Robust Gesture-Based Communication for Underwater Human-Robot Interaction in the context of Search and Rescue Diver Missions

Arturo Gomez Chavez[1] Christian A. Mueller[1] Tobias Doernbach[1] Davide Chiarella[2] and Andreas Birk[1].

***Abstract*— We propose a robust gesture-based communication pipeline for an Autonomous Underwater Vehicle (AUV) for divers to instruct the robot to assist them in performing high-risk tasks and helping in case of emergency. A gesture communication language (CADDIAN) is developed, based on consolidated and standardized diver gestures, including an alphabet, syntax and semantics, ensuring a logical consistency. Next, a hierarchical classification approach is introduced for hand gesture recognition based on stereo imagery and multi-descriptor aggregation to specifically cope with image artifacts that occur underwater, e.g. due to light backscatter or color attenuation. Once the classification task is finished, a syntax check step is performed to filter out invalid command sequences sent by the diver or generated by errors in the classifier. Furthermore, the command is displayed in an underwater tablet for the diver to acknowledge before the corresponding mission is started. The objective is to prevent the AUV from executing unnecessary, infeasible or potentially harmful motions. Experimental results under different environmental conditions in archaeological exploration and bridge inspection applications show that the system performs well in the field.**

I. INTRODUCTION

Underwater environments pose a great number of technological challenges for robotics in the areas of navigation, communication, autonomy, manipulation and others. Although progress has been quickly made in the last years, the unstructured and dynamic nature of these environments makes human intervention indispensable for applications such as acquisition of relevant biological data, monitoring of marine areas, exploration of archaeological sites, inspection of damaged infrastructure and search and rescue operations.

For this reason, the EU-funded Project CADDY (Cognitive Autonomous Diving Buddy) was developed with the aim of transferring robotic technology into the diving world to improve the safety levels during the underwater missions with divers in the loop. The main objective is the development of a pair of companion/buddy robots (Fig. 1) - an Autonomous Underwater Vehicle (AUV) and an Autonomous Surface Vehicle (ASV) - to monitor and support human operations and activities during the dive.

[*]The research leading to the presented results has received funding from the German Federal Ministry of Education and Research within the project "Vulnerability of Transportation Structures Warning and Evacuation in Case of Major Inland Flooding" (FloodEvac) and EU FP7 project "Cognitive Autonomous Diving Buddy (CADDY)".

[1] Robotics Group, Computer Science & Electrical Engineering, Jacobs University Bremen, Germany. {a.gomezchavez,chr.mueller, t.fromm,a.birk}@jacobs-university.de.

[2] Institute of Intelligent Systems for AutomationNational Research Council of Italy, Genova, Italy. davide.chiarella@cnr.it

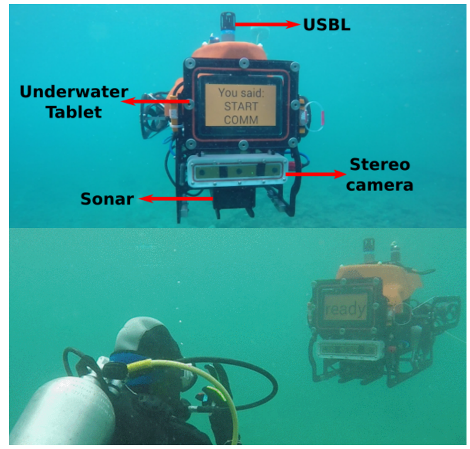


Fig. 1: (Top) BUDDY-AUV equipped with sensors to monitor the diver - Blueprint Subsea X150 USBL, ARIS 3000 Multibeam Sonar, BumbleBeeXB3 Stereo Camera, Underwater Tablet. (Bottom) Diver issuing gesture command to the AUV, reprinted from [1]

In underwater scenarios, there are unique sensor problems due to the attenuation of high-frequency electromagnetic waves: WiFi and radio signals become quickly unreliable at $0.5\,\mathrm{m}$ depth, no GPS and optical devices and camera imagery suffer artifacts due to light backscattering. Acoustic sensors are mostly used, although they offer limited bandwidth. Thus, the adopted solution is a communication framework letting the diver communicate close to the AUV using an extension of the commonly used diver-gestures (CADDIAN language [2]). The diver gesture is captured using a stereo camera and process through a multi-descriptor classifier robust to underwater distortions. Then, when a complete message has been relayed, a syntax check is done to corroborate logical coherence and an approval request is issued to the diver via an underwater tablet (Fig. **??**fig:buddy-auv).

