Implicit Robot Selection for Human Multi-Robot Interaction in Search and Rescue Missions

Jonathan Cacace[1], Alberto Finzi[1] and Vincenzo Lippiello[1]

***Abstract*— We present a system suitable for human multi-robot interaction that supports the operator in the robot selection process. The proposed framework allows a human to issue commands to a robotic team without an explicit robot selection. This work is framed in the operative context of the SHERPA project [1], which proposes the deployment of a robotic platform for Search & Rescue in an alpine scenario and assumes the presence of a human rescuer that can orchestrate the robots operations with multimodal commands. In this context, implicit robot selection is mainly motivated by fast communication and the difficulties to distinguish different robots of similar shape in a hazardous environment and in adverse weather conditions. In the proposed approach, each robot of the team can evaluate the probability to be referred in an incomplete command, considering its actual capabilities along with geometrical and contextual information. We describe the overall system architecture focusing on the human intention recognition process. The proposed framework is trained and evaluated in a simulated case study.**

I. INTRODUCTION

In order to facilitate the collaboration between a human operator and a multi-robot system, the robot ability to interpret non verbal cues and react accordingly is a crucial issue, in particular when complex interactive tasks are to be executed [2]. In this paper, a framework for implicit selection of robots in human multi-robot interaction is presented. Specifically, we will address a *robot selection problem*, in which the human operator must designate a particular robot of interest as the selected one in order to assign a task. We propose an approach where the human operator can omit to explicitly indicate the intended robot in task assignment, because the robotic team is able to infer the candidate that best match the operator's intention. Our work is framed in the context of SHERPA project [1] whose goal is to develop a mixed ground and aerial robotic platform for search and rescue operations in an alpine environment. A sketch of SHERPA scenario is depicted in Figure 1. In this domain a human rescuer leads a team of heterogeneous robots in the search of survivors after an avalanche. The robotic team is mainly composed of aerial vehicles (in this paper we will assume quadcopters) equipped with different types of sensors in order to assist the human operator in the rescue mission. In this context, the human operator is not fully dedicated to the control of the robots, but he is involved in the rescue operations too. In order to enable an effective and natural interaction with the robots, the human operator is endowed with wearable devices (*gesture control armband*, *headset*, *etc...*) that allows him to orchestrate the team operations using voice and gestures commands, and a mobile devices (*tablet*) in order to receive back relevant video information about the mission retrieved by the robotic team. In addition, we endow the operator's headset of a pose estimation module composed by a *GPS* and a *IMU* in order to track his position and orientation. As for the robotic platforms, we will mainly refer to electrical flying robots whose main limitation is the battery duration. In order to address this issue, the robotic team includes a ground rover that works as a docking station for the drones supporting both landing and battery recharging. In this scenario, depending on the battery charge, we have a continuous reconfiguration of the active members of the robotic team and their actual capabilities. Selecting

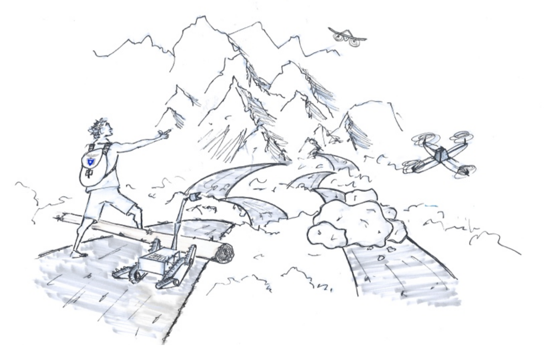

Fig. 1. A sketch of the SHERPA scenario [1]

and commanding individual robots in this setting, without the support of specialized user interface, could represent a challenge. For this reason, we propose a method in which the human operator is able to select the desired robot in a non verbal and implicit manner. Specifically, we propose a method in which each available robot can evaluate the probability to be the one designated by the human for the execution of a command. The single 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 a learned contextual and geometrical information. Finally, in order to train the system and test its effectiveness, we defined a simulated interaction scenario where a human operator is to lead a group of drones in a search mission in alpine scenario.

