Multimodal Interaction with Co-located Drones for Search and Rescue

Jonathan Cacace, Alberto Finzi, and Vincenzo Lippiello

Università degli Studi di Napoli Federico II
{jonathan.cacace,alberto.finzi,lippiello}@unina.it

Abstract. We present a multimodal interaction framework that allows a human operator to interact with co-located drones during search and rescue missions. In contrast with usual human-multidrones interaction scenarios, in this case the operator is not fully dedicated to the control of the robots, but directly involved in search and rescue tasks, hence only able to provide fast, although high-value, instructions to the robots. This scenario requires a framework that supports intuitive multimodal communication along with an effective and natural mixed-initiative interaction between the human and the robots. In this work, we describe the domain along with the designed multimodal interaction framework.

Introduction

We present a multimodal interaction framework suitable for human-multidrone interaction in search and rescue missions. This work is framed in the context of the SHERPA project [1, 13] whose goal is to develop a mixed ground and aerial robotic platform supporting search and rescue (SAR) activities in a real-world alpine scenario. One of the peculiar and original aspects of the SHERPA domain is the presence of a special rescue operator, called the *busy genius*, that is to cooperate with a team of aerial vehicles in order to accomplish the rescue tasks. In contrast with typical human-multidrones interaction scenarios [8, 4, 14], in place of an operator fully dedicated to the drones, we have a rescuer which might be deeply involved in a specific task, hence only able to provide incomplete, although high-value, inputs to the robots. In this context, the human should focus his cognitive effort on relevant and mission-critical activities (e.g. visual inspection, precise maneuvering, etc.), while relying on the robotic autonomous system for more stereotypical tasks (navigation, scan procedures, etc.). In this work, we illustrate the multimodal and mixed-initiative interaction framework we are designing for this domain. The multimodal interaction system should allow the human to communicate with the robots in a natural, incomplete, but robust manner exploiting gestures, vocal-, or tablet-based commands. The interpretation of the multimodal communication is based on a late fusion classification method that permits a flexible and extensible combination of multiple interaction modalities [18]. In order to communicate with the robots, we assume the human equipped with a headset and *Myo Gesture Control Armband*[1]

[1] https://www.thalmic.com/en/myo/

for speech-based and gesture-based interaction respectivelly. In this domain, we introduced a set of multimodal communication primitives suitable for the accomplishment of cooperative search tasks. We consider both *command-based* and *joystick-based* interaction metaphors, which can be exploited and smoothly combined to affect the robots behavior. In order to test the framework and the associated interaction modalities, we developed a test-bed where a human operator is to orchestrate the operations of simulated drones while searching for lost people in an alpine scenario.

Multimodal Interaction with Multiple Drones

We designed a modular architecture suitable for supervising and orchestrating the activities of both groups of robots and single robots (see Fig. 1). The operator should be capable of interacting with the system using different modalities (joystick, gestures, speech, tablet, etc.) at different levels of abstraction (task, activity, path, trajectory, motion, etc.). These continuous human interventions are to be suitably and reactively interpreted and integrated in the robotics control loops providing a natural and intuitive interaction. Feedback from the drones, beyond line-of-sight control, can be provided via a tablet interface, the headphones, and the armband.

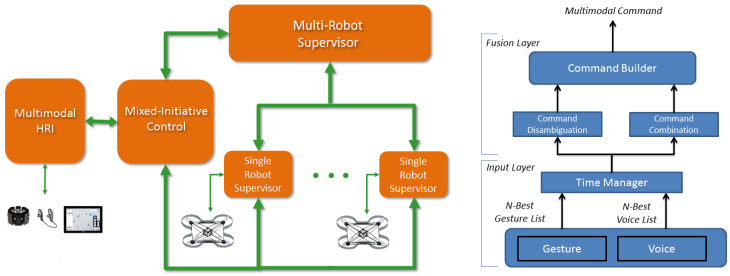

Fig. 1. The overall HRI architecture (left) and multimodal framework (right)

Multimodal Interaction. The multimodal interaction block in Fig. 1 (right) allows the human operator to naturally interact with the drones while interpreting the underlying intention. For instance, while the voice commands concern movement, selection, and exploration commands, gesture-based communication can be used to specify navigational commands with deictic communication (e.g. "*go-there*") with the co-located drones. We consider multimodal commands for drone *selection* and *navigation* (e.g. "*you go down*", "*red hawks land*"), *search* (e.g. *search-expanding*, *search-parallel-track*, according to helicopter search [15, 16, 2] standards), *switch* modality, to change the interaction mode and metaphor, e.g.

from command-based to joystick-like interactive control. Indeed, gestures can be also used as joystick-like control commands to manually guide the robots or to adjust the execution of specific tasks.

