A Human Multi-Robot Interaction Framework for Search and Rescue in the Alps

Jonathan Cacace, Riccardo Caccavale, Alberto Finzi, and Vincenzo Lippiello

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

Abstract. In this work, we present a framework that allows a single operator to monitor and control the operations of multiple robots during Search & Rescue missions in an alpine environment. This work is framed in the context of the SHERPA project whose goal is to develop a mixed ground and aerial robotic platform to support search and rescue activities in alpine scenario. In this context, the human operator is not fully dedicated to the robot control, but involved in the search and rescue mission, hence only able to partially monitor and command the robotic team. In this paper, we briefly illustrate the overall framework and describe on-the-field tests with two drones equipped with on-board cameras.

Introduction

We present an architecture suitable for human multi-robot interaction for search and rescue missions in an alpine environment. This work is framed in the context of the SHERPA project [8] whose goal is to develop a mixed ground and aerial robotic platform supporting search and rescue (SAR) activities in an alpine scenario. In this context, a special rescue operator, called *busy genius*, can monitor and interact with a team of robots during the mission operations. In particular, we focus on the interaction of the operator with a set of drones. In contrast with typical human-UAVs interaction scenarios [6, 1], in place of a fully dedicated operator, we have a rescuer which can be deeply involved in the SAR mission, hence only able to provide incomplete and sparse inputs to the robots. In this setting, the operator should focus his/her cognitive effort on relevant and critical activities (e.g. visual inspection, precise maneuvering, etc.), while relying on the robotic autonomous system for specialized tasks (navigation, scan, etc.). Moreover, the human should operate in proximity with the drones in hazardous scenarios (e.g. avalanche), hence the required interaction is substantially dissimilar to the one considered in other works where the human and co-located UAVs cooperate in controlled indoor conditions [9]. The interaction framework illustrated in this paper allows the operator to interact in a natural and efficient way with the flying robots. We assume the human equipped with light and wearable devices, such as a headset, a gesture control armband, a health monitor bracelet, and a mobile device endowed with a touch user interface. In addition, these devices are used by the operator to get information from the robots, monitoring

their state and retrieving relevant information about the search mission visualizing the data acquired with the on-board robot sensors. In order to track the human pose the headset of the operator has been equipped with standard *GPS* and *IMU* sensors. Finally, a bracelet permits to track the operator status (GSR, heart-rate, temperature, etc.). The proposed framework allows the operator to monitor and control the team of robots in an adaptive manner, ranging from a minimally supervised autonomous robotic team, when the human is busy, to direct and docile teleoperation, when the operator takes the direct control of one of the robots. In order to test the effectiveness of the presented architecture, initial on-the-field experiments have been performed in which a co-located human operator controls two drones equipped with an on-board cameras in order to explore an alpine area.

System Architecture

The proposed architecture is depicted in Figure 1. The operator is equipped with light wearable devices to interact with the robotic system. The output of these devices is continuously sent to the Multimodal Human-Robot-Interaction (*MHRI*) module in order to generate new commands. If the command provided by the human operator does not explicitly assign tasks to the robots, the Distributed Multi-Robot Task Allocation (*DMRTA*) module is to decompose abstract tasks into primitive action finding a valid allocation for the robotic team members; when the task is already completely specified and allocated this is validated with respect to resources and mission constraints. In the following we describe the associated modules.
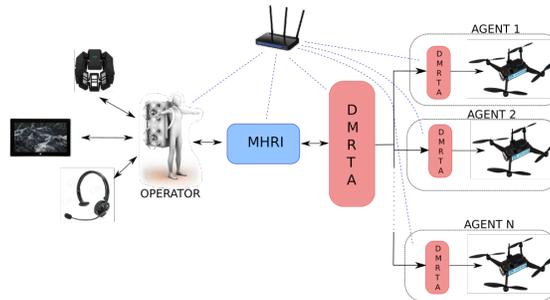

Fig. 1. The overall control architecture. The Multimodal Human-Robot Interaction (MHRI) module allows the human to interact with the robots, which are supervised by a Distributed Multi-Robot Task Allocation System (DMRTA).

