frontiers

High-Level Teleoperation System For Aerial Exploration Of Indoor Environments

W. Alexander Isop[1*], **Christoph Gebhardt**[2], **Tobias Nägeli**[2], **Friedrich Fraundorfer**[1], **Otmar Hilliges**[2], **Dieter Schmalstieg**[1,3]

[1] *Institute of Computer Graphics and Vision, Department of Computer Science, Graz University of Technology, 8010 Graz, Austria*
[2] *AIT Lab, Department of Computer Science, ETH Zürich, Zurich 8092, Switzerland*
[3] *VRVis Research Center, 1220 Vienna, Austria*

Correspondence*:
W. Alexander Isop
isop@icg.tugraz.at

Number of Words: 11483, **Number of Figures:** 11

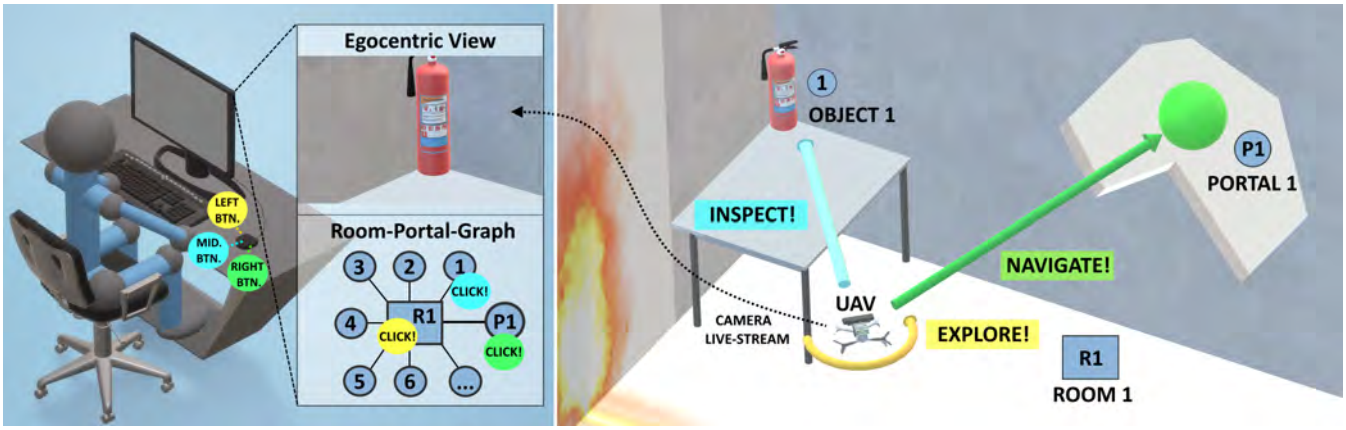

Figure 1. High-Level teleoperation system. (left) The room-portal graph displays an interactive topological view of an indoor environment, created in real-time, to facilitate automation. (right) Conceptual illustration of the aerial telerobot, implemented as unmanned aerial vehicle (UAV), and its according high-level tasks.

ABSTRACT

Exploration of challenging indoor environments is a demanding task. While automation with aerial robots seems a promising solution, fully autonomous systems still struggle with high-level cognitive tasks and intuitive decision making. To facilitate automation, we introduce a novel teleoperation system with an aerial telerobot that is capable of handling all demanding low-level tasks. Motivated by the typical structure of indoor environments, the system creates an interactive scene topology in real-time that reduces scene details and supports affordances. Thus, difficult high-level tasks can be effectively supervised by a human operator. To elaborate on the effectiveness of our system during a real-world exploration mission, we conducted a user study. Despite being limited by real-world constraints, results indicate that our system better supports operators with indoor exploration, compared to a baseline system with traditional joystick control.

Keywords: teleoperation system, telerobotics, system design, scene abstraction, unmanned aerial vehicles, indoor exploration

1 INTRODUCTION

Teleoperation of small-sized aerial robots in indoor environments is important for applications like search-and-rescue or exploration missions. A recurring problem in such applications is lack of situation awareness and consequently decreasing overall task performance (Stubbs et al., 2007; Burke and Murphy, 2004).

One important aspect is that with an increasing amount of scene details, operators struggle to comprehend the visualized information of the teleoperation system (Atherton and Goodrich, 2009). While it is required that the system presents the information in a way that does not overwhelm the operator, also the levels of autonomy (LOA) play a crucial role. Increasing autonomy of the system can improve operators task performance by reducing their mental load. The goal is to free up the operators to be engaged in other important high-level tasks (Goodrich et al., 2007), such as navigation or identification of victims or hazards. However, related work has shown that true full autonomy is still hard to accomplish for complex missions (MahmoudZadeh et al., 2018). This emphasizes difficulty of an optimal level of autonomy for a teleoperation system. As a tradeoff, approaches were introduced in which operators can explicitly adjust the autonomy of the system to the desired level (Bruemmer et al., 2002; Muszynski et al., 2012). Unfortunately, such approaches typically consider a handover to low-level demanding tasks (Reese, 2016). While trading off task automation and manual control is task-specific and remains non-trivial to date, our system on one hand suggests to automate all low-level tasks. On the other hand, high-level tasks can be accessed via an interactive scene with reduced details. Yet, the question remains how such a system effects aerial exploration missions in a real-world setting.

To investigate this question, we introduce a fully working teleoperation system. The system uses a small-sized aerial telerobot to perform the challenging task of indoor exploration. In particular, our system is capable of: i) indoor navigation in the presence of narrow spaces and wind-turbulence without collision; ii) automated repetitive exploration and navigation of structured spaces; iii) detection and navigation of constrained spaces, like narrow gateways or windows, without collision; iv) and detection of objects of interest (OOIs), like victims or fire extinguishers. To relieve the operator, the system automates all low-level mission-critical tasks (see Figure 5). However, we allow the operator to override non-mission-critical, high-level objectives. This results in a design where the system usually runs at the highest LOA (*highest automation*), but can be effectively supervised at collaborative level (*high automation*) if necessary (see Table 2)[1].

The operator supervises the teleoperation system using a multi-view GUI which consists of a traditional egocentric scene view, an exocentric scene view, and a complementary topological graph view of the scene, called the room-portal graph (RPG) (see Figure 2). The RPG represents the scene as a subdivision of rooms connected by portals, creating a clearly distinguishable spatial structure of a typical indoor environment. The RPG reduces scene details and allows fast comprehension of important scene structure and objects during an exploration mission. It is interactive and lets the operator improve time-performance and resolve system failures, for example false detection of OOIs.

To understand the task effectiveness of our teleoperation system in a real-world setting, we conducted a user study. Participants accomplished an exploration mission using our proposed system and, in comparison, using a baseline system with traditional joystick control. While results indicate increased task performance and comfort with the outcome of our experiments, our findings provide evidence that our system can better support operators with challenging aerial indoor exploration missions.

[1] The reported levels of automation (LOA) are based on the ALFUS framework of Huang et al. (2005a,b).

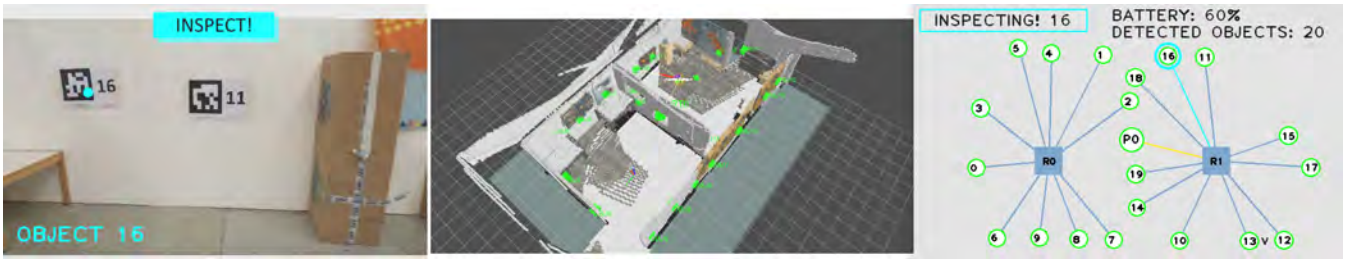

Figure 2. Implementation of the RPG as part of our high-level teleoperation system. It is presented during an inspection task, after full exploration of two connected rooms. (left) Egocentric virtual live view from the on-board camera of the UAV, highlighting an inspected object. (middle) Exocentric virtual 3D view of the reconstructed scene. (right) Interactive, topological RPG of the same scene, with two rooms (represented as squares), detected objects including a portal. Objects are shown as round labels with leader lines. Inspected objects are highlighted.

In summary, we contribute: i) a fully working teleoperation system for aerial indoor exploration missions, including a real-time interactive scene topology that enables supervisory control of high-level tasks, and ii) the empirical evidence that such a system effectively supports human operators during realistic exploration missions. In the remainder of the paper, we provide an extensive overview on related work, including a brief history of teleoperation and current state-of-the-art systems. We present details of design rationals and implementation of our system, as well as limitations. Followed by reporting on experimental design and results of our user study, we finally conclude our paper and propose interesting directions for future work.

2 RELATED WORK

The research conducted in the area of robotic teleoperation is extensive and was explored starting from the mid of the 20th century. The research is highly interdisciplinary and addresses a rich variety of aspects in human-robot-interaction, mobile robotics and visualization. The purpose of this section is to help the reader to understand the complexity of the field and to provide an overview of state-of-the-art systems. It starts with a short historical summary of teleoperation. It continues with a broad range of related work regarding telerobotics and ends with addressing specific work on high-level indoor exploration with aerial teleoperation systems. Thus, we provide an extensive literature survey and discussion about origin and history of robotic teleoperation, its motivation, typical use cases and changing demands to the system over time. Finally, it adds a discussion about limitations and potential future work to relevant sections. Ultimately, our goal is to help the reader to better understand the motivation of our system and its novelty in the presented form.

2.1 History Of Teleoperation - From The Poking Of Fire To Telerobotics in Space

Interestingly, an extensive survey by Lichiardopol (2007) suggests that the first form of teleoperation was the poking of fire in the early age of human mankind. By utilizing the stick to set fire, the human was actually teleoperating (or telemanipulating) the fire place. A million years later, in the early 1900‘s, for the first time teleoperation appeared for piloting unmanned aircrafts (Fong and Thorpe, 2001). Related work continues in the mid of the past century, but in the context of remote handling hazardous materials. According to an extensive summary of Vertut and Coiffet (1986), R. Goertz worked on a pantograph as telemanipulation device for radioactive materials inside a nuclear reactor. The obvious purpose was to enable safe handling of the otherwise dangerous materials by human operators. In the following decades,

the need for robotic teleoperation significantly increased, also highlighted as part of a comprehensive survey of Alexander (1972). In his work, Alexander (1972) declares the terms teloperation and telemanipulation under ”telemation” and investigates on which use cases in the civil sector development on teleoperation systems would have greatest impact. Amongst nuclear reactor fuel handling, mining, oceanographic and medical teleoperation systems, additionally a strong need for teleoperation was growing in aeronautics and space (Corliss and Johnson (1967), Alexander III (1973)). At this point, the original purpose of robotic teleoperation becomes apparent, which is to enable a human operator to work at a (larger) distance in hazardous environments. According to Lichiardopol (2007), this requires the following essential definitions:

- **(autonomous) Teleoperator:** The automatically operated machine that replaces human effort, though it may not resemble human beings in appearance or perform functions in a humanlike manner.
- **Operator:** A human operator is the person who monitors the operated machine and takes control of the required actions.
- **Teleoperation:** The task or activity itself to ”operate a vehicle or a system over a distance” (Fong and Thorpe, 2001).

More recent work about teleoperation (Cui et al., 2003) gives a clearer overview of the required main components, data-flow and interfaces. These are in particular:

- Remote System (**teleoperator or telerobot**): which is a machine or robot that enables a human operator to move about, sense and mechanically manipulate objects at a distance.
- Human Machine Interface (**human-robot interface**): A software framework, including a graphical user interface (GUI) software, control software and interface software to process sensory data.
- Manual Control Unit (**input device**): Including an input device for manual control inputs.

By definition, the human operator manually interacts with the teleoperation system via the input device, whereas the telerobot senses and performs actions in the far distant environment.

2.2 Telerobots - Reducing Human Effort

According to Cui et al. (2003), telerobots are a subclass of teleoperators. They are typically implemented as mobile robots that accept instructions from a human operator at a distance. Moreover, they perform live actions at a distant environment through the use of sensors or other control mechanisms. Usually they have sensors and/or effectors for manipulation and/or mobility, and moreover provide mechanisms for the human operator to communicate with both. Due to the rich variety of applications and remote environments, related work addressed various types and sizes of mobile robots during the past decades. Such are ranging from space exploration-, mining-, medical- and underwater-applications (Alexander III (1972), Sheridan and Verplank (1978)) in the early years of modern teleoperation, to humanoid telepresence robots with increasing popularity in the early 2000s (Lichiardopol, 2007) and more recent telerobots for bomb defusal (Technology, 2014). Amongst others, important types of telerobots include stationary robots like arm manipulators (Murray (2017), Rakita et al. (2018)), underwater robots (Costa et al., 2018), ground based search and rescue robots (Stepanova et al., 2017), humanoid robots for telepresence (Cortellessa et al., 2018) and aerial telerobots for surveillance (Jha, 2016) or exploration and mapping (Papachristos et al., 2017).

