Pose and Paste - An Intuitive Interface for Remote Navigation of a Multi-robot System

Michael Lichtenstern[1], Michael Angermann[1], Martin Frassl[1], Gunther Berthold[1],
Brian J. Julian[2,3] and Daniela Rus[2]

***Abstract*— We present *Pose and Paste* (*P&P*) - an intuitive interface designed to facilitate interaction between a single user and a number of robots equipped with cameras. With this interface, a user wearing a head-mounted display is able to cycle through the real-time video streams originating from the robots' cameras. The user is also able to select a robot and remotely position it by simply walking or turning his/her head, i.e., control the robot's motion in a master/slave-type fashion. We report the results of an initial hardware experiment where a user located in the USA is tasked to position two quadrotor robots within a motion capture laboratory located in Germany. These results suggest that *P&P* is a feasible approach to remotely inspect disaster affected sites. Lastly, we conduct a user study to compare *P&P* with a *baseline* interface composed of a traditional computer monitor and a video game controller. The quantitative results and qualitative discussions resulting from this user study highlight how such multi-robot interfaces can be further improved.**

I. Introduction

We consider the task of positioning a multi-robot system, for example to find persons in a collapsed building, inspect an underwater oil well, or perform exploration of lunar environments. Our goal is to enable a single user to easily interact with all of the system's robots regardless of his/her previous experience operating such platforms. To accomplish this goal, we propose an intuitive interface that allows the user to focus on one robot at a time, receive a first-person perspective from the robot's camera, and position the robot via simple body movements, such as the turning of his/her head. The result is an interaction that better supports novice operators, provides sufficient capabilities to more experienced ones, and is intuitive to use.

The user of the *Pose and Paste* (*P&P*) interface is able to cycle through the visual perspective of each robot, acquire control of a robot, and map the user's body movements into robot motion. For example, walking forwards while a robot's motion is activated results in an equivalent forward motion with respect to its first-person (camera) perspective, which is displayed to the user using a head-mounted display (Fig. 1).

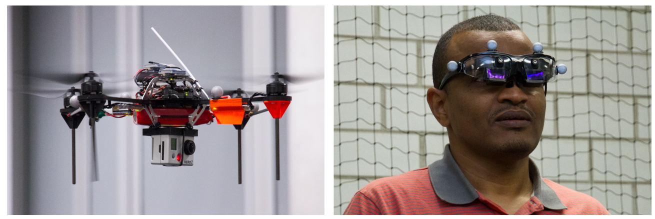


Fig. 1. Aerial robot with camera and head-mounted display equipped user.

In order to maintain situational awareness and expert control of a fielded multi-agent system, a single user needs to constantly switch his/her attention between the robots [1]. In an attempt to increase overall control performance, much recent work has been focused on increasing robot autonomy to reduce the demand on human attention [2]. However, fully autonomous systems (with no human-in-the-loop) may not be the answer for all tasks, especially when such autonomy is infeasible with today's technology. Particularly in the unpredictable and dynamic environments of disaster response, humans possess unmatched characteristics to perceive a situation, react to unexpected events, and make decisions based on context.

A. Application Domain

One of the earliest examples of the usage of teleoperated robots for disaster management was in response to the attack on the World Trade Center on 11 September 2001, when unmanned ground vehicles (UGVs) were used to search for victims and examine inaccessible voids in the building rubble [3]. More recently, unmanned underwater vehicles (UUVs) were used to track oil plumes from the blowout at Deepwater Horizon on 20 April 2010 [4], and unmanned micro aerial vehicles (MAVs) were used to survey building infrastructure after the Cyprus ammunition explosion on 11 July 2011 [5]. During MAV assessments on Cyprus, semi-autonomous robots providing real-time video streams were operated by three authors of this paper acting as European Commission technical experts for low altitude aerial reconnaissance. These first responders noted that current joystick-based control of hovering MAVs in combination with onboard camera feedback displayed on the remote control is lacking intuitive and accurate control, in addition to insufficient situation awareness. Another result out of this

This work is partially funded by the Helmholtz Foundation and the Project SOCIETIES, co-funded by the European Commission within FP7. This work is also sponsored by the Department of the Air Force under Air Force contract number FA8721-05-C-0002. The opinions, interpretations, recommendations, and conclusions are those of the authors and are not necessarily endorsed by the United States Government.

We are thankful to Iris Wieser, Andre Ryll, Thomas Seidel and Tobias Meyer for their help in the experiments, all participants of our user study as well as the anonymous reviewers of this paper for their valuable suggestions.

[1]M. Lichtenstern, M. Angermann, M. Frassl and G. Berthold are with the Institute of Communications and Navigation of the German Aerospace Center (DLR), 82234 Wessling, Germany, m.lichtenstern@dlr.de, michael.angermann@dlr.de and martin.frassl@dlr.de

[2]B. J. Julian and D. Rus are with the Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA 02139, USA, bjulian@mit.edu and rus@csail.mit.edu.

[3]B. J. Julian is also with MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420, USA.

deployment is that humans are more capable than current robots at responding to anomalies and unexpected events, which is consistent with the discussions of Rodriguez et al. [6]. Furthermore, since the mapping between operator's mental model and the robot's behavior (i.e., understanding the dynamics of the human-robot interface) is a critical factor for system performance [7], we wish to improve this mapping through more intuitive interfaces.

B. Problem Statement

From our perspective, teleoperated multi-robot systems are lacking intuitive control in 3D space. Thus, we strive to empower a single user to control many robots through the intelligent kinematic coupling of natural human movements with clutching in a shared control architecture [8]. This will enable a user to interpret robot navigation as the "pasting" of his/her own pose onto the robots. We hypothesize that coupling visual and proprioceptive inputs to robot motion allows a user to draw from his/her own experience on how to plan and navigate in a physical space. We believe this will result in more intuitive human-robot interactions compared to conventional control interfaces. In our approach, we consider holonomic robots that autonomously maintain their pose by lower-level control loops, whereas the user is responsible for higher-level path generation, including obstacle avoidance.