In this work, we show the application of the developed CADDY system into bridge inspection. Particularly, we triggered the inspection procedures on [3] using the CADDY system. Bridges have an important role during disaster scenarios for evacuation and humanitarian supply. Hence, the remaining load capacity of bridges after a disaster is a crucial factor to analyze. In order to obtain data about structural components and the situation of foundations, the current state-of-the-art procedure relies on diver operations.

Workshop on Human-Aiding Robotics
International Conference on Intelligent Robots and Systems 2018 (IROS), Madrid, Spain

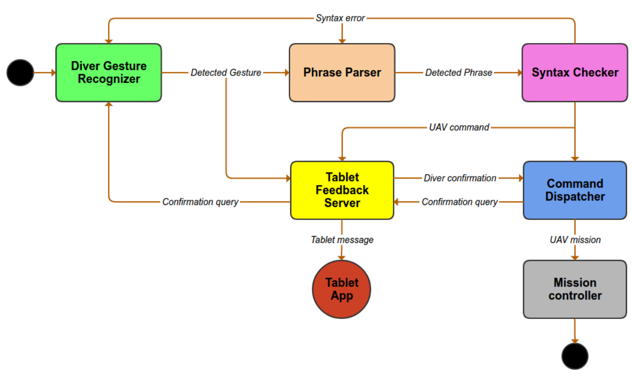


Fig. 2: System architecture of AUV mission generator based on diver hand gestures.

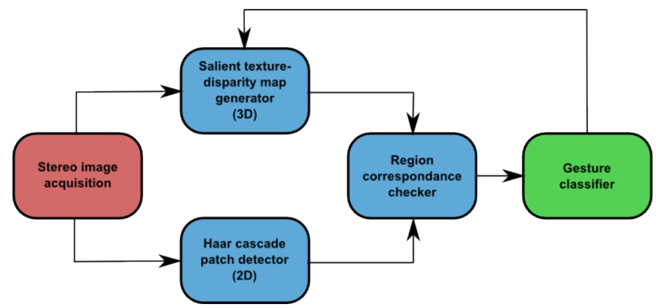


Fig. 3: General framework diagram for hand gesture recognition. The different stages of the process are color coded: (Red) Image acquisition, (Blue) Hand detection, (Green) Gesture classification.

However, this is an undertaking high risk in flooded areas, and because of poor visibility divers often can only perform haptic surveys of the damage areas.

II. System Overview

The two main objectives of the system is the easy generation of AUV missions consisting of multiple commands derived from hand gestures and to be highly fault-tolerant. The latter goal is of great importance in field underwater applications since recovery or maintenance of the equipment during a mission is not an easy task; and when there is a close human-robot interaction, the safety of the user is top priority. For these reasons, the system architecture presented in Figure 2 consists of several modules that on the high level ensure that the issued commands by the diver are logically consistent and scalable (stack up to form more complex commands), and that the diver receives useful feedback about in case there is a need for corrections. On the lower level, these modules ensure that the recognition of the gestures are robust against diver movement, light backscattering or poor lightening (image distortions), and that the commanded missions are carried out by the AUV successfully by following the correct motion primitives.