In particular, small sized UAVs or micro aerial vehicles (MAVs) soared popularity in the past decades (Cai et al., 2014). While their small size and low weight make them attractive for research and education (Wu et al., 2018), they are also extensively used outdoors for industrial purposes. Typical use cases involve

agriculture (Tripicchio et al., 2015), medical transportation (Scalea et al., 2019) and moreover search and rescue or exploration missions (Silvagni et al., 2019). However, such use cases put different demands on the UAV, compared to indoor flights. In outdoor environments spatial constraints play a subsidiary role and very often GPS-based localization and navigation is possible. In contrast, exploration of indoor environments requires small UAV dimensions to avoid problems with reaching narrow passages. Also, typical limitations remain like onboard computational power and flight times. Furthermore, low-thrust propulsion is important, since turbulences could strongly affect stability and overall flight performance of the UAV. State of the art off-the-shelf UAVs which could be potentially used for such purposes are the DJI Mavic or the Intel Aero platform. However, on one hand, such solutions are more bulky and heavy and are not easily customizeable (software interfaces, sensors). On the other hand, smaller and more lightweight solutions (Kushleyev et al. (2013); Loianno et al. (2016)) are more limited regarding payload and flight times. Very recent UAV designs with similar all-up-weight and dimensions (Sanket et al. (2018); Falanga et al. (2019)) also discuss navigation and exploration of portals. Though, they either provide shorter hover flight times or do not carry Kinect-like RGB-D sensors. Such are used for creating dense mapping data (Henry et al., 2012), while this can be beneficial for indoor exploration. In general, to achieve an optimal UAV design for aerial exploration and online mapping of indoor environments still remains difficult. A potential future design, also reflecting the room for improvement of our presented UAV, may include the following challenging specifications if combined in one system:

- **Small-scale dimension** with tip-to-tip diameter under $100mm$ (Giernacki et al., 2017) (approx. size of the palm of a humans hand) which would enable it to reach very narrow passages and make it easy to hold and carry.
- **Low all-up-weight** below $100g$ (Kushleyev et al., 2013) to make transportation easier and reduce produced thrust and turbulences during flight.
- **Low-weight sensors** including kinect-like or omnidirectional vision sensors.
- **Powerful computing units** (e.g. Tegra K1, Kyristsis et al. (2016)), including GPUs to execute all tasks which are important for exploration of indoor environments onboard (robust localization, exploration, navigation, motion planning, online mapping and object recognition).
- **Increased flight times** of more than $30mins$, which is typical for state of the art commercial UAVs in this size- and weight-category (Robotics, 2019), or extendable flight times by autonomous wireless recharging technology (Junaid et al. (2016), Al-Obaidi et al. (2019)).

2.3 Remote Connection - Coping With The Issues Of Time Delay

Although the remote connection between telerobot and human-robot interface is not listed as an individual component by Cui et al. (2003), it has major impact on the overall task performance. According to a survey of Sheridan (1993), time delay is a "serious problem for teleoperation". Issues with time delays were for the first time addressed in the early 1960's by (Adams, 1961), whereas later work by Ferrell (1965) found that human operators can cope with time delays by using a simple "move and wait strategy". His experiments also showed that task performance is linearly dependent and predictable on the time-delay during teleoperation. Remarkbly, Held et al. (1966) found that sensory-motor adaptation is essentially impossible for delays as small as $0.3s$, and that human operators dissociate the teleoperator movements from those of their own in the presence of such delays (Held and Durlach, 1991). Especially if direct control methods are used, this could lead to a nonoptimal task performance and Sheridan (1992) explicitely states that direct control in the presence of delay (transmission or otherwise) is tedious, fatiguing, and error prone. Since our teleoperation system includes a wireless remote connection with potential higher time-delays

($> 0.3s$), this also affected our design decisions about the control approach (Section 3.2.1). More recent related work discusses impact of time delays during teleoperation of small-sized UAVs (Riestock et al., 2017a,b), whereas they elaborate on effects of limited bandwidth on the GUI. They compare operators performance during collision avoidance tasks and use traditional egocentric live camera views and grid-map representations of the scene. Interestingly, their results indicate that the operator performance suffered less under a change of communication quality using grid-maps, compared to the egocentric live camera views. Consequently this was considered in the design of our GUI (Section 3.2.2).

2.4 Human-Robot Interfaces - Facilitating Control And Cognition

Based on the summary of Cui et al. (2003), the human-robot interface processes data of sensors and actuators of the telerobot for control, typically at different LOA. It further visualizes information about the remote system and the remote scene with a GUI. Finally, it is responsible for processing the operators manual inputs of the input device.

2.4.1 Levels of Automation And Approaches For Control

The LOA of a teleoperation system is important, since it could have great impact on the overall design of the teloperation system (Save et al. (2012), Endsley (2018)). Moreover, the LOA greatly effects the operators overall task performance during teleoperation (Materna et al., 2017). Subsequently, also a variety of taxonomies for LOA were introduced for robotic teleoperation. While the idea of LOA was introduced by(Sheridan and Verplank, 1978) in the late 1970s for underwater teleoperation applications, more recent work broadened this concept (Endsley, 1999; Parasuraman et al., 2000). Depending on the application, related work discusses various models with more or less fine grained LOA for flight traffic control (Council et al., 1998), manufacturing (Frohm et al., 2008) and, most recently, autonomous driving (SAE, 2014). Most interesting four our work is a LOA-taxonomy specifically designed for unmanned systems. Introduced by Huang et al. (2005b,a), it was successfully adapted for indoor exploration missions by Valero-Gomez et al. (2011).

On one hand, low LOA typically imply that the human operator has direct *manual control* over the telerobot and can also directly access detailed sensory data (Sheridan, 1992). However, the operator must be also able to process and interpret this data. During challenging tasks under time constraints, this could overwhelm the operator and lead to decreased task performance or even mission failure. On the other hand, so called ”fully automated systems” without any control of an operator are still hard to put into practice. At least their overall performance can be still significantly improved if they are teamed with a human (Johnson and Vera, 2019). In order, also the design of our human-robot interface is motivated by the capabilities and moreover limitations of the telerobot. The presented GUI builds on top, enabling the human to supervise with difficult high-level tasks. Automation at higher levels means that the telerobot is able to accomplish certain low-level tasks independently and could relieve the operator based on *supervisory control* (Sheridan, 1992). If the telerobot fails, on demand switching the system to lower LOA could be helpful (*adjustable or sliding autonomy*), whereas an extensive survey is presented by (Mostafa et al., 2019). Muszynski et al. (2012), Bruemmer et al. (2005), Chen et al. (2013) and Leeper et al. (2012) propose teleoperation systems with different LOA for control of ground-based robots from classical egocentric and exocentric views. These works consistently report on improved operator performance with increasing autonomy. Extensive research has been conducted concerning the concept of switching between LOA. Valero-Gomez et al. (2011) suggest two autonomy models to interact with robot teams during exploration missions and enable low-level operation on demand. Fong et al. (2003) explored semi-autonomous systems that query users for input, in case of uncertain decision. Both papers suggest that system failures should be handled manually by the

operator. However, their design focuses on ground based navigation or grasping. They also do not provide a minimum LOA to the operator, avoiding mission-critical tasks during flight missions. To cope with the issues of direct control, Gebhardt et al. (2016), Nägeli et al. (2017b) and Nägeli et al. (2017a) suggest optimized planning of constrained quadrotor paths. They also avoid passing low-level tasks to the operator and instead introduce indirect, high-level flight goals. They allow inexperienced operators to control the UAV without deeper knowledge of the underlying methods for quadrotor control or the target domain. An important prerequisite for such ease of use is that the UAV can move along a collision-free path. More recent work combining supervisory control and adjustable autonomy is presented by Lin and Goodrich (2015), Szafir et al. (2017) and Lan et al. (2017). Remarkable limitations with all supervisory control approaches are the *lumberjack effect* (Onnasch et al., 2014) and the *automation conundrum* (Endsley, 2017). These effects summarize a tradeoff between high LOA improving task performance and problems with sudden passing of low-level tasks if problems occur. Moreover, a general concept that provides an optimum LOA and level of operator control for all applications and tasks seems impossible and will remain difficult in the future. Such limitations must be considered in the design of the teleoperation system. Consequently, also our system avoids sudden passing of low-level functions to the operator and only allows for overriding functions at high LOA (supervisory control). In contrast to related work, the LOA design of our work considers challenging tasks that occur during aerial indoor exploration missions. Design details can be found in Section 3.2.

2.4.2 Graphical User Interfaces

Various types of GUIs, with different combination of scene views, have been investigated to improve task efficiency for teleoperation of small-sized UAVs. Examples range from interfaces with traditional egocentric live camera views (Cho et al., 2017), combined with direct joystick based control, to fully immersive interfaces utilizing the operators upper body for control (Rognon et al., 2018).

As an alternative to UAV navigation from egocentric views, direct commands can be issued in an adaptive exocentric perspective (Saakes et al., 2013; Thomason et al., 2017, 2019) or from a 3D map view (Materna et al., 2017). The exocentric view can improve the operator's understanding of the environment and further increase safety and task performance. Additionally, concepts of switching between ego- and exocentric views is discussed by Baudisch et al. (2002) whereas Ferland et al. (2009) suggest to switch between egocentric and exocentric views for robot teleoperation. Following the overview- and detail paradigm, their overall goal is to improve task performance by providing information details on demand. According to Gebhardt et al. (2016) and Nägeli et al. (2017a), pure exocentric planning views can be beneficial for applications such as robotic light painting, drone-racing and aerial cinematography. Also the work of Lan et al. (2017) combines exocentric scene views with a high-level GUI for photo taking with a semi-autonomous UAV. However, these applications do not generate an interactive scene topology from 3D data. They either require that a 3D map is already pre-generated or use mapping for localization and path planning only. Importantly, they do not consider challenging tasks that occur during indoor exploration missions, like flight through narrow passages. In contrast, the design of our system focuses on flight missions in challenging indoor environments. Most importantly, we generate an interactive scene topology in real-time and thus facilitate automation.

Remarkably, only providing exocentric 3D map views can lead to problems. For example, Chellali and Baizid (2011) state that on one hand the third dimension is an additional degree of freedom that helps to add constraints and information to disambiguate location of objects. On the other hand, they report on significantly decreased task performance when localizing objects in 3D maps, compared to localization in 2D. They suggest that the additional dimension provided within 3D maps leads to a greater space to

explore and thus the operator needs more time. This tradeoff was also considered in the design of our GUI, which is outlined in Section 3.2.2.

By even more reducing scene details, abstract topological views prevent the operator from being overwhelmed and typically rely on 2D or 3D data (Richtsfeld et al., 2012; Yang and Worboys, 2015; Wang et al., 2018). Bormann et al. (2016) and Ambrus et al. (2017) introduce segmentation of rooms in indoor environments. The goal of their work is to provide the segmented data to mobile robots for autonomous navigation. In contrast, Kortenkamp (1993), Choset and Nagatani (2001), Schröter et al. (2003), Vasudevan et al. (2007) and Angeli et al. (2008) represent the environmental understanding of a mobile robot in a way that facilitates human comprehension. They suggest a topological representation of places visualized as object graphs. The visualization of the environment is hierarchical, and strongly motivates usage for navigation. However, highlighting OOIs in real-time during flight missions is not investigated. Yang and Worboys (2015) also supports structuring of indoor spaces into rooms and portals from offline generated data. Kun et al. (2017) report on an ontology-based navigation model as part of an indoor positioning framework, introducing basic categories of abstract 2D objects (right Figure 3). All these approaches strongly support design of an abstract comprehensive representation of the scene to compute interactive navigation graphs for an indoor space (Section 3.2.2). However, none of these authors evaluate real-time generation of an interactive scene topology as part of a teleoperation system for aerial indoor exploration under real-world constraints.

2.5 Input Devices - Enabling Manual Control

Extensive research on teleoperation of small sized aerial robots with various types of input-devices and according interactions, has been conducted in the past decade. Despite the popularity of novel interaction paradigms, like hand-gesture (Yu et al., 2017; Duan et al., 2018), body-gesture (Rognon et al., 2018), gaze(Yu et al., 2014; Erat et al., 2018; Yuan et al., 2019) and language (Huang et al., 2010, 2019)), more recent work still focuses on aspects of traditional joystick based teleoperation of small-sized UAVs, for example avoiding collisions during navigation (Cho et al., 2017). Interestingly, Sanders et al. (2018) report that operators still prefer joystick control over indirect gaze based steering, whereas findings of Herrmann and Schmidt (2018) indicate that a traditional input device is more efficient than their extensive and carefully designed system based on natural interactions. In conclusion, if task-efficiency is preferred over user experience (fun to use, increasing enjoyment during teleoperation) traditional input devices are still hard to outperform. Remarkably, joystick controls can be still considered as state-of-the-art input device and are commonly chosen as baseline for performance evaluations. As a consequence, also our work aimed for selecting a traditional input device and according interaction design which would be able to compete against conventional joystick controls. Experimental results are detailed in Section 6.1. Other important aspects were required pre-training, complexity and cost-effectiveness. Details of the according design rationales can be found in Section 3.3.

2.6 High-Level Teleoperation Systems For Exploration Of Indoor Environments

Very recent work on fully working teleoperation systems for indoor environments is for example discussed by Valner et al. (2018). In their work they introduce high-level interaction methods based on gesture and language, but for ground-based robots. While they also suggest seamless switching between navigation, inspection and manipulation tasks, they use traditional egocentric 2D views and a 3D map to improve task performance. Recent work on fully working teleoperation systems, but with aerial telerobots is discussed by Hedayati et al. (2018), Walker et al. (2019), Huang et al. (2019) and Paterson et al. (2019). All systems use state-of-the-art AR, MR or VR input devices, whereas they also design high-level interactions for their

Figure 3. Example map views of complex office environments with gradual loss of details. left) Full 3D map view. middle) Floorplan in 2D. right) Topological view (Kun et al., 2017).

human-robot interface. Their overall goal is to improve task-efficiency when commanding aerial telerobots in indoor environments. Remarkably, they all compare their teleoperation systems against baseline systems (using traditional joystick or keyboard controls) and their independent variable in the study corresponds to what type of teleoperation interface the participants used. Though, their systems are based on natural gaze or language commands and do not refer to an interactive 2D scene topology created in real-time. Further, they do not consider aerial exploration missions in challenging indoor environments where simple and robust input devices can be beneficial to improve task performance. Related work, which might be closest to ours, is presented by Szafir et al. (2017). The work presents three prototypes with different methods to control an aerial telerobot. Interestingly, they also make use of an abstract floor-plan representation of the scene. However, this plan is static and not autonomously created in real-time. Although related work already proposed abstract topological views for the control of teleoperation systems, to our understanding we are the first who introduce a fully working system that refers to an interactive scene topology, created in real-time during flight. This raises the interesting question, if the performance of our teleoperation system is also preserved when put into practice. Compared to a variety of related teleoperation systems with similar mission complexity (Cho et al., 2017; Riestock et al., 2017b; Thomason et al., 2017), we evaluate the performance of our fully working system with a user study under real-world constraints (Section 6.1).