II. Related Work

Our inspiration for mapping the natural motion of the user to the controlled movements of robots comes from numerous studies concerning immersive virtual environments. The early experiments of Usoh et al. investigated the impact of normal walking versus walking-in-place for exploring virtual environments [9]. They concluded that normal walking is preferred by the user, which is consistent with Bowman et al. suggesting that facilitating physical movement for 3D tasks leads to higher subjective sense of user presence [10]. These findings have since influenced much work in first-person augmented reality applications, e.g., [11].

In addition, we are interested in research that investigates human interaction with multiple robots. Many works focused on minimizing human intervention [12], developing collaborative control models [13], or using select and command actions [14]. To compare the interaction effort across these and other varying approaches, Olsen and Wood formalized the concept of neglect tolerances for multi-robot systems [15]. In earlier work, we investigated the possibilities of controlling a multi-agent system of aerial robots in a 3D space [16] without the user carrying any input device [17].

Stramigioli et al., Lee et al. and Franchi et al. studied haptic navigation of multiple aerial agents considering time delays [18], [19], [20]. These haptic interactions utilizing stationary input devices do not require lower body movement of the user, and therefore are usually carried out in a sitting position. In contrast, our approach builds on the human ability to localize and position oneself within a 3D environment by explicitly exploiting proprioception of lower body movements. Recent research by Herdocia et al. [21] introduces an approach for kinematic coupling of poses. Compare to our proposed system a haptic force feedback device is used to state the user's pose. Moreover the controlled robot is a one-arm mobile manipulator mounted on a nonholonomic mobile platform.

We recognize the relevant research in deploying multi-robot systems for disaster management. Nourbakhsh et al. proposed an agent-based architecture for human-robot teams and a methodology for mixing real-world and simulation-based testing for search and rescue applications [22]. Our case and user studies are similar to their development model, and our focus is related to that of Murphy et al., who proposed crew roles and operational protocols for deploying aerial robots in disaster management [23]. Even though we wish to enable a single user to operate multiple robots, we acknowledge that the increase in the number of robots will most likely cause a need for additional humans due to other mission-specific concerns, e.g., human safety. To this effect, much work has focused on the scalability of multi-robot deployments, most notably the findings of Wang et al. concerning potentially diminishing returns for large multi-robot systems [24].

III. Technical Approach

Consider a number of robots deployed in a configuration space $\mathcal{P}$ and a single user moving in a different space $\mathcal{X}$. Our technical approach is to map the motion of the user (i.e., master) to the motion of one selected robot (i.e., slave). Each robot, denoted i, knows its current configuration $p_i \in \mathcal{P}$ by some means of measurement (e.g., GPS, visual localization). By pressing a button on an input device, the user has the ability to incrementally cycle through the visual perspectives $\mathbb{P}_i$ of all robots on a provided head-mounted display (HMD), where $\mathbb{P}_i$ is a function of p_i. With a second button, the motion state $m_i \in \{0, 1\}$ of the robot currently providing its visual perspective is either enabled ($m_i = 1$) or disabled ($m_i = 0$).

We assume the capability of accurately tracking the user's configuration $x \in \mathcal{X}$, whether via proprioceptive (e.g., inertial measurement units), exteroceptive (e.g., rangefinders), or external (e.g., GPS, motion capture system) sensing. For the selected robot i, its velocity $\dot{p}_i \in \dot{\mathcal{P}}$ is a predefined mapping from the user's velocity $\dot{x} \in \dot{\mathcal{X}}$ when the motion state is enabled, otherwise, $\dot{p}_i$ is set to the zero vector $\mathbf{0}$. More formally, we have that

$$\dot{p}_i = \begin{cases} f_i(\dot{x}), & m_i = 1, \\ \mathbf{0}, & m_i = 0, \end{cases}$$

where $f_i : \dot{\mathcal{X}} \to \dot{\mathcal{P}}$. For all non-selected robots $j \neq i$, the motion $\dot{p}_j$ is set to the zero vector.

For our studies in Section IV, we map the motion of a user wearing a HMD to the motion of multiple flying robots. Within our laboratory, a motion capture system tracks the configuration of the HMD. The velocities $\dot{x}$ needed for our mapping f_i are approximated in real-time using the six degree-of-freedom poses in $\mathcal{X} \subset \mathbb{R}^3 \times \mathbf{SO}(3)$, where $\mathbb{R}^3$ and $\mathbf{SO}(3)$ represent the three-dimensional Euclidean space and three-dimensional Euler rotation group, respectively.

Since the robots' configuration space is also a subset of $\mathbb{R}^3 \times \mathbf{SO}(3)$, one may initially want to directly map the user's velocities such that $\dot{p}_i = \dot{x}$ when $m_i = 1$. However, this approach is not feasible since standard quadrotor aerial robots are underactuated and can only be controlled in four degrees-of-freedom. In addition, we are interested in mapping motion with respect to the first person perspective $\mathbb{P}_i$, especially since the real-time video originates from cameras rigidly mounted to the robots' frames. Hence, relative forward motion of the user should result in relative forward motion of a motion enabled robot with respect to $\mathbb{P}_i$.

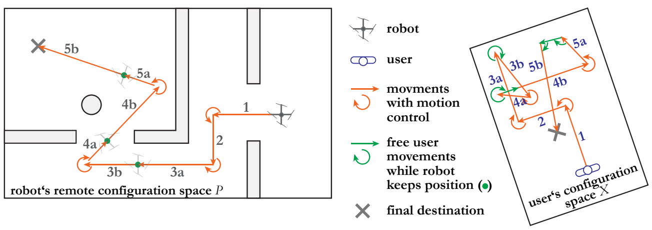


Fig. 2. Illustration of 2D trajectories of a user controlling one robot. Orange arrows indicate movements of the user in the motion control mode repositioning the robot in its remote configuration space P, whereas green arrows denote realignments within the limited configuration space X of the user. At these points in time the robot maintains its pose (green dots).

