vehicle for diver recognition [30]. And for a practical and safe solution, we can include a surface vehicle that communicates between the diver and robot and the boat [35]. Work is continuing in this area.

But for reliable diver robot interaction, the gesture-based motion capture glove presented here has opened the door to direct communication for intuitive and timely control of assistive UW robots.

Acknowledgment

This research was approved by the University of Auckland Human Participants Ethics Committee application number UAHPEC-024190.

References

[1] L. Behnke, *Scuba Diving Hand Signals: Underwater Communication Pocket Companion for Recreational Scuba Divers*. Scotts Valley, CA, USA: CreateSpace, 2015.

[2] C. W. Shilling, M. F. Werts, and N. R. Schandelmeier, "Underwater communications," in *The Underwater Handbook*, C. W. Shilling, M. F. Werts, and N. R. Schandelmeier, Eds. Boston, MA, USA: Springer, 1976, doi: 10.1007/978-1-4684-2154-5_10.

[3] B. Woodward and H. Sari, "Digital underwater acoustic voice communications," *IEEE J. Ocean. Eng.*, vol. 21, no. 2, pp. 181–192, Apr. 1996.

[4] P. J. Denoble, J. L. Caruso, G. de L Dear, C. F. Pieper, and R. D. Vann, "Common causes of open-circuit recreational diving fatalities," *Undersea Hyperbaric Med.*, vol. 35, no. 6, pp. 393–406, 2008.

[5] F. Chen et al., "WristCam: A wearable sensor for hand trajectory gesture recognition and intelligent human–robot interaction," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8441–8451, Oct. 2019.

[6] C. Georgiades et al., "AQUA: An aquatic walking robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 4, Sep. 2004, pp. 3525–3531.

[7] R. Altendorfer et al., "Rhex: A biologically inspired hexapod runner," *Auton. Robots*, vol. 11, no. 3, pp. 207–213, 2001.

[8] J. Sattar et al., "Enabling autonomous capabilities in underwater robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 3628–3634.

[9] D. Chiarella et al., "Gesture-based language for diver-robot underwater interaction," in *Proc. IEEE OCEANS Conf.-Genova*, May 2015, pp. 1–9.

[10] A. G. Chavez, C. A. Mueller, T. Doernbach, D. Chiarella, and A. Birk, "Robust gesture-based communication for underwater human-robot interaction in the context of search and rescue diver missions," in *Proc. IROS Workshop Human-Aiding Robot.*, 2018, pp. 1–4.

[11] H. Buelow and A. Birk, "Gesture-recognition as basis for a human robot interface (HRI) on a AUV," in *Proc. IEEE/MTS OCEANS KONA*, Sep. 2011, pp. 1–9.

[12] J. Yang, J. P. Wilson, and S. Gupta, "Diver gesture recognition using deep learning for underwater human-robot interaction," in *Proc. IEEE/MTS OCEANS SEATTLE*, Oct. 2019, pp. 1–5.

[13] A. Gomez Chavez, A. Ranieri, D. Chiarella, E. Zereik, A. Babić, and A Birk, "CADDY underwater stereo-vision dataset for human–robot interaction (HRI) in the context of diver activities," *J. Mar. Sci. Eng.*, vol. 7, no. 1, 2019, Art. no. 16.

[14] M. A. M. Martija, J. I. S. Dumbrique, and Prospero C. Naval Jr., "Underwater gesture recognition using classical computer vision and deep learning techniques," *J. Image Graph.*, vol. 8, no. 1, 2020, pp. 9–14.

[15] K. J. DeMarco, M. E. West, and A. M. Howard, "Underwater human-robot communication: A case study with human divers," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2014, pp. 3738–3743.

[16] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.

[17] S. Jiang, P. Kang, X. Song, B. P. Lo, and P. B. Shull, "Emerging wearable interfaces and algorithms for hand gesture recognition: A survey," *IEEE Rev. Biomed. Eng.*, vol. 15, pp. 85–102, 2021.

[18] A. Tairych and I. A. Anderson, "Capacitive stretch sensing for robotic skins," *Soft Robot.*, vol. 6, no. 3, pp. 389–398, 2019.