3 SYSTEM DESIGN RATIONALES

The design of our teleoperation system is highly task-oriented for the purpose of aerial exploration. It focuses on exploration of civil buildings with constrained indoor spaces and repeating room geometry. Example representations of an office building are shown in left and middle Figure 3 (3D map and 2D floorplan). Typically, an exploration mission would require to navigate inside the building and detect OOIs (fire extinguishers or trapped victims). For such applications teleoperation systems can be helpful, if disaster relief forces are not able to reach inside such buildings and assessment of the situation is required (Lichiardopol, 2007). In order, our teleoperation system uses the same main components as state-of-the-art systems (Figure 4):

- Aerial Telerobot: Our telerobot is a small-sized UAV, holding various sensors (cameras and inertial sensors) and actuators to perform the challenging task of aerial indoor exploration. Additionally it is equipped with an onboard computer to transfer sensor and actuator data to the human robot interface via the remote connection.
- Human Robot Interface: Our human robot interface includes all software components for processing the sensory data and flight-control of the telerobot. Further it holds the interactive scene topology (RPG) which is enabled by the underlying system components (Section 4.3).

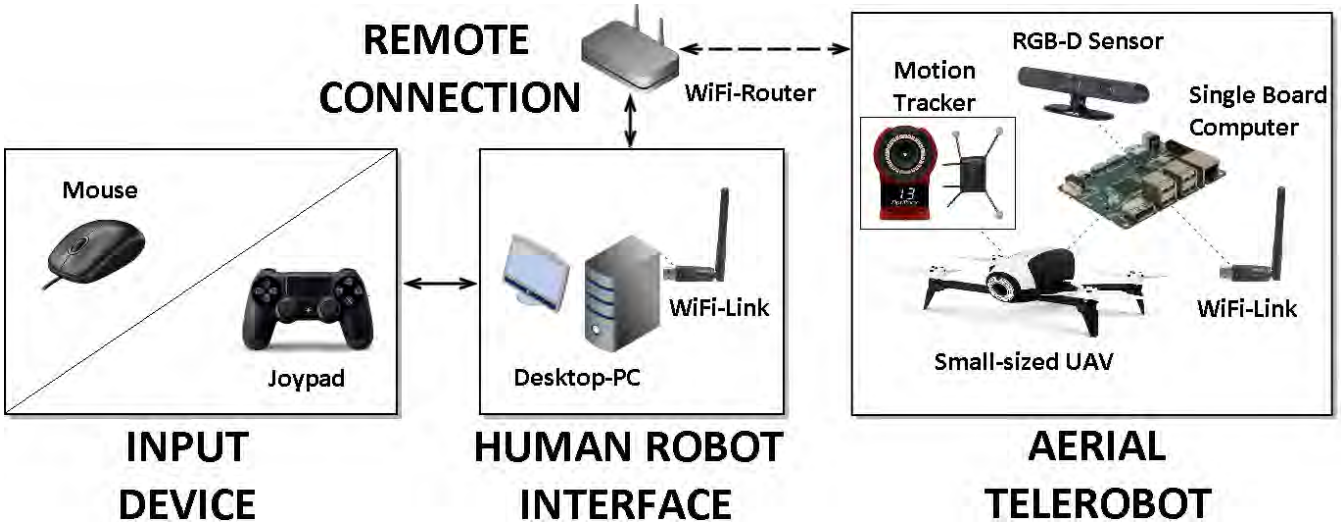

Figure 4. Overview of the main components of the teleoperation system design: The aerial telerobot which is a small-sized UAV and localized by a motion tracker (Optitrack, 2019). A remote connection between the telerobot and the human-robot interface which runs on a desktop PC. The input devices used for manual control of our teleoperation system.

- Input Device: The design of our system considers a simple and cost effective input device sending manual high-level commands to the human-robot interface.

3.1 Teleoperated Aerial Exploration Of Indoor Environments

Indoor space is typically limited and room exploration may require passing through narrow passages or so called portals, which can be hallways or windows. As a consequence, for our teleoperation system we designed a highly mobile and rather small sized UAV as telerobot. While important aspects are mentioned in Section 2.2, the design of our telerobot focuses on core functionalities which are vital for indoor exploration. On a higher task-level, our telerobot provides functions for **room exploration**, **object inspection** and **navigation of portals**. However, such high-level tasks entail a variety of low-level functionalities with increased complexity (Figure 5). Also, it is important to distinguish between **mission-critical tasks** and **non-mission-critical tasks**, whereas mission-critical tasks have to be solved by the teleoperation system under all circumstances and at all time. If the system fails with a mission-critical task this could lead to serious damage of the telerobot and potentially end the overall mission. For our system design we define the following low-level mission-critical tasks which are vital for indoor exploration:

- Localization: The telerobot has to be able to localize itself against the environment at all time. A failure in self-localization would typically result in that the telerobot collides with its environment.
- Landing/Take-off: Based on a robust localization and proper control of speed and acceleration, the telerobot provides assistive features like take-off and landing.
- Hold Position: Due to the turbulences that occur in the indoor environment, our design has to consider methods for stabilizing the telerobot while in-air and rejecting disturbances. Disturbances can occur due to flying close to obstacles or passing through portals.
- Collision-free Path Planning: Path- and motion-planning ensures collision-free navigation inside the indoor environment. It is vital if navigation between objects is required (waypoint-based navigation).

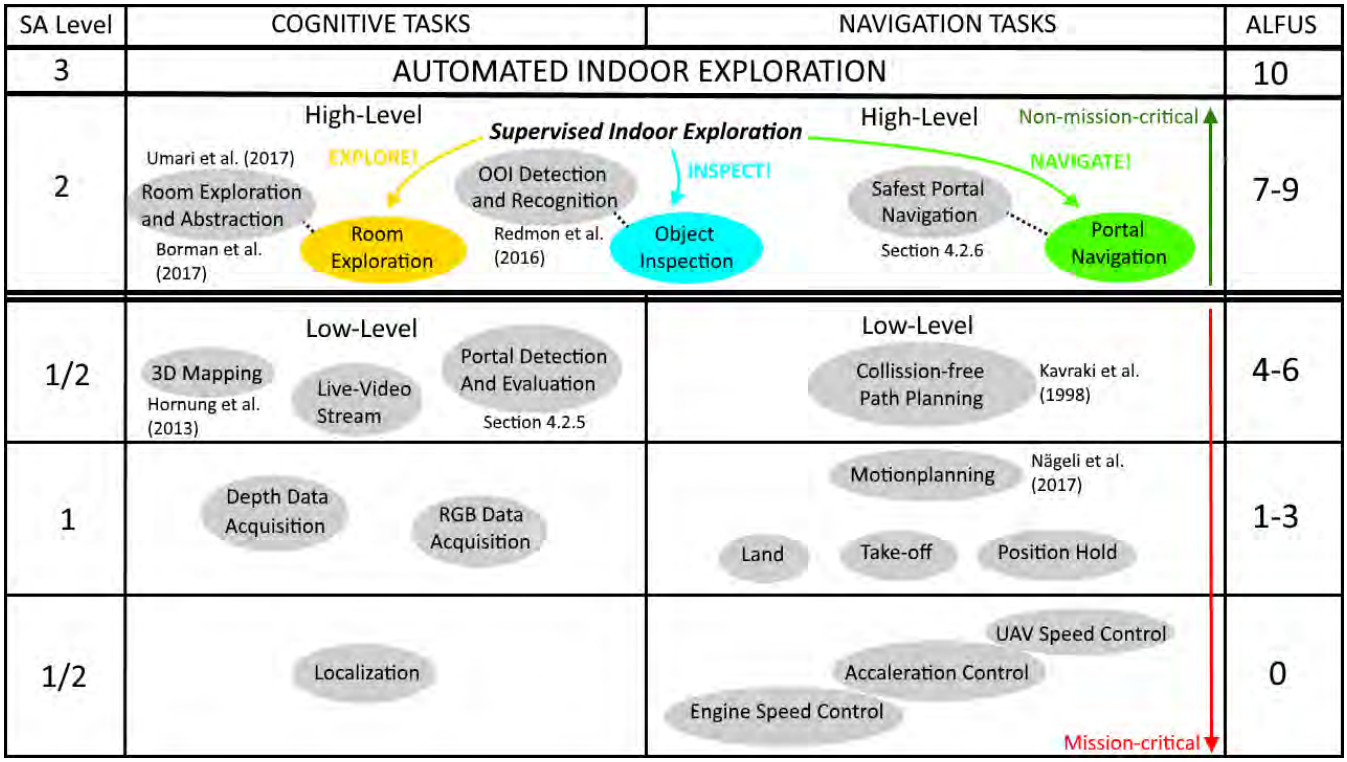

Figure 5. Overview of high-level and according low-level sub-tasks that have to be solved by our teleoperation system. Noteworthy is the separation between cognitive- and navigation-tasks, definition of mission-critical tasks, levels of situation awareness (Burke and Murphy, 2004), relation to the ALFUS (Huang et al., 2005b) and the recovery behaviours, triggered by high-level interactions of the operator (Explore!, Inspect! and Navigate!).

- Live-Video Stream And 3D-Mapping: It is based on a robust acquisition of sensory data, whereas abstraction into a topology requires the 3D data. Since it also provides minimum understanding of the remote scene to the human operator, these tasks must not fail.
- Portal Detection And Evaluation: We detect portals by analyzing single depth images, whereas we recognize the contour of the portal in 3D and estimate size (minimum diameter) and the normal vector of the contour at the geometric center. Once a potential portal is detected, it must be evaluated correctly.

On top of the low-level tasks we introduce high-level non-mission-critical tasks. These are difficult cognitive tasks, where a human operator can still improve overall performance. The high-level tasks can be summarized to one ”automated indoor exploration” task, autonomously executed by the system at highest LOA. In particular, the system uses an automated search strategy to explore one single room, identifies objects and portals on the fly and is able to navigate the safest portal. Implementation details are given in Section 4. Non-mission-critical tasks are considered as the following:

- Room Exploration and Abstraction: Room exploration is based on a state-of-the-art RRT exploration algorithm. On lower level this requires collision-free navigation. In parallel the system has to tackle the challenging tasks of detecting portals for navigation and objects of importance (OOIs). Once a full room is explored it is abstracted into a node and added to the scenes topology.
- Object Detection And Recognition: For object detection our system design aims for using state-of-the-art real-time detection algorithm.
- Safest Portal Navigation: After room exploration the system navigates the safest detected portal.

However, if the system fails with one of the high-level tasks, the operator can intervene by commanding high-level recovery behaviours in the GUI of the human-robot interface (Figure 1). In detail the operator can: trigger a simplistic but robust search strategy (**Explore!**), select a preferred portal over the other (**Navigate!**) or correct for object detection by close inspecting the object and/or registering the object manually (**Inspect!**) (Figure 7). Noteworthy is that our Inspect! command is motivated by the overview and detail paradigm, also used in the work of Seo et al. (2017) to improve effectiveness of teleoperation.

3.2 Human-Robot Interface

Our human-robot interface is designed to support the human operator during teleoperation. Core design aspects are typical essential tasks during aerial indoor exploration, limitations of the telerobot and usage of an untethered remote connection.

3.2.1 Levels of Automation And Approaches For Control

Our proposed scenario for aerial indoor exploration involves rather complex tasks, like object recognition and path planning. Such tasks have to be executed at the same time and involve mission-critical tasks like collision-free navigation. Due to the complexity of the tasks, the design of our system assumes that true full autonomy is not feasible. For our scenario a human operator is necessary to support the system with complex cognitive tasks on higher level. However, these tasks are non-mission-critical. The purpose is to avoid the lumberjack-effect and avoid sudden passing of control to the operator. If tasks fail on higher level, the telerobot is not damaged and able to continue with the overall mission. As a consequence, in accordance to related work Valero-Gomez et al. (2011), we design a supervisory control approach for our system which adapts the ALFUS framework (Huang et al., 2005b). Details about task definitions, high-level interactions to supervise the system with recovery behaviours and relation to LOA are presented in Figure 5 and Table 2. Importantly, hazardous regions in challenging indoor environments require the usage of an untethered remote connection. Consequently, potential sudden network dropouts and time delays during control support supervisory control.

3.2.2 Graphical User Interface

The user interface is one vital design aspect of our full high-level teleoperation system. Moreover, its design is based on the complex interplay with the underlying system components, whereas the overall goal is to improve teleoperation during aerial exploration missions. Yanco et al. (2004) summarizes core design aspects to improve overall task performance which are 1) using a map; 2) fusing sensor information; 3) minimizing the use of multiple windows; and 4) providing more spatial information to the operator. In addition, (Nielsen et al., 2007) discusses several window layouts in a standard paradigm. Besides of the rich variety of designs found in related work, a very common window layout is placement of exocentric map views on the bottom half of the screen whereas egocentric live camera views are placed on top.

The design of the GUI is also based on a standard layout, whereas we keep all view windows at equal size. It includes a traditional egocentric live view on top and a 3D map view on the bottom half of the screen. The purpose is to provide a minimum of spatial understanding to the operator. For the 3D view we use grid-map representations as they are a more robust in the presence of network delays and sudden dropouts (Riestock et al., 2017b). We place the view of the interactive scene topology (RPG) side by side to the traditional views to avoid occlusions or switching. The RPG is motivated by exploration of structured human environments, which can have complex and repetitive geometry (e.g. office buildings). While the structure of such environments motivates a topological representation of the environment, related work (Section 2.4.2) clearly supports the use for navigating robots. Other motivational aspects are extensively

discussed by Johnson (2018). Amongst other benefits, the work states that a topological representation is suitable for telerobots which have to navigate reliably from one place to another without the need of building an accurate map. While this is not valid during exploration of the environment, clear benefits occur for repeated navigation from one object to another after exploration. Johnson (2018) also points out that the topology supports affordances (opportunities for interactions) and poses a human-like representation of the environment. Based on the concept of an ecological interface (Nielsen et al., 2007), we designed visualization of objects that support affordances, but do not overwhelm the operator (Atherton and Goodrich, 2009). Consequently, we define general OOIs, which are detected during the exploration mission and highlighted in the topological scene view. Based on these considerations and avoiding to overwhelm the operator with too rich scene details in the traditional views (left and middle Figure 3), our design leads to the RPG which poses an interactive topological 2D map of the indoor environment. Implementation details can be found in Section 4.2.