The paper is organized as follows: in Section II we introduce the main motivation of our work. In Section III a brief overview of related works is presented. In Section IV the multi-layer architecture for the probability evaluation

[1] Universita' degli Studi di Napoli Federico II, via Claudio 21, 80125, Naples, Italy. {jonathan.cacace,alberto.finzi,lippiello}@unina.it

is described. In Section V the whole architecture for the interaction between the human operator and the multi-robot system is presented. Finally, in Section VI the system training and testing phases are described.

II. Motivation

We consider the following motivational scenario. The human operator is involved in a rescue mission after an avalanche in adverse weather conditions and decides to exploit the fast flying capabilities of the *UAVs* to retrieve important information by scanning quickly a large area. In order to accomplish this mission, the robot responsible of the scanning must be properly selected and this is usually performed by selecting the desired robot using vocal keywords such as the *code-name* or the *id* of the desired robot (e.g *Red Wasp* or *Wasp Zero*). However, the hard operative domain caused by the adverse weather conditions or the hazardous environment along with the limited time to accomplish the rescue mission affect the psychophysical state of the rescuer and his cognitive capacity. Moreover, even thought the human operator is co-located with the robotic team, the similar shape of different robots could make difficult for him to select the desired robot using verbal communication without running into misunderstanding on the correct robot *id*, in so provoking the selection of robot with different capabilities or resources or in different location. Furthermore, the operative environment presents different unmodelled sound noise sources such as the wind, the propeller of the robots or the helicopters used for the transportation of the rescuer or the victims. This could even increase the failure rate of the automatic speech recognition algorithm affecting the effectiveness of the overall interaction system.

III. Related Work

Unmanned Aerial Vehicles (*UAVs*) are extensively employed in several service applications such as in industrial building inspection [3][4], surveillance, remote sensing and many others. Related to our domain, [5] and [6] demonstrate how Search & Rescue operations can greatly benefit from the use of autonomous UAVs to survey the environment and collect evidence about the position of a missing person. In this context, different works have been focussed on the problem of controlling and commanding UAVs in such scenarios. In [7] a mixed-initiative system for the supervision of multiple drones suitable for Search & Rescue mission in alpine scenario is presented, while in [8] test bed and algorithms for collaborative human and autonomous decision making within the context of Wilderness Search & Rescue is presented.

In the literature different techniques who aim to semplify the interaction between a human operator and a multi-robot system have been proposed. In particular, in [9] an intuitive interface to map the natural motion of the user to the controlled movements of flying robots equipped with onboard camera is designed in order to provide an intuitive control in the 3D space during teleoperation. Considering the *robot selection problem*, in [10] authors designed an *ad-hoc* user interface to help the human operator to select and command a single robot in a multi-robot system composed of very large groups of robots, while in [11] and [12] a solution that rely on face engagement is proposed where eye contact between the human and the robot is used to select the robot to control. These kind of approaches are not applicable to our noisy and outdoor domain because of the distance between the operator and the robot and the low visibility conditions. In addition, in our scenario the human operator can command a robot even beyond line-of-sight fly. From a different perspective, in *Multi Robot Task Allocation* (*MRTA*) the robot selection is obtained considering a feasible assignment of tasks to the robot that optimizes some objective [13]. Notice, that our aim instead is to select the robot that best matches the human operator intention, even when this selection is far from an optimal choice.

Finally, the solution proposed in this work partially realies on probabilistic graphical models like *Dynamic Bayesian Network* (*DBN*) [14][15]. The use of *DBN* has been widely explored in human robot interaction applications mainly in the field of task and activity recognition in order to allow the robot to estimate the plan of the human and proactively react to operators intentions[16][17]. In contrast, in our approach the *DBN* is used to provide an estimation of the human intention regarding the selection of a robot among a group of possible candidates.