In our implementation we achieve this agreement between the user and motion enabled robot by the following mapping f_i for all robots. Firstly, the selected robot's *body-fixed* translational velocities in Euclidean space are set to the user's *body-fixed* translational velocities multiplied by constant scaling factors. Note that with respect to all configuration spaces, f_i includes trigonometric functions dependent on the angular configuration of both the robot and the user. For the rotation of the robot, the angular velocity about the vertical direction (i.e., yaw) is set to the corresponding velocity of the user. As a result, if the user's head rotates 90° clockwise over a period of time, a motion enabled robot will do the same, e.g., the rotation from segment 2 to segment 3a in Fig. 2. All other angular velocities of the robot are controlled to yield the desired translational motion of the underactuated platform. Hence, pitching and rolling of the user's head do not influence the robot's motion.

IV. Case Study

For the initial case study, we designed and employed a real hardware system. *Pose and Paste* (*P&P*) is constructed as a master-slave telerobotic system that can accommodate random communication delays over the internet, similar to the single-operator-multi-robot teleoperation system introduced by Jia et al. [25].

Using the Robotic Operating System (ROS) [26], we implemented two independent subsystems: the multi-robot control and the user interface (see Fig. 3). The robots and the head-mounted display (HMD) are equipped with infrared-reflecting markers, enabling a motion capture system to measure the positions and attitudes of these objects at a rate of 100 Hz. For the reported experiment, we used two such motion capture systems: one at DLR in Germany for the aerial robots and the other at MIT in the USA for the user.

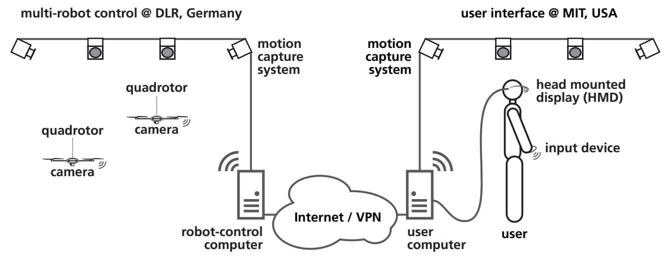


Fig. 3. System overview of the two motion capture systems: one at DLR in Germany with two camera equipped quadrotors and a robot-control computer, and the other at MIT in the USA with an user wearing the HMD and an input device connected to the user computer.

The robotic platforms used in the case study are quadrotor-type hovering MAVs [27] with controllers described in [28]. Each robot is equipped with a bottom mounted front facing GoPro Hero 2 camera (see Fig. 1). We set the onboard cameras to a medium wide angle of 127° for the experiments. We found this to be a good compromise between maximizing the field of view (FOV) and minimizing distortion created by the wide-angle lenses of the cameras (see also Pazuchanics [29]). The camera video stream is transmitted via an analog wireless link to a control computer. This computer forwards the stream over the internet using a virtual private network (VPN) to the user computer that displays the robot's first-person perspective.

The user is given a multi-button wireless hand-held device to (i) switch between the robots and (ii) enable/disable motion of the selected robot. A tethered Vuzix STAR 1200 HMD with a resolution of 1280×720 pixels (see Fig. 1) provides the user with a see-through display that covers a diagonal 23° FOV. The display is used as a feedback channel to show the first-person perspective $\mathbb{P}_i$ and the motion state m_i of the selected robot i. The input device is connected using IEEE 802.15.1 (Bluetooth) and pose commands are transmitted to the robot via IEEE 802.15.4 (ZigBee).

A. Tasks

With this hardware setup, we expect that a user is able to remotely position flying robots to clearly identify the status of three different panels (see Fig. 8) fixed to colored boxes and placed within a motion capture system's tracking volume. In an initial experiment, the user was tasked to fly to and view each of theses panels using a single flying robot. In a followup experiment, the user was task to simultaneously view two of these panels (which faced away from one another) using two flying robots. The tasks were considered completed when the user correctly identified the numeric values of these panels. Fig. 4 shows the tracking areas at DLR in Germany and at MIT in the USA, and the first-person perspectives of the cameras mounted on two MAVs during an experimental run for the second task.

B. Results and Discussion

Four different individuals used the *P&P* interface during the development for extended periods of time. All these users had no prior training on the *P&P* interface, and some of

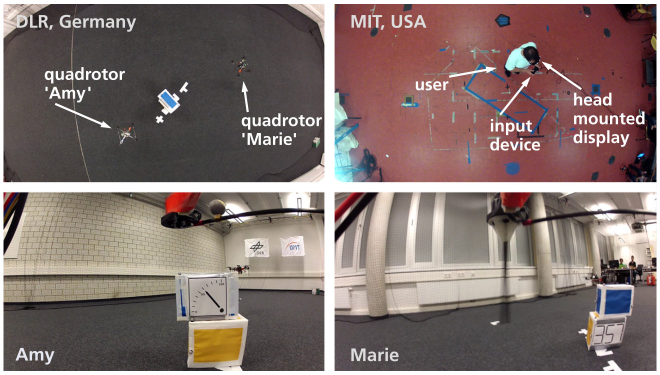


Fig. 4. The upper left image shows the two robots "Amy" and "Marie" (with not visible bottom mounted cameras (see Fig. 3)) flying at DLR in Germany, operated by a user in the upper right image at MIT in the USA. The lower images display the first-person perspective of the two robots.

them had no experience in manually piloting a quadrotor robot. For two one-robot and two two-robot experiment runs, we recorded the real-time streaming videos from the robots' onboard cameras as well as third-person video perspectives from within both motion capture laboratories. We also logged all robot positions and communication data. With the addition of five preliminary experiment runs, a total of nine runs by four individuals were carried out using the actual aerial robot platforms and interface hardware.

