2024 IEEE International Conference on Robotics and Automation (ICRA)
May 13-17, 2024. Yokohama, Japan

Spatial Assisted Human-Drone Collaborative Navigation and Interaction through Immersive Mixed Reality

Luca Morando and Giuseppe Loianno

2024 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3503-8457-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRA57147.2024.10611351

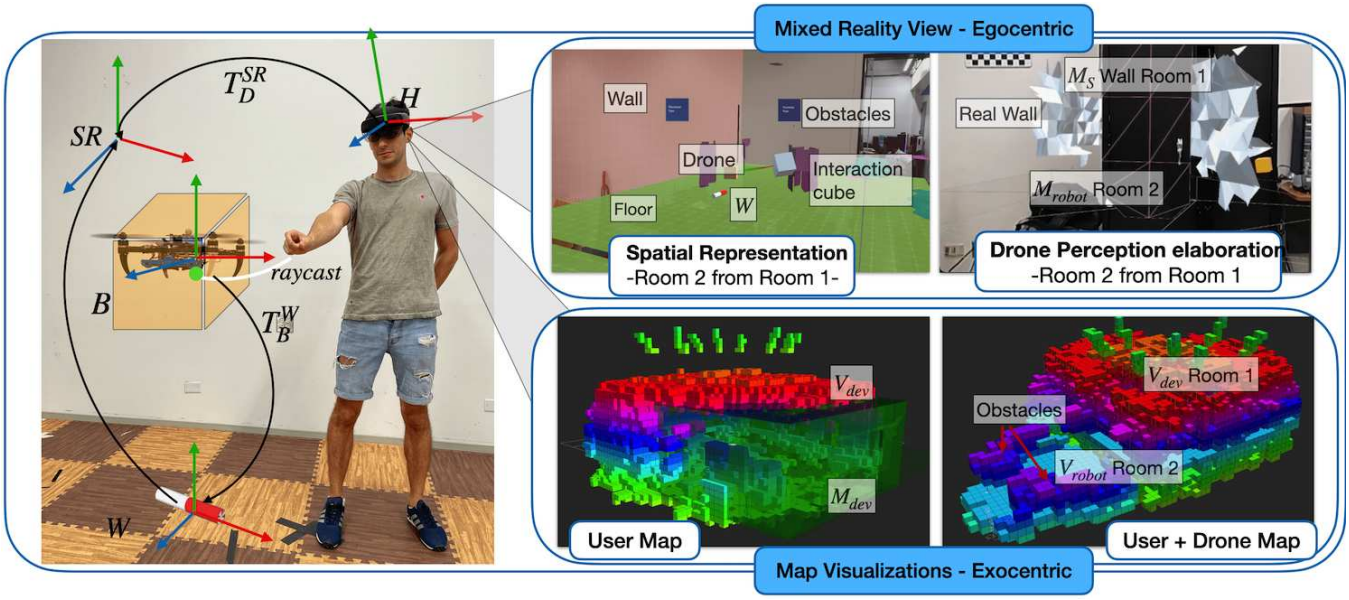

Fig. 1: a.) *Human-Drone interaction*: Pinch and drag gestures are available to manipulate and to interact with the hologram of the robot. b.) *Spatial-Awareness*: the robot flying space (left image) and the drone generated mesh (right image) are spatially represented with a through-the-wall visualization, c.) *Maps Merging*: visualization of the merged user and drone maps of the two different rooms.

***Abstract*— Aerial robots have the potential to play a crucial role in assisting humans with complex and dangerous tasks. Nevertheless, the future industry demands innovative solutions to streamline the interaction process between humans and drones to enable seamless collaboration and efficient co-working. In this paper, we present a novel tele-immersive framework that promotes cognitive and physical collaboration between humans and robots through Mixed Reality (MR). This framework incorporates a novel bi-directional spatial awareness and a multi-modal virtual-physical interaction approaches. The former seamlessly integrates the physical and virtual worlds, offering bidirectional egocentric and exocentric environmental representations. The latter, leveraging the proposed spatial representation, further enhances the collaboration combining a robot planning algorithm for obstacle avoidance with a variable admittance control. This allows users to issue commands based on virtual forces while maintaining compatibility with the environment map. We validate the proposed approach by performing several collaborative planning and exploration tasks involving a drone and an user equipped with a MR headset.**

SUPPLEMENTARY MATERIAL

Github: https://github.com/arplaboratory/mri_ros_public.git
Video: https://youtu.be/oWdqLuXe1Zw

The authors are with the New York University, Tandon School of Engineering, Brooklyn, NY 11201, USA. email: {luca.morando, loiannog}@nyu.edu.

This work was supported by the NSF CPS Grant CNS-2121391, the NSF CAREER Award 2145277, the DARPA YFA Grant D22AP00156-00, Qualcomm Research, Nokia, and NYU Wireless.

I. INTRODUCTION

The increased use of robotics systems in our everyday life raises the need to design novel solutions that can facilitate the interaction between humans and robots decreasing the physical and cognitive human workload especially when executing complex or dangerous tasks. Human-Robot Collaboration and Interaction (HRI) is a well-studied problem and being integrated in multiple areas, such as Industry 4.0, construction, and emergency responses [1], [2]. Traditionally, human-centered robotic systems focus on collaboration between humans and grounded robots. In aerial robotics, due to the additional complexity introduced by the navigation requirements in 3D unstructured environments, current state of the art approaches still limit the collaboration to be mostly unidirectional (i.e., with commands or information sent from the human to the robot) using classic teleoperation [3] with representation of the output signals projected on flat surface monitors. Aerial robotics will become essential in supporting humans in multiple tasks including, but not limited to warehouses' operations, vertical farming inspection, space missions, and search and rescue. In these settings, robots' role can no longer be limited to simply assist humans. Instead, they will need to co-exist, collaborate, and co-work with them, sharing the same workspace and participating

Authorized licensed use limited to: New York University. Downloaded on December 01,2024 at 09:13:12 UTC from IEEE Xplore. Restrictions apply.

in shared tasks' execution. Therefore, these requirements call for the novel approaches that enable a high level of mutual and seamless human-robot interaction. We envision aerial robots to be elevated to the role of humans' teammates allowing a bi-directional flux of information to concurrently augment both human and robot situational awareness.