IV. Implicit Robot Selection

In this section we describe the implicit robot selection process. Given the set of all the available robots $R = \{r_1, r_2, ..., r_n\}$, $A_R(t) \subseteq R$ represents the set of the active robots in the rescue mission at time t. The robot $i \in R$ is endowed with the set of capabilities $K_i \subseteq K$, where $K = \{k_1, k_2, ..., k_j\}$ is the set of all the available capabilities, and its state is represented by the pair $s_i =< b_i, f_{s_i} >$, where b is the battery level and f_s is the flying status of the robot. Let $C = \{c_1, c_2, ..., c_k\}$ be the set of all possible commands the human operator can invoke. Given a command $c_j \in C$ that requires the set K_j of capabilities to be properly executed, the probability for the robot $i \in A_R(t)$ to be the referred in the command c_j is $P(r_i) = P(r_i \mid c_j)$. As previously stated, in our approach each robot of the team is able to estimate this probability when a command is requested by the operator. In order to perform this estimation process, we endow the robots with the *Intention Estimation System* depicted in Figure 2 that is to estimate the probability $P(r)$ fusing different kind of data. Specifically, upon receiving the new command c_j, the estimator calculates three different factors α, β and γ associated with, respectively, *Capability*, *Geometrical* and *Contextual* information. The final probability value is defined as $P(r \mid c_j) = w_1\alpha + w_2\beta + w_3\gamma$, where w_1, w_2 and w_3 are empirically estimated weights.

In the scenario considered in this paper, we mainly focus on navigation and multimedia commands, such as *take-off*, *land*, *go* along a direction, *stop*, *rotate*, *explore*, *take-a-picture* and *start/stop recording* a video. In addition, in order to assist

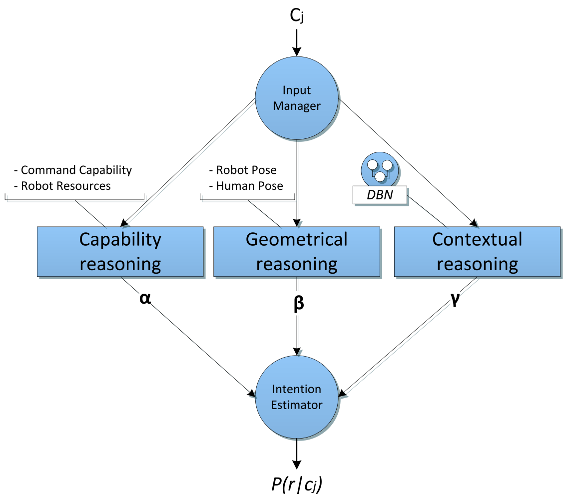

Fig. 2. The *Intention Estimation System* architecture

the human operator in the rescue mission, the robot can be equipped with a *camera*, a *camcorder* or an *ARVA*[1] sensor. Finally, we define three different operative states of the robot: *idle*, *hovering* and *flying* if the robot is, respectively, hold on the ground, flying but not employed in any mission or navigating toward a destination point.

A. Capability Reasoning

Each robot should be capable of evaluating its own ability to perform a possible command (e.g. a take-a-picture command should be referred a robot equipped with a camera). For this purpose, the robot estimates an α value taking into account its equipment (capabilities) and the available power (resources). As for the capabilities, each robot r estimates a probability $P(K_r \mid c)$ that a given command c refers to its set of capabilities K_r. In the second case, the robot estimates a probability $P(B_r \mid c)$ that a given command refer to robot r with an estimated power charge B_r after the execution of c, this allows the robot to evaluate the command assuming that the human is aware about the associated energy consumption.

B. Geometrical Reasoning

This module assesses geometrical relationships between the commands and the poses (position and orientation) of the human and the robot. For instance, the robots closer to the operator or in his field of view could have a higher probability to be selected for specific commands. In order to discover these relationships, we collected a domain specific training set (see Section VI), where, for each command we consider the orientation of the operator with respect of the selected robot and the absolute distance between the two. Specifically, two different probabilistic values $\beta_1 = P(dist(h,r) \mid c)$ and $\beta_2 = P(fov(h,r) \mid c)$ are evaluated by the robot, where $dist(h,r)$ and $fov(h,r)$ are, respectively, the distance and the orientation of the human with respect to the robot. The overall β value is calculated as a weighted and normalized sum of β_1 and β_2. In this context, the absolute distance and the orientation between the human operator and the robots are calculated considering *GPS* and *IMU* information provided by the standard equipment of the quadrotors and the sensorized headset of the operator.

[1]The *Appareil de Recherche de Victimes en Avalanche* is an instrument commonly used to find victims of avalanches.