Multimodal Interaction. The *MHRI* module interprets the operator commands integrating inputs from multiple communication channels. For instance, either speech- or gesture-based commands may be used to stop a drone, while vocal commands in combination with deictic gestures can be used to specify navigational commands (e.g. "*go-there*") to the co-located drones. We mainly focus on

commands suitable for interacting with the set of co-located drones during navigation and search tasks. More specifically, we are concerned with multimodal communication with the drones suitable for the following purposes: robot *selection* commands (e.g. "*all wasps*", "*red wasp*", "*you wasp*"); *motion* commands (e.g. "*go there*", "*land*", etc.); *search* primitives (to scan an area with a specific search pattern); *multimedia* commands used to acquire data through on-board sensors (e.g. pictures, video, environment mapping, etc.); finally, *switch* meta level commands allow the operator to change the interaction mode (e.g. high-level commands, interactive control, teleoperation). We rely on the *Julius* framework for continuous speech recognition. As for gesture recognition, we recognize harm, hand, and tablet-based gestures. We exploit the *Thalmic Myo Armband* to detect and distinguish several poses of the hand (from the electrical activity of the muscles) and the arm (the band is endowed with a 9 *DOF* IMU that permits motion capture). The results form these multiple inputs (gestures, voice, hand, tablet, etc.) are combined into a unique interpretation of the operator intention exploiting a late-fusion approach: single modalities are first classified and synchronized, then fused into a single interpretation using the confidence value associated with each input data (see Figure 2, left).
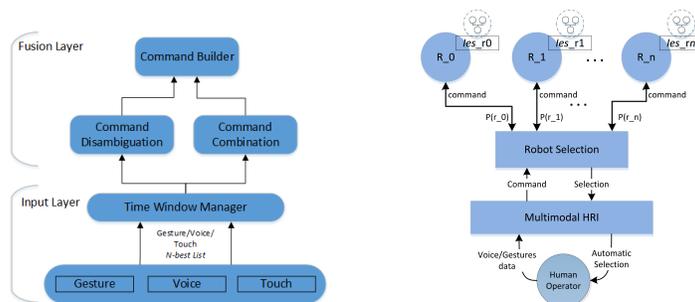

Fig. 2. Multimodal framework (left) and implicit task assignment system (right)

Implicit Drone Selection. In order to simplify the interaction with the robotic team, we permit both implicit and explicit robot selection. Namely, we propose an approach where, each available robot can evaluate the probability to be the one designated by the human for the execution of a command when the target is not explicitly assigned by a selection command, but it can be inferred by the context (see Figure 2, right). The robot evaluation process relies on a multi-layered architecture in which a *Dynamic Bayesian Network* is deployed to infer the human intentions form the state of the robots and learned contextual and geometrical information. More details can be found in [3].

Distributed Multi-Robot Task Allocation. The *DMRTA* module is responsible for multi-robot task allocation. Specifically, given a task to perform, this module should find a suitable decomposition and assignment to the robotic team. This process takes into account different constraints, such as resources and capabilities

needed to perform the task (e.g. *take-a-picture* needs a robot equipped with a camera), the state of the robotic system and time constraints. Moreover, the *DMRTA* is also to validate the feasibility of the operator's request (see [7] for additional details).

Multi-robot Supervision and Adaptive Attentional Interface. A suitable interface is need to filter and adapt the information presented to the busy genius through different communication channels: head-set (audio), tablet (video), band (vibrotactile). We assume that information filtering and adaptive presentation of data are managed by an attentive interface modeled as a supervisory attentional system that regulates contentions among multiple bottom-up stimuli (alerts, periodic information, warnings, state changes, etc.) depending on the mission state and the human constraints. This attentional regulation process [5] takes into account the operative tasks of the human and the drones, the limited human workload for each communication mode (visual, audio, vibro-tactile) along with timing and task switching constraints. In [2] a detailed description of this framework is provided.