- *Diver Gesture Recognizer* contains all the image processing and classification algorithms to correctly detect the diver's hands and determine if they correspond to a gesture associated with a command for the AUV. The output of this module is a label indicating the name of this command.
- *Phrase Parser* receives the previous labels and parse them according to some delimiters - special gestures that group a sequence of gesture based commands. The resultant sequences, named *phrases*, pass to the *Syntax Checker* module for validation to check if they follow the syntax rules of the CADDIAN language.
- *Syntax Checker* evaluates the previous *phrases* according to the CADDIAN language syntax rules (Section **??**). In this way, only *complete* and *logically consistent* commands will be saved in memory for future execution, after receiving the diver's confirmation. For example the command "Do a Map and Go Down" will not be accepted since both actions need parameters to be executed i.e. "Do a Map of 10 by 12 meters and Go Down 1 meter."
- *Command Dispatcher* receives the validated AUV commands from the *Syntax Checker* and stores them until the diver approves them. Then, it feeds these commands sequentially to the *Mission Controller* as each of them are executed and completed.
- *Mission Controller* maps the received AUV commands into high-level tasks and monitors the activation and execution of the robotic functionality primitives to complete them.
- *Tablet Feedback Server* prepares the message that is going to be displayed in the Tablet for the diver to see according the input of the rest of the modules; this message is sent to the tablet where a custom application displays it in the given format (referenced as *Tablet App* in Figure 2).

III. Diver Gesture Recognizer

Hand gesture recognition applications have achieved high precision and performance through the use of RGB-D cameras that generate dense point clouds from which 3D features can be computed and reasoned about. However, this technology is not best suited for underwater applications due to light backscatter in water [4]. The use of stereo cameras allows to retrieve more accurate 3D information, but the generated point clouds are sparse and the current state of the art gesture recognition methods do not perform well on these data.

The different physical phenomenons that affect light in water, absorption (which removes light energy) and scattering (which changes the direction of the light path), lead to image feature's degradation. This and the fact that underwater environments do not often possess diverse textures as onland scenery, make finding and matching good distinctive features between stereo images difficult; hence the lack of denser point clouds. Thus, a hybrid-method using 3D

Workshop on Human-Aiding Robotics
International Conference on Intelligent Robots and Systems 2018 (IROS), Madrid, Spain

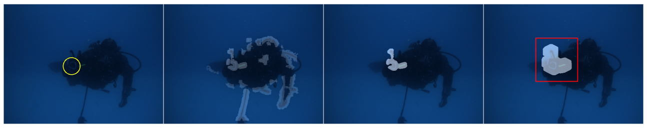


Fig. 4: General

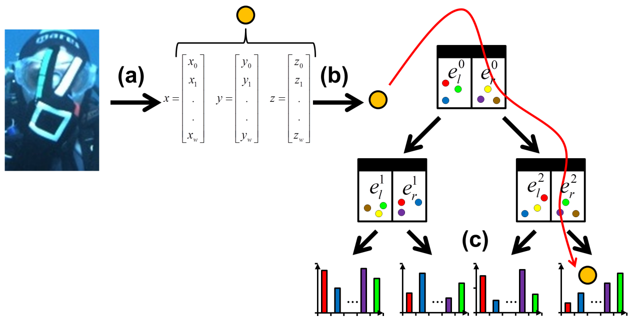


Fig. 5: Classification of an image by a Multi-Descriptor NCM tree (MD-NCM); each class centroid is represented by a colored dot and the traversed path of the sample while being classified is shown in red.

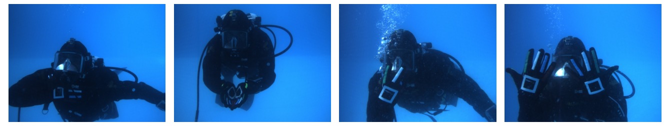
and 2D information was implemented as it was found to yield the best results for object detection and recognition tasks in the highly variant underwater imagery. Figure 3 shows the diagram of the overall hand gesture recognition framework; there are two main stages: *hand detection* and *gesture classification*.

For *detection*, we take a hybrid approach combining both 3D information through disparity maps and cascade classifiers to ensure robustness against motion and light attenuation. Segmentation of the disparity maps based on distance and density offers reliable hand detection, however, it fails in the presence of bubbles and other texture-prominent areas. Thus, 2D cascade classifiers can help filter these false positive regions. Figure **??**fig:gesture detection shows a successful example of hand detection with very low light conditions.

Once the system has detected possible locations of the diver's hands within the image, all of these candidate patches pass through a final classifier: Multi-Descriptor Nearest Class Mean Forests (NCMFs), which was first introduced for diver localization in our previous work in [5]. This classifier filters out all the false positives generated by the previous modules and maps the true positives (hands) to a specific gesture within the CADDIAN language. The main purpose of this proposed variant of Random Forest is to aggregate multiple descriptors that encode different representations of the objects of interest, each of them robust to different types of image distortion (Fig 5).