C. Contextual Reasoning

This layer represents the core of the *Intention Estimation System*. At this level, the probability value is calculated exploiting of a Dynamic Bayesian Network (*DBN*). A Bayesian Networks N is a triplet (V, A, P) where V is a set of random variables, A is a set of arcs and $P = P\{(v \mid \pi_v) : v \in V\}$, $G = (V, A)$ and P represents, respectively, a directed acyclic graph and the set of conditional probabilities of all variables given their parents. Similarly, a *DBN* captures the development of the network over time steps. The proposed *DNB*, illustrated in Figure 3, allows the robot to infer on the operator's intention given the contextual information represented in the nodes of the network. In particular, the probability density of the robot r over the command c is calculated considering only the contextual information $cont(r)$ as $P(cont(r) \mid c)$ from the $P(cont(r)_t \mid cont(r)_{t-1}, c)$ provided by the network. The proposed network (see Figure 3) in composed of the following nodes:

- *Status node*: the robot operative state, its battery level, and the time elapsed from last received command.
- *Command node*: the last received command;
- *Robot node*: This node represents the probability that the robot is selected by the operator.

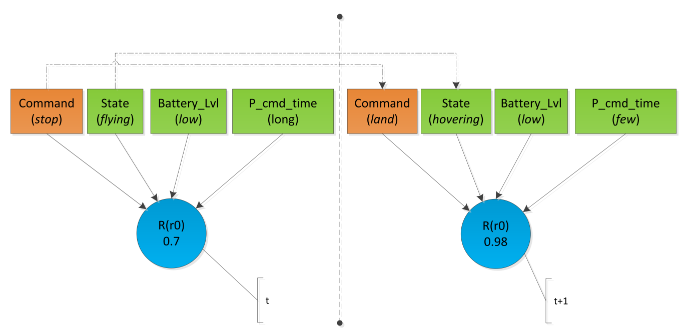

Fig. 3. Dynamic Bayesian Network. In *orange* the command node, in *green* the status nodes and in *blue* the robot node.

Here, the contextual information is defined by status and command nodes, while the *Robot* node estimates the probability of a robot to be the selected. Notice, that the contextual nodes assume discretized values.

V. Architecture

The overall system architecture is depicted in Figure 4 and defines the interface between a heterogeneous team of robots and the human operator. The *Human Operator* interacts with the *Robot Selection* module via a set of multimodal commands generated by the *Multimodal HRI* module on the base of the input provided by the human operator via a combination of voice and gestures. In order to fuse different type of information sources this module adopts a *late fusion* approach in which different signals are first separately classified and then combined. In this context, the human operator exploits voice and gestures in order to combine or disambiguate single commands, as already proposed in [18]. Basically, the human operator can choose to issue commands using just voice, just gesture or a combination of voice and gestures. In the last case an information channel act as a *redundant* or a *complementary* command to the other in which the redundancy is used by the operator to make more robust the command recognition (*i.e.* the same command is given both in gesture and in speech in order to consider the one with higher confidence) while the *complementary* is used to create more complex commands (*e.g. Go there*: saying Go and pointing in a specific direction *There*). We considered different classes of recognizable commands typical of the aerial robotic domain. In particular, *Primitive* commands are used to invoke atomic action such as *take-off*, *land* or *brake*. *Motion* commands are used to allow the motion of the robot along a desired direction (e.g *Go-ahead* for a certain distance), or the quadrotor rotation. *Search* commands allow the robot to perform a search pattern and scanning an area. *Multimedia* commands allow the robot to use its on-board sensors to take pictures or recording a video.

Upon received a new command, the *Robot Selection* module is responsible for contacting all the active robots of the team in order to collect their probability estimation of being selected and then select the one with the higher value. Finally, to produce the estimation of the operator's intention, each robot exploits the *Intention Estimation System* described in the previous section. Notice that, in our approach, we assume that each robot evaluates its own probability of being selected, neglecting the status of the other robots. We deliberately decided this simplified setting to better handle the continuous change of robots involved in the rescue mission. Notice that in our framework, implicit selection is not mandatory, and the human operator can always directly refer to a particular robot in an explicit manner using its code name. Moreover, in order to minimize the errors, once all the estimated probabilities have been collected from the robots, the *Selector* module uses two parameters τ_1 and τ_2 to identify and manage ambiguous selections. In particular, τ_1 provides a threshold on the probabilities for selectable robots, while τ_2 represents a minimum difference between the two best generated results needed to define a selection. When τ_1 and τ_2 are not satisfied, the selection is considered ambiguous and an explicit interaction with the operator can be started to disambiguate the human intention. Using *text-to-speech* technology, the *Multimodal HRI* can ask the human operator to choose among the most probable robots. The values of τ_1 and τ_2 are set by means of the learning process described in the next section.

