Design and Evaluation of a Mixed Reality-based Human-Robot Interface for Teleoperation of Omnidirectional Aerial Vehicles

Mike Allenspach, Till Kötter, Rik Bähnemann, Marco Tognon, Roland Siegwart

Preprint version, final version at http://ieeexplore.ieee.org/ 2023 International Conference on Unmanned Aircraft Systems (ICUAS)

Design and Evaluation of a Mixed Reality-based Human-Robot Interface for Teleoperation of Omnidirectional Aerial Vehicles

Mike Allenspach[1], Till Kötter[1], Rik Bähnemann[1], Marco Tognon[2], and Roland Siegwart[1]

***Abstract*— Omnidirectional aerial vehicles are an attractive solution for visual inspection tasks that require observations from different views. However, the decisional autonomy of modern robots is limited. Therefore, human input is often necessary to safely explore complex industrial environments. Existing teleoperation tools rely on on-board camera views or 3D renderings of the environment to improve situational awareness. Mixed-Reality (MR) offers an exciting alternative, allowing the user to perceive and control the robot's motion in the physical world. Furthermore, since MR technology is not limited by the hardware constraints of standard teleoperation interfaces, like haptic devices or joysticks, it allows us to explore new reference generation and user feedback methodologies. In this work, we investigate the potential of MR in teleoperating 6DoF aerial robots by designing a holographic user interface (see Fig. 1) to control their translational velocity and orientation. A user study with 13 participants is performed to assess the proposed approach. The evaluation confirms the effectiveness and intuitiveness of our methodology, independent of prior user experience with aerial vehicles or MR. However, prior familiarity with MR improves task completion time. The results also highlight limitation to line-of-sight operation at distances where relevant details in the physical environment can still be visually distinguished.**

I. Introduction

Regular visual infrastructure inspection remains crucial in many commercial and industrial settings. Robotic systems have become increasingly common, offering to enhance the capabilities of skilled human workers in complex environments or replace them in dangerous or laborious situations [1]. Micro aerial vehicles (MAVs) have gained particular interest since their mobility is not restricted by the available infrastructure, such as walkways, and thus provide more flexibility compared to ground mobile robots [2], [3].

However, standard underactuated rotary-wing flying systems have limited maneuverability. This restriction is particularly constraining for inspections that require observations from different angles and viewpoints or even contacts. In this case, omnidirectional micro aerial vehicles (OMAVs) offer an intriguing solution by being able to hover at arbitrary orientations, thanks to omnidirectional thrust generation [4], [5]. Despite recent advances in the autonomous control of these vehicles [6]–[8], the decisional autonomy of modern robots is still limited. Therefore, safety considerations and existing regulations often dictate a human operator in the loop, especially when performing inspection tasks in uncertain or initially unknown environments. Such real-time inclusion of human knowledge, decision-making, and supervision skills requires the careful design of teleoperation interfaces. Regarding omnidirectional micro aerial vehicles, this entails decoupled 6 degrees of freedom (DoF) motion control and appropriate user feedback to deduce the vehicle's full $\mathsf{SE}(3)$[1] pose in space. Inspired by related works, we propose a Mixed-Reality (MR)-based user interface (see Fig. 1) to address these challenges.

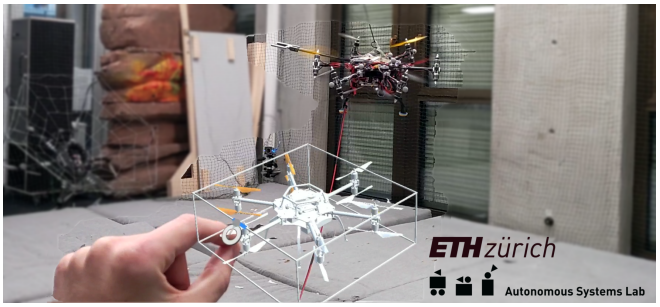

Fig. 1: Operator view when interacting with the developed MR-based holographic teleoperation interface (foreground) to control the 6 DoF motion of an OMAV (background) and receive feedback of its attitude.

A. Related Work

Most MAV teleoperation interfaces rely on dedicated hardware, taking one of two forms: i) haptic devices or generic joysticks with first-person view visual feedback, allowing the human to control the motion of the robot manually [9], [10]; ii) higher-level supervisory systems to specify path and trajectory waypoints in third-person perspective on a virtual map [11]. However, previous work has already concluded that extending these standard methodologies to OMAVs is not straightforward [12]. Given the large number of controllable DoF, the operator is physically unable to decouple them and command motion along a single axis only. Although haptic feedback can support the user in preventing unintended input, online intention detection and careful control stability treatment are necessary [13].

Aside from being able to take full control authority of an OMAV's motion, effective teleoperation requires the operator to maintain spatial awareness of the robot's surroundings [14], [15]. Here, visual feedback from an on-board camera is often inconclusive regarding general location

[1]Autonomous Systems Lab, ETH Zürich, Zürich 8092, Switzerland amike@ethz.ch, tkoetter@ethz.ch, brik@ethz.ch, mtognon@ethz.ch, rsiegwart@ethz.ch

[2]Inria, Univ Rennes, CNRS, IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France. marco.tognon@inria.fr

This research was in part supported by Microsoft Swiss Joint Research Center and in part by the EU H2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant 953454, project AERO-TRAIN.