Total latency, mostly originating from video compression and network latency, added up to approximately two seconds. Despite communications over the internet, there was no significant variance or burst effects on the delay. Prewett et al. [30] state that practice and system learning can overcome constant latency issues, a property we also observed during the hardware experiments.

The horizontal 127° FOV of each robot's forward facing camera gave the operator a good overview of the forward space but also resulted in an acceptable amount of distortion. Increasing the FOV to larger angles was reported to impair the user's depth perception. We note that this observation is in contrast to Arthur [31], who reported that increasing the FOV results in no significant effect with respect to distance estimation. Although of great interest, we have not made any attempts to further investigate this issue.

In addition, users frequently requested supplementary situation information concerning the distances to obstacles outside the selected robot's FOV. This issue can be addressed with additional sensors, for example, by equipping the robot with laser rangefinders or ultrasonic sensors to presenting depth information in combination with the camera imagery [32]. Instead of a see-through HMD, one may also use a fully enclosed optical HMD. Nevertheless, we prefer the see-through HMD for *P&P*, which is known to avoid motion sickness of users. See [33] for a detailed discussion on the synchronization of visual and proprioceptive information.

The case study has shown that our proposed concept to sequentially control multiple robots by tracking the natural user's motion and showing the robots first-person view on a HMD is a feasible interaction method in 3D space.

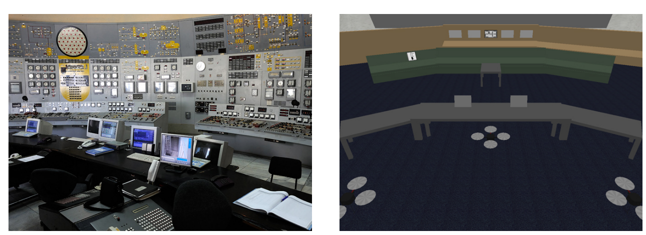
V. User Study

After showing technical feasibility of *Pose and Paste* (*P&P*), we now ask (i) if users would adopt such an interface, and (ii) how it performs in comparison with a conventional state-of-the-art interface. In order to relate performance of the *P&P* interface to existing interfaces, we implemented for the latter the combination of standard computer monitors and joystick-based input modality (e.g., remote control, gamepad controller) for controlling the robots using a first-person perspective. In the following, we denote this conventional interface combination as the *baseline* interface.

For the comparison of the two interfaces, the input provided by each interface needs to be considered. Since the employed aerial robots accept velocity commands as inputs, we configured the *baseline* interface to generate roll, pitch, yaw, and thrust commands. In contrast, *P&P* generates target poses, which are converted into these velocity commands using lower-level control loops. We note that this additional conversion in *P&P* should be kept in mind when comparing the two interfaces.

For the multi-robot system component of the user study, we used the open-source 3D multi-robot environment simulator Gazebo [34] in combination with simulated quadrotors developed by Meyer et al. [35]. This enabled us to create a simulation environment consisting of three robots, their corresponding sensors (onboard cameras), and a shared robot configuration space. The resulting video imagery is displayed in real-time on the HMD hardware used for the study (see Section IV). The dynamics of the simulated robots closely match those of the actual hardware platforms, a claim that was qualitatively verified by several experienced pilots and quantitatively in [35].

The modeled virtual environment loosely resembles a typical control room of a power plant, such as the control room of the Kozloduy nuclear power plant shown in Fig. 5(a). The virtual control room has a surrounding console (brown), a main console (green) and three supervisor tables (grey) with two display screens (see Fig. 5(b)).



(a) Control room of the Kozloduy nuclear power plant. (Yovko Lambrev / Wikipedia) (b) Three simulated aerial robots in virtual control room displayed in Gazebo.

Fig. 5. Users were tasked to navigate aerial robots in a virtual control room that was modeled to resemble a typical control room of a power plant.

Collisions of the robots between each other or with environment objects are modeled and detected with the Gazebo environment. Minor collisions are logged during the study but do not result in any adversarial consequence. In the case of a severe crash, the involved robots are restarted at their corresponding initial positions. The additional time resulting from this restart indirectly penalizes the user's overall performance. Since the case study showed noticeable delays, in particular of video feeds, we replicated this property and implemented a constant delay of 2 seconds in the simulation.

The *baseline* interface employs conventional stick-based radio remote control input and screen-based output. Each of the two sticks on a remote control allows to control two degrees of freedom. Typically, this is yaw and thrust on the left stick and roll and pitch on the right stick. For the user study we integrated a Sony Playstation 3 (PS3) wireless gamepad device (see Fig. 6) for this setup and used a standard 23" computer screen (see Fig. 7(a)) to display the simulated video feeds and status information to the user.

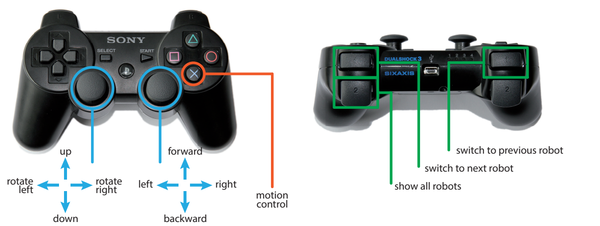
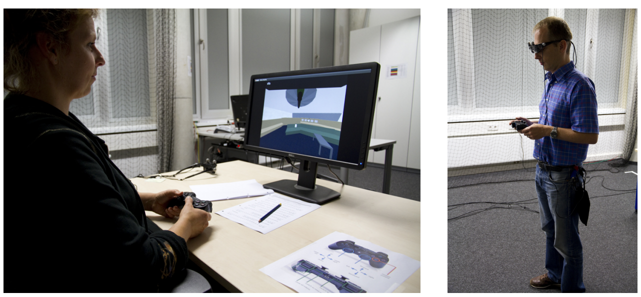


Fig. 6. Wireless input device (PS3 controller) with button assignments.



