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

Arturo Gomez Chavez¹, Christian A. Mueller¹, Tobias Doernbach¹, Davide Chiarella² and Andreas Birk¹

Abstract—We propose a robust gesture-based communication pipeline for divers to instruct an Autonomous Underwater Vehicle (AUV) 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. A hierarchical classification approach is introduced for hand gesture recognition based on stereo imagery and multi-descriptor aggregation to specifically cope with underwater image artifacts, e.g. light backscatter or color attenuation. Once the classification task is finished, a syntax check is performed to filter out invalid command sequences sent by the diver or generated by errors in the classifier. Throughout this process, the diver receives constant feedback from an underwater tablet to acknowledge or abort the mission at any time. 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 using robotic technology to improve the safety levels during 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 a dive.

In this work, we show the application of the developed CADDY system into bridge inspection. Particularly, we triggered the inspection procedures on [3] as part of the FloodEvac BMBF research project*. Bridges have an important role during disaster scenarios for evacuation and humanitarian supply; 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. However, this is an undertaking high risk in

*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 Automation, National Research Council of Italy, Genova, Italy. davide.chiarella@cnr.it

Workshop on Human-Aiding Robotics
International Conference on Intelligent Robots and Systems 2018 (IROS), Madrid, Spain
flooded areas, and commonly divers can only perform haptic surveys of the damaged areas due to poor visibility.

II. SYSTEM OVERVIEW

The two main objectives of the system are: generation of AUV missions consisting of multiple commands derived from hand gestures and high fault-tolerance. 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 Fig. 2 consists of several modules that ensure that the issued diver commands are logically consistent and scalable (stack up to form more complex missions), and that useful feedback is produced to make corrections on the fly if necessary. At the algorithmic level, these modules ensure that the recognition of the gestures are robust against sudden movements, light backscattering and poor contrast i.e., image distortions; and that the commanded missions are carried out by the AUV successfully by following the correct motion primitives. We describe the functionality of each of these modules as follows:

- **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 known command for the AUV. The output is a label indicating the command’s name.
- **Phrase Parser** receives the previous labels and parse them according to some delimiters i.e., special gestures that group basic commands into sequences named phrases (Fig. 7).
- **Syntax Checker** evaluates the previous phrases according to the CADDIAN language syntax rules (Section IV). In this way, only complete and logically consistent commands are saved in memory for future execution. 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 gives a final approval command. Then, it feeds these commands sequentially to the Mission Controller as each of them is successfully completed.
- **Mission Controller** maps the received AUV commands into low-level tasks and monitors the activation and execution of the necessary action primitives to complete them.
- **Tablet Feedback Server** prepares the message that is going to be displayed in the Tablet according the input from the previous the modules.

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.

Thus, a hybrid-method using 3D 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 the hybrid approach combining disparity maps and cascade classifiers to ensure robustness against motion and light attenuation. Segmentation of disparity maps based on distance and density offers reliable hand detection, however, it fails in the presence of bubbles and texture-prominent areas. Hence, we use 2D cascade classifiers to filter out these false positive regions.

Once the system has detected possible locations of the
dive’s hands within the image (Fig. 4), 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).

It is important to note that this hybrid method was used and developed throughout the duration of the CADDY and FloodEvac projects due to its robustness given small sample datasets from underwater environments. After several field trials and data collection efforts, a detailed dataset with stereo images from the divers communicating with the AUV in different environmental conditions has been assembled (CADDY-Dataset [6]). Based on this, we hope to further investigate underwater image artifacts and enable data-driven methods.

IV. CADDIAN LANGUAGE

As described before, the underwater environment is an hazardous scenario, where communication between peers is 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 messages efficiently. However, diver’s hand signals vary from region to region and between organizations, hence the most common ones were selected from [7] [8] [9]. 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.

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 ∀; 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, ∀) 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 Fig. 2. It constantly saves the detected gestures until it detects among them one of the mentioned delimiter pairs (A, A), (A, ∀); then it sends these gestures - commands - 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
Fig. 7: Mission segmentation into single messages which are processed and validated by the Syntax Checker.

the diver through the underwater tablet or a system of lights, Fig. 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 is 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 (Adaptive-Resolution Imaging Sonar) 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 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 verdict about the state and usability of the bridge can be done.

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. Egí, K. Grammer, A. Marroni, A. Vasiljevic, 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] 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,” ArXiv e-prints, Jul. 2018.
[7] Confédération Mondiale des Activités Subaquatiques. Segni convenzionali CMAS. (retrieved May 5, 2018). [Online]. Available: http://www.cmas.ch/docs/it/downloads/codici-comunicazione-cmas/it-Codici-di-comunicazione-CMAS.pdf
[8] Recreational Scuba Training Council. Common Hand Signals for Recreational Scuba Diving. (retrieved May 5, 2018). [Online]. Available: http://www.neadc.org/CommonHandSignalsforScubaDiving.pdf
[9] Scuba diving fan club. Most common diving signals. (retrieved May 5, 2018). [Online]. Available: http://www.scubadivingfancub.com/Diving_Signals.html
[10] 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.