[1]$\mathsf{SE}(3)$ denotes the special Euclidean group of direct isometries.

in space and especially the robot's orientation due to low visibility and poor image quality [16]. Virtual maps allow only a simplified representation of the environment, are restricted to known objects at known locations and are thus highly susceptible to changes in the surroundings [10], [17]. Additionally, user feedback is provided via 2D displays in both cases, which are inherently unable to represent 3D space [18]. As such, understanding localization information on a global scale for full $\mathsf{SE}(3)$ pose situational awareness proves difficult. In summary, standard hardware- and 2D display-based interfaces are generally unsuitable for teleoperating OMAVs.

Alternatively, recent investigations have started exploiting MR as a tool to control underactuated MAVs in the context of the physical environment itself [19], [20]. Studies have already confirmed the potential of MR technology over standard visual feedback, demonstrating visualization of the commanded position [21], the camera feed [22], or the robot motion intent [23] for improved operator situational awareness. However, the cited works still relied on physical joysticks for piloting the MAVs. MR can completely replace interfaces based on specialized hardware for a more immersive and lightweight setup. Exemplary applications have been developed for sending high-level commands to the robots by placing waypoints [18], [24] or drawing the desired path [25] directly in the physical world. Similarly, [26] demonstrates hand gesture-based manual control of aerial vehicles using only MR interfaces, although an in-depth evaluation through user studies is missing.

The literature presented thus far is focused exclusively on underactuated platforms and is naturally limited to controlling the position and yaw angle of the vehicle only. Nevertheless, the results suggest that replacing traditional hardware- and 2D display-based interfaces with virtual MR ones is a promising approach for teleoperating OMAVs as well. A holographic interface can be of any desired shape at an arbitrary position and orientation in space without being subject to physical constraints or requiring special safety considerations. The emerging design freedom allows exploring new approaches for enhanced situational awareness and practical motion control. In summary, MR offers a compelling solution for the teleoperation of omnidirectional aerial robots in an immersive and environment-independent setup. However, the complex $\mathsf{SE}(3)$ motion dynamics of OMAVs require special design consideration even for standard teleoperation methods [12]. Therefore, the question arises whether similar conditions also apply to MR interfaces. We provide a corresponding implementation, thorough evaluation, and general discussion about the technology's suitability.

B. Contributions

Aiming to extend MR technology for the teleoperation of OMAVs, the main contributions of this work are:

- The design of an MR interface for 6 DoF manual control of an OMAV's motion in $\mathsf{SE}(3)$;
- Evaluation of the proposed application's intuitiveness by assessing the effect of user experience on their ability to infer and control the pose of an OMAV;
- Discussion about limitations of the presented approach and MR technology in general, namely being constrained to line-of-sight operation.

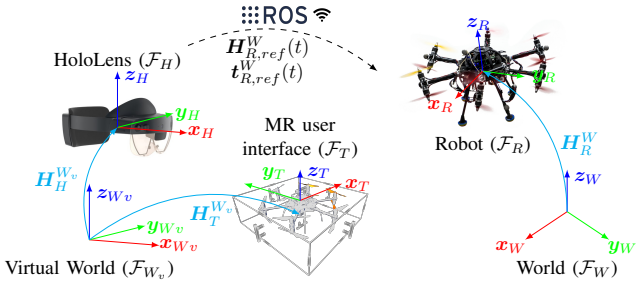

Fig. 2: System setup: dataflow (top) and coordinate system convention (bottom).

As such, this study serves as a first step towards MR-based remote control of omnidirectional aerial systems by developing and evaluating a corresponding human-robot interface, as well as identifying potential issues when applying MR technology to flying robots.

II. System

The MR-based teleoperation system should be as compact and lightweight as possible for convenient use in generic environments. In this case, the proposed setup consists of an *omnidirectional aerial robot* to perform the visual inspection task and an *MR interface* for human interaction. The OMAV presented in [4], and the HoloLens 2 MR headset by Microsoft[2] are considered for the remainder of the paper. However, the developed methodology is generally applicable. Data is transmitted from the HoloLens to the robot through wireless ROS-based communication, as shown in Fig. 2.

The following coordinate frames capture the configuration of the robot and MR system components:

$\mathcal{F}_W$	inertial frame located at arbitrary origin point;
$\mathcal{F}_R$	body-fixed frame rigidly attached to OMAV;
$\mathcal{F}_{W_v}$	inertial frame initialized at HoloLens startup pose;
$\mathcal{F}_H$	body-fixed frame rigidly attached to HoloLens;
$\mathcal{F}_T$	body-fixed frame rigidly attached to the interactable holographic MR user interface (see Sec. III-B).

The x-, y- and z-axis of an arbitrary frame $\mathcal{F}_\star$ are denoted as $\boldsymbol{x}_\star, \boldsymbol{y}_\star, \boldsymbol{z}_\star \in \mathbb{R}^3$, respectively. Given the position $\boldsymbol{p}^A_B \in \mathbb{R}^3$ and attitude $\boldsymbol{R}^A_B \in \mathsf{SO}(3)$[3] of a generic frame $\mathcal{F}_B$ with respect to another frame $\mathcal{F}_A$, we define the corresponding translational and angular velocities as $\boldsymbol{v}^A_B \in \mathbb{R}^3$ and $\boldsymbol{\omega}^A_B \in \mathbb{R}^3$, respectively. The full pose $\boldsymbol{H}^A_B \in \mathsf{SE}(3)$ and twist $\boldsymbol{t}^A_B \in \mathbb{R}^6$ are then expressed as:

$$\boldsymbol{H}^A_B = \begin{pmatrix} \boldsymbol{R}^A_B & \boldsymbol{p}^A_B \\ \boldsymbol{0} & 1 \end{pmatrix} \qquad \boldsymbol{t}^A_B = \begin{pmatrix} \boldsymbol{v}^A_B \\ \boldsymbol{\omega}^A_B \end{pmatrix}. \tag{1}$$

[2] https://www.microsoft.com/en-us/hololens/

[3] $\mathsf{SO}(3)$ denotes the 3D rotation group.