We employ a multimodal interaction framework that exploits a late fusion approach [18] where single modalities are first separately classified and then combined. Speech recognition is based on Julius[12], a two-pass large vocabulary continuous speech recognition (LVCSR) engine. The proposed *gesture recognition* system exploits the *Thalmic Myo Armbend*. This device permits to detect and distinguish several poses of the hand from the electrical activity of the muscles of the arm where the band is weared. In addition, the band is endowed with a 9 *DOF* IMU for motion capture. In our framework, the position of the hand is used to enable/disable control modalities/metaphors (*switch* commands), while the movements of the hand are to be interpreted as gestures. The *fusion module* combines the results of vocal and/or gesture recognition into a uniform interpretation. We deploy a late fusion approach that exploits the confidence values associated with the single modalities, i.e. speech and gestures. First of all, the two channels are to be synchronized. We assume that the first channel that becomes active (speech or gesture) starts a time interval during which any other activity can be considered as synchronized. When this is the case, contextual rules are used to *disambiguate* the conflicting commands or to *combine* vocal and gesture inputs exploiting the information contained in both the channels.

Mixed-Initiative Interaction. The operator is allowed to interact with the drones at any time at different levels of abstraction, while using different interaction metaphors [17]. The abstract inputs of the human are further interpreted by the mixed-initiative module (MI) that interacts with the single robot supervisor (SRS) and multirobot supervisors (MRS) mediating between the human and the robotic initiative. For instance, commands can be provided to both single or multirobots, and sketchy instructions are to be completed and instantiated by the robots supervisory systems. In our framework, both the MRS and the SRS are endowed with a BDI executive engine [10] and task/path/motion planners are to support mixed-initiative control and sliding autonomy. Following a mixed-initiative planning approach [9, 6, 7] the human interventions are integrated into continuous planning and execution processes that reconfigure the robotic plans of activities according to the human intentions and the operative state. At a low level of interaction, the operator can continuously interact with path/trajectory planners with multimodal inputs. In particular, we distinguish between *command-based* and *joystick-based* interaction using hand position to switch between these two modalities. We defined an intuitive switching strategy based on the armband: if the hand is closed the command-based gesture interpretation is enabled, otherwise, if the hand is open the joystick-like control is active. In a *command-based* interaction the human can issue commands to the drones using voice, tablet, gestures, while in the *joystick-based* the planned trajectory can be directly modified during the execution. For example, the operator can start the execution of a task (e.g. "blue hawks go there"), and during the execution, use the *hand-open* mode to correct the trajectory. A full hand-open

teleoperation can be reached once the current command execution has been stopped with a *brake*. Specifically, we deploy a RRT^* algorithm [11] for path planning, while trajectory planning is based on a 4-*th* order spline concatenation preserving continuous acceleration. Following the approach in [5], during the execution, the human is allowed to on-line adjust the robot planned trajectory without provoking replanning. Moreover, similarly to [5, 3], we assume that the human operator can move the robot within an adaptive workspace, while a replanning process is started when the human operator moves the robot outside this area.

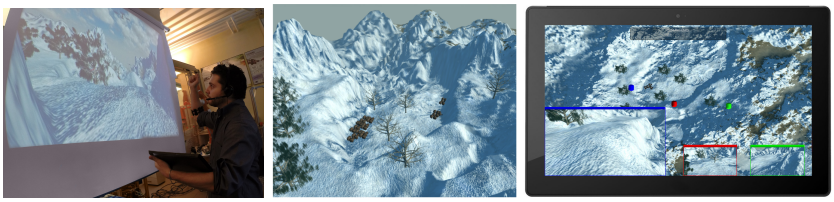

Fig. 2. Simulated alpine scenario testbed (left, center), tablet interace (right).