VI. Training & Testing

In order to train and test the proposed system, different simulated interaction scenarios have been set up (see Figure 6). The simulated environment reproduces typical scenarios of our domain, while the operator can navigate the scene in

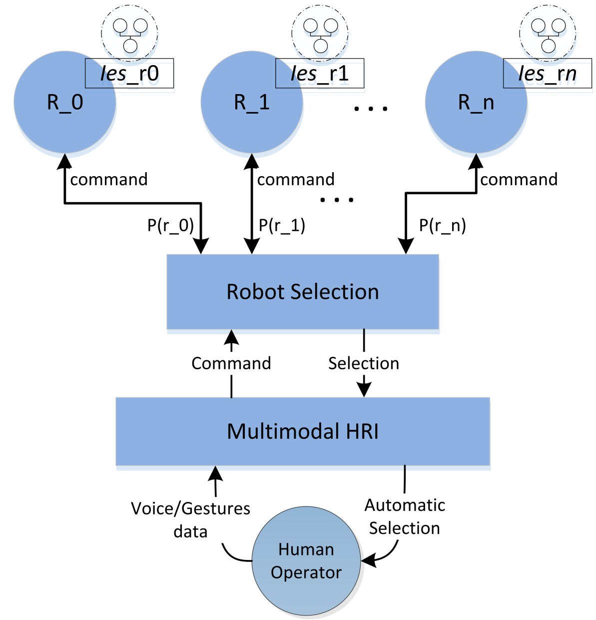

Fig. 4. System architecture

a first person perspective and control a group of maximum 6 drones. In this context, a tester can perform different tasks interacting with the simulator with *command-line* interface to send commands using a joypad to navigate the environment. The drones are differently coloured and the user can refer to them using code names based on colors (*e.g. Red Drone*). In order to get information about the status of the robot and the environment, the user can exploit a basic *user interface* shaped as a non-invasive map displayed in a corner of the monitor. This interface is intended to substitute the human operator's tablet used in real world scenario. An example of this interface is depicted in Figure 5, where the battery level, the flying state and the capabilities of the drones are illustrated for the user situation awareness while the drones are represented on the map as coloured spheres. Moreover, the streams of the drones' on-board camera can be inspected. In the following, we discuss the system training and evaluation.

A. Training

In the training session, we generate the training sets needed by the *Intention Estimation System*. For this purpose, three mission scenarios has been defined considering different situations. In the first scenario, represented in Figure 6*a*, the user was asked to land the drones on different landing zones represented by the red flat squares in the figure. Both the initial position of the drones and the landing zones are located close to the operator's initial position in order to encourage the user to use commands for line-of-sight navigation. In the second scenario, depicted in Figure 6*b*, the drones are all equipped with an on-board camera and the mission of the users is bring the robots in a landing zone located in a farther place with respect to the previous scenario. The aim here is to force the operator to use the on-board camera of the drones in order to avoid obstacles, such as wood or mountains, and follow them during the navigation. Finally in the last scenario, depicted in Figure 6*c*, the operator must command the robots over a mountain and use the drone on-board camera to acquire pictures or videos of a predefined area. In this context, the end of a training session is determined by the discharging of the batteries of the drones. In the training phase, we involved 7 different users already expert of the system. Once the *Intention Estimation System* trained, another training session is needed to adapt the τ_1 and τ_2 thresholds. This is obtained by asking the testers to execute another training session in which they validate the framework by accepting or rejecting the results of the selection process.

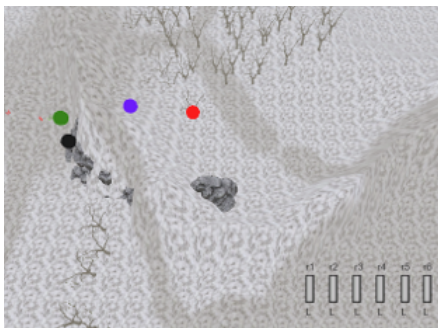

Fig. 5. User interface for the user situation awareness.