A. OMAV

The omnidirectional aerial robot considered in this work is the platform shown in Fig. 2. Independently tiltable propeller groups at the end of each arm enable omnidirectional force-torque generation and, thus, hovering at arbitrary orientations. The impedance control law presented in [6] is used to render the closed-loop dynamics as the one of a mass-spring-damper system

$$\boldsymbol{M}\dot{\tilde{\boldsymbol{v}}}_R^W + \boldsymbol{D}(\boldsymbol{t}_R^W - \boldsymbol{t}_{R,ref}^W) + \boldsymbol{K}\boldsymbol{e}(\boldsymbol{H}_R^W, \boldsymbol{H}_{R,ref}^W) = \boldsymbol{0}. \quad (2)$$

Hereby, $\boldsymbol{M}, \boldsymbol{D}, \boldsymbol{K} \in \mathbb{R}^{6\times6}$ are the virtual inertia, damping and stiffness tuning parameters of the controller, respectively. The error function

$$\boldsymbol{e}(\boldsymbol{H}_R^W, \boldsymbol{H}_{R,ref}^W) = \begin{pmatrix} \boldsymbol{p}_R^W - \boldsymbol{p}_{R,ref}^W \\ \frac{1}{2}\left(\boldsymbol{R}_{R,ref}^{W\top}\boldsymbol{R}_R^W - \boldsymbol{R}_R^{W\top}\boldsymbol{R}_{R,ref}^W\right)^\vee \end{pmatrix}, \quad (3)$$

is based on [27] to ensure globally asymptotic tracking on $\mathsf{SE}(3)$. The *vee*-map $(\cdot)^\vee : \mathfrak{so}(3) \to \mathbb{R}^3$ is the inverse of the skew-symmetric operator $[\cdot]_\times : \mathbb{R}^3 \to \mathfrak{so}(3)$.

State estimation is ensured by an on-board EKF, fusing IMU data (accelerometer and gyroscope) with external position measurements (e.g., motion capture system, GNSS, VIO, Leica total station) to obtain the current vehicle pose $\boldsymbol{H}_R^W \in \mathsf{SE}(3)$ and twist $\boldsymbol{t}_R^W \in \mathbb{R}^6$. The respective reference values $\star_{ref}$ are directly generated by the user according to the methodology described in Sec. III.

Frame Initialization: Without loss of generality, we place the inertial frame $\mathcal{F}_W$ at the robot's take-off position. Therefore $\boldsymbol{H}_R^W(t=0) = \boldsymbol{I}$, $\boldsymbol{t}_R^W(t=0) = \boldsymbol{0}$ at the beginning of an inspection mission. The initialization simplifies the HoloLens rotation alignment.

B. HoloLens 2

Unlike Virtual-Reality where the entire surroundings need to be digitally replicated, MR technology superimposes holograms on the real world. Aside from holographic rendering of various objects, the HoloLens 2 headset also supports fully articulated hand tracking for real time interaction and manipulation of the MR environment, as well as spatial mapping for headset pose estimation. App development is conducted in Unity, using the Mixed Reality Toolkit[4] for accelerated development and the ROS-TCP-Endpoint library[5] for TCP-based communication between the application and ROS server. During teleoperation, the user commands the current translational velocity and attitude of the omnidirectional vehicle, thereby generating a trajectory of desired poses $\boldsymbol{H}_{R,ref}^W(t) \in \mathsf{SE}(3)$ and twists $\boldsymbol{t}_{R,ref}^W \in \mathbb{R}^6$ to be tracked by the flight controller.

Frame Initialization: Note that all MR holograms are rendered with respect to the inertial virtual world frame of the HoloLens $\mathcal{F}_{W_v}$, initialized at its startup pose. Whenever spatial information is shared between multiple devices in MR applications, $\mathcal{F}_{W_v}$ must be properly calibrated and localized with respect to $\mathcal{F}_W$ [15]. In the presented application, we specifically require $\boldsymbol{R}_W^{W_v} = \boldsymbol{I}$. Under the stated assumption $\boldsymbol{H}_R^W(t=0) = \boldsymbol{I}$, sufficient accuracy ($\approx \pm10°$) is achieved by simply starting the app while standing behind the robot (along negative $\boldsymbol{x}_R$) and mirroring its heading ($\boldsymbol{x}_H$ aligned with $\boldsymbol{x}_R$). Alternatively, more advanced co-localization methods could also be used, such as Azure Spatial Anchors [15] or point cloud matching [24], [28]. However, the presented calibration process not only improves ease of use but also makes the setup independent of the available infrastructure and sensor suite of the robot.

[4] https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05

[5] https://github.com/Unity-Technologies/ROS-TCP-Endpoint

III. Teleoperation

The developed teleoperation framework has two main parts, namely i) a *reference generation* method for commanding full $\mathsf{SE}(3)$ pose trajectories; ii) an MR *user interface* that ensures intuitive interaction and situational awareness.