[19] G. Frediani, F. Vannetti, L. Bocchi, G. Zonfrillo, and F. Carpi, "Monitoring flexions and torsions of the trunk via gyroscope-calibrated capacitive elastomeric wearable sensors," *Sensors*, vol. 21, no. 20, 2021, Art. no. 6706.

[20] G. Frediani, L. Bocchi, F. Vannetti, G. Zonfrillo, and F. Carpi, "Wearable detection of trunk flexions: Capacitive elastomeric sensors compared to inertial sensors," *Sensors*, vol. 21, no. 16, 2021, Art. no. 5453.

[21] C. Walker and I. Anderson, "From land to water: Bringing dielectric elastomer sensing to the underwater realm," *Proc. SPIE*, vol. 9798, Apr. 2016, Art. no. 97982B.

[22] C. R. Walker and I. A. Anderson, "Monitoring diver kinematics with dielectric elastomer sensors," *Proc. SPIE*, vol. 10163, Apr. 2017, Art. no. 1016307.

[23] H. Böse, M. Thuy, and S. Stier, "Wearable operation device with different types of dielectric elastomer sensors," *Proc. SPIE*, vol. 10594, Mar. 2018, Art. no. 105940X.

[24] H. Böse, S. Stier, and S. Muth, "Glove with versatile operation tools based on dielectric elastomer sensors," *Proc. SPIE*, vol. 10966, pp. 242–254, Mar. 2019.

[25] S. Gratz-Kelly, A. Meyer, P. Motzki, S. Nalbach, G. Rizzello, and S. Seelecke, "Force measurement based on dielectric elastomers for an intelligent glove providing worker assessment in the digital production," *Proc. SPIE*, vol. 11375, Apr. 2020, Art. no. 1137525.

[26] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 3, pp. 311–324, May 2007.

[27] H. Kaushal and G. Kaddoum, "Underwater optical wireless communication," *IEEE Access*, vol. 4, pp. 1518–1547, 2016.

[28] J. Sticklus, P. A. Hoeher, and R. Röttgers, "Optical underwater communication: The potential of using converted green LEDs in coastal waters," *IEEE J. Ocean. Eng.*, vol. 44, no. 2, pp. 535–547, Apr. 2019.

[29] D. W. O. Antillon, C. Walker, S. Rosset, and I. A. Anderson, "The challenges of hand gesture recognition using dielectric elastomer sensors," *Proc. SPIE*, vol. 11375, May 2020, Art. no. 1137524.

[30] Đ. Nađ et al., "Towards advancing diver-robot interaction capabilities," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 199–204, 2019.

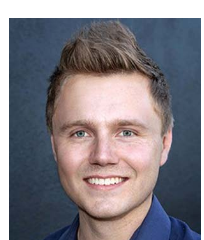
[31] M. J. Islam, J. Hong, and J. Sattar, "Person-following by autonomous robots: A categorical overview," *Int. J. Robot. Res.*, vol. 38, no. 14, pp. 1581–1618, 2019.

[32] J. Sattar and G. Dudek, "Underwater human-robot interaction via biological motion identification," in *Proc. Robot., Sci. Syst.*, Jun. 2009, pp. 1–8.

[33] J. Sattar and G. Dudek, "A vision-based control and interaction framework for a legged underwater robot," in *Proc. Can. Conf. Comput. Robot Vis.*, May 2009, pp. 329–336.

[34] J. Sattar and G. Dudek, "Where is your dive buddy: Tracking humans underwater using spatio-temporal features," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 3654–3659.

[35] N. Stilinović, Đ. Nađ, and N. Mišković, "AUV for diver assistance and safety—Design and implementation," in *Proc. IEEE OCEANS Conf.-Genova*, May 2015, pp. 1–4.



Christopher R. Walker received the bachelor's degree (Hons.) in engineering and Ph.D. degree in biomedical engineering from the University of Auckland, Auckland, New Zealand, in 2015 and 2019, respectively.

He is currently a Postdoctoral Researcher with the Auckland Bioengineering Institute, University of Auckland. He has been working on sensor-based human motion capture of scuba divers along with real-time physiological monitoring using noninvasive sensing systems. He has authored and coauthored six conference proceedings, five journal articles, and one U.S. patent.