In this work, we propose a novel tele-immersive framework for human-drone cognitive and physical collaboration through Mixed Reality (MR) as illustrated in Fig. 1 to facilitate seamless interaction between humans and robots. The emergence of innovative spatial computing techniques, such as Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), presents a technological opportunity in robotics. These techniques facilitate enhanced collaboration between humans and robots through multi-modal information sharing within the human-robot team based on vision [4], gestures [5], natural languages [6], and gaze [7], [8].

The contributions of this paper can be summarized as follows. First, we present a novel bi-directional spatial awareness concept that enables co-sharing of spatial information between a human and an aerial robot at different levels of abstractions, making them co-aware of the surrounding environment. Second, we propose a novel virtual-physical interaction that further facilitates bi-directional collaboration. This approach uniquely combines a Variable Admittance Control (VAC) [9], [10], implemented on top of the drone's guidance loop, with a planning algorithm [11] responsible for robot's obstacle avoidance. The proposed VAC allows through gesture recognition to input a user virtual external force as an external command to the robot while ensuring compatibility with the environment map. The user force is further combined with a repulsive force field acting on the robot when located in proximity of obstacles increasing the overall system safety.

To the best of our knowledge, the proposed framework is the first to enable continuous spatial virtual-physical navigation and interaction with an aerial robot via MR. We open source the framework to the community.

II. Related Works

Teleoperation. Drone teleoperation, while arguably one of the simplest interaction methods, remains an open research area especially in constrained robot environments [12], [13]. Current approaches still rely on simple joysticks, keyboards, and pointers receiving the feedback from the robots on monitors. Overall, state of the art autopilots can enable drones to follow aggressive trajectories [14], but they do not offer an easy and intuitive communication paradigm between human and robot. Recent solutions try to overcome these drawbacks by employing hand gesture [15] or impedance control with vibrotactile feedback [16] to control multiple drones or impedance control for human-machine [17], [18] or machine-environment interaction [19].

Extended Reality. Innovative body-machine interfaces or wearable devices (e.g., head mounted displays) utilize body motion to control robots and incorporate haptic feedback, along with VR or AR [20]–[24]. These interfaces aim to address the constraints of basic teleoperation control. For instance, AR proves to be a valuable interaction modality that facilitates collaborative information visualization in robotics tasks, therefore enabling efficient communication of robots' intentions to human co-workers [25], [26]. In [27], the authors study how to maximize the user ability to tele-operate a particular drone projecting its camera field of view camera in the user headset. In [28]–[30], an exocentric view of the drone pose in a virtual scene that is built a-priori, helps the user to interact with the drone via a pick and place gesture, commanding a new goal position to the robot. A gesture based trajectory completion in MR is proposed in [31], where an automated technique auto-complete the user intended motion. Conversely, the authors in [32] show a collaborative MR interface for immersive planning where the user is able, using simple gestures to draw 3D trajectories for quadrotors in its workspace. While in [32] the main focus is on drone teleoperation via simple gestures, in [33] the authors employ MR to analyze the information sharing problem, in order to simplify the human-robot communication during an exploration task. Information about the robot's future actions and predicted trajectory is visualized using AR. Finally, a recent work [34] proposes a fascinating instance of enhancing human situational awareness in MR during human-robot interaction. The MR experience merges a downsized satellite map with the robot's local environment representation. Users have the ability to choose a location on the map and stream the locally captured robot's point cloud into the MR headset, providing them with a first-person view of the scene.

Novelties. Our tele-immersive approach goes beyond simple tele-operation and extended reality for robotics. It facilitates a bi-directional flux of information to enhance the spatial awareness of both the human and the drone. This aspect is further enhanced incorporating a novel human-robot virtual-physical interaction in 3D based on a VAC [35]. Compared to [35], our work goes beyond the manipulator case, it eliminates the need for prior knowledge about the environment's structure and seamlessly and uniquely designs and deploys the VAC within the online 3D navigation framework.

III. Methodology

As visible in Fig. 1.a, the robot body frame is denoted with B and the robot world frame, located at the drone take-off position, is denoted with W and it is represented in the MR environment as a red and white cylinder. Referring to Fig. 1.a, we denote the spatial transformation between two frames i and j as with $\mathbf{T}^i_j$ that represents the relative pose of the j frame with respect to the i.

A. Spatial Awareness

We introduce the concept of human-drone spatial awareness. This cutting-edge technique facilitates an improved representation of the robot's perceived environment on the user side, utilizing advanced mapping and visualization methods. Moreover, on the robot side, the user's perception of the

Authorized licensed use limited to: New York University. Downloaded on December 01,2024 at 09:13:12 UTC from IEEE Xplore. Restrictions apply.