(a) User performing the task by using the *baseline* interface (joystick-based gamepad input and monitor).

(b) User controlling the system with *P&P*.

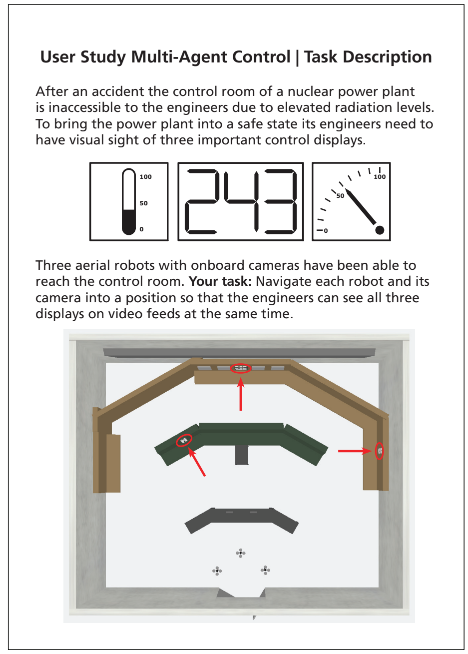
Fig. 7. Users completing the experimental task. After each experiment run the interface method is alternated between the *baseline* and *P&P* interface.

For both interfaces (*P&P* and *baseline*), the upper buttons of the wireless gamepad are used to choose between robots' perspectives and an overview of all camera outputs at the same time (see Fig. 6 green rectangles). In the *baseline* experiment, the two sticks of the pad are used to steer the aerial robot (see Fig. 6 blue circles), whereas in the *P&P* mode these are deactivated and only the X-button (see Fig. 6 red circle) is used to enable the motion state of the robot.

A. *Participant Introduction and Task*

Each trial started with a preliminary introduction to the participant given by the experiment supervisor. Each participant began the evaluation session with a familiarization to the NASA Task Load Index (TLX) questionnaire, as suggested by Hart in [36]. Then both interfaces and how the aerial robots would react to them were explained. No participant had been exposed to the simulation or the case study before. In particular none of the participants had used an HMD before or operated a robot using position tracking. After explaining the interfaces, the following task description was handed out (see Fig. 8).

Participants were made aware of the two second delay on the camera feedback for both input methods. Tasks were



User Study Multi-Agent Control | Task Description

After an accident the control room of a nuclear power plant is inaccessible to the engineers due to elevated radiation levels. To bring the power plant into a safe state its engineers need to have visual sight of three important control displays.

Three aerial robots with onboard cameras have been able to reach the control room. **Your task:** Navigate each robot and its camera into a position so that the engineers can see all three displays on video feeds at the same time.

Fig. 8. Task description handed out to all participants.

	never (--)	once a year (-)	once a month	once a week (+)	daily (++)
How often do you play games consoles using a control pad, such as PS3 or WII remote controllers?	o	o	o	o	o
How often do you play 3D first-person shooters?	o	o	o	o	o
How often do you play flight simulators?	o	o	o	o	o
How often do you fly remote controlled quadrotors / helicopters / airplanes?	o	o	o	o	o

Fig. 9. Requested prior experiences of the user in controlling aerial robots.

not subjected to any time constraints, but users were told that it is better to fulfill the task in shorter time. At the beginning each participant filled out an initial questionnaire giving name, date, age and gender. To identify the users' prior experiences four questions were asked (see Fig. 9).

If participants still had questions, further explanation of the scenario and task was given. Each user was informed that he or she had to interact during the task with a study observer who evaluates the robots' view of each panel based on a caliber-like transparent mask that is overlayed on the control screen to objectify the observer's decisions (see Fig. 10).

Fig. 10. The left image shows robots' views after the user finished a run. The right image shows the study observer giving feedback to the user by checking the view dimensions on the screen with a transparent mask.

Users had to perform the task 6 times in total. After each run, the interaction mode was alternated between the *baseline* interface (see Fig. 7(a)) and the *P&P* interface (see Fig. 7(b)). In addition, the starting interface for the first run of each user was also alternated with every new user.

Upon each task completion a NASA TLX [36] questionnaire was filled out and the user switched to the next interface and the simulation system was restarted. After completion of all tasks the users were asked which interface they would prefer. Finally users were encouraged to make comments or suggestions for improvements using free text.

B. User Statistics

A total number of 11 participants (8 male, 3 female), all students and employees at the German Aerospace Center (DLR) participated in the user study without compensation. None of the participants had been involved in any form in conception, design or implementation of the prototype or experiment. Mean age of participants at the time of the study was 36 years (ranging from 26 years to 46 years). Educational background was academic in the engineering disciplines. All 11 participants were able to complete the given task 6 times, resulting in a total number of 66 runs. The average session duration including time for briefing, questions, filling of questionnaires and actual flight time per participant was 52 minutes (ranging from 42 minutes to 76 minutes). About 55% of the participants had never used a gaming control pad and 36% use such a device once a year. Also 64% of all participants never remote-controlled a real quadrotor, helicopter or airplane and only 18% do so once a year. Of all participants, 27% stated that they never play 3D first-person shooters.

C. Quantitative Results

During all 66 runs empirical data on performance and task load was collected. Our performance indicators are time to completion (TTC) and number of collisions (NoC). For both interfaces the time the user needed to position all three robots to see all three panels simultaneously was measured. Fig. 11 shows the statistics (medians, first and third quartiles) of the TTCs of all users for each run and input method. As expected, it reveals that the TTC decrease overtime for both interfaces, presumably due to learning effects, and that there is little difference between the two interfaces. One user was able to complete the task exceptionally well (83 seconds for run # 2) using the novel *P&P* interface.