3.3 Input Device

The design of our high-level teleoperation system includes a topological scene view which is represented in 2D. Because the topology supports affordances, we make OOIs explicit for interaction during flight. Motivated by the 2D representation and also considering the design aspects discussed in Section 2.5 we consequently selected a 2D mouse as input device. Besides of being robust and simple to setup (e.g., no need for calibration), other advantages are shorter pre-training phases and cost-effectiveness (Espingardeiro, 2012). The mouse holds three buttons, whereas the operator can trigger three high-level recovery behaviours (Figure 1 and Figure 9) of the aerial telerobot (Section 4.1).

4 SYSTEM IMPLEMENTATION

To solve the challenging tasks that occur during aerial indoor exploration missions, we implemented the following components as part of our high-level teleoperation system:

- **Aerial telerobot** represented as small-sized UAV. The UAV is equipped with a sensory setup to acquire RGB-video and depth data. The data is transfered to a desktop PC via the remote connection.
- **Human-robot interface** to facilitate control of the aerial telerobot by providing different views of the remote scene. Based on these views the operator controls the telerobot in a supervisory manner via high-level interactions. It further holds the underlying system components that are responsible for flight control, 3D mapping, abstraction, detection of portals and object detection in real-time. Remarkably, the components are vital for enabling the interactive scene topology of the human-robot interface. Thus, they are also essential to enable the high-level interactions (Explore!, Inspect!, Navigate!).
- **Input devices** that sends manual inputs to the human-robot interface. It is implemented as a simple and cost-effective mouse to interact with the RPG. To compare our system against traditional controls, we use a joypad controller for our user study (Section 6.1).

While the physical setup of the teloperation system is shown in Figure 6, we give an overview of the software implementation represented as state diagram in Figure 7.

4.1 Aerial Telerobot (UAV)

For the aerial telerobot (UAV) of our system we use a modified Parrot Bebop 2 (Parrot, 2015). It is compact, suitable for narrow indoor spaces and offers open-source drivers (Monajjemi, 2015) for low-level flight control. For reliable experimentation, we attach retro-reflective markers for outside-in localization using

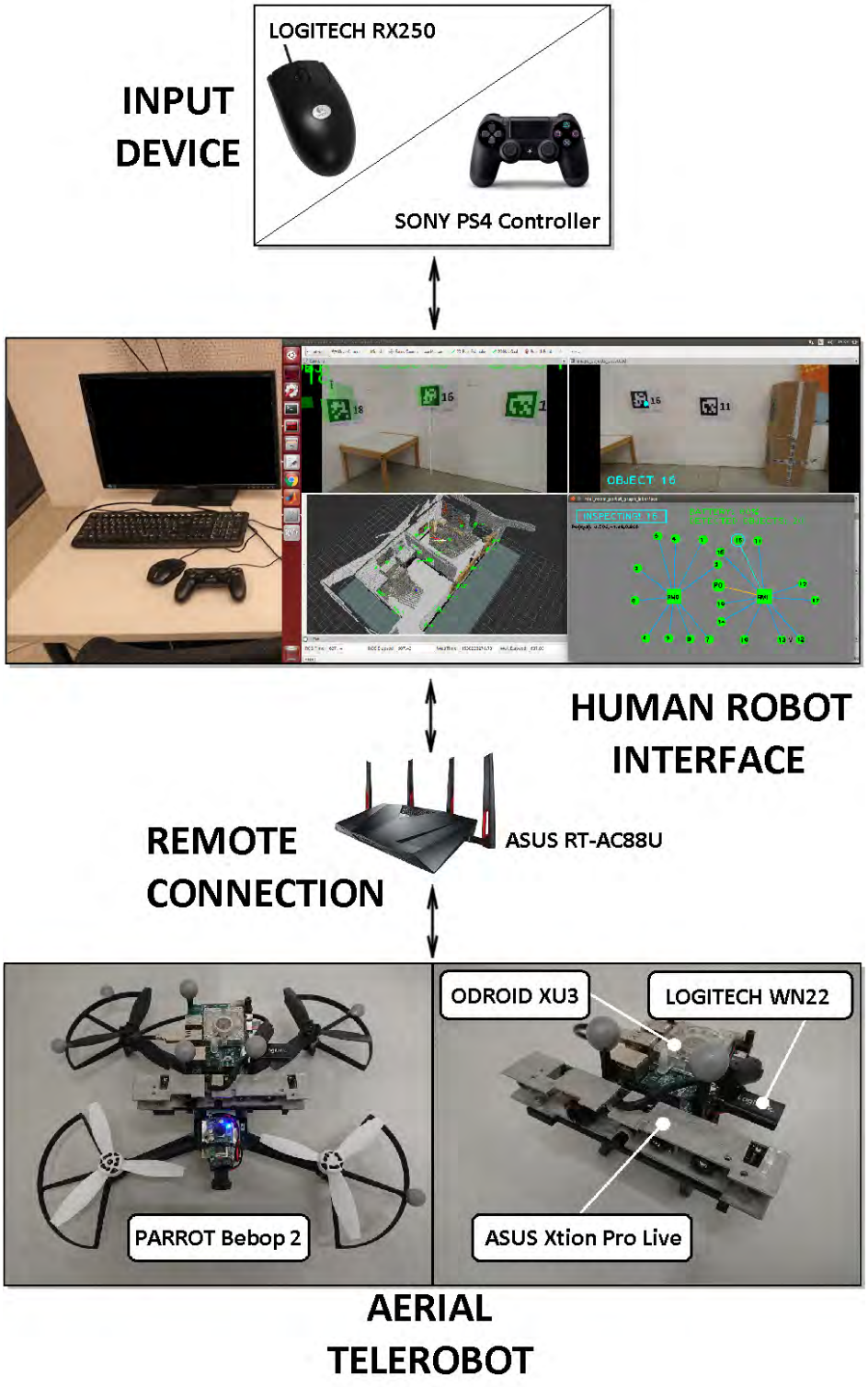

Figure 6. Implementation of the main components of the high-level teleoperation system: The UAV based on a Parrot Bebop 2 with onbaord single-board-computer and sensory setup. The remote connection implemented as ASUS RT-AC88U wireless router. The GUI, including the RPG, implemented on a desktop PC in ROS. The input devices implemented as a Logitech RX250 mouse and a PS4 controller.

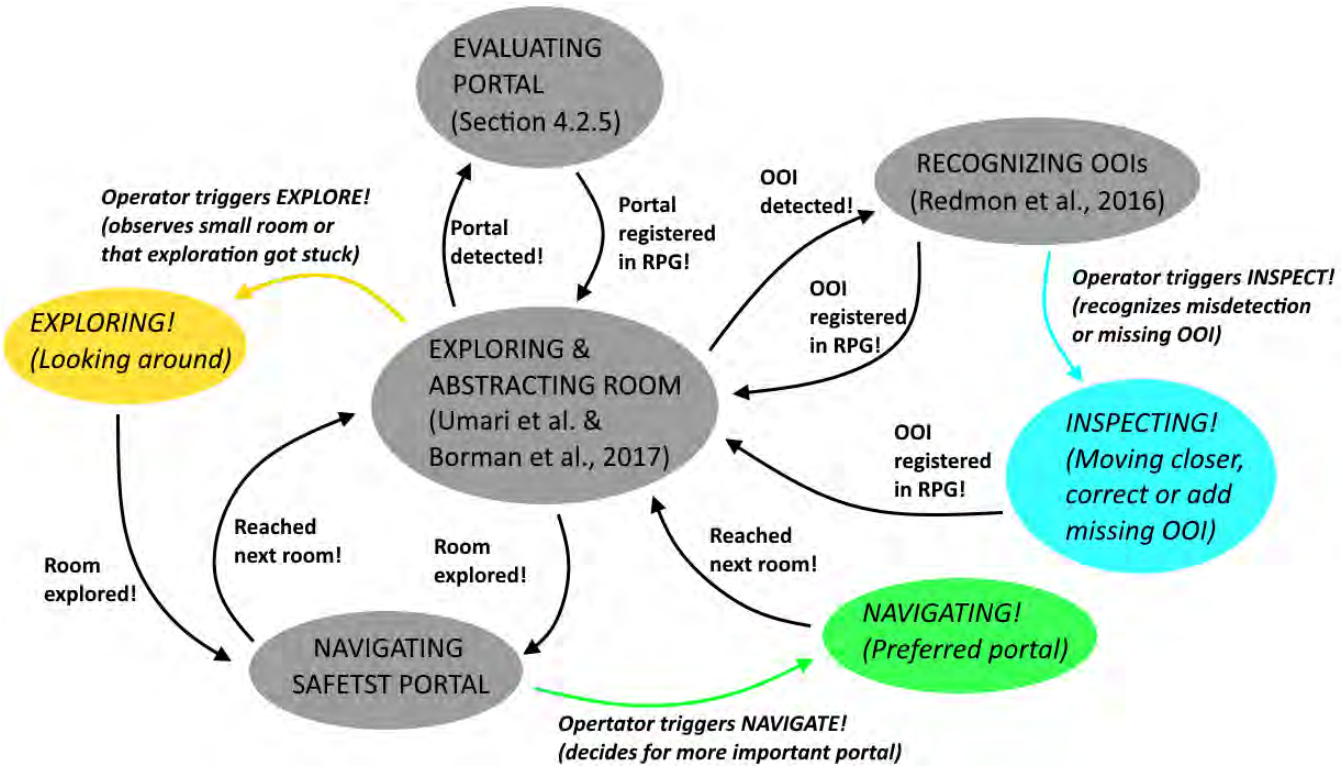

Figure 7. State diagram of the software framework of our teleoperation system.

an Optitrack Flex 13 infrared motion capturing system. An overview of the physical setup is shown in Figure 6.

With all on-board sensors attached, the outer dimensions of the UAV are $32.8 \times 38.2 \times 25cm$, and it weighs $700g$, with flight times of up to approx. 10 minutes. On top of our UAV, we mount a customized RGBD sensor rig ($250g$), consisting of an ASUS Xtion Pro Live sensor ($FOV_{hor.} = 58°$, $FOV_{vert} = 45°$) and a Logitech WN22 WiFi stick, connected via USB to an ODROID XU3 single-board computer. During our experiments, the UAV was navigating at a default flight height of $z_{takeoff} = 1.25m$.

4.2 Human-Robot Interface And Input Devices

In this section, we give details about the human-robot interface which enables the operator to high-level control our teleoperation system. As one vital component it holds the RPG as interactive scene topology which is created, based and, thus, strongly dependent on the complex interplay of its underlying system components (Section 4.3). While we motivate a supervisory control approach in Section 3.2.1, in the following we discuss implementation details and the correspondence to the LOA. Furthermore, we give details about the traditional baseline system with a joystick as input device in Section 4.2.2. It runs on low automation level, and the operator has manual control over the system. We compare the two different systems against each other and report on results in Section 6.1.

4.2.1 High-Level Teleoperation (RPG Condition)

High-level teleoperation of our system is enabled by the RPG to let an operator effectively supervise our system on high LOA (Table 2). We intentionally do not provide low-level access, so that the operator is not burdened with demanding mission-critical tasks (ALFUS 1-6). This also means that the system must achieve all mission-critical tasks even in a challenging indoor environment. The system usually operates on

highest LOA (ALFUS 10), but we let the operator switch to a lower collaborative level (ALFUS 7-9), if supervision is required. This is particularly relevant if the underlying system components components do not perform satisfactorily, e.g., when object recognition fails (Materna et al., 2017).

For the RPG (Figure 2), we combine a traditional egocentric view (on-board camera of the UAV) with an exocentric 3D map view. The views include visual aids for current pose of the UAV, view frustum of the onboard camera, the online reconstructed 3D environment and invalid flight zones. The purpose is provide a basic spatial understanding of the scene. We extend the ego- and exocentric views with an interactive topological view, the RPG. It consists of rooms (nodes) and portals (edges) to other rooms or OOIs (e.g., a fire extinguisher or a victim). OOIs registered in the RPG are highlighted in real-time. Once an interactive OOI is highlighted, the operator can use 2D mouse inputs to supervise the system by a reduced set of high-level interactions (**Explore!**, **Navigate!** and **Inspect!**). This triggers recovery behaviours (Figure 9 and implies switching from highest LOA (ALFUS 10) to collaborative level (ALFUS 7-9).

The Explore! command lets the system more effectively uncover smaller rooms. During this task, our system autonomously detects OOIs and adds them as interactive nodes to the RPG topology. If a false detection occurs, the operator can use the Inspect! command to move closer. If one of the detected objects is selected, a safe path is generated between the current location of the UAV and the object. After the system navigates close to the false detection, the operator can inspect the situation in a close-up egocentric view and determine further action. During exploration of a room, also portals which are safe for navigation are detected and highlighted (Section 4.3.4) automatically. Detected portals add a new node and edge to the RPG. In case of multiple detections, the operator is able to select a preferred portal to trigger navigation into the adjacent room (Navigate!). A picture sequence of the recovery behaviours is shown in Figure 9, whereas we present real-world flights in our supplementary video (Section 7). Details about the physical setup of the aerial telerobot are discussed in Section 4.1.

The goal of the RPG is to provide a topological map that is a human-like representation of the environment. Since it provides natural interactions for commanding the system and describing the environment, it facilitates and eases human-robot-interaction Johnson (2018). Moreover, its purpose is to reduce scene details in the presence of cluttered traditional views (left and middle Figure 3). However, the concept of the RPG has also limitations which we detail in Section 5.