Testing Scenarios. We are currently testing the effectiveness of the multimodal interaction framework in a simulated alpine scenario (see Fig. 2, left). We used *Unity 3D* to simulate a set of dornes equipped with an onboard camera. A tablet-based user interface (see Fig. 2, right) allows the operator to monitor the robots position on a map, while receiving video streams for the cameras of the drones on multiple windows, the one associated with the selected drone has a bigger size. In this scenario, a set of victims is randomly positioned within the environment and the mission goal is to find a maximum number of missed persons within a time deadline. This setting allows us to compare the user performace and behavior in different conditions, changing the number of drones, the available modalities, the tablet interface, the time pressure, etc.. We are currently testing the overall system, initial results show that the interaction system allows the human to naturally interact with two or three drones with a satisfactory distribution of the search effort.

Acknowledgement

The research leading to these results has been supported by the SHERPA and SAPHARI projects, which have received funding from the European Research Council under Advanced Grant agreement number 600958 and 287513, respectively. The authors are solely responsible for its content. It does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of the information contained therein.

References

1. Sherpa, eu collaborative project ict-600958, http://www.sherpa-project.eu/sherpa/
2. Bernardini, S., Fox, M., Long, D.: Planning the behaviour of low-cost quadcopters for surveillance missions. In: Proc. of ICAPS 2014. pp. 446–453 (2014)
3. Bevacqua, G., Cacace, J., Finzi, A., Lippiello, V.: Mixed-initiative planning and execution for multiple drones in search and rescue missions. In: Proc. of ICAPS. pp. 315–323 (2015)
4. Bitton, E., Goldberg, K.: Hydra: A framework and algorithms for mixed-initiative uav-assisted search and rescue. In: Proc. of CASE 2008. pp. 61–66 (2008)
5. Cacace, J., Finzi, A., Lippiello, V.: A mixed-initiative control system for an aerial service vehicle supported by force feedback. In: Proc. of IROS 2014. pp. 1230–1235 (2014)
6. Cacace, J., Finzi, A., Lippiello, V., Loianno, G., Sanzone, D.: Aerial service vehicles for industrial inspection: Task decomposition and plan execution. In: Proc. of EIA-AIE 2013 (2013)
7. Cacace, J., Finzi, A., Lippiello, V., Loianno, G., Sanzone, D.: Aerial service vehicles for industrial inspection: task decomposition and plan execution. Appl. Intell. 42(1), 49–62 (2015)
8. Cummings, M., Bruni, S., Mercier, S., Mitchell, P.J.: Automation architecture for single operator, multiple uav command and control. The International Command and Control Journal 1(2), 1–24 (2007)
9. Finzi, A., Orlandini, A.: Human-robot interaction through mixed-initiative planning for rescue and search rovers. In: Proc. of AI*IA 2005. pp. 483–494 (2005)
10. Ingrand, F., Georgeff, M., Rao, A.: An architecture for real-time reasoning and system control. IEEE Expert: Intelligent Systems and Their Applications (1992)
11. Karaman, S., Frazzoli, E.: Incremental sampling-based algorithms for optimal motion planning. In: Robotics: Science and Systems (RSS) (2010)
12. Lee, A., Kawahara, T., Shikano, K.: Julius–an open source real-time large vocabulary recognition engine. In: Proceedings of EUROSPEECH-2001 (2001)
13. Marconi, L., Melchiorri, C., Beetz, M., Pangercic, D., Siegwart, R., Leutenegger, S., Carloni, R., Stramigioli, S., Bruyninckx, H., Doherty, P., Kleiner, A., Lippiello, V., Finzi, A., Siciliano, B., Sala, A., Tomatis, N.: The sherpa project: Smart collaboration between humans and ground-aerial robots for improving rescuing activities in alpine environments. In: Proceedings of SSRR-2012. pp. 1–4
14. Nagi, J., Giusti, A., Gambardella, L., Di Caro, G.A.: Human-swarm interaction using spatial gestures. In: Proc. of IROS-14. pp. 3834–3841 (2014)
15. NATO: ATP-10 (C). Manual on Search and Rescue. Annex H of Chapter 6 (1988)
16. NATSAR: Australian National Search and Rescue Manual. Australian National Search and Rescue Council (2011)
17. Pfeil, K., Koh, S.L., Jr., J.J.L.: Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In: Proc. of IUI 2013. pp. 257–266 (2013)
18. Rossi, S., Leone, E., Fiore, M., Finzi, A., Cutugno, F.: An extensible architecture for robust multimodal human-robot communication. In: Proc. of IROS 2013. pp. 2208–2213 (2013)