IV. CADDIAN LANGUAGE

As described before, the underwater environment is an hazardous scenario, where communication between peers is



(a) Out of air (b) Boat (c) Start comm (d) Do a mosaic

Fig. 6: Examples of CADDIAN gestures. (a) and (b) are extracted from already established gestures. (c) and (d) were defined to be easily related to the actions they represent.

crucial for the success and safety of the dive. The developed CADDIAN gestures [2] were chosen and/or defined from those common to divers to be as intuitive as possible, reduce learning time and transmit messThe developed CADDIAN gestures [2] were chosen and/or defined from those common to divers to be as intuitive as possible, reduce learning time and transmit messages efficiently. However, diver's hand signals vary from region to region and between organizations, hence the most common ones were selected from [6] [7] [8] [9] [10]. Figure 6 shows examples of CADDIAN gestures, the first two gestures are used in recreational diving while the last two were defined to be easily related to the action they are representing: the *Start Communication* gesture consists of pointing the index and middle finger to the mask, which symbolizes *"Look at me!"*; likewise, the *Do a mosaic/map* gesture mimics a person holding a map in front of him and reading it.

A. SYNTAX CHECKER

We described the methodology and algorithms used to detect and classify single gestures. However, to exploit the full potential of the application, the system must be able to understand complex commands - sequences of gestures - which can be also aggregated to form missions composed of several tasks. To achieve this and allow the diver to understand the status and progress of a mission, a communication protocol was established after defining the CADDIAN language and its syntax.

To allow synchronization between the diver and the AUV feedback messages, the CADDIAN syntax defines boundaries in order to ensure correct interpretation of commands and missions. These boundaries are associated with the gestures "START_COMMUNICATION" or A and "END_COMMUNICATION" or $\forall$; thus, commands are sequences of individual gestures delimited by (A, A) which commonly represent a single task e.g. "Take a photo at 3 meters altitude", and missions are delimited by $(A, \forall)$ which consist of aggregated commands e.g. "Take a photo here, go to the boat and carry the equipment back here". Of course, a mission composed of a single command is valid.

The gesture sequence interpretation is implemented through the *Phrase Parser* module introduced in Figure 2. It constantly saves the detected gestures until it detects among them either one of the mentioned delimiter pairs (A, A), $(A, \forall)$; then it sends these gestures - commands -

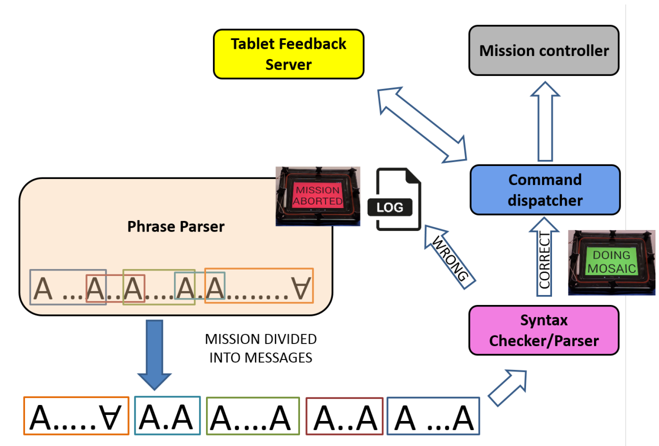


Fig. 7: Mission segmentation into single messages which are processed and validated by the Syntax Checker.

to the *Syntax Checker* for validation. If the command is syntactically correct, it passes to the *Command Dispatcher* module where it is saved until a complete mission is received i.e. an "END_COMMUNICATION" gesture issued by the diver. Finally, these commands are sequentially passed to the *Mission Controller* for execution. On the other hand, if the command is not valid, the error is logged and a warning is issued to the diver through the underwater tablet or a system of lights, Figure 7 depicts this process.

V. BRIDGE INSPECTION APPLICATION

In flood scenarios, disaster management, e.g. led by government authorities, demands accurate assessment of the vulnerability of infrastructures like bridges in order to plan safe evacuation. Given a bridge inspection request, the vehicle in Fig. 1 its deployed along with a diver. In our test deployment scenario the Karl Carstens bridge in Bremen, Germany (Fig. 8) has been selected.