A. Reference Generation

Reference generation addresses the issue of converting the user's manipulation of the MR interface (see Sec. III-B) into motion commands for the actual system. The goal is an intuitive interface where the user does not have to understand the physics of the drone to perform meaningful tasks. We derive our methodology from standard teleoperation approaches considering the OMAV's $\mathsf{SE}(3)$ control space topology.

Rate control [10] is utilized to command translational motion on the open vector space $\mathbb{R}^3$. Our method generates translational velocity commands by displacing the hologram from an idle position. The mapping from displacements to velocities allows the user to cover $\mathbb{R}^3$ through small-scale interactions.

The *Leader-Follower* principle [13] is utilized to command rotational motion in $\mathsf{SO}(3)$. Here, the robot directly mirrors the interface configuration, allowing intuitive feedback on the vehicle's attitude. In contrast to translational motion, the complete $\mathsf{SO}(3)$ workspace can be displayed due to the closed orientation isometry.

Combining both approaches enables the full exploitation of the OMAV's capabilities without requiring excessive physical or cognitive user effort (see Sec. IV). Formally, translational velocities and attitude references are therefore given by

$$\boldsymbol{v}_{R,ref}^W = \alpha \boldsymbol{R}_{R,ref}^W \boldsymbol{R}_T^{W_v\top}\left(\boldsymbol{p}_T^{W_v} - \boldsymbol{p}_{T,idle}^{W_v}\right) \quad (4)$$

$$\boldsymbol{R}_{R,ref}^W = \boldsymbol{R}_T^{W_v}, \quad (5)$$

where $\alpha \in \mathbb{R}_{>0}$ is a scaling factor and $\boldsymbol{p}_{T,idle}^{W_v} \in \mathbb{R}^3$ is the idle position of the user interface. The full pose $\boldsymbol{H}_{R,ref}^W$ and twist $\boldsymbol{t}_{R,ref}^W$ trajectories are then constructed according to (1) by numerical integration and differentiation

$$\boldsymbol{p}_{R,ref}^W = \int_0^t \boldsymbol{v}_{R,ref}^W(\tau)d\tau \quad (6)$$

$$\boldsymbol{\omega}_{R,ref}^W = \boldsymbol{R}_{R,ref}^W\left(\boldsymbol{R}_{R,ref}^{W\top}\dot{\boldsymbol{R}}_{R,ref}^W\right)^\vee. \quad (7)$$

Note that all values required for reference computation are completely independent of the robot state and readily available on the HoloLens.

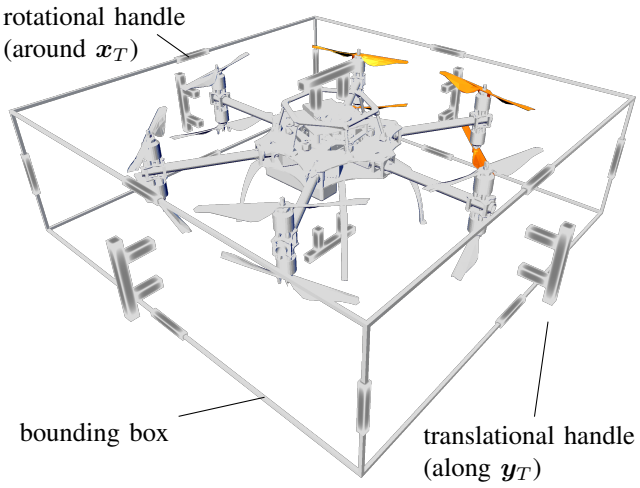

Fig. 3: Holographic user interface with translational and rotational interaction handles on the bounding box's sides and edges, respectively.

B. User Interface

As shown in Fig. 3, the holographic user interface is a model of the physical robot (approximately 1:5 ratio) to enhance the operator's situational awareness. Since the interface should always be reachable while not obstructing the user's vision, it is placed at the bottom of their field-of-view (see Fig. 1). In other words, $\boldsymbol{p}^{H}_{T,idle}$ is kept constant by continuously updating the idle position $\boldsymbol{p}^{W_v}_{T,idle}$ based on the on-board state estimation of the headset.

Intuitively, the attitude of the hologram directly represents the orientation reference for the real OMAV in $\mathcal{F}_W$ according to (5), thanks to the proper calibration of $\mathcal{F}_{W_v}$ described in Sec. II-B. Note that this is essentially identical to the robot's current attitude, as the state will eventually converge to the reference value (see Sec. III-C). Intuition is further improved by commanding translational velocities in the body frame $\mathcal{F}_R$ (translation along $\boldsymbol{x}_T$ causes velocity along $\boldsymbol{x}_R$), following standard practice in remote control and manned rotary-wing vehicle piloting.

The operator can directly grasp and manipulate the appropriate translational or rotational handle on the bounding box to interact with the interface. Here, each handle only changes a single DoF at a time for improved precision and to address the axes decoupling issue mentioned in [12]. Alternatively, one can interact with the center of the bounding box to simultaneously change all 6 DoF for fast and large-scale manipulation. The OMAV tracks changes in the hologram immediately. The live feedback further improves intuitiveness and user experience [29], [30].