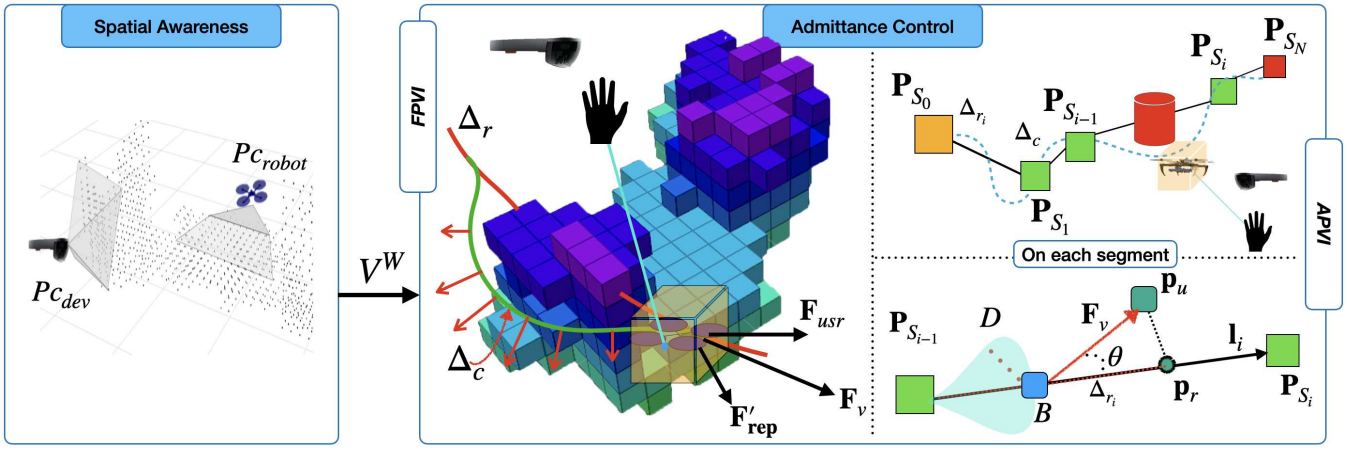

Fig. 2: Key framework's components. The pointcloud set P^W is collected and sent to the SAM where a collective octomap is generated (left). A VAC coupled with a planner provides assistance to the user during the drone navigation (right).

surrounding environment is translated in a map representation that is interpretable by the robot navigation layer. The raw data obtained from the drone and the wearable device sensors are processed through the Spatial Awareness Module (SAM) as presented in Fig. 1. Specifically, the robot point cloud P_{crobot} (in our case a stereo disparity as detailed in Section IV) and the wearable device point cloud P_{cdev} are forwarded as input to the robot SAM which provides as output two types of geometric representations namely an octomap and a mesh visualization shown in Fig. 1.b and Fig. 1.c respectively.

The set of the robot and device pointclouds ${P_c}^W = \{P_{cdev}, P_{crobot}\}^W$ can be defined and expressed in W using the following transformations: i) $\mathbf{T}_B^W$ for P_{crobot} and ii) $\mathbf{T}_{SR}^W$ for P_{cdev}, where the frame SR is initialized with the same pose of the device frame H, when the MR application is launched. Leveraging the Voxblox framework [36], it is possible to obtain an octomap set from the pointcloud ${P_c}^W$, denoted as $V^W = \{V_{robot}, V_{dev}\}^W$, where V_{robot} is the voxel representation on the robot side while V_{dev} is similarly obtained from the user perspective. The set V^W define a unified global voxel representation visible in Fig. 1.c where the merged maps of two adjacent rooms (one explored by the robot and the other one explored by the user) are depicted. An example of the representation instantiated by the SAM is depicted in Fig.1.b and Fig.1.c, showcasing a dual visualization of the environments perceived by the user and the robot, respectively.

In Fig. 1.b, we show the user Ego-centric visualization that allows the MR layer to translate the drone perceived environment in a spatial language easily understandable by the human. The Exo-centric visualization instead, depicted in Fig. 1.c, merges the user and the drone maps captured from two different rooms or perspectives, therefore augmenting the drone perception capabilities. A further level of abstraction and functionality is provided by the instantiation of the MR environment in the user field of view, where both the wearable device frame H and the world frame W, are referred to the SR frame via $\mathbf{T}_W^{SR}$ and $\mathbf{T}_H^{SR}$ respectively, as shown in Fig. 1.a. The wearable device can extract information in the perceived environment about various surfaces (i.e., walls, ceilings, floors etc.) or artificially create additional elements in form of holograms around the user (see Fig. 1.b). This produces a seamless mix of real and virtual worlds allowing a customized and reconfigurable spatial interaction with increased configuration complexities between the human and/or the robot even in an empty physical space. An example is provided in Fig. 1.b, where the drone mesh M_{robot}, obtained by P_{crobot} and visible in white, is merged with other spatially defined holograms in the MR framework.

Finally, the drone pose is always represented in the MR environment through an orange cube C referred to the frame W (see Fig. 1.a).

B. Virtual-Physical Human-Drone Interactive Navigation

We propose a VAC to link the physical and virtual worlds producing two types of virtual-physical interaction

- **Assisted Physical and Virtual Interaction (APVI)**: In this modality, a planning algorithm (in our case a Rapidly exploring Random Tree in its optimal formulation (RRT*) [11]) is built on top of the generated octomap and combined with a VAC algorithm. Concurrently, the VAC provides a modulated damping feedback based on a desired function. This can modify the drone compliance depending the user intention to modify its path from the original one proposed by the planner.
- **Free Physical and Virtual Interaction (FPVI)**: This is a specific case of the APVI modality when the RRT* planner is disabled by the user. This modality provides additional freedom when driving the robot. Since the robot is not anymore constrained to follow a specific path, an additional safety layer is required. We propose the generation of an obstacle force field that is injected in the admittance controller as an additional input.