In this case the sonar-based camera ARIS is used to map the foundations of the bridge due to poor visibility. If the diver is unsure whether it is safe to approach the bridge, he can issue a command *Do a map of X and Y dimensions* under the bridge. Then, after visualizing the images via the underwater tablet or in an on-shore center (where another expert is located), he can decide to approach the bridge for a more in detailed inspection. Frequently, analysts are in the look out for *scouring* (soil erosion) and *log-jams* (piled-up wood). In our situation, since no debris was found, the diver decided he could safely analyze the structure up-close (Fig. 8).

Based on this first-hand assessment and the gathering of other measurements such as *flow velocity* through a discharge sensor to analyze hydrodynamic forces; a final veredict about the state and usability of the bridge can be done.

Fig. 8: (Top) Karl Carstens Bridge (Bottom) Registration sample of lakeshore section beneath the bridge using *Fourier-Mellin Invariant* (FMI) method [11].

VI. CONCLUSIONS

Despite the high risk that involves disaster scenarios in underwater environments, full autonomy has not been achieved by ROVs and/or UAVs; thus, human (diver) experts are needed to perform high precision tasks or assess the stability of submerged structures. We think the major challenge consists in the development of a communication protocol between robot and diver to actively interact and cooperate and, above all, ensure the safety of the diver. Based on this and previous work, and on the bridge inspection scenario presented, we believe that symbiosis between human experts and robots is the trigger to spread the usage of this technology; particularly in these applications, where not all possible cases can be predicted.

REFERENCES

[1] N. Miskovic, A. Pascoal, M. Bibuli, M. Caccia, J. A. Neasham, A. Birk, M. Egi, K. Grammer, A. Marroni, A. Vasilijevic, d. Nad, and Z. Vukic, "Caddy project, year 3: The final validation trials," in *OCEANS 2017 - Aberdeen*, June 2017, pp. 1–5.

[2] D. Chiarella, M. Bibuli, G. Bruzzone, M. Caccia, A. Ranieri, E. Zereik, L. Marconi, and P. Cutugno, "Gesture-based language for diver-robot underwater interaction," in *OCEANS 2015 - Genova*, May 2015, pp. 1–9.

[3] C. A. Mueller, T. Fromm, H. Buelow, A. Birk, M. Garsch, and N. Gebbeken, "Robotic bridge inspection within strategic flood evacuation planning," in *OCEANS 2017 - Aberdeen*, June 2017, pp. 1–6.

[4] F. Sun, J. Yu, and D. Xu, "Visual measurement and control for underwater robots: A survey," in *2013 25th Chinese Control and Decision Conference (CCDC)*, May 2013, pp. 333–338.

[5] A. G. Chavez, M. Pfingsthorn, A. Birk, I. Renduli, and N. Miskovi, "Visual diver detection using multi-descriptor nearest-class-mean random forests in the context of underwater human robot interaction (hri)," in *OCEANS 2015 - Genova*, May 2015, pp. 1–7.

[6] 2003, confédération Mondiale des Activités Subaquatiques, retrieved May 5, 2016, from http://www.cmas.ch/downloads/it-Codici di comunicazione CMAS.pdf.

[7] 2005, recreational Scuba Training Council, retrieved May 5, 2016, from http://www.neadc.org/CommonHandSignalsforScubaDiving.pdf.

[8] 2005, scuba diving fan club, retrieved May 5, 2016, from http://www.scubadivingfanclub.com/Diving_Signals.html.

[9] 2013, fordivers.com, retrieved May 5, 2016, from http://www.fordivers.com/en/blog/2013/09/12/senales-de-buceo-que-tienes-que-conocer/.

[10] 2013, www.dive-links.com, retrieved May 5, 2016, from http://www.dive-links.com/en/uwzeichen.php.

[11] H. Bulow, A. Birk, and V. Unnithan, "Online generation of an underwater photo map with improved fourier mellin based registration," in *OCEANS 2009-EUROPE*, May 2009, pp. 1–6.