B. Case Study

The effectiveness of the proposed system has been evaluated involving a group of 10 testers (8 men), who were asked to perform a mission in the simulation scenario depicted in Figure 6*d* that combine the scenarios introduced in the training phase. Indeed, here, both navigation by sight and exploration with multimedia data acquisition tasks are considered. Specifically, the goal of the user is to both bring an arbitrary number of robots to a landing zone and to use the capability of the drones to explore a predefined searching area. During the test, the user is asked to confirm the correctness of the selection process results for each command. This way, we can collect the *True Positive* (*tp*), *False Positive* (*fp*) and *True Negative* (*tn*). The performances of the system are then reported in terms of *Precision*, *Recall*, *Accuracy*, *Sensitivity*, and *Specificity*, with the standard definition, i.e. $Precision = \frac{tp}{tp+fp}$, $Recall = \frac{tp}{tp+fn}$ while *Sensitivity* and *Specificity* are the *tp* and *tn* rates, respectively. Moreover, our aim is to assess both the *Intention Estimation System* and the overall system. Therefore, we designed two different test cases, with or without the thresholds check. The results of both tests are reported in Table I and in Table II, respectively. In the first test case, in the case of a wrong robot selection, tester can also assess the system error as fair alternative of the intended selection, i.e. the selected drone is different form the intended one, but the user considers it as an equivalent choice. The percentage of these mistakes is reported under the sm entry in Table I. Instead, in the second case, we consider in the percentage interactions needed to disambiguate a selection (*dialogue* entry in Table II).

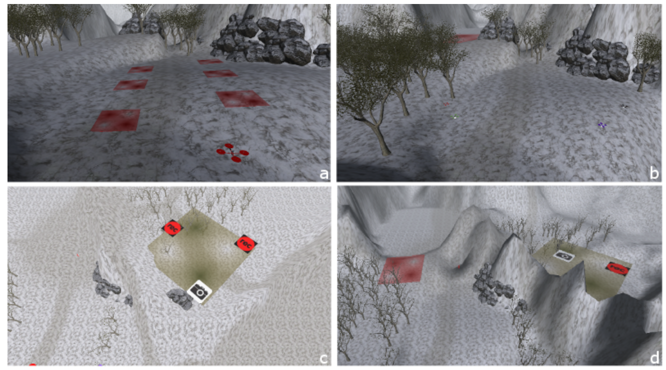

Fig. 6. Simulated environment for system training and testing.

TABLE I

INTENTION ESTIMATION SYSTEM RESULTS

Precision	*Recall*	*Accuracy*	*Sensitivity*	*Specificity*	*sm*
79%	79%	88%	79%	20%	31%

TABLE II

SYSTEM RESULTS

Precision	*Recall*	*Accuracy*	*Sensitivity*	*Specificity*	*dialogue*
91%	91%	95%	91%	8%	36%

The presented results shows that the intention estimator is able to correctly select the intended robot with a satisfactory performance. The enhanced performances in Table II, show that, as expected, the dialogue system used in the second test case enhances the precision of the selection process. On the other hand, a high percentage of the selection errors in the first test case are considered not relevant by the user.

VII. CONCLUSIONS

In this paper, we presented a system that supports human multi-robot interaction during the execution of collaborative interactive tasks by facilitating the robot selection process. In our approach the human operator is able to implicitly designate the robot responsible for the execution of a command, because each robot of the team can estimate the probability of being involved in the task from its current status and the operative context. This work is framed in the context of SHERPA [1] project in which a human operator is employed in Search & Rescue missions in winter alpine scenario leading a heterogeneous multi-flying-robot team using multimodal commands. In this context, the human operator interacts with the robots with fast and incomplete commands, while the hard operative conditions impairs a proper selecting of different robots of similar shape using code name keywords. For this purpose, an automatic selection process mechanism that uses human intention estimation for the robot selection has been designed. To estimate the human intention a three layer architecture has been proposed to fuse different kind of information such as the robot capabilities, geometrical and contextual information. The system has been trained and tested in a proposed simulation scenario that reproduces the typical situation of a Search & Rescue mission in alpine scenario. The obtained results seems to confirm the potential of the approach and encourages us towards more extensive evaluation tests in real-world scenarios.