Testing Scenario

We now illustrate an initial on-the-field experimentation with the proposed framework. In the experimental setting the human operator is located in a real alpine scenario (i.e. *Pordoi Pass* in Alps: 2200 meters altitude) endowed with his/her wearable devices to control two drones (called textitGreen and *Red* wasp), each equipped with a standard camera. In this set-up human monitoring and the adaptive interface were not enabled. The goal of the operator is to inspect two different regions of the area depicted in Figure 3 (*Left*). The operator exploits the on-board camera of the drones to visualize real-time video streaming and collect pictures of the terrain [4]. At the beginning of the mission the drones are landed on the ground. The user can select the robots either via the user interface or via vocal commands. Different communication channels are used to drive the *Green Wasp* towards a desired location and then request the scanning of an area. During this operation, several pictures of the terrain are autonomously taken by the robot and provided to the operator via the user interface as show in Figure 3 (b - e, *Right*). At the same time, while the *Green Wasp* is autonomously executing the scanning mission, the user can focus on the *Red Wasp* navigating in *Direct Control* mode towards a different area not covered by the *Green wasp*, while receiving the video streaming from the on-board camera. An example of the interaction in *Direct Control* mode is shown in Figure 3 (f - h, *Right*), when this control mode is active the operator can directly generate velocity data from the orientation of its arm. At the end of the test, all the drones are driven back to their initial position. In Figure 3 (*Left*) we illustrate the trajectories of the green and red drone along with the associated inspected zones. The duration of the reported flying mission is about 3 minutes, with a covered area of about 4200 m^2. The two zones are adjacent and non-overlapping with a satisfactory
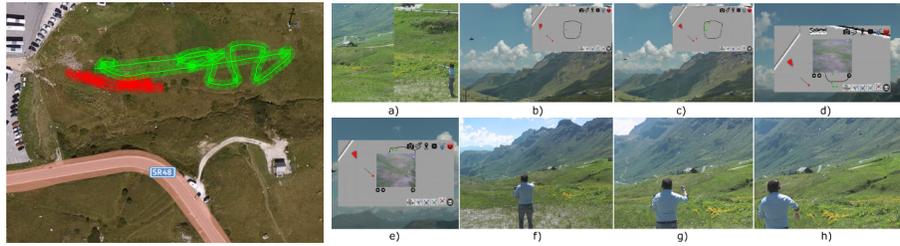
Fig. 3. (Left) Areas covered by the *red* and *green* drones. (Right) Operator during the on-the-field experiment:(*a*) command for the green wasp, (*b - e*) *User Interface* with pictures taken during a drone search, (*f - h*) *Direct Control* mode.

coverage of the area; the red zone is smaller because inspected be the human the *Direct Control* mode.

References

1. 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)
2. Cacace, J., Caccavale, R., Finzi, A., Lippiello, V.: Attentional multimodal interface for multidrone search in the alps. In: Proc. of SMC-16 (2016)
3. Cacace, J., Finzi, A., Lippiello, V.: Implicit robot selection for human multi-robot interaction in search and rescue missions. In: Proc. of Ro-Man-16 (2016)
4. Cacace, J., Finzi, A., Lippiello, V., Furci, N., Mimmo, N., Marconi, L.: A control architecture for multiple drones operated via multimodal interaction in search & rescue mission. In: Proc. of SSRR-16 (2016)
5. Caccavale, R., Finzi, A.: Flexible task execution and attentional regulations in human-robot interaction. IEEE Trans. Cognitive and Developmental Systems 9(1), 68–79 (2017)
6. 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)
7. Doherty, P., Heintz, F., Kvarnström, J.: High-level Mission Specification and Planning for Collaborative Unmanned Aircraft Systems using Delegation. Unmanned Systems 1(1), 75–119 (2013)
8. 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
9. Szafir, D., Mutlu, B., Fong, T.: Communication of intent in assistive free flyers. In: Proc. of HRI '14. pp. 358–365 (2014)