Đula Nađ received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Zagreb, Zagreb, Croatia, in 2008 and 2017, respectively.

He is currently an Assistant Professor with the Laboratory of Underwater Systems and Technologies, Faculty of Electrical Engineering and Computing, and the Department for Control Engineering and Computing, University of Zagreb. He participated in multiple European projects over the past decade. His research interests include navigation, guidance, and control of marine vehicles, localization aiding, and diver–robot cooperation.

Derek W. Orbaugh Antillon was born in Guatemala City, Guatemala, in 1990. He received the B.S. degree in electronics engineering from the Universidad del Valle de Guatemala, Guatemala City, Guatemala, in 2015, and the M.S. degree in smart systems from Furtwangen University, Furtwangen, Germany, in 2017. He is currently working toward the Ph.D. degree in bioengineering with the Biomimetics Laboratory, Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand.

He has authored and coauthored three conference proceedings and four journal articles.

Igor Kvasić received the M.Sc. degree in electrical engineering and computing in 2013 from the Laboratory for Underwater Systems and Technologies, Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, where he is currently working toward the Ph.D. degree in the field of sonar image processing.

He is currently a Researcher with the Faculty of Electrical Engineering and Computing, University of Zagreb and a member of the Laboratory for Underwater Systems and Technologies, University of Zagreb. He is currently involved in the ONR NICOP project called ROADMAP – Robot Aided Diver Navigation in Mapped Environments. His research interests include underwater vision, sonar image processing, and human–robot interaction.

Mr. Kvasić is the Vice-Chair of the IEEE Oceanic Engineering Society Student Branch Chapter of the University of Zagreb.

Samuel Rosset received the M.Sc. and Ph.D. degrees in microengineering from the Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland, in 2004 and 2009, respectively.

He is currently a Research Fellow with the Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand. Since 2005, he has been working on soft transducers, with a focus on optical and biomedical applications and the fabrication processes of these devices. He has authored and coauthored more than 40 peer-reviewed papers.

Nikola Mišković (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from University of Zagreb, Zagreb, Croatia, in 2010.

He is currently a Full Professor in control and marine robotics and the Head of the Laboratory for Underwater Systems and Technologies, Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia. He participated in 20 European projects (coordinated five), six Office of Naval Research Global projects, two NATO projects, one Schmidt Ocean Institute project, and 16 national projects. He authored/coauthored more than 130 papers in journals and conference proceedings in the area of navigation, guidance, and control, as well as cooperative control in marine robotics.

Dr. Mišković was President of the Chapter for Robotics and Automation of the IEEE Croatian Section (2016–2019), member of the IEEE Oceanic Engineering Society Administrative Committee (2023–2025), and member of the IFAC Technical Committee on Marine Systems and Centre for Underwater Systems and Technologies.

Iain A. Anderson received the Ph.D. degree in engineering science from the University of Auckland, Auckland, New Zealand, in 1996.

He is a Group Leader with the Biomimetics Laboratory of the Auckland Bioengineering Institute, University of Auckland. He was a Whiteware Product Designer with Fisher and Paykel Ind., a Vibrations Consulting Engineer with the NZ Department of Scientific and Industrial Research, and a Research Scientist associated with hip and knee implant design and surgery with the Industrial Research Ltd., New Zealand. In 2000, he returned to the Department of Engineering Science as a staff member, and was one of the founding members of the Auckland Bioengineering Institute. His interest in artificial muscles led to the formation of the Biomimetics Laboratory in 2004. The lab's research is currently focused on the control and self-sensing of artificial muscles and artificial muscle energy harvesting. He is involved in two Biomimetics Lab spin-out companies: StretchSense (2012) and PowerOn Ltd. (2019). StretchSense produces motion-capture gloves using soft elastomer dielectric elastomer sensors. PowerOn is commercializing the lab's electroactive polymer actuator technology.

Dr. Anderson was the recipient of the Royal Society of New Zealand's 2016 Pickering Medal for his efforts in developing and commercializing electroactive polymer technology.