4.2.2 Traditional Direct Teleoperation (JOY Condition)

To compare the effect of our high-level teleoperation system against a state-of-the-art baseline system, we implemented traditional joystick controls. For our study we define it as condition JOY. In this condition, the operator uses a joypad to command the UAV at lower ALFUS (Table 2) with a high-level of interaction on sides of the human operator (ALFUS 1-3). At this level, the system takes care of automatic take-off, position stabilization and landing. Besides, the operator is also responsible for mission-critical tasks.

To achieve fair comparison against the RPG, we added a visualization of flight zone boundaries to help the operator prevent collisions. The boundaries are displayed in the horizontal plane via a color-coded surface at the height of the UAV. Operators must not exceed this indicated boundary and get color-coded feedback, if they are close to exceed the maximum flight height. The surface turns orange, if the UAV is close to the height boundary, which means the distance of the geometric center of the UAV to the upper boundary is smaller than the height of the UAV. The surface turns red, if the distance is smaller than half of the height of the UAV, indicating that the operator has to steer the UAV downwards immediately.

The joypad is used in MODE-2 configuration, allowing the operator to give direct motion commands. In this configuration, the left rudder stick controls translational and rotational velocity of the z-axis of the UAV, and the right rudder stick gives acceleration inputs along the x-axis and y-axis of the UAV.

4.3 Underlying System Components

This sections summarizes the underlying components of our high-level teleoperation system. They are implemented as part of the human-robot interface on a Desktop PC and responsible for exploration, flight planning and navigation, 3D mapping of the environment and highlighting of OOIs. Since they even enable the RPG as interactive scene topology, the effectiveness of our full system strongly depends on their performance. Thus, they must be emphasized as core for interaction. The aerial telerobot supports with automated indoor exploration and the human operator can trigger recovery behaviours via the RPG. In order, if a non-mission-critical task fails, time performance could be improved. The recovery behaviours are designed in a supervisory manner so that the human operator can effectively supervise the system with difficult tasks on higher-level. Their purpose is emphasized throughout the following use cases:

- Explore: After take-off, the UAV autonomously starts exploring the current room using an RRT-based exploration method (Umari and Mukhopadhyay, 2017). If the operator decides that the room seems rather small or the exploration fails to fully explore the room, the operator can on demand trigger a simple recovery behaviour. In that case the UAV explores the local environment by flying a simple circular trajectory. Once a room is fully explored we use the implementation of Bormann et al. (2016) for room-segmentation.
- Inspect: During exploration of a room, the telerobot autonomously detects portals and OOIs, like victims or fire-extinguisher. However, if the operator feels that an object was misdetected, the operator can command the telerobot to move closer to a detected OOI or portal for verification.
- Navigate: During room exploration, the telerobot detects portals which are safe to navigate. However, if multiple safe portals are detected, the human operator might intuitively prefer one portal over the other for navigation. In such cases the operator can manually trigger portal navigation.

4.3.1 Room Exploration

At the beginning of every mission, the UAV ascends to a default flight height (Section 6.1). After reaching the default height, the system starts to autonomously explore the local environment (Figure 8, Step 1). For local exploration of a single room, we use a frontier detection algorithm, based on rapidly-exploring random trees (Umari and Mukhopadhyay, 2017). If no failure cases occur, we consider the system to work on highest LOA (ALFUS 10).

Once the UAV takes off, we start detection of local frontiers by taking into account the occupancy map constructed online. First, we project 3D occupancy information into 2D, since this helps to clearly define boundaries of a single room. We project occupied cells into the 2D map. Second, we let a local frontier detector discover valid navigation points, which are derived from a rapidly growing random tree biased towards unexplored regions of the occupancy map. Third, we directly steer the UAV towards the detected point, incrementally exploring the local environment. These steps are repeated, until no new frontier points are detected and the room is locally fully explored. To abstract the local room and to further obtain room information we make use of the segmentation approach presented by Bormann et al. (2016). Note that we assume the range and FOV of our depth sensor to be wide enough to cover the close environment and detect potential obstacles, when navigating at default height. For simplicity, we assume that there are no additional obstacles between the UAV and the detected room boundaries. The operator is further able to

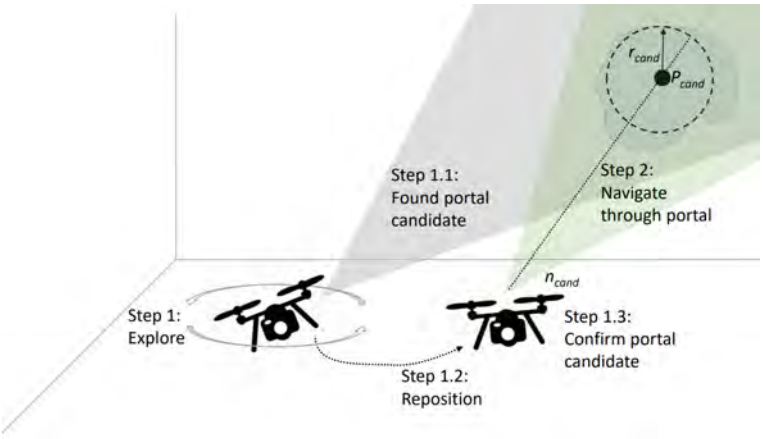

Figure 8. Method for detecting and evaluating potential safe portals directly from depth-image data. The UAV first explores the close environment and, if a safe-portal candidate is detected, positions itself to confirm that the portal candidate is safely traversable.

manually override frontier detection by selecting the abstract room representation of the RPG (triggering Explore! and switching from highest- to collaborative LOA). This prompts the system to execute a more efficient circular trajectory.

4.3.2 Room Navigation

To enable collision-free navigation through portals from one room to another, we use a global path planning approach based on probabilistic road maps (PRM) (Kavraki and Latombe, 1998). The global path planner generates a PRM based on the occupancy grid map (Hornung et al., 2013). The PRM is represented as a set of 3D points given in world coordinates. For an example of generated paths, please refer to Figure 9c.

The PRM is passed to a real-time motion planning framework Gebhardt et al. (2016); Nägeli et al. (2017b,a). The motion planner involves a model predictive controller (MPC), which produces smooth motion trajectories for the UAV when moving along the global path. Following a receding-horizon MPC formulation, at each timestep Δt, a locally optimal path with N steps and a duration of $N\Delta t$ is computed. This optimization problem is re-evaluated at every sampling instance T_s, leading to a closed-loop behavior. Thus, we make use of the disturbance rejection characteristics of the MPC to stabilize the UAV during the mission. Stabilization against turbulence is necessary when flying close to objects or passing through portals. The real-time motion planner is implemented in MATLAB (Robotics Toolbox), utilizing the FORCES Pro real-time solver (Domahidi and Jerez, 2013).

4.3.3 Environmental Reconstruction

To provide the operator with basic environmental understanding during navigation (Section 4.3.1), we make use of the RTABMap reconstruction framework (Labbe and Michaud, 2013, 2014). It represents the reconstructed geometry as a colored occupancy grid map and is capable of loop-closure detection. The grid map is created by fusing depth- and RGB-data from the onboard sensor setup of the UAV (Figure 6) and visualized in the exocentric view.

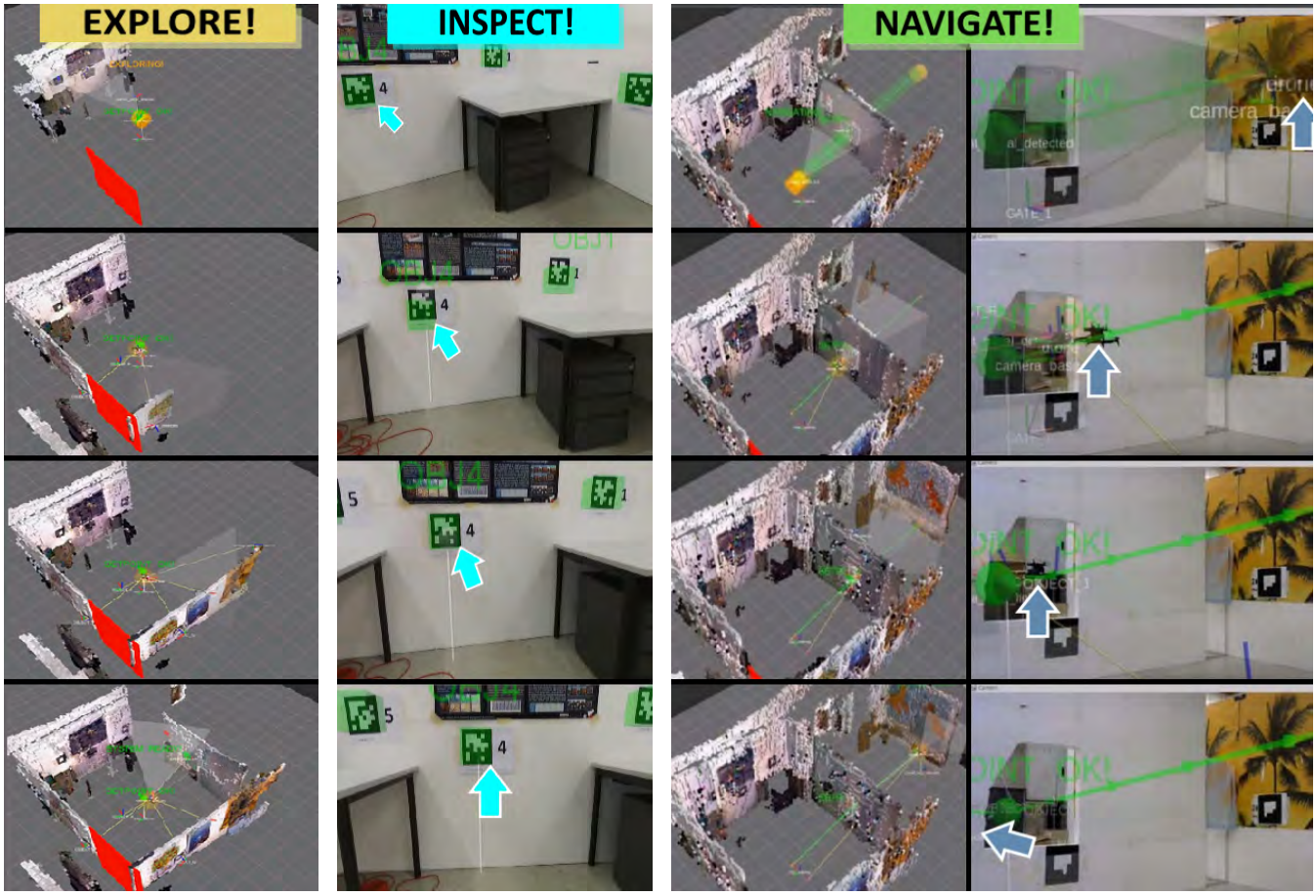

Figure 9. Exocentric virtual views of the aerial robotic system during execution of the recovery behaviours. a) The UAV is commanded to explore its surrounding by flying a circular trajectory and simultaneously builds a 3D model of its environment from RGBD data. b) The UAV is commanded to close inspect a detected object for verification. c) The UAV is commanded to navigate through a safe portal to the adjacent room along an autonomously planned path, shown in green. The UAVs position is marked with blue arrows.

4.3.4 Detecting And Highlighting Objects Of Interest

For our experimental setup, we introduce different types of OOIs which are commonly present in exploration scenarios. These objects can be hazardous areas (location of fire extinguishers, broken power lines, gas leaks), human victims (embodied by human-like mannequins) or portals (potentially narrow passages), which connect adjacent rooms. The OOIs are automatically highlighted as virtual overlays in the GUI to direct the operators attention towards them. This requires automatic object detection and registration of the observed object positions in world coordinates. Noteworthy, we use a true relative scale between objects in the current design of the RPG. We detect objects either using the YOLO framework for object detection (Redmon et al., 2016) or by simply marking them with Apriltag markers (Olson, 2011) during the user study (Section 6.1).

Indoor environments can typically be structured into wider open areas (rooms) and more narrow spaces (portals) connecting rooms (Kun et al., 2017). During the exploration task, our goal is to detect and visualize potential portals. Making rooms and portals explicit is vital in our scenario, since they support navigation. Interactive highlighting, helps operators to get a clearer understanding of the environment and make an educated decision on which part of the environment to explore next. The portal detection proceeds

as follows (Figure 8): During exploration of the close environment (Step 1), we detect discontinuities in the depth images captured by the RGBD sensor. If the discontinuities form a closed, bounded region with minimum radius r_{cand} and depth d_{min} (measured from the centroid P_{cand} of the entry surface), the region is selected as a portal candidate (Step 1.1). This intermediate step is necessary to ensure the portal can be safely traversed, as looking at portals from larger offset angles would result in shadowed regions behind the portals. Based on the surface geometry of the portal candidate, we derive P_{cand} and the corresponding normal vector $\vec{n}_{cand}$. The normal $\vec{n}_{cand}$ is oriented perpendicularly to this entry surface and has its origin in P_{cand}. In Step 1.2, the UAV is commanded to align the x-axis of its local coordinate frame F_{UAV} with $\vec{n}_{cand}$. The distance to the portal candidate d_{cand} is calculated based on the minimum radius r_{cand} and the narrower vertical field of view of the depth sensor FOV_{vert}. d_{cand} can be expressed as $d_{cand} = r_{cand} / \tan(FOV_{vert})$.