The position $\boldsymbol{p}^{W_v}_{T}$ of the user interface is limited to a given radius around its idle state and snaps back as soon as the interface is released. Given the rate control nature of the translational reference generation, this method ensures that the commanded velocity remains bounded at all times and defaults to zero without user interaction.

C. Stability Considerations

Although no formal proof of stability is provided in this paper, some stability-related aspects of the developed teleoperation system are briefly discussed here. Firstly, the presented controller is based on one-way communication where no data is being transmitted from the robot back to the user (see Fig. 2). Therefore, no special considerations regarding latency, package loss, or other imperfections in the communication are required [31]. Secondly, it can be shown that the closed-loop robot dynamics (2) rendered by impedance control remain passive under bounded reference velocities $\boldsymbol{t}^{W}_{R,ref}$ [32]. Given the saturated reference generation in (4) and (5), the robot will thus remain stable at all times and asymptotically converge to a zero-velocity equilibrium in an idle state. Finally, flight experiments and simulations conducted during development and evaluation have verified the practical stability of the presented teleoperation policy.

IV. Results

To evaluate the efficacy of the proposed approach, we conducted a human subject study, where the subjects controlled an OMAV through the developed MR interface. The goal is to: i) assess the intuitiveness of the methodology in terms of the user's ability to infer and control the pose of an OMAV; ii) identify possible limitations of the specific implementation and MR technology in general.

Note that the methodology is incomparable with state-of-the-art, since traditional teleoperation interfaces are ineffective in controlling OMAVs.

A. Setup and Task Definition

The scenario considered in this study is designed for the operator to sequentially perform flight maneuvers of increasing complexity in terms of the required precision, number of DoF to control, and situational awareness. By continuously advancing the difficulty, we aim to meticulously detect failure cases of the developed application and accurately determine their source, e.g., cognitive overload or interface visualization. The task definitions are listed in Tab. I and Fig. 4 annotates their respective location in the simulated scene. Note that any visual or contact inspection mission is a combination of these basic maneuvers. Therefore, the successful completion of all tasks demonstrates the suitability of the developed methodology for generic real-world applications.

For safety and repeatability reasons, the evaluation is conducted in simulation. The environment is constructed in the Gazebo robotics simulator [33], depicting the native gas station world extended with planes according to the specified tasks. The state estimation and control algorithms of the simulated OMAV are implemented in ROS and identical to code running on the physical system, thus rendering realistic flight performance. The operator has an eye-level third-person perspective of the scene, displaying a view similar to real applications when using MR technology. However, the simulated viewpoint is fixed for better comparison among

Task #	Description	Required precision	Required number of DoF	Required situational awareness
1	Take off and land on a horizontal plane	Precise manipulation	1 translational	vehicle altitude
2	Large-scale obstacle parkour	Large-scale manipulation	2 translational	vehicle position
3	Reach and land on a horizontal plane	Precise manipulation	3 translational	vehicle position
4	Reach and land on a single-angle inclined plane	Precise manipulation	3 translational, 1 rotational	full vehicle pose
5	Reach and land on a multi-angle inclined plane	Precise manipulation	3 translational, 3 rotational	full vehicle pose
6	Land on a horizontal plane	Precise manipulation	3 translational, 3 rotational	full vehicle pose

TABLE I: User study tasks with increasing complexity in terms of required precision, number of DoF to control and situational awareness.

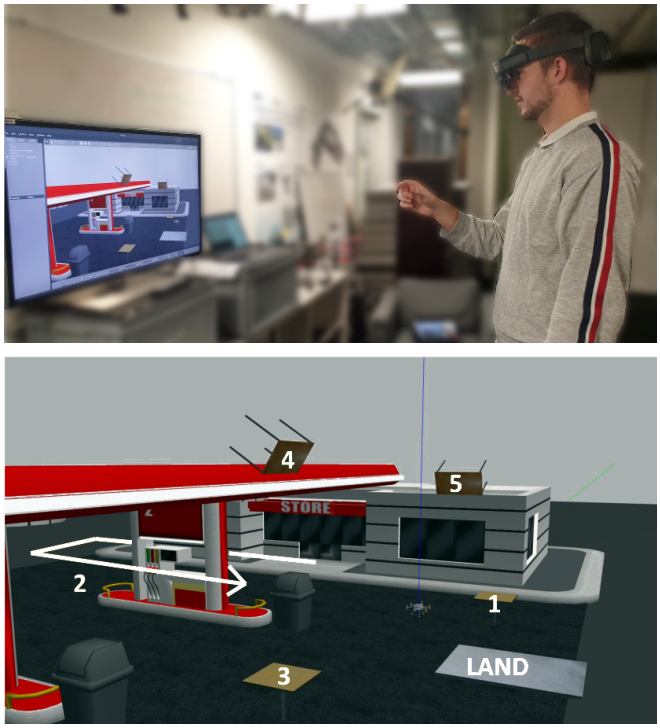

Fig. 4: Experimental setup for human subject study (top) and simulation environment (bottom).

the subjects, while a physical environment could be further explored to enhance spatial perception.

B. Subjects

Thirteen subjects voluntarily took part in the experiments. They reported no eyesight problems or general deficiencies in perceptual and motor abilities that would negatively bias their performance during the study. Two participants had previous experience with MR technology and MAVs, and six had previous experience with aerial robots only. The remaining participants had no previous experience with either topic. The subjects were given a short description of the experimental setup, robot, conditions, and tasks. To not influence the intuitiveness evaluation, explanations about how to interact with the hologram to control the OMAV's motion were omitted. Similarly, no test trial to become acquainted with the robot and user interface was admitted.