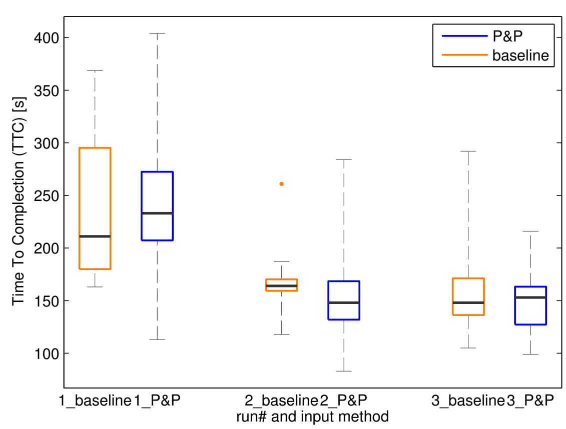


Fig. 11. Time to completion (TTC) of individual runs. Almost no significant difference between *P&P* (orange) and *baseline* interface (blue) is visible.

Fig. 12 shows that both interfaces enable users to navigate with only very few collisions. Note that – assuming standard blade protection of the aerial robots – such collisions are not considered to be fatal. Nevertheless, more collisions occurred using the *P&P* interface. From our observations during the trials we believe that users tend to explore the *P&P* capabilities and limits once they feel familiar with it.

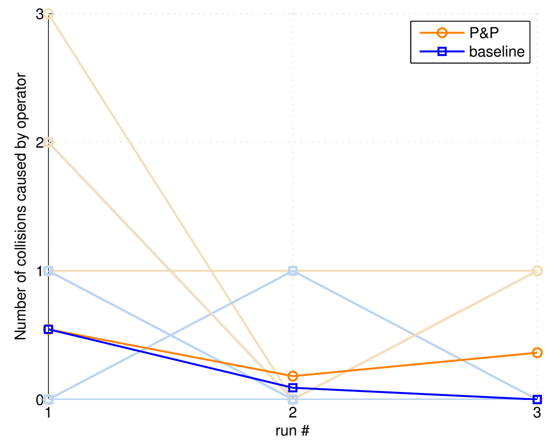


Fig. 12. Number of collisions (NoC) users caused. All users were able to navigate the robots with only few collisions for both interfaces. Note that thick lines show means not medians.

We apply NASA Task Load Index (TLX) to obtain user subjective indications [36]. Overall TLX values in Fig. 13 show that users initially stated a higher task load for *P&P*. In subsequent runs the TLX decreased and reached similar values as the *baseline* interface.

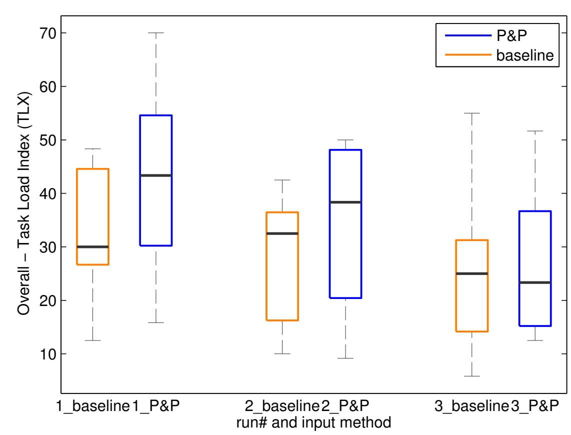
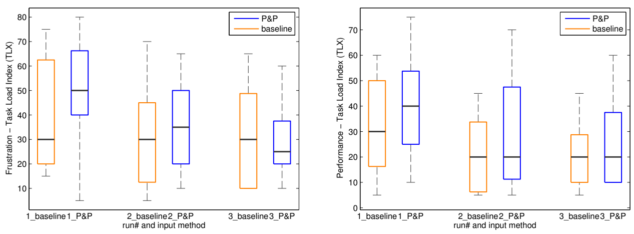


Fig. 13. The overall Task Load Index (TLX) reveals that in the beginning users were more challenged with the *P&P* interface, whereas it converges to similar values over the course of three runs.

Fig. 14(a) reveals a significant contribution of the frustration indicator to the higher overall TLX of *P&P* in the beginning. Frustration decreases over the course of the runs, whereas for the *baseline* interface it stays almost constant.
A surprising result is the users' estimation of their own performance. We expected a high rating for the *baseline* interface, whereas Fig. 14(b) shows that only in run # 1 the users rated their performance higher for *P&P*. It also shows that after the following runs the users' performance estimation is almost equal for both interfaces.



(a) Frustration (b) Performance

Fig. 14. Two interesting components of the TLX. The frustration indicator shows the convergence of the interfaces, whereas the performance estimation of the users apparently is constant after three runs.

Fig. 15 relates the final preference stated by each user after the third run to the actual performance using both interfaces in terms of average time to completion over all three runs for *P&P* (horizontal) and *baseline* (vertical). A dot in the upper left triangle shows that on average the user was quicker using *P&P*, dots in the lower right triangle indicate that these users were quicker with the *baseline* interface. It can be seen that users do not necessarily choose the interface they performed best with. It can also be seen that the performance is strongly user specific, i.e. users who were quick with one interface were also more likely to be quick with the other.

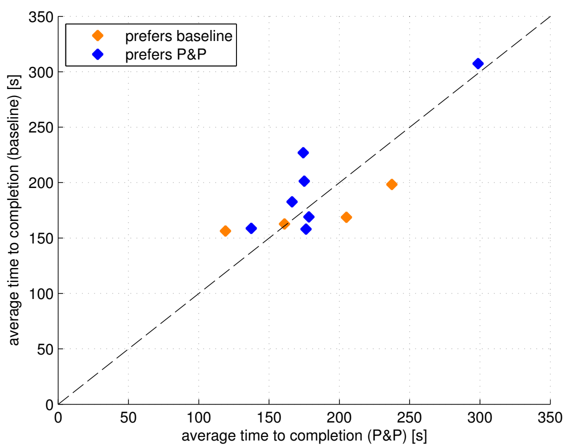


Fig. 15. Each dot in this plot represents one of 11 participants. Color represents the users final preference for the *P&P* (blue) or the *baseline* (orange) interface.