5 SYSTEM LIMITATIONS

The teleoperation system presented in this work has also several limitations. Besides of typical issues related to telerobotics as such, most important limitations are discussed in the following:

- Telerobot: Besides of there is room for improvement of our physical design (weight, size and computational onboard power), also the ability to morph and adapt to challenging environments could be added. Speaking of passing narrow portals or gaps, also highly dynamic maneuvers (Falanga et al., 2017) are currently not possible but could be interesting for future work. Another limitation of the telerobot is the exploration algorithm. While we make use of an RRT-based approach to explore a single room but at constant flight height, a more powerful approach would involve full 3D exploration. Additionally, a gimbal could help to resolve constraints with the cameras limited FOV, making room exploration more efficient.
- Wireless Remote Connection: Due to the usage of an untethered remote connection between the telerobot and the human-robot interface, typical problems could occur like limited bandwidth and sudden connection dropouts. While in-field applications would require a much more sophisticated (and expensive) setup, in our implementation we considered commodity hardware only. However, it must be stated that due to usage of a powerful WiFi router, comparably short ranges, and non-overlapping/non-populated channels the effects during the user study could be reduced to a minimum.
- Supervisory Control of high-level tasks: The supervisory control approach of our system aims for effectively resolving failures of high-level tasks. However, this is only valid if the telerobot is capable of handling all low-level mission-critical tasks without failure.
- Human-Robot Interface: An essential component of our human-robot interface is the RPG, serving as interactive scene topology. The focus of its design is to supplement traditional views by supporting affordances and reducing scene detail. Thus, overwhelming the operator should be avoided. However, several aspects could not be elaborated on as part of this work. While in our current RPG design we use a true relative scale of rooms, portals and objects, we did not elaborate on different layouts of the objects inside the RPG view or adapting its orientation relative to the 3D view. We also did not yet investigate on proper placement of the simplistic 2D objects in case they overlap or on altering their shapes and size. Future work would also include elaboration on a zooming function for wider areas and adding important details on demand. Such helper functions could display size and volume of the selected room or distance between the telerobot and according OOI if selected with the input device.
- Input device: The design of our teleoperation system supports a robust and simple-to-use input device which is also cost effective. As a consequence we utilize a traditional 2D mouse with 3-buttons. These

are dedicated to our three high-level interactions (Figure 1 to trigger recovery behaviours. However, the design of interactions and button mappings could be still improved by evaluating different layouts towards optimum usability. Further, utilizing a mouse with more degrees of freedom (Razor, 2015) could improve support for multi-floor exploration or manual steering of a camera gimbal with the attached joystick.

- Multi-floor environments: To be able to explore multi-floor environments, our system would require further components. For instance, the system would need to be able to detect stairways (Delmerico et al., 2013). In addition, the robustness of the untethered remote connection would have to be improved. While the implementation of our current system uses commodity hardware, systems with increased power and higher penetration of structures are for example presented by Patra et al. (2007). Additionally, like introduced for nuclear power plant inspection (Nagatani et al., 2013; Zhang et al., 2018), one or multiple additional telerobots could be used as mobile wireless transmission relays, retaining reliable communication.

6 USER STUDY

The purpose of our study is to investigate the effect of our high-level teleoperation system on operator performance during real-world exploration missions. We considered the different teleoperation systems as strongest baseline for our study conditions, whereas we compare our high-level teleoperation system, including the RPG (Section 4.2.1), against a traditional baseline system with direct controls (Section 4.2.2). Table 1 gives an overview of the experimental conditions, type of systems, and view modes.

A core aspect of our study is that, despite a variety of related work has shown semi-autonomous systems positively effecting task performance, however it is unclear if this holds in a realistic setting where a system has to generate an interactive abstract topological scene view in real-time during flight missions. While operators with traditional direct controls can issue commands based on their quasi-instantaneous human cognition, operators of the semi-autonomous system need to wait until it processes, abstracts and outputs (visualizes) the abstracted information. This raises the question if such systems are still able to improve task performance over traditional control approaches in a realistic setting, where operators potentially need to wait until information is available in the topological view.

6.1 Experimental Design

In the following sections we summarize the experimental design of our user study, including study conditions and tasks. Besides, we give details about study procedure, participants and accordance of the study to the local legislation.

6.1.1 Conditions

The main objective of our study was to assess the effect of the two user interface conditions, RPG and JOY, on operators task times, mental load and general comfort during a real-world indoor exploration mission. We based our study on **within-subjects design** and varied the order of the conditions using **full counterbalancing**. We defined task completion time, mental load and general comfort of the operator as main task performance metrics. We formulated the following hypothesis for the user study and report on results in Section 6.2:

- H1: The operator's task time decreases in RPG.
- H2: The operator's mental load decreases in RPG.

•H3: The operator's general comfort increases in RPG.

6.1.2 Tasks

According to Bhandari et al. (2015), typical indoor exploration tasks involve investigation of the unknown environment and evaluation of hazardous areas to minimize human risk. We designed our study so that participants had to fulfill similar tasks in our experimental setup (Figure 11). We assumed a situation where the operator is far from the indoor space and has no prior knowledge of it. To ensure a basic degree of validity, we discussed the design of our experimental task-design with a local fire brigade. As a conclusion, the members confirmed validity of our task design and further emphasized usefulness of our system for assessment of a stable but still potentially dangerous situation. As an example use case, they specified the on-site inspection of a damaged chemical recovery boiler where an imminent explosion cannot be ruled out.

The indoor exploration task of our study comprises three subtasks, which had to be completed by each participant in each of the conditions. During this task, participants had to fully explore the environment and find all OOIs. In particular, participants were told to:

•Find all 19 hazardous areas marked with fiducial markers.
•Find the safe portal.
•Find the victim.

The placement of objects was altered in a controlled fashion to avoid learning effects. An overview of the experimental indoor environment is given in Figure 11.

6.1.3 Procedure

Before each experimental session, an introduction to the teleoperation system was given to the participant by the experimenter. Preliminary questions were asked to identify eye-sight restrictions. The evaluation procedure of each experimental condition can be split into three phases. In a training phase, participants learned to use the system of the specific condition. This phase ended when participants reported to be comfortable in using the system. In the second phase, participants had to accomplish the indoor exploration task as fast as possible. For each participant, we captured screen recordings and measured the task completion time using a stop watch. The task was considered to be completed when the system detected all safe portals, hazardous areas and victims (RPG-condition) or users verbally confirmed to the experimenter that they found all of those objects (JOY-condition). In both conditions, users were aware of the number of objects they already identified as well as of the objects they still need to find. Finally, participants were asked to fill out a NASA-TLX (Hart and Staveland, 1988) task-load questionnaire (Scale: 0-100) as well as a custom questionnaire with respect to their experience in the respective condition. The custom questionnaire contained 8-point Likert items (ranging from 1, "strongly disagree", to 8, "strongly agree") asking participants about accuracy and smoothness of control as well as their perception of control over the system and their general comfort during the task.

6.1.4 Participants

A total of 23 participants were invited, 20 of them successfully finished the given tasks in all conditions. 3 participants had to stop the study due to technical problems and their results have been excluded. We invited 17 male participants and 3 female participants which were either students or researchers in the field of computer science or electrical engineering at Graz University of Technology (age: M=27.6, SD=3.98).

6.1.5 Ethics Statement

The presented study was carried out in accordance with the World Medical Association's Declaration of Helsinki, as revised in 2013 (HELSINKI, 2013). The study did not involve any medical experiments and further, no biometric data was taken from participants. We did not take any personal data from participants besides of participants age, whereas all taken data was fully anonymized. In general, the study was conducted in accordance with the local General Data Protection Regulation (GDPR) in Austria and all participants gave written informed consent via an IRB consent-form. In order, as per the local legislation, no IRB approval was required for our particular study type.

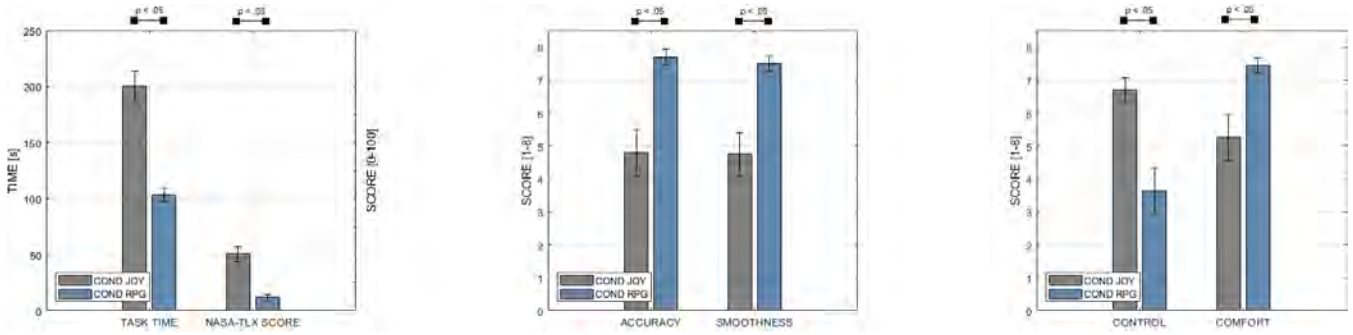

Figure 10. Our study results indicate significantly decreasing task times (Scale: 0-250s) and decreasing NASA-TLX score (Scale: 0-100) with our high-level teleoperation system (condition RPG). Based on an even 8-point Likert scale (Agreement-Scale: 1 Strongly Disagree - 8 Strongly Agree), we managed to retain general comfort during operation, compared to our baseline system with traditional joystick controls (condition JOY). In addition, participants reported increasing accuracy of control and smoothness of control.

6.2 Results

In each of our 20 sessions, we tested the teleoperation system in both conditions, JOY and RPG. This resulted in a total of 40 valid runs. For each participant, we took screen recordings and measured the task completion time during the flight. After finishing the flight for one condition, participants were asked to fill out the NASA-TLX score as well as a custom questionnaire. This questionnaire contained several 8-point Likert items asking participants about the accuracy of control, the smoothness of control, their perception of control over the system and their comfort in general during the task. We report mean, standard deviation and interval estimates for a 95% confidence interval. For significance testing, we use a paired samples t-test for task execution time as the data is normally distributed. All other measures are compared using Wilcoxon signed-rank test as questionnaire responses are non-parametric.

The main findings of our study are summarized in Figure 10. Statistical testing revealed that the task completion time was significantly lower for the RPG- ($M = 103.7s, SD = 13.7s$) compared to the JOY-condition ($M = 200.1s, SD = 30.58$, $t(19) = 12.01, p < 0.001$). In addition, a significant effect of conditions on mental load, as determined by NASA-TLX, has been revealed ($Z = 210.0, p < 0.001$). Again, RPG ($M = 11.75, SD = 6.43$) caused a significantly lower mental load than JOY ($M = 50.71, SD = 16.41$).

In our custom questionnaire, we asked participants about their perception of the tested user interface. Unsurprisingly, the perceived level of control in conditions decreased with the increasing LOA from JOY ($M = 6.7, SD = 0.87$) to RPG ($M = 3.65, SD = 1.63$). Wilcoxon signed-rank test showed that

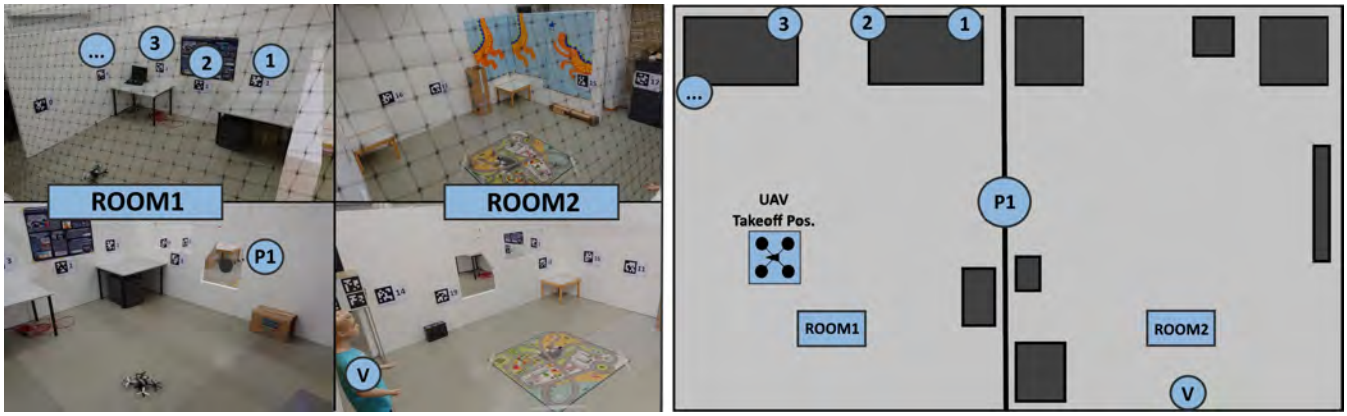

Figure 11. (left) Physical setup for our experimental evaluation. Note the two rooms, connected via a safe portal and the objects of interest including a victim. (right) The same environment, represented as a floor-plan in 2D.

these differences are significant ($Z = 169.0, p < 0.001$). In contrast, participants perceived RPG ($M = 7.7, SD = 0.57$) to be significantly more accurate than JOY ($M = 4.8, SD = 1.67, Z = 0.0, p < 0.001$). Similarly, perceived smoothness of control was higher for RPG ($M = 7.5, SD = 0.51$) compared to JOY ($M = 4.75, SD = 1.55$). Again, differences are significant ($Z = 0.0, p < 0.001$). Finally, perceived general comfort was significantly higher in the RPG condition ($M = 7.45, SD = 0.51$), compared to JOY ($M = 5.25, SD = 1.62$), with ($Z = 0.0, p < 0.001$). This lets us accept H3, which is supported by a significantly higher task completion confidence in RPG ($M = 7.8, SD = 0.41$), compared to JOY ($M = 6.8, SD = 1.06, Z = 0.0, p < 0.001$).

6.3 Discussion

Overall, we were able to support all of our three hypotheses, implying that our high-level teleoperation system is successful in supporting the operator during aerial exploration missions in challenging indoor environments. Remarkably, our teleoperation system reduced task execution times by 48.28% and task load by 76.82% compared to the JOY condition. Moreover, results indicate an increase in general comfort by 41.90%. We attribute the significant differences between conditions to the interplay of the RPG-view and the autonomous system. However, further research is necessary to differentiate the influence of the autonomous system and the topological scene view on results.