C. Evaluation Metrics

Time is a reasonable metric to evaluate the efficacy of the proposed approach. In the case of a non-intuitive interface or inherent methodological limitations, the total required time will naturally increase because the user will spend time figuring out how to perform each task, leading to excessive trial and error. We chose total mission time and normalized time per task to evaluate and compare the subject's performance. The results are split according to previous user experience with MR and MAVs, MAVs only or neither. Figure 5 shows the statistics related to the total time T as the sum of all tasks $T_i \in \mathbb{R}_{>0}$, i.e., $T = \sum_{i=1}^{6} T_i \in \mathbb{R}_{>0}$. Figure 6 displays the normalized time per task, i.e., each task's ratio of the total time $T_i/T \in \mathbb{R}_{>0},\ i \in 1, ..., 6$.

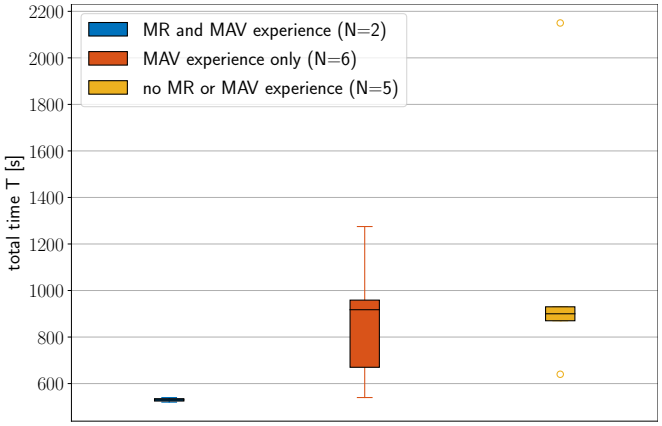

Fig. 5: Total completion time statistics, grouped according to previous user experience. Median, first quartile and third quartile range are indicated by the black line, box and whiskers, respectively. Values outside these ranges are considered as outliers, indicated by circles.

Aside from the objective time measure, participants reported the main difficulties faced during the experiment. We summarized the responses into three categories, namely: i) lacking *spatial understanding* of the OMAV's position in the simulated environment due to missing depth information and fixed viewpoint; ii) unintended or unrecognized *MR interaction*, due to lacking HoloLens experience; iii) erroneous *OMAV motion*, due to misunderstanding of the reference generation methodology. Figure 7 gives a summary of the replies for each experience group.

D. Discussion

All the subjects completed the entire mission, independent of their previous experience with MR technology or MAVs. This suggests that the proposed methodology is generally suitable for real-world visual or contact inspection missions. A thorough analysis of the timing performance and user feedback further allows introspection into the approach's intuitiveness.

First, familiarity with aerial robots is unnecessary. Figure 5 shows that the median total completion time $med(T)$ of people experienced with MAVs ($med(T) = 900\,\mathrm{s}$) is only

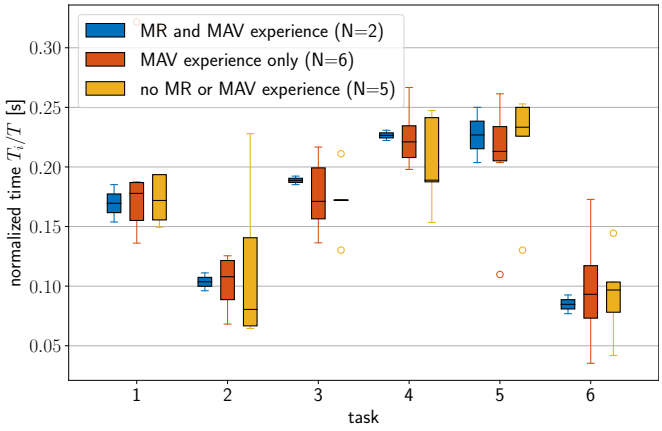

Fig. 6: Normalized time per task statistics, grouped according to previous user experience. Median, first quartile and third quartile range are indicated by the black line, box and whiskers, respectively. Values outside these ranges are considered as outliers, indicated by circles.

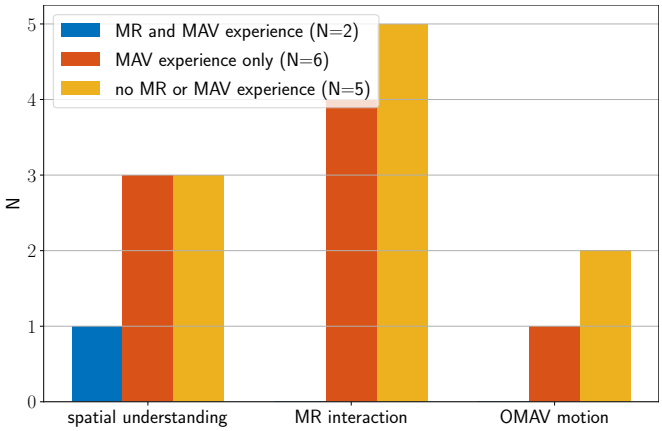

Fig. 7: Number of users reporting a specific difficulty, grouped according to previous user experience. Note that participants can report multiple types of difficulties.