ACKNOWLEDGMENT

The research leading to these results has been supported by the FP7-ICT-600958 SHERPA project.

REFERENCES

[1] L. Marconi, C. Melchiorri, M. Beetz, D. Pangercic, R. Siegwart, S. Leutenegger, R. Carloni, S. Stramigioli, H. Bruyninckx, P. Doherty, A. Kleiner, V. Lippiello, A. Finzi, B. Siciliano, A. Sala, and N. Tomatis, "The sherpa project: Smart collaboration between humans and ground-aerial robots for improving rescuing activities in alpine environments," in *Safety, Security, and Rescue Robotics (SSRR), 2012 IEEE International Symposium on*, Nov 2012, pp. 1–4.

[2] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug 2005, pp. 708–713.

[3] J. Cacace, A. Finzi, V. Lippiello, G. Loianno, and D. Sanzone, "Aerial service vehicles for industrial inspection: task decomposition and plan execution," *Applied Intelligence*, vol. 42, no. 1, pp. 49–62, 2015. [Online]. Available: http://dx.doi.org/10.1007/s10489-014-0542-0

[4] V. Lippiello and B. Siciliano, "Wall inspection control of a vtol unmanned aerial vehicle based on a stereo optical flow," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2012, pp. 4296–4302.

[5] S. Waharte and N. Trigoni, "Supporting search and rescue operations with uavs," in *Emerging Security Technologies (EST), 2010 International Conference on*, Sept 2010, pp. 142–147.

[6] M. A. Goodrich, J. L. Cooper, J. A. Adams, C. Humphrey, R. Zeeman, and B. G. Buss, "Using a mini-uav to support wilderness search and rescue: Practices for human-robot teaming," in *2007 IEEE International Workshop on Safety, Security and Rescue Robotics*, Sept 2007, pp. 1–6.

[7] G. Bevacqua, J. Cacace, A. Finzi, and V. Lippiello, "Mixed-initiative planning and execution for multiple drones in search and rescue missions," 2015.

[8] E. Bitton and K. Goldberg, "Hydra: A framework and algorithms for mixed-initiative uav-assisted search and rescue," in *2008 IEEE International Conference on Automation Science and Engineering*, Aug 2008, pp. 61–66.

[9] M. Lichtenstern, M. Angermann, M. Frassl, G. Berthold, B. J. Julian, and D. Rus, "Pose and paste. an intuitive interface for remote navigation of a multi-robot system," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 1632–1639.

[10] J. McLurkin, J. Smith, J. Frankel, D. Sotkowitz, D. Blau, and B. Schmidt, "Speaking swarmish: Human-robot interface design for large swarms of autonomous mobile robots." in *AAAI Spring Symposium: To Boldly Go Where No Human-Robot Team Has Gone Before*. AAAI, 2006, pp. 72–75.

[11] A. Couture-Beil, R. T. Vaughan, and G. Mori, "Selecting and commanding individual robots in a multi-robot system," in *Computer and Robot Vision (CRV), 2010 Canadian Conference on*, May 2010, pp. 159–166.

[12] V. M. Monajjemi, J. Wawerla, R. Vaughan, and G. Mori, "Hri in the sky: Creating and commanding teams of uavs with a vision-mediated gestural interface," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 617–623.

[13] G. A. Korsah, A. Stentz, and M. B. Dias, “A comprehensive taxonomy for multi-robot task allocation,” vol. 32, no. 12, pp. 1495–1512, 2013.

[14] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” *Machine Learning*, vol. 29, no. 2, pp. 131–163, 1997.

[15] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[16] O. C. Schrempf, U. D. Hanebeck, A. J. Schmid, and H. Worn, “A novel approach to proactive human-robot cooperation,” in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, Aug 2005, pp. 555–560.

[17] D. W. Albrecht, I. Zukerman, and A. E. Nicholson, “Bayesian models for keyhole plan recognition in an adventure game,” *User Modeling and User-Adapted Interaction*, vol. 8, no. 1, pp. 5–47, 1998.

[18] J. Cacace, A. Finzi, and V. Lippiello, “Multimodal interaction with co-located drones for search and rescue,” in *CEUR Workshop Proceedings*, vol. 1544, 2015, pp. 54–67.