D. User Comments

Most participants used the final questionnaire for remarks that helped us to better understand and interpret their performance and reported task load. Comments are mostly translated from German.

A number of comments dealt with inconveniences of the specific hardware implementation of the HMD:

Comment	User
The cable distracts.	P3
Image small / contrast of image.	P4
The lamp was distracting using HMD.	P7
Glasses are uncomfortable and the distances in the virtual image are not the same as they are in reality. That's the reason I feel walking doesn't really make sense.	P8

Despite these inconveniences, a number of users apparently have perceived good controllability with *P&P*:

The control with HMD feels more natural.	P5
... it [the HMD] was considerably easier [to control] as the camera movement is oriented on body movements. Overall it takes a little getting used to, but then it is definitely the better solution.	P9
Unfortunately the HMD have been very heavy and uncomfortable, the movements in contrast are way more natural and as a result controlling is more convenient.	P10

Comments were less conclusive with respect to the suitability of the interfaces for dealing with the purposely introduced delay:

Delay especially with HMD a problem	P7
The delay is very annoying – although you don't notice the delay that much using the HMD.	P10

While 63% of the users stated an overall preference for *P&P*, comments showed that preferences were not particularly pronounced:

Initially I preferred using the Joystick, but you are getting used to handle the HMD very fast. I liked both devices about the same in the end.	P1
I have the feeling that I can control better with the HMD, but would rather trust the joystick.	P3

VI. Conclusions

We have conceived and investigated a novel interface for controlling the pose of robots in a multi-agent system. Experiments with real aerial robots have shown the technical feasibility of this new approach. In these experiments operators at MIT in the USA successfully controlled two aerial robots flying in a lab at DLR in Germany. We then combined a simulation of aerial robots in a synthetic environment with the same interface hardware and carried out a user study with 11 participants. We have measured performance metrics and NASA Task Load Index for *P&P* and a *baseline* interface. Our empirical results indicate that the performance of both interfaces in terms of task completion time is almost identical. Users stated a higher task load for the *P&P* interface in their first runs. In subsequent runs the task load of *P&P* declined more quickly than the *baseline* interface, reaching about the same task load for the third run.

We believe that these results are encouraging, since almost all users had significant prior exposure to the hand-held remote control device but nevertheless performed very similar or even slightly better with *P&P* in terms of task completion time. In addition several users stated the controlling from a first-person perspective with *P&P* feels more natural. We therefore think that they would adopt the interface for such tasks, which supports our hypothesis.

We intend to build upon these encouraging results in numerous ways. In a first step we want to remove the tether from the head-mounted display which will remove a considerable source of distraction from users. A natural extension will be to equip the aerial robots with a pan/tilt unit and couple its movement to the user's head motion. This will introduce a fifth degree of freedom allowing the user to "look" up or down. In addition we want to apply this approach to other robot types, such as underwater robots [37] with fully actuated six degrees of freedom.

A further task will be to improve the efficiency of the interface. Since *P&P* frees the user's hands we want to investigate how we can make use of this fact and investigate if the ability to tune scaling factors, possibly by giving them the freedom to dynamically adjust the scaling factors along different axes is beneficial. This way users would be able to adjust between a "microscopic" mode for close inspection of small objects and "seven-league boots" for covering distances between multiple points of interest.

REFERENCES

[1] P. Velagapudi, P. Scerri, K. Sycara, H. Wang, M. Lewis, and J. Wang, "Scaling effects in multi-robot control," in *Intelligent Robots and Systems, 2008. IEEE/RSJ International Conference on*, Sept. 2008.

[2] J. Crandall, M. Goodrich, D. Olsen, and C. Nielsen, "Validating human-robot interaction schemes in multitasking environments," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 35, no. 4, pp. 438 – 449, Jul. 2005.

[3] J. Casper and R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 33, no. 3, pp. 367 – 385, Jun. 2003.

[4] R. Camilli, C. M. Reddy, D. R. Yoerger, B. A. S. V. Mooy, M. V. Jakuba, J. C. Kinsey, C. P. McIntyre, S. P. Sylva, and J. V. Maloney, "Tracking Hydrocarbon Plume Transport and Biodegradation at Deepwater Horizon," *Science*, vol. 330, pp. 201–204, 2010.

[5] M. Angermann, M. Frassl, and M. Lichtenstern, "Mission review of aerial robotic assessment - ammunition explosion Cyprus 2011," in *Safety, Security, and Rescue Robotics (SSRR), 2012 IEEE International Symposium on*, College Station, Texas, USA, 2012.

[6] G. Rodriguez and C. Weisbin, "A new method to evaluate human-robot system performance," *Autonomous Robots*, vol. 14, pp. 165–178, 2003.

[7] A. Steinfeld, "Interface lessons for fully and semi-autonomous mobile robots," in *Robotics and Automation, 2004 IEEE International Conference on*, New Orleans, LA, USA, Apr.-1 May 2004.

[8] G. Niemeyer, C. Preusche, and G. Hirzinger, "Telerobotics," in *Springer Handbook of Robotics*. Springer Berlin Heidelberg, 2008.

[9] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks, Jr., "Walking > walking-in-place > flying, in virtual environments," in *26th annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '99. ACM, 1999.

[10] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, "An introduction to 3-D user interface design," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 1, pp. 96–108, Feb. 2001.

[11] B. Thomas, B. Close, J. Donoghue, J. Squires, P. D. Bondi, and W. Piekarski, "First person indoor/outdoor augmented reality application: Arquake," *Personal Ubiquitous Comput.*, vol. 6, no. 1, 2002.

[12] T. Balch and R. Arkin, "Behavior-based formation control for multi-robot teams," *Robotics and Automation, IEEE Transactions on*, vol. 14, no. 6, pp. 926 –939, Dec. 1998.