2% faster than those of inexperienced operators ($med(T) = 917.5\,\mathrm{s}$). In this context, it should be noted that the upper outlier in Fig. 5 originates from the respective participant initially struggling to interact with the hologram and not from their limited flight skills. The normalized time per task in Fig. 6 also supports that prior knowledge of aerial robot operation is not required. All user groups required a similar time ratio per task. If prior knowledge in aerial robotic teleoperation were an advantage, we would have expected experienced users to outperform in complex tasks, like task 4 and 5, which require complicated flight maneuvers and multiple DoF manipulation.

Second, subjects familiar with the HoloLens had a major advantage. Figure 5 shows a considerable reduction in completion time of more than 40% when participants had prior experience with the technology ($med(T) = 530\,\mathrm{s}$). Almost all users unfamiliar with the HoloLens reported MR manipulation as a major difficulty (see Fig. 7). Since the device only recognizes specific gestures when grabbing holograms, grasp and release were often not properly registered, which in turn made mission execution more time-consuming. As mentioned above, however, the matching time ratios per task still suggest that any user can infer and effectively control the pose of an OMAV, despite possible technological challenges when commanding the planned motion. Therefore, simply training users in the specifics of HoloLens grasp detection (e.g., hand and finger poses) promises to yield significant performance improvements.

Third, lacking depth perception and a perceivable level of detail weaken situational awareness. Participants across all experience levels identified problems with spatial understanding of the simulated environment. Specifically, their ability to comprehend the robot's position in general and relative to other objects was affected. The issue reportedly became prominent with increasing distance between the OMAV and operator, as the perceivable level of detail is reduced. Such degrading resolution of the human eye at range highlights a fundamental limitation of MR-based teleoperation. To ensure the user's situational awareness, the proposed methodology is constrained to line-of-sight operation at short and medium distances (maximum $\approx 30\,\mathrm{m}$ in simulation). In other words, the size of the MAV's operational workspace directly depends on the operator's mobility on the ground. Note that the fixed viewpoint setup of the simulation and the notorious challenge of accurately representing a 3D environment on a 2D screen might also contribute towards the degraded spatial understanding. However, this is not a limitation of the developed interface per se but rather the evaluation setup.

V. Conclusion

This work investigates the potential of MR technology for the teleoperation of omnidirectional aerial robots. The designed application enables a user to directly control the translational velocity and orientation of an OMAV through the manipulation of a small-scale holographic interface. This implementation allows efficient coverage of large areas with minimal user effort while ensuring intuitive perception and command of the vehicle's attitude. We assess the intuitiveness of the proposed approach and identify potential limitations by performing a user study with 13 participants.

Independent of prior experience with MAVs or MR, all subjects could successfully interpret and control the pose of an OMAV during complex flight maneuvers. These results confirm the general effectiveness and intuitiveness of the approach. However, prior familiarity with MR technology improves task completion time since the interface gesture detection still needs refinement.

Nevertheless, MR is limited to line-of-sight operation at distances where the operator can still visually distinguish relevant details in the physical environment. Future work will focus on mitigating this constraint, such as increasing the MAV's admissible operation workspace. Furthermore, we aim to extend our analysis with a more extensive user study in real-world flight experiments, seeking to confirm the results of this preliminary investigation.

REFERENCES

[1] D. Lattanzi and G. Miller, "Review of robotic infrastructure inspection systems," *Journal of Infrastructure Systems*, vol. 23, no. 3, p. 04017004, 2017.

[2] A. Harris, "Rising to the challenge," *Engineering & Technology*, vol. 7, no. 10, pp. 56–58, 2012.

[3] A. Ollero, M. Tognon, A. Suarez, D. Lee, and A. Franchi, "Past, present, and future of aerial robotic manipulators," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 626–645, 2021.

[4] M. Allenspach, K. Bodie, M. Brunner, L. Rinsoz, Z. Taylor, M. Kamel, R. Siegwart, and J. Nieto, "Design and optimal control of a tiltrotor micro-aerial vehicle for efficient omnidirectional flight," *The International Journal of Robotics Research*, vol. 39, no. 10-11, pp. 1305–1325, 2020.

[5] M. Hamandi, F. Usai, Q. Sablé, N. Staub, M. Tognon, and A. Franchi, "Design of multirotor aerial vehicles: A taxonomy based on input allocation," *The International Journal of Robotics Research*, vol. 40, no. 8-9, pp. 1015–1044, 2021.

[6] K. Bodie, M. Brunner, M. Pantic, S. Walser, P. Pfandler, U. Angst, R. Siegwart, and J. Nieto, "Active Interaction Force Control for Contact-Based Inspection With a Fully Actuated Aerial Vehicle," *IEEE Transactions on Robotics*, vol. 37, no. 3, pp. 709–722, 2021.

[7] S. Park, J. Lee, J. Ahn, M. Kim, J. Her, G. Yang, and D. Lee, "Odar: Aerial manipulation platform enabling omnidirectional wrench generation," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 4, pp. 1907–1918, 2018.

[8] M. Tognon, H. A. T. Chávez, E. Gasparin, Q. Sablé, D. Bicego, A. Mallet, M. Lany, G. Santi, B. Revaz, J. Cortés, and A. Franchi, "A truly-redundant aerial manipulator system with application to push-and-slide inspection in industrial plants," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1846–1851, 2019.

[9] A. Coelho, H. Singh, K. Kondak, and C. Ott, "Whole-body bilateral teleoperation of a redundant aerial manipulator," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 9150–9156.