A schematic representation of both modalities with their integration with the SAM is visualized in Fig. 2. The two modalities offer varying levels of intuitiveness and environmental safety during the human-robot interaction, allowing the user to depend more on the robot's perception of its surroundings for effective exploration, especially in

Authorized licensed use limited to: New York University. Downloaded on December 01,2024 at 09:13:12 UTC from IEEE Xplore. Restrictions apply.

unfamiliar or obstacle-filled environments. In the following, we describe the key components of the APVI and FPVI namely the RRT*, the VAC and how this is couple with the RRT* for the APVI modality, and the obstacle force field.

1) RRT assisted Human-Robot Navigation:* The RRT* is enabled in the APVI mode, expanding its tree based on a local robot map defined within a circular horizon centered on the drone body frame origin $\mathcal{O}_{\mathcal{B}}$. This avoids the problem to know a-priori the environment map and local minima planning issues can be taken care by the user intervention with the proposed VAC. The output of the RRT* is a sequence of setpoints $\mathbf{P}_{S_i}$, for $i = 0, \cdots N$, which are computed only within the drone horizon. The path is updated on-line, while the drone is moving along the proposed direction. As visible in Fig. 2 (right), the robot reference trajectory is composed by a sequence of reference positions $\mathbf{p}_r$ that lies on the line each defined by a segment between two consecutive setpoints $\mathbf{P}_{S_i}$ and $\mathbf{P}_{S_{i+1}}$ generated by the RRT*. The sequence of reference position $\mathbf{p}_r$ compose the trajectory $\mathbf{\Delta}_r$.

During the random nodes exploration phase embedded in the RRT* , the total cost of the new node c^* is obtained by adding the linking cost c_l to the parent cost c_p.

2) Coupling Planning and VAC for assisted virtual-physical HRI: We combine the RRT* planning algorithm, presented in Section III-B.1 with a VAC. Prior to describing the VAC, we design an admittance controller on top of the drone inner control position pipeline. This controller, in its classic formulation visible in eq. (1), is driven by the 3D virtual input force $\mathbf{F}_v$ generating the position perturbation $\mathcal{X} = \mathbf{p}_c - \mathbf{p}_r$ between a commanded position $\mathbf{p}_c$ and a reference position $\mathbf{p}_r$. The dynamical relationship between the input force $\mathbf{F}_v$ and the position perturbation $\mathcal{X}$ mimics a mass spring damper system as

$$\mathbf{M}\left(\ddot{\mathbf{p}}_c - \ddot{\mathbf{p}}_r\right) + \mathbf{D}\left(\dot{\mathbf{p}}_c - \dot{\mathbf{p}}_r\right) + \mathbf{K}\left(\mathbf{p}_c - \mathbf{p}_r\right) = \mathbf{F}_v. \quad (1)$$

The terms $\mathbf{M}$, $\mathbf{D}$ and $\mathbf{K}$ are 3×3 (diagonal matrices in our case) representing the mass, the damping and the stiffness acting on each axis of the world frame W. The right input term to the equation is the virtual force $\mathbf{F}_v = \mathbf{F}_{usr} + \mathbf{F}_{rep}$ resulting from the sum of the user interaction force $\mathbf{F}_{usr}$ with the obstacle repulsive force $\mathbf{F}_{rep}$, which will be defined later in Section III-B.3, while $\mathbf{F}_{usr} = \mathbf{K}_p(\mathbf{p} - \mathbf{p}_u) - \mathbf{K}_d(\dot{\mathbf{p}} - \dot{\mathbf{p}}_u)$ with $\mathbf{p}$ and $\mathbf{p}_u$ are the drone position and the user interactive marker position respectively. These are estimated using a linear Kalman Filter (KF) for motion estimation considering a constant acceleration model which takes as input the acceleration $\ddot{\mathbf{p}}_u$. The terms $\mathbf{K}_p$ and $\mathbf{K}_d$ represent respectively the proportional and the derivative terms. Compared to the classic admittance control, the damping value $\mathbf{D}$, in the VAC definition, varies according to the direction of the input force $\mathbf{F}_v$. A variable damping enables to change the feeling of the user during the interaction since the drone is constrained to follow the direction of minimum resistance.

Starting from eq. (1) for the classic admittance controller and referring to the RRT* segment in Fig. 2, let $\theta = [0, \pi]$ be the angle determined by the human force direction $\mathbf{F}_v$ with respect to $\mathbf{l}_i$: $\theta = \cos^{-1}(\mathbf{l}_i^\top \mathbf{F}_v / ||\mathbf{F}_v||)$ where $\mathbf{l}_i$ is the vector describing the direction, expressed in W, of the segment between two consecutive setpoints $\mathbf{P}_{S_{i-1}}$ and $\mathbf{P}_{S_i}$ traversed by the drone. Focusing on a generic segment in Fig. 2, $\mathbf{p}_u$ is projected on the segment defining the position $\mathbf{p}_r$, forwarded as input to the Variable Admittance Control (VAC). The sequence of reference positions $\mathbf{p}_r$ defines the reference trajectory $\mathbf{\Delta}_r$. Inspired by [35], we define the VAC with three functions $D_k = f_k(\theta)$, with D_k a diagonal element of $\mathbf{D}$, as $linear$, $squared$ and $exponential$ with

- $D_{k,min} = f_k(\theta = 0)$,
- $D_{k,max} = [f_k(\theta = \pi) \wedge f_k(\theta = -\pi)]$,
- $f_k(\theta)$ is continuous in $[0, \pi]$ and $[0, -\pi]$.

The three damping functions $D_k = f_k(\theta)$ are integrated within the APVI mode. The functions modify the damping behaviour (visualized as the blue area around the drone position in Fig. 2) depending on the pointing direction of the virtual force vector $\mathbf{F}_v$ that is forwarded as input to the admittance controller as shown in eq. (1).

The set of commanded positions $\mathbf{p}_c$ compose the trajectory $\mathbf{\Delta}_c$, visible as a blue dashed line in Fig. 2 (right), compliant with the direction of the force $\mathbf{F}_v$.