Although, participants conducted real-world flights to solve the posed exploration task, the study took place in a controlled environment. For instance, localization of the UAV was achieved with a motion capture system. However, on-board localization methods like SLAM have proven to be sufficiently accurate and fast to be used for UAV position estimation (Weiss et al. (2011), Mur-Artal and Tardós (2017)). In addition, due to limited lab space, the environment of our study did only comprise two rooms. Nonetheless, we believe that differences between conditions further evolve in favor of our system in wider- or multi-floor environments. The reason is that it is evidently harder to gain a good spatial understanding of larger compared to smaller environments. Thus, operators will benefit more from the RPG-view in larger spaces as it abstracts the environment in an easy-to-understand manner. Furthermore, the task of our study was a simplification of complex real-world search and rescue missions. However, it is likely that our system even better supports operators in more complex task scenarios. For instance, research has shown that topological views, like the RPG, are beneficial if an environment is fully explored and operators are required to repetitively navigate between OOIs Johnson (2018). With regards to our system, the reinspection of

an OOI could easily be performed by triggering its visualization in the RPG. The telerobot would then autonomously renavigate to the specific room and object. Due to mentioned reasons, we argue that, despite limitations, our experimental setting is an ecologically valid approximation of a real-world exploration mission.

Summarizing, our study has shown that high-level teleoperation systems with an on-the-fly created interactive scene topology are still able to better support operators in real-world settings, compared to systems using traditional controls.

7 CONCLUSION AND OUTLOOK

In our work, we demonstrate a fully working teleoperation system for aerial exploration missions. It improves task performance by using an interactive scene topology, whereas related work motivates using topological representations for robotic teleoperation. However, in contrast to related work, we for the first time investigate on how task performance is effected if the topology is created in real-time during actual flight missions. The overall goal of our system was to reduce task times and mental load of the operator while conserving general comfort. To elaborate on the expected improvement, we evaluated our teleoperation system with a user study under real-world conditions. We compared our high-level teleoperation system against a traditional baseline system with joystick control. Results indicate that our system positively effects task performance and operators comfort during aerial exploration of challenging indoor environments.

In future work we would like to address the limitations of our system (Section 5) and conducted user study (Section 6.3). Further we would like to evaluate our system in more complicated or even multi-floor environments, for which abstraction has a potentially larger benefit in terms of overall task performance.

Table 1. User Study Conditions

	JOY Condition	**RPG Condition**
Type of Control	Traditional Direct	High-Level Supervisory
LOA	1-3	7-10
RPG View	No	Yes
EXO View	Yes	Yes
EGO View	Yes	Yes

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments and funding

The authors wish to thank the collaborators all study participants, the fire brigade Gössendorf, Austria, and the members of TEDUSAR (Technology and Education for Search and Rescue Robots) group at Graz University of Technology. This work was funded by FWF grant I1681.

Supplemental Data

We provide a supplementary video to support the concept of our high-level teleoperation system.

Table 2. Relation between the ALFUS and operating the UAV of our teleoperation system.

ALFUS	Level Definitions	Operating The UAV
[10] Approaching 0% HRI	Highest level of automation, high complexity, all missions, extreme environment	Maximum automation level. Full autonomous exploration of the environment including object detection. The user can still improve task performance by switching to the collaborative levels (e.g., navigation of one preferred portal over the other or inspection of an OOI for verification).
[7-9] Low-level HRI	Collaborative level of automation, high complexity missions, difficult environments	High automation level involving non-mission-critical tasks. Functioning of repetitive low-level tasks is guaranteed (collision-free navigation). The operator can switch to this level, to supervise the system if it fails with complex tasks on highest level. **Minimum level that the operator can access in the RPG condition.**
[4-6] Medium-level HRI	Medium level of automation and complexity of missions, multi-functional missions, moderate environment	No operator access at this level for RPG condition.
[1-3] High-level HRI	Low level of automation, low level tactical behaviour, simple environments	No operator access at this level for RPG condition. **We use that range of levels in the JOY condition.**
[0] 100% HRI	Lowest level of automation, full manual remote control by the operator	No operator access at this level for RPG condition.

REFERENCES

Adams, J. L. (1961). *An investigation of the effects of the time lag due to long transmission distances upon remote control. phase i. tracking experiments*. Tech. rep., NATIONAL AERONAUTICS AND SPACE ADMINISTRATION WASHINGTON DC

Al-Obaidi, M., Mustafa, M., Hasan, W. Z. W., Azis, N. B., Sabry, A., Ang, S., et al. (2019). Efficient charging pad for unmanned aerial vehicle based on direct contact. In *2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)* (IEEE), 1–5

Alexander, A. D. (1972). Impacts of telemation on modern society. In *On Theory and Practice of Robots and Manipulators* (Springer). 121–136

Alexander III, A. (1972). Survey study of teleoperators. *Robotics, and Remote Systems Technology, Remotely Manned Systems Exploration and Operation in Space* , 449–458

Alexander III, A. D. (1973). A survey study of teleoperators, robotics, and remote systems technology.

Ambrus, R., Claici, S., and Wendt, A. (2017). Automatic room segmentation from unstructured 3-d data of indoor environments. *IEEE Robotics and Automation Letters* 2, 749–756

Angeli, A., Doncieux, S., Meyer, J.-A., and Filliat, D. (2008). Incremental vision-based topological slam. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on* (Ieee), 1031–1036

Atherton, J. A. and Goodrich, M. A. (2009). Supporting remote manipulation with an ecological augmented virtuality interface. In *Proc. of the AISB Symposium on New Frontiers in Human-Robot Interaction, Edinburgh, Scotland*. 381–394

Baudisch, P., Good, N., Bellotti, V., and Schraedley, P. (2002). Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (ACM), 259–266

Bhandari, S., Viska, S., Shah, H., Chen, C., Tonini, G., and Kline, S. (2015). Autonomous navigation of a quadrotor in indoor environments for surveillance and reconnaissance. In *AIAA Infotech@ Aerospace* (AIAA SciTech Forum). 0717

Bormann, R., Jordan, F., Li, W., Hampp, J., and Hägele, M. (2016). Room segmentation: Survey, implementation, and analysis. In *Robotics and Automation (ICRA), 2016 IEEE International Conference on* (IEEE), 1019–1026

Bruemmer, D. J., Dudenhoeffer, D. D., and Marble, J. L. (2002). Dynamic-autonomy for urban search and rescue. In *AAAI mobile robot competition*. 33–37

Bruemmer, D. J., Few, D. A., Boring, R. L., Marble, J. L., Walton, M. C., and Nielsen, C. W. (2005). Shared understanding for collaborative control. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 35, 494–504. doi:10.1109/TSMCA.2005.850599

Burke, J. L. and Murphy, R. R. (2004). Situation awareness and task performance in robot-assisted technical search: Bujold goes to bridgeport. *citeseer. ist. psu. edu/burke04situation. html*

Cai, G., Dias, J., and Seneviratne, L. (2014). A survey of small-scale unmanned aerial vehicles: Recent advances and future development trends. *Unmanned Systems* 2, 175–199

Chellali, R. and Baizid, K. (2011). What maps and what displays for remote situation awareness and rov localization? In *Symposium on Human Interface* (Springer), 364–372

Chen, T., Ciocarlie, M., Cousins, S., Grice, P. M., Hawkins, K., Hsiao, K., et al. (2013). Robots for humanity: A case study in assistive mobile manipulation. *IEEE Robotics & Automation Magazine, Special issue on Assistive Robotics* 20

Cho, K., Cho, M., and Jeon, J. (2017). Fly a drone safely: Evaluation of an embodied egocentric drone controller interface. *Interacting with computers* 29, 345–354

Choset, H. and Nagatani, K. (2001). Topological simultaneous localization and mapping (slam): toward exact localization without explicit localization. *IEEE Transactions on robotics and automation* 17, 125–137

Corliss, W. and Johnson, E. (1967). Teleoperators and human augmentation. an aec-nasa technology survey

Cortellessa, G., Fracasso, F., Sorrentino, A., Orlandini, A., Bernardi, G., Coraci, L., et al. (2018). Robin, a telepresence robot to support older users monitoring and social inclusion: development and evaluation. *Telemedicine and e-Health* 24, 145–154

Costa, D., Palmieri, G., Palpacelli, M.-C., Panebianco, L., and Scaradozzi, D. (2018). Design of a bio-inspired autonomous underwater robot. *Journal of Intelligent & Robotic Systems* 91, 181–192

Council, N. R. et al. (1998). *The future of air traffic control: Human operators and automation* (National Academies Press)

Cui, J., Tosunoglu, S., Roberts, R., Moore, C., and Repperger, D. W. (2003). A review of teleoperation system control. In *Proceedings of the Florida Conference on Recent Advances in Robotics* (Florida Atlantic University Boca Raton, FL), 1–12

Delmerico, J. A., Baran, D., David, P., Ryde, J., and Corso, J. J. (2013). Ascending stairway modeling from dense depth imagery for traversability analysis. In *2013 IEEE International Conference on Robotics and Automation* (IEEE), 2283–2290

Domahidi, A. and Jerez, J. (2013). Forces pro: code generation for embedded optimization. Homepage. Retrieved November 1, 2018 from `https://www.embotech.com/FORCES-Pro`

Duan, T., Punpongsanon, P., Iwai, D., and Sato, K. (2018). Flyinghand: extending the range of haptic feedback on virtual hand using drone-based object recognition. In *SIGGRAPH Asia 2018 Technical Briefs* (ACM), 28

Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics* 42, 462–492

Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research. *Human factors* 59, 5–27

Endsley, M. R. (2018). Level of automation forms a key aspect of autonomy design. *Journal of Cognitive Engineering and Decision Making* 12, 29–34

Erat, O., Isop, W. A., Kalkofen, D., and Schmalstieg, D. (2018). Drone-augmented human vision: Exocentric control for drones exploring hidden areas. *IEEE transactions on visualization and computer graphics* 24, 1437–1446

Espingardeiro, A. (2012). Human performance in telerobotics operations. In *Advanced Materials Research* (Trans Tech Publ), vol. 403, 772–779

Falanga, D., Kleber, K., Mintchev, S., Floreano, D., and Scaramuzza, D. (2019). The foldable drone: A morphing quadrotor that can squeeze and fly. *IEEE Robotics and Automation Letters* 4, 209–216

Falanga, D., Mueggler, E., Faessler, M., and Scaramuzza, D. (2017). Aggressive quadrotor flight through narrow gaps with onboard sensing and computing using active vision. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE), 5774–5781

Ferland, F., Pomerleau, F., Le Dinh, C. T., and Michaud, F. (2009). Egocentric and exocentric teleoperation interface using real-time, 3d video projection. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (ACM), 37–44

Ferrell, W. R. (1965). Remote manipulation with transmission delay. *IEEE Transactions on Human Factors in Electronics* , 24–32

Fong, T. and Thorpe, C. (2001). Vehicle teleoperation interfaces. *Autonomous robots* 11, 9–18

Fong, T., Thorpe, C., and Baur, C. (2003). Multi-robot remote driving with collaborative control. *IEEE Transactions on Industrial Electronics* 50, 699–704

Frohm, J., Lindström, V., Winroth, M., and Stahre, J. (2008). Levels of automation in manufacturing. *Ergonomia*

Gebhardt, C., Hepp, B., Nägeli, T., Stevšić, S., and Hilliges, O. (2016). Airways: Optimization-based planning of quadrotor trajectories according to high-level user goals. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (ACM), 2508–2519

Giernacki, W., Skwierczyński, M., Witwicki, W., Wroński, P., and Kozierski, P. (2017). Crazyflie 2.0 quadrotor as a platform for research and education in robotics and control engineering. In *2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)* (IEEE), 37–42

Goodrich, M. A., McLain, T. W., Anderson, J. D., Sun, J., and Crandall, J. W. (2007). Managing autonomy in robot teams: observations from four experiments. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction* (ACM), 25–32

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Human Mental Workload*, eds. P. A. Hancock and N. Meshkati (North-Holland), vol. 52 of *Advances in Psychology*. 139 – 183. doi:https://doi.org/10.1016/S0166-4115(08)62386-9

Hedayati, H., Walker, M., and Szafir, D. (2018). Improving collocated robot teleoperation with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (ACM), 78–86

Held, R. and Durlach, N. (1991). Telepresence, time delay and adaptation. *Pictorial communication in virtual and real environments* , 232–246

Held, R., Efstathiou, A., and Greene, M. (1966). Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology* 72, 887

HELSINKI, W. D. O. (2013). Ethical principles for medical research involving human subjects. Homepage. Retrieved November 1, 2018 from `https://www.wma.net/policy/current-policies/`

Henry, P., Krainin, M., Herbst, E., Ren, X., and Fox, D. (2012). Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *The International Journal of Robotics Research* 31, 647–663

Herrmann, R. and Schmidt, L. (2018). Design and evaluation of a natural user interface for piloting an unmanned aerial vehicle. *i-com* 17, 15–24

Hornung, A., Wurm, K. M., Bennewitz, M., Stachniss, C., and Burgard, W. (2013). Octomap: An efficient probabilistic 3d mapping framework based on octrees. *Auton. Robots* 34, 189–206. doi:10.1007/s10514-012-9321-0

Huang, A. S., Tellex, S., Bachrach, A., Kollar, T., Roy, D., and Roy, N. (2010). Natural language command of an autonomous micro-air vehicle. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE), 2663–2669

Huang, B., Bayazit, D., Ullman, D., Gopalan, N., and Tellex, S. (2019). Flight, camera, action! using natural language and mixed reality to control a drone (ICRA)

Huang, H.-M., Pavek, K., Albus, J., and Messina, E. (2005a). Autonomy levels for unmanned systems (alfus) framework: An update. In *Unmanned Ground Vehicle Technology VII* (International Society for Optics and Photonics), vol. 5804, 439–449

Huang, H.-M., Pavek, K., Novak, B., Albus, J., and Messin, E. (2005b). A framework for autonomy levels for unmanned systems (alfus). *Proceedings of the AUVSI's Unmanned Systems North America* , 849–863

Jha, A. (2016). Unmanned aerial vehicles for military applications. In *Theory, Design, and Applications of Unmanned Aerial Vehicles* (CRC Press). 69–124

Johnson, C. (2018). Topological mapping and navigation in real-world environments

Johnson, M. and Vera, A. (2019). No ai is an island: The case for teaming intelligence. *AI Magazine* 40, 16–28

Junaid, A. B., Lee, Y., and Kim, Y. (2016). Design and implementation of autonomous wireless charging station for rotary-wing uavs. *Aerospace Science and Technology* 54, 253–266

Kavraki, L. E. and Latombe, J.-C. (1998). Probabilistic roadmaps for robot path planning

Kortenkamp, D. M. (1993). *Cognitive maps for mobile robots: A representation for mapping and navigation.* Ph.D. thesis, TRACLabs Inc.