[10] J. Lee, R. Balachandran, Y. S. Sarkisov, M. D. Stefano, A. Coelho, K. Shinde, M. J. Kim, R. Triebel, and K. Kondak, "Visual-Inertial Telepresence for Aerial Manipulation," 2020.

[11] S. N. Young and J. M. Peschel, "Review of Human–Machine Interfaces for Small Unmanned Systems With Robotic Manipulators," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 2, pp. 131–143, 2020.

[12] M. Allenspach, N. Lawrance, M. Tognon, and R. Siegwart, "Towards 6dof bilateral teleoperation of an omnidirectional aerial vehicle for aerial physical interaction," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 9302–9308.

[13] F. Ferraguti, C. T. Landi, L. Sabattini, M. Bonfè, C. Fantuzzi, and C. Secchi, "A variable admittance control strategy for stable physical human–robot interaction," *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 747–765, 2019. [Online]. Available: https://doi.org/10.1177/0278364919840415

[14] M. Tranzatto, T. Miki, M. Dharmadhikari, L. Bernreiter, M. Kulkarni, F. Mascarich, O. Andersson, S. Khattak, M. Hutter, R. Siegwart, and K. Alexis, "Cerberus in the darpa subterranean challenge," *Science Robotics*, vol. 7, no. 66, p. eabp9742, 2022.

[15] J. Delmerico, R. Poranne, F. Bogo, H. Oleynikova, E. Vollenweider, S. Coros, J. Nieto, and M. Pollefeys, "Spatial computing and intuitive interaction: Bringing mixed reality and robotics together," *IEEE Robotics & Automation Magazine*, vol. 29, no. 1, pp. 45–57, 2022.

[16] S. Islam, R. Ashour, and A. Sunda-Meya, "Haptic and Virtual Reality Based Shared Control for MAV," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 5, pp. 2337–2346, 2019.

[17] G. A. Yashin, D. Trinitatova, R. T. Agishev, R. Ibrahimov, and D. Tsetserukou, "Aerovr: Virtual reality-based teleoperation with tactile feedback for aerial manipulation," in *2019 19th International Conference on Advanced Robotics (ICAR)*, 2019, pp. 767–772.

[18] D. Szafir, B. Mutlu, and T. Fong, "Designing planning and control interfaces to support user collaboration with flying robots," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 514–542, 2017.

[19] B. Hoppenstedt, T. Witte, J. Ruof, K. Kammerer, M. Tichy, M. Reichert, and R. Pryss, "Debugging quadrocopter trajectories in mixed reality," in *Augmented Reality, Virtual Reality, and Computer Graphics*, L. T. De Paolis and P. Bourdot, Eds. Cham: Springer International Publishing, 2019, pp. 43–50.

[20] M. Walker, T. Phung, T. Chakraborti, T. Williams, and D. Szafir, "Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy," 2022. [Online]. Available: https://arxiv.org/abs/2202.11249

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

[22] H. Hedayati, M. Walker, and D. Szafir, "Improving collocated robot teleoperation with augmented reality," in *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 78–86.

[23] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 316–324.

[24] M. Ostanin, R. Yagfarov, D. Devitt, A. Akhmetzyanov, and A. Klimchik, "Multi robots interactive control using mixed reality," *International Journal of Production Research*, vol. 59, no. 23, pp. 7126–7138, 2021.

[25] A. Angelopoulos, A. Hale, H. Shaik, A. Paruchuri, K. Liu, R. Tuggle, and D. Szafir, "Drone brush: Mixed reality drone path planning," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022, pp. 678–682.

[26] K. Konstantoudakis, K. Christaki, D. Tsiakmakis, D. Sainidis, G. Albanis, A. Dimou, and P. Daras, "Drone control in ar: An intuitive system for single-handed gesture control, drone tracking, and contextualized camera feed visualization in augmented reality," *Drones*, vol. 6, no. 2, 2022.

[27] T. Lee, M. Leok, and N. H. McClamroch, "Geometric tracking control of a quadrotor uav on se(3)," in *49th IEEE Conference on Decision and Control (CDC)*, 2010, pp. 5420–5425.

[28] M. Ostanin, S. Mikhel, A. Evlampiev, V. Skvortsova, and A. Klimchik, "Human-robot interaction for robotic manipulator programming in mixed reality," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 2805–2811.

[29] Y. Liu, N. Yang, A. Li, J. Paterson, D. McPherson, T. Cheng, and A. Y. Yang, "Usability evaluation for drone mission planning in virtual reality," in *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry*, J. Y. Chen and G. Fragomeni, Eds. Cham: Springer International Publishing, 2018, pp. 313–330.

[30] E. Triantafyllidis, W. Hu, C. McGreavy, and Z. Li, "Metrics for 3d object pointing and manipulation in virtual reality: The introduction and validation of a novel approach in measuring human performance," *IEEE Robotics & Automation Magazine*, vol. 29, no. 1, pp. 76–91, 2022.

[31] J. Artigas, R. Balachandran, C. Riecke, M. Stelzer, B. Weber, J.-H. Ryu, and A. Albu-Schaeffer, "Kontur-2: Force-feedback teleoperation from the international space station," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 1166–1173.

[32] F. Caccavale, C. Natale, B. Siciliano, and L. Villani, "Six-dof impedance control based on angle/axis representations," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 2, pp. 289–300, 1999.

[33] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, vol. 3, 2004, pp. 2149–2154 vol.3.