3) Obstacle Perception as Force Field: The FPVI modality employs the same spatial representation of the APVI case as visible in Fig. 2 (left). Given multiple repulsive forces per voxel $\mathbf{F}_{r_i}$ with $i \in \{0, ..., N_v\}$ with N_v is the total number of voxels $v_i \in V^W$ within the robot horizon h, it is possible to define the direction of the resulting force $\mathbf{F}_{rep}$ acting on the robot frame B and expressed in the frame W. The normalized vector $\mathbf{F}_{rep} = \mathbf{F}_r / \|\mathbf{F}_r\|$ defines the direction of the repulsive force acting on the robot, where $\mathbf{F}_r = \sum_{i=0}^{Nv} \mathbf{F}_{r_i}$ is composed by the sum of all the force vectors $\mathbf{F}_{r_i}$ generated by each of the voxel v_i inside the robot horizon h. Each term $\mathbf{F}_{r_i}$ is therefore defined following the exponential decay behaviour in function of the euclidean distance d from the voxel v_i as

$$\|\mathbf{F}_{r_i}\| = \frac{F_s}{k} e^{(-\lambda d)}(1 - e^{(h-d)}), \quad (2)$$

where $k = 1 - e^h$, $\|\mathbf{F}_{r_i}\| = F_s$ when $d = 0$ and λ is the exponential decay constant. Finally, at each iteration, the magnitude of the vector $\mathbf{F}_{rep}$ is defined as $\|\mathbf{F}_{rep}\| = \|\mathbf{F}_r\|$.

IV. Experimental Results

We validate the components of the proposed framework and show their application in joint human-drone exploration tasks through multiple real-world experiments. These are conducted at the Agile Robotics and Perception Lab (ARPL) at the New York University in the flying arena with size of 60 m^2 as well as in an adjacent room to show the fully capabilities of the tele-immersive framework in collaborative tasks, as proposed in Fig. 1.b and Fig. 1.c.

The quadrotor platform used in the experiments is equipped with a Qualcomm®Snapdragon™ board for on-board computing where an embedded stereo camera is used for a disparity representation, obtaining the point cloud ${P_{crobot}}^B$. Our algorithm provides the state of the robot (pose and perception output) directly in the W frame via

Authorized licensed use limited to: New York University. Downloaded on December 01,2024 at 09:13:12 UTC from IEEE Xplore. Restrictions apply.

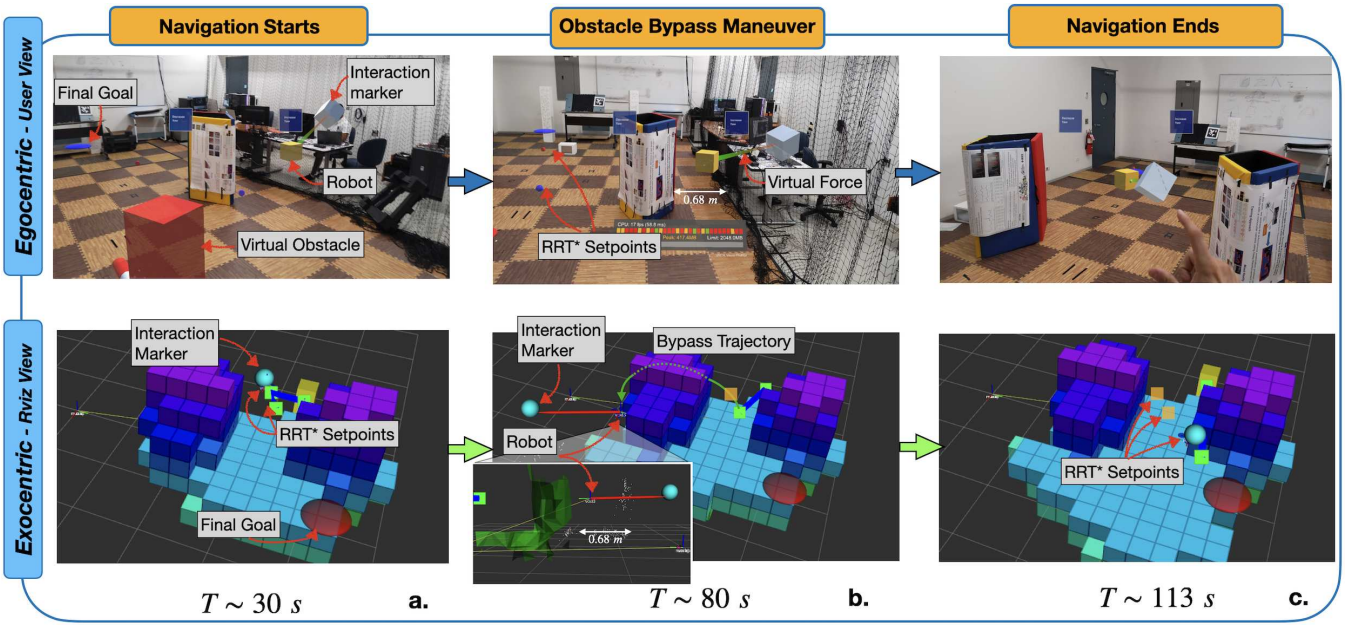

Fig. 3: Ego and Exo-centric visualization of the proposed experiment captured at different salient moments. An enhanced safety layer is provided by the robot and the SAM, with the generation of a safe trajectory through the obstacles.

the transformation $\mathbf{T}_B^W$. In addition, a mutual co-localization mechanism between the MR device and the aerial robots can be used [37]. Both the robot and the AHRMD use onboard self-localization mechanism. The Mixed Reality framework is implemented in C# and it is executed on the Microsoft® Hololens 2.0. Combining different features, the proposed framework offers four types of deployment modalities

- **Full Simulation (FS)** The complete interaction is based on ARPL custom quadrotor simulator and visualized in ROS RViz.
- **Mixed Simulation (MS)**: It is the same as **FS**, but the user can interact with the robot using the MR headest.
- **Real Interaction (RI)**: The user employs ROS Rviz to send commands to the real flying drone. In this case, the mapping and navigation functionalities are enabled.
- **Real Interaction in MR (RIMR)**: In this case the framework is fully enabled.