[13] T. Fong, C. Thorpe, and C. Baur, "Multi-robot remote driving with collaborative control," *Industrial Electronics, IEEE Transactions on*, vol. 50, no. 4, pp. 699 – 704, Aug. 2003.

[14] C. Jones and M. J. Mataric´, "Sequential task execution in a minimalist distributed robotic system," in *Proceedings of the seventh international conference on simulation of adaptive behavior on From animals to animats*. Cambridge, MA, USA: MIT Press, 2002.

[15] D. R. Olsen, Jr. and S. B. Wood, "Fan-out: measuring human control of multiple robots," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2004.

[16] M. Lichtenstern, M. Angermann, and M. Frassl, "IMU- and GNSS-assisted single-user control of a MAV-swarm for multiple perspective observation of outdoor activities," in *ION ITM 2011*, San Diego, CA, USA, 2011.

[17] M. Lichtenstern, M. Frassl, B. Perun, and M. Angermann, "A prototyping environment for interaction between a human and a robotic multi-agent system," in *Proceedings of the 7th ACM/IEEE international conference on Human-Robot Interaction*, Boston, MA, USA, 2012.

[18] D. Lee, A. Franchi, P. Giordano, H. I. Son, and H. Bulthoff, "Haptic teleoperation of multiple unmanned aerial vehicles over the internet," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, Shanghai, China, May 2011.

[19] S. Stramigioli, R. Mahony, and P. Corke, "A novel approach to haptic tele-operation of aerial robot vehicles," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, Anchorage, Alaska, USA, May 2010.

[20] A. Franchi, C. Secchi, H. I. Son, H. Bulthoff, and P. Giordano, "Bilateral teleoperation of groups of mobile robots with time-varying topology," *Robotics, IEEE Transactions on*, vol. 28, no. 5, pp. 1019 –1033, Oct. 2012.

[21] A. Herdocia, A. Shademan, and M. Jagersand, "Unimodal asymmetric interface for teleoperation of mobile manipulators: A user study," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Vilamoura, Portugal, Oct. 2012.

[22] I. Nourbakhsh, K. Sycara, M. Koes, M. Yong, M. Lewis, and S. Burion, "Human-robot teaming for search and rescue," *Pervasive Computing, IEEE*, vol. 4, no. 1, pp. 72 – 79, Jan.-Mar. 2005.

[23] R. R. Murphy, K. S. Pratt, and J. L. Burke, "Crew roles and operational protocols for rotary-wing micro-UAVs in close urban environments," in *Proceedings of the 3rd ACM/IEEE international conference on Human-Robot Interaction*, Amsterdam, The Netherlands, 2008.

[24] H. Wang, M. Lewis, P. Velagapudi, P. Scerri, and K. Sycara, "How search and its subtasks scale in N robots," in *Proceedings of the 4th ACM/IEEE international conference on Human-Robot Interaction*, La Jolla, California, USA, 2009.

[25] Y. Jia, N. Xi, and J. Buether, "Design of single-operator-multi-robot teleoperation systems with random communication delay," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, San Francisco, CA, USA, Sept. 2011.

[26] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source robot operating system," in *ICRA Workshop Open Source Software*, Kobe, Japan, 2009.

[27] D. Gurdan, J. Stumpf, M. Achtelik, K.-M. Doth, G. Hirzinger, and D. Rus, "Energy-efficient autonomous four-rotor flying robot controlled at 1 kHz," in *Robotics and Automation (ICRA), 2007 IEEE International Conference on*, Roma, Italy, Apr. 2007.

[28] M. W. Achtelik, S. Weiss, S. Lynen, M. Chli, and R. Siegwart, "Vision-based MAV navigation," in *RSS Workshop on Integration of perception with control and navigation*, Sydney, Australia, 2012.

[29] S. L. Pazuchanics, "The effects of camera perspective and field of view on performance in teleoperated navigation," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 16, pp. 1528–1532, 2006.

[30] M. S. Prewett, R. C. Johnson, K. N. Saboe, L. R. Elliott, and M. D. Coovert, "Managing workload in human-robot interaction: A review of empirical studies," *Comput. Hum. Behav.*, vol. 26, no. 5, pp. 840–856, Sept. 2010.

[31] K. Arthur, "Effects of field of view on performance with head-mounted displays," Ph.D. dissertation, Department of Computer Science, University of North Carolina, Chapel Hill, 2000.

[32] B. Ricks, C. Nielsen, and M. Goodrich, "Ecological displays for robot interaction: a new perspective," in *Intelligent Robots and Systems (IROS), 2004 IEEE/RSJ International Conference on*, Sendai, Japan, Sept.-2 Oct. 2004.

[33] J. P. Rolland and H. Fuchs, "Optical versus video see-through head-mounted displays in medical visualization," *Presence: Teleoper. Virtual Environ.*, vol. 9, no. 3, pp. 287–309, Jun. 2000.

[34] N. Koenig and A. Howard, "Design and use paradigms for Gazebo, an open-source multi-robot simulator," in *Intelligent Robots and Systems (IROS), 2004 IEEE/RSJ International Conference on*, Sendai, Japan, Sept.-2 Oct. 2004.

[35] J. Meyer, A. Sendobry, S. Kohlbrecher, U. Klingauf, and O. Stryk, "Comprehensive simulation of quadrotor uavs using ROS and Gazebo," in *Simulation, Modeling, and Programming for Autonomous Robots*, Tsukuba, Japan, 2012.

[36] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, ser. Advances in Psychology, 1988, vol. 52, pp. 139 – 183.

[37] I. Vasilescu, C. Detweiler, M. Doniec, D. Gurdan, S. Sosnowski, J. Stumpf, and D. Rus, "AMOUR V: A hovering energy efficient underwater robot capable of dynamic payloads," *The International Journal of Robotics Research*, vol. 29, no. 5, 2010.