Kun, D. P., Varga, E. B., and Toth, Z. (2017). Ontology based navigation model of the ilona system. In *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)* (IEEE), 000479–000484

Kushleyev, A., Mellinger, D., Powers, C., and Kumar, V. (2013). Towards a swarm of agile micro quadrotors. *Autonomous Robots* 35, 287–300

Kyristsis, S., Antonopoulos, A., Chanialakis, T., Stefanakis, E., Linardos, C., Tripolitsiotis, A., et al. (2016). Towards autonomous modular uav missions: The detection, geo-location and landing paradigm. *Sensors* 16, 1844

Labbe, M. and Michaud, F. (2013). Appearance-based loop closure detection for online large-scale and long-term operation. *IEEE Transactions on Robotics* 29, 734–745

Labbe, M. and Michaud, F. (2014). Online Global Loop Closure Detection for Large-Scale Multi-Session Graph-Based SLAM. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2661–2666

Lan, Z., Shridhar, M., Hsu, D., and Zhao, S. (2017). Xpose: Reinventing user interaction with flying cameras. In *Robotics: Science and Systems*

Leeper, A. E., Hsiao, K., Ciocarlie, M., Takayama, L., and Gossow, D. (2012). Strategies for human-in-the-loop robotic grasping. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (ACM), HRI '12, 1–8. doi:10.1145/2157689.2157691

Lichiardopol, S. (2007). A survey on teleoperation

Lin, L. and Goodrich, M. A. (2015). Sliding autonomy for uav path-planning: Adding new dimensions to autonomy management. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (International Foundation for Autonomous Agents and Multiagent Systems), 1615–1624

Loianno, G., Brunner, C., McGrath, G., and Kumar, V. (2016). Estimation, control, and planning for aggressive flight with a small quadrotor with a single camera and imu. *IEEE Robotics and Automation Letters* 2, 404–411

MahmoudZadeh, S., Powers, D. M., and Zadeh, R. B. (2018). *Autonomy and Unmanned Vehicles: Augmented Reactive Mission and Motion Planning Architecture* (Springer)

Materna, Z., Španěl, M., Mast, M., Beran, V., Weisshardt, F., Burmester, M., et al. (2017). Teleoperating assistive robots: A novel user interface relying on semi-autonomy and 3d environment mapping. *Journal of Robotics and Mechatronics* 29, 381–394

Monajjemi, M. (2015). Bebop autonomy driver. Homepage. Retrieved November 1, 2018 from `https://github.com/AutonomyLab/bebop_autonomy`

Mostafa, S. A., Ahmad, M. S., and Mustapha, A. (2019). Adjustable autonomy: a systematic literature review. *Artificial Intelligence Review* 51, 149–186

Mur-Artal, R. and Tardós, J. D. (2017). ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *IEEE Transactions on Robotics* 33, 1255–1262. doi:10.1109/TRO.2017.2705103

Murray, R. M. (2017). *A mathematical introduction to robotic manipulation* (CRC press)

Muszynski, S., Stückler, J., and Behnke, S. (2012). Adjustable autonomy for mobile teleoperation of personal service robots. In *RO-MAN, 2012 IEEE* (IEEE), 933–940

Nagatani, K., Kiribayashi, S., Okada, Y., Otake, K., Yoshida, K., Tadokoro, S., et al. (2013). Emergency response to the nuclear accident at the fukushima daiichi nuclear power plants using mobile rescue robots. *Journal of Field Robotics* 30, 44–63

Nägeli, T., Alonso-Mora, J., Domahidi, A., Rus, D., and Hilliges, O. (2017a). Real-time motion planning for aerial videography with dynamic obstacle avoidance and viewpoint optimization. *IEEE Robotics and Automation Letters* 2, 1696–1703. doi:10.1109/LRA.2017.2665693

Nägeli, T., Meier, L., Domahidi, A., Alonso-Mora, J., and Hilliges, O. (2017b). Real-time planning for automated multi-view drone cinematography. *ACM Trans. Graph.* 36, 132:1–132:10. doi:10.1145/3072959.3073712

Nielsen, C. W., Goodrich, M. A., and Ricks, R. W. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics* 23, 927–941

Olson, E. (2011). Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*. 3400–3407. doi:10.1109/ICRA.2011.5979561

Onnasch, L., Wickens, C. D., Li, H., and Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human factors* 56, 476–488

Optitrack (2019). Flex 13. Homepage. Retrieved March 1, 2019 from `www.optitrack.com/products/flex-13/`

Papachristos, C., Khattak, S., and Alexis, K. (2017). Uncertainty-aware receding horizon exploration and mapping using aerial robots. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE), 4568–4575

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30, 286–297

Parrot (2015). Parrot bebop 2. Homepage. Retrieved November 1, 2018 from `https://www.parrot.com/us/drones/parrot-bebop-2`

Paterson, J., Han, J., Cheng, T., Laker, P., McPherson, D., Menke, J., et al. (2019). Improving usability, efficiency, and safety of uav path planning through a virtual reality interface. *arXiv preprint arXiv:1904.08593*

Patra, R. K., Nedevschi, S., Surana, S., Sheth, A., Subramanian, L., and Brewer, E. A. (2007). Wildnet: Design and implementation of high performance wifi based long distance networks. In *NSDI*. vol. 1, 1

Rakita, D., Mutlu, B., and Gleicher, M. (2018). An autonomous dynamic camera method for effective remote teleoperation. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (ACM), 325–333

Razor (2015). Orbweaver. Homepage. Retrieved March 1, 2019 from `https://www.razer.com/gaming-keyboards-keypads/razer-orbweaver-chroma`

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788

Reese, H. (2016). Autonomous driving levels 0 to 5: Understanding the differences. Homepage. Retrieved November 1, 2018 from `https://www.techrepublic.com/article/autonomous-driving-levels-0-to-5-understanding-the-differences/`

Richtsfeld, A., Morwald, T., Prankl, J., Zillich, M., and Vincze, M. (2012). Segmentation of unknown objects in indoor environments. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 4791–4796. doi:10.1109/IROS.2012.6385661

Riestock, M., Engelhardt, F., Zug, S., and Hochgeschwender, N. (2017a). Exploring gridmap-based interfaces for the remote control of uavs under bandwidth limitations. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (ACM), 263–264

Riestock, M., Engelhardt, F., Zug, S., and Hochgeschwender, N. (2017b). User study on remotely controlled uavs with focus on interfaces and data link quality. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE), 3394–3400

Robotics, A. (2019). Evo drone. Homepage. Retrieved March 1, 2019 from `https://auteldrones.com/products/evo`

Rognon, C., Mintchev, S., Dell'Agnola, F., Cherpillod, A., Atienza, D., and Floreano, D. (2018). Flyjacket: An upper body soft exoskeleton for immersive drone control. *IEEE Robotics and Automation Letters* 3, 2362–2369

Saakes, D., Choudhary, V., Sakamoto, D., Inami, M., and Lgarashi, T. (2013). A teleoperating interface for ground vehicles using autonomous flying cameras. In *2013 23rd International Conference on Artificial Reality and Telexistence (ICAT)* (IEEE), 13–19

SAE (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Homepage. Retrieved November 1, 2018 from `https://saemobilus.sae.org/content/j3016_201401`

Sanders, B., Vincenzi, D., Holley, S., and Shen, Y. (2018). Traditional vs gesture based uav control. In *International Conference on Applied Human Factors and Ergonomics* (Springer), 15–23

Sanket, N. J., Singh, C. D., Ganguly, K., Fermüller, C., and Aloimonos, Y. (2018). Gapflyt: Active vision based minimalist structure-less gap detection for quadrotor flight. *IEEE Robotics and Automation Letters* 3, 2799–2806

Save, L., Feuerberg, B., and Avia, E. (2012). Designing human-automation interaction: a new level of automation taxonomy. *Proc. Human Factors of Systems and Technology* 2012

Scalea, J. R., Restaino, S., Scassero, M., Bartlett, S. T., and Wereley, N. (2019). The final frontier? exploring organ transportation by drone. *American Journal of Transplantation* 19, 962–964

Schröter, D., Beetz, M., and Radig, B. (2003). Rg mapping: Building object-oriented representations of structured human environments. In *6th Open Russian–German Workshop on Pattern Recognition and Image Understanding (OGRW)*

Seo, S. H., Rea, D. J., Wiebe, J., and Young, J. E. (2017). Monocle: Interactive detail-in-context using two pan-and-tilt cameras to improve teleoperation effectiveness. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE), 962–967

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control* (MIT press)

Sheridan, T. B. (1993). Space teleoperation through time delay: Review and prognosis. *IEEE Transactions on robotics and Automation* 9, 592–606

Sheridan, T. B. and Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Tech. rep., Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab

Silvagni, M., Tonoli, A., Zenerino, E., and Chiaberge, M. (2019). Multipurpose uav for search and rescue operations in mountain avalanche events mario silvagni1, andrea tonoli1, enrico zenerino1, marcello chiaberge2

Stepanova, E. R., von der Heyde, M., Kitson, A., Schiphorst, T., and Riecke, B. E. (2017). Gathering and applying guidelines for mobile robot design for urban search and rescue application. In *International Conference on Human-Computer Interaction* (Springer), 562–581

Stubbs, K., Hinds, P. J., and Wettergreen, D. (2007). Autonomy and common ground in human-robot interaction: A field study. *IEEE Intelligent Systems* 22

Szafir, D., Mutlu, B., and Fong, T. (2017). Designing planning and control interfaces to support user collaboration with flying robots. *The International Journal of Robotics Research* 36, 514–542

Technology, A. (2014). Bomb disposal robots - evolution and revolution. Homepage. Retrieved March 1, 2019 from `https://www.army-technology.com/features/featurebomb-disposal-robots-evolution-and-revolution/`

Thomason, J., Ratsamee, P., Kiyokawa, K., Kriangkomol, P., Orlosky, J., Mashita, T., et al. (2017). Adaptive view management for drone teleoperation in complex 3d structures. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (ACM), 419–426

Thomason, J., Ratsamee, P., Orlosky, J., Kiyokawa, K., Mashita, T., Uranishi, Y., et al. (2019). A comparison of adaptive view techniques for exploratory 3d drone teleoperation. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9, 17

Tripicchio, P., Satler, M., Dabisias, G., Ruffaldi, E., and Avizzano, C. A. (2015). Towards smart farming and sustainable agriculture with drones. In *2015 International Conference on Intelligent Environments* (IEEE), 140–143

Umari, H. and Mukhopadhyay, S. (2017). Autonomous robotic exploration based on multiple rapidly-exploring randomized trees. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 1396–1402. doi:10.1109/IROS.2017.8202319

Valero-Gomez, A., De La Puente, P., and Hernando, M. (2011). Impact of two adjustable-autonomy models on the scalability of single-human/multiple-robot teams for exploration missions. *Human factors* 53, 703–716

Valner, R., Kruusamäe, K., and Pryor, M. (2018). Temoto: Intuitive multi-range telerobotic system with natural gestural and verbal instruction interface. *Robotics* 7, 9

Vasudevan, S., Gächter, S., Nguyen, V., and Siegwart, R. (2007). Cognitive maps for mobile robots-an object based approach. *Robot. Auton. Syst.* 55, 359–371. doi:10.1016/j.robot.2006.12.008

Vertut, J. and Coiffet, P. (1986). Teleoperations and robotics: Evolution and development,"robot technology," vol. 3a

Walker, M. E., Hedayati, H., and Szafir, D. (2019). Robot teleoperation with augmented reality virtual surrogates. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (IEEE), 202–210

Wang, L., Zhao, L., Huo, G., Li, R., Hou, Z., Luo, P., et al. (2018). Visual semantic navigation based on deep learning for indoor mobile robots. *Complexity* 2018

Weiss, S., Scaramuzza, D., and Siegwart, R. (2011). Monocular-slam–based navigation for autonomous micro helicopters in gps-denied environments. *Journal of Field Robotics* 28, 854–874

Wu, Y., Du, X., Duivenvoorden, R., and Kelly, J. (2018). Hummingbird: An open-source dual-rotor tail-sitter platform for research and education. *arXiv preprint arXiv:1810.03196*

Yanco, H. A., Drury, J. L., and Scholtz, J. (2004). Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. *Human-Computer Interaction* 19, 117–149

Yang, L. and Worboys, M. (2015). Generation of navigation graphs for indoor space. *International Journal of Geographical Information Science* 29, 1737–1756

Yu, M., Lin, Y., Schmidt, D., Wang, X., and Wang, Y. (2014). Human-robot interaction based on gaze gestures for the drone teleoperation. *Journal of Eye Movement Research* 7, 1–14

Yu, Y., Wang, X., Zhong, Z., and Zhang, Y. (2017). Ros-based uav control using hand gesture recognition. In *2017 29th Chinese Control And Decision Conference (CCDC)* (IEEE), 6795–6799

Yuan, L., Reardon, C., Warnell, G., and Loianno, G. (2019). Human gaze-driven spatial tasking of an autonomous mav. *IEEE Robotics and Automation Letters* 4, 1343–1350

Zhang, Q., Zhao, W., Chu, S., Wang, L., Fu, J., Yang, J., et al. (2018). Research progress of nuclear emergency response robot. In *IOP Conference Series: Materials Science and Engineering* (IOP Publishing), vol. 452, 042102