We provide an experimental overview of the FPVI and APVI interaction modalities proposed in Section III-B and shown in Fig. 2. These modalities are tested in a lab environment by four individuals, all of whom provided positive feedback regarding the projected amount of visual information and the ease of interaction with an aerial robot. They consistently achieved the assigned task to reach a target in a densely populated obstacle environment, relying on both the robot perception feedback represented by the computed planner path and the obstacle repulsive force. A visualization of two of the possible tasks assigned to the subjects are described in the following sections.

A. VAC with RRT* exploration and Mapping

The interaction behaviour during the APVI mode is represented in Fig. 3 showing a sequence of frames with an Egocentric and Exocentric views at different time steps during the same experiment whose results are in Fig. 4.

During the experiments the height of the robot is kept constant at $p_{r,z} = 0.8$ m. The black dashed line at time instant $t_{APVI} = 19.5$ s, in Fig. 4, represents the switches from (FPVI) to (APVI) mode which happens when the user places a final goal to the robot to reach, denoted as $\mathbf{G}^W$ and visualized in Fig. 3 with a red circle. Once activated, the RRT* obstacle aware planning algorithm, coupled with the VAC initialized with $M_k = 2.4$ Kg, $K_k = 20$ Kg $\cdot$ s^{-2} and 1 Kg $\cdot$ s$^{-1} \leq D_k \leq 70$ Kg $\cdot$ s^{-1}, where the subscript k represents the diagonal element of $\mathbf{M}$, $\mathbf{K}$, and $\mathbf{D}$ respectively, constraints the drone to follow the proposed path. Nonetheless, the user can still pull away the robot from the initial trajectory despite this safety condition, applying an increasing virtual Force $\mathbf{F}_v$ as input to the system, in order to define a maneuver to bypass virtual and real obstacles. This makes the robot movement compliant to the user intentions but still constrained to follow the reference trajectory $\mathbf{\Delta}_r$, relying on the planner proposed path. The maneuver and the human-drone assisted navigation are visible in detail also in the sequence of frames in Fig. 3.

To show the system's compliance to the human intentions, in Fig. 3.b, we observe that the user's action can pull the robot away from the RRT* planned path and drive it along a new trajectory passing through a narrow gap measured in only 0.68 m (schematized as a green arrow). The estimated obstacle Truncated Signed Distance Field (TSDF) [38], [39], represented by the green mesh M_{robot}, visualizes the detected surface of the obstacle as perceived by the drone sensors during the bypass maneuver. Once this maneuver terminates, in Fig. 3.c, the RRT* restarts its expansion towards $\mathbf{G}^W$, only when the drone is attracted back to the last generated setpoint. Moreover, in Figs. 3.a, 3.b, and 3.c the interaction marker (white cube), the drone camouflage (orange cube) and the generated force (colorful string) are easily identifiable by the user. The generated

Authorized licensed use limited to: New York University. Downloaded on December 01,2024 at 09:13:12 UTC from IEEE Xplore. Restrictions apply.

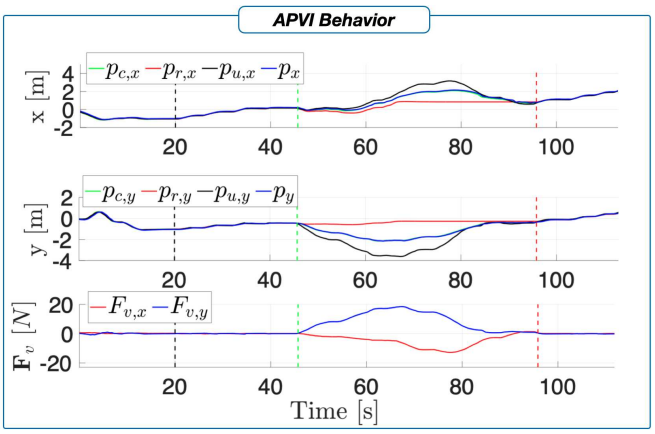

Fig. 4: VAC behavior during the APVI interaction.

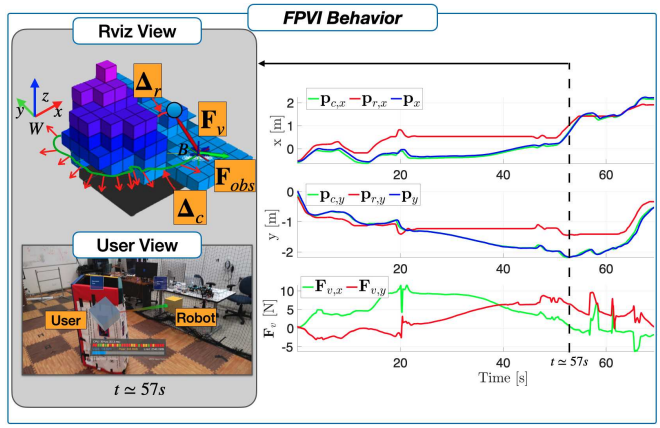

Fig. 5: Quantitative visualization of the obstacle safety layer acting on the system visualized in User and Rviz view

trajectory depending on the Force $\mathbf{F}_v$ is shown in Fig. 4. The Root Mean Squared Error between $\mathbf{\Delta}_r$ and $\mathbf{\Delta}_c$ during the FPVI and APVI modalities is respectively of 0.0373 m and 2.0142 m. The higher RMSE value obtained during the APVI phase is due to the obstacle bypass maneuver, where the commanded trajectory is compliant with the user intentions rather than following the original desired path.

B. FPVI with Obstacle Force Field

We propose a second experiment which provides a deeper insight of the behavior of the system when subject to the action of the obstacle generated force field, defined in Section III-B.3 when the system is set to the FPVI mode. Along the overall experiment, the gains of the admittance controller, except for $D_k = 20\ \mathrm{Kg} \cdot \mathrm{s}^{-1}$, are not modified with respect the previous scenario shown in Section III-B.3. The exponential decay behaviour of the force $\mathbf{F}_{rep}$ is defined respectively with $Fs = 8$ N, $\lambda = 1$ and horizon $h = 1.5$ m. Referring to Section 2 and following the exponential decay physical behavior, the magnitude of the force $\|\mathbf{F}_{rep}\| = 0$ when $d = h$ and $\|\mathbf{F}_{rep}\| = F_s$ when $d = 0$. The results of this experiment are provided in Fig. 5, with a detailed Ego ("User view") and Exo-centric ("Rviz view") representations of the interaction captured at time $t = 57$ s. As visible from the location of the white interaction cube, visualized as a sphere in the map, the user tries to impose an unsafe desired trajectory $\mathbf{\Delta}_r$ passing through an obstacle. As the robot perceives the environment a repulsive force $\mathbf{F}_{rep}$, defined as in III-B.3, is generated and visualized in the Exo-centric representation in Fig. 5 as a green arrow. The user pulling force and the obstacle repulsive force are forwarded as input to the admittance controller which generates new safer commanded trajectory $\mathbf{\Delta}_c$, visualized in Fig. 5 as a green line. The complete evolution of the trajectories $\mathbf{\Delta}_c$, $\mathbf{\Delta}_r$ and the robot actual sequence of positions $\mathbf{p}$, are represented in the plot in Fig. 5 where the components of the trajectories along the x, y axis are visualized. The "User View" in Fig. 5 provides an enhanced visualization of the human robot interaction at the time of the representation in the "Rviz view" as seen through the head mounted device.

C. Results Discussion

The primary objective of this work is to validate the technical modules of proposed framework. The presented results validate each component and the overall framework's functionality, showcasing potential for natural and intuitive interaction beyond traditional tele-operation and basic MR solutions. Tests in a cluttered environment show that safety measures introduced through the planner algorithm (Section IV-A) and obstacle force field (Section IV-B) enable secure, gesture-based interaction with the drone, adaptable to user needs. Tests performed by multiple users further validate the methodologies and the benefits of the proposed solution. This approach serves as a flexible building block, open to incorporation with other modalities for enhanced bi-directional data interpretation and interaction.

V. Conclusion

In this paper, we presented a tele-immersive human-robot collaborative framework. The key features include a novel bi-directional spatial awareness approach that provides seamless human-robot spatial awareness through MR and a multi-modal virtual-physical interaction that further enhances this collaboration experience. We open-source the framework with the goal to promote research in this area and incorporate feedback from the community.

Despite four users already validated the usability of the system in the lab, in the future, we expect to conduct a complete user case study, leveraging the NASA Task Load Index [40], subjective questions and quantitative data collected on an inspection target reaching based task, in order to have a complete performance evaluation on multiple users. This will help to further validate the presented approaches and results. Specifically, we would like to compare the users' feelings during a 2D based interaction and full MR experience. We believe the results will guarantee an easier manipulation and a continuity in the interaction between the robot and non expert pilots when both agents share the same environment. Finally, we aim to extend the proposed framework by allowing the incorporation of heterogeneous teams consisting of multiple users and robots.

Authorized licensed use limited to: New York University. Downloaded on December 01,2024 at 09:13:12 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the human–robot collaboration," *Autonomous Robots*, vol. 42, no. 5, pp. 957–975, Jun 2018.

[2] L. Morando, C. T. Recchiuto, and A. Sgorbissa, "Social drone sharing to increase the uav patrolling autonomy in emergency scenarios," in *29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2020, pp. 539–546.

[3] S. Kim, "Working with robots: Human resource development considerations in human–robot interaction," *Human Resource Development Review*, vol. 21, no. 1, pp. 48–74, 2022.

[4] R.-J. Halme, M. Lanz, J. Kämäräinen, R. Pieters, J. Latokartano, and A. Hietanen, "Review of vision-based safety systems for human-robot collaboration," *Procedia CIRP*, vol. 72, pp. 111–116, 2018, 51st CIRP Conference on Manufacturing Systems.

[5] H. Liu and L. Wang, "Gesture recognition for human-robot collaboration: A review," *International Journal of Industrial Ergonomics*, vol. 68, pp. 355–367, 2018.

[6] S. Pate, W. Xu, Z. Yang, M. Love, S. Ganguri, and L. L. S. Wong, "Natural language for human-robot collaboration: Problems beyond language grounding," *CoRR*, vol. abs/2110.04441, 2021. [Online]. Available: https://arxiv.org/abs/2110.04441

[7] L. Yuan, C. Reardon, G. Warnell, and G. Loianno, "Human gaze-driven spatial tasking of an autonomous mav," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1343–1350, 2019.

[8] M. Pavliv, F. Schiano, C. Reardon, D. Floreano, and G. Loianno, "Tracking and relative localization of drone swarms with a vision-based headset," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1455–1462, 2021.

[9] A. Lecours, B. Mayer-St-Onge, and C. Gosselin, "Variable admittance control of a four-degree-of-freedom intelligent assist device," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 3903–3908.

[10] 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.

[11] X. Wang, X. Luo, B. Han, Y. Chen, G. Liang, and K. Zheng, "Collision-free path planning method for robots based on an improved rapidly-exploring random tree algorithm," *Applied Sciences*, vol. 10, no. 4, 2020.

[12] W. A. Isop, C. Gebhardt, T. Nägeli, F. Fraundorfer, O. Hilliges, and D. Schmalstieg, "High-level teleoperation system for aerial exploration of indoor environments," *Frontiers in Robotics and AI*, vol. 6, 2019.

[13] B. Rojas, H. Abaunza, P. Castillo, I. Thouvenin, and R. Lozano, "Teleoperation of a drone based on a virtual environment," in *Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS)*, 2019, pp. 160–165.

[14] A. Saviolo, G. Li, and G. Loianno, "Physics-inspired temporal learning of quadrotor dynamics for accurate model predictive trajectory tracking," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 256–10 263, 2022.

[15] J. Nagi, A. Giusti, L. M. Gambardella, and G. A. Di Caro, "Human-swarm interaction using spatial gestures," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014, pp. 3834–3841.

[16] E. Tsykunov, R. Agishev, R. Ibrahimov, L. Labazanova, A. Tleugazy, and D. Tsetserukou, "Swarmtouch: Guiding a swarm of micro-quadrotors with impedance control using a wearable tactile interface," *IEEE Transactions on Haptics*, vol. 12, no. 3, pp. 363–374, 2019.

[17] A. Tagliabue, M. Kamel, S. Verling, R. Siegwart, and J. Nieto, "Collaborative transportation using mavs via passive force control," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 5766–5773.

[18] F. Augugliaro and R. D'Andrea, "Admittance control for physical human-quadrocopter interaction," 07 2013, pp. 1805–1810.

[19] M. Car, A. Ivanovic, M. Orsag, and S. Bogdan, "Position-based adaptive impedance control for a uav," in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2018, pp. 957–963.

[20] M. Macchini, F. Schiano, and D. Floreano, "Personalized telerobotics by fast machine learning of body-machine interfaces," *IEEE Robotics and Automation Letters*, vol. 5, no. 1, pp. 179–186, 2019.

[21] J. Bruce, J. Perron, and R. Vaughan, "Ready—Aim—Fly! Hands-Free Face-Based HRI for 3D Trajectory Control of UAVs," in *14th Conference on Computer and Robot Vision*, 2017.

[22] J. Nagi, A. Giusti, G. A. Di Caro, and L. M. Gambardella, "Human control of uavs using face pose estimates and hand gestures," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2014, pp. 252–253.

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

[24] C. Reardon, K. Lee, and J. Fink, "Come see this! augmented reality to enable human-robot cooperative search," in *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2018.

[25] M. D. Coovert, T. Lee, I. Shindev, and Y. Sun, "Spatial augmented reality as a method for a mobile robot to communicate intended movement," *Computers in Human Behavior*, vol. 34, pp. 241–248, 2014.

[26] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," *The International Journal of Robotics Research*, vol. 38, no. 12-13, pp. 1513–1526, 2019.

[27] P. Fletcher, A. Luther, C. Detweiler, and B. Duncan, "Predicting visual differentiability for unmanned aerial vehicle gestures," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 799–11 806, 2022.

[28] O. Erat, W. A. Isop, D. Kalkofen, and D. Schmalstieg, "Drone-augmented human vision: Exocentric control for drones exploring hidden areas," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1437–1446, 2018.

[29] C. Liu and S. Shen, "An augmented reality interaction interface for autonomous drone," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 11 419–11 424.

[30] J. Betancourt, B. Wojtkowski, P. Castillo, and I. Thouvenin, "Exocentric control scheme for robot applications: An immersive virtual reality approach," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2022.

[31] M. K. Zein, M. Al Aawar, D. Asmar, and I. H. Elhajj, "Deep learning and mixed reality to autocomplete teleoperation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 4523–4529.

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

[33] C. Reardon, K. Lee, J. G. Rogers, and J. Fink, "Augmented reality for human-robot teaming in field environments," in *Virtual, Augmented and Mixed Reality. Applications and Case Studies*, J. Y. Chen and G. Fragomeni, Eds. Cham: Springer International Publishing, 2019, pp. 79–92.

[34] M. Walker, Z. Chen, M. Whitlock, D. Blair, D. A. Szafir, C. Heckman, and D. Szafir, "A mixed reality supervision and telepresence interface for outdoor field robotics," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 2345–2352.

[35] D. Bazzi, G. Priora, A. M. Zanchettin, and P. Rocco, "Rrt* and goal-driven variable admittance control for obstacle avoidance in manual guidance applications," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1920–1927, 2022.

[36] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, "Voxblox: Incremental 3d euclidean signed distance fields for on-board mav planning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 1366–1373.

[37] Microsoft, "azure_spatial_anchors_ro," https://github.com/microsoft/azure_spatial_anchors_ros, 2020.

[38] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: Association for Computing Machinery, 1996, p. 303–312.

[39] D. Werner, A. Al-Hamadi, and P. Werner, "Truncated signed distance function: Experiments on voxel size," in *Image Analysis and Recognition: 11th International Conference, ICIAR 2014*. Springer-Verlag, 2014, p. 357–364.

[40] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Human mental workload*, vol. 1, no. 3, pp. 139–183, 1988.

Authorized licensed use limited to: New York University. Downloaded on December 01,2024 at 09:13:12 UTC from IEEE Xplore. Restrictions apply.