Enhancing Remote Collaboration Through Drone-Driven Agent and Mixed Reality

Shihui Xu(✉), Like Wu, Wenjie Liao, and Shigeru Fujimura

Graduate School of Information, Production, and Systems, Waseda University, Kitakyushu, Japan
shxu@toki.waseda.jp, wulike@fuji.waseda.jp, jie3040@akane.waseda.jp, fujimura@waseda.jp

Abstract. Mixed Reality (MR) and Augmented Reality (AR) technologies have been used to improve remote collaboration. However, existing MR- or AR-based remote collaboration systems lack of a fully independent view sharing between the local user and remote user. This research propose a novel approach to enhance the remote collaboration using a drone and MR technology. By augmenting a virtual 3D avatar on the drone in the local environment, we propose the drone-driven agent to embody the remote user. And the view sharing between local and remote user is achieved by sending a real-time video stream of the local environment captured with the drone. There are three novelties including 1) fully independent view sharing, 2) augmenting virtual character on the drone to embody remote user, and 3) 3D AR sketching o facilitate communication between local and remote users. We implemented a proof-of-concept prototype to illustrate our design using a see-through type head-mounted display and a small-size drone. In addition, we provide discussion and implication for the future work to design drone-based remote collaboration systems.

Keywords: Remote collaboration · Mixed reality · Augmented reality · Drone · Embodiment using character

1 Introduction

Mixed Reality (MR) and Augmented Reality (AR) technologies have been verified to be effective in remote collaborative work. Many researchers proposed architectures, models, methods, and systems to improve task performance and user experience of remote collaboration [1–5,21]. However, in existing MR- or AR-based remote collaboration systems, the remote users heavily rely on the local users. For instance, most systems often employ cameras worn by local users to transmit the local environment, thereby limiting remote users to the first-person perspective of the local users. How to provide remote users with greater freedom remains unsolved.

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2024, LNCS 14707, pp. 116–127, 2024.
https://doi.org/10.1007/978-3-031-61044-8_9

A drone, also known as an unmanned aerial vehicle (UAV), is an aircraft that operates without any humans on board [6]. Drones are increasingly used to support industrial and agricultural tasks, aerial telepresence, remote collaborative work, and social interactions, especially events that involve dangerous or costly tasks [7]. For example, companies like Amazon and Google are testing drones for delivering packages, in particular in remote or hard-to-reach areas to reduce delivery times and costs [8,9]. Firefighters employed drones to monitor the progression of the Notre Dame fire and to identify optimal locations for directing fire hoses [10].

To address the problem that remote users are heavily depended on the local users in the remote collaboration systems, we propose to combine Drone and MR technology, enabling the remote user to have a drone-driven agent in the local environment. The remote user can have a independent view of local environment through drone. Besides, we embodied the drone-driven agent using MR to enhance the collaboration.

In this research, we explore how the MR technology and a drone can enhance remote collaboration, especially for the physical tasks. Compared with prior remote collaboration system using MR or drones, the novelties of our system locate in three aspects:

1. View Independence: The perspective of remote users typically relies on the viewpoint of local users. For instance, current systems often employ cameras worn by local users to transmit the local environment, thereby limiting remote users to the first-person perspective of the local users. Remote user can detach themselves from the viewpoint of local users and utilize camera-equipped unmanned aerial vehicles (UAVs) to explore the local environment. This enables remote users to have enhanced mobility and greater freedom in navigation.
2. Embodiment of Remote User: In previous remote collaboration systems, it was challenging to depict remote users due to the difficulty of representing their actions. Recent research has explored the use of augmented reality (AR) technology to convey the head movements of remote users to local users. However, such embodied representation is limited to rotational movements in the original position and fails to depict displacement or movement. This study employs unmanned aerial vehicles (UAVs) as a medium for instantiating remote users. We utilize AR technology and UAVs as carriers to represent the navigation and manipulation of remote users within the local environment, aiming to enhance the sense of presence between users.
3. Communication Cues: Sharing 3D sketching is challenging in remote collaboration because of spatial mismatching. With the drone as a spatial medium, our system implements AR 3D sketching as communication methods in addition to auditory and visual channels. Remote users can convey information to local users through sketching, facilitating improved collaboration between them.

The main contributions of our work include:

1. The design and implementation details of a remote collaboration system that combines MR with drone technologies, including the framework and specific interaction methods.
2. Providing discussions for designing future remote collaboration systems using MR and drone technologies.

In the following parts of this paper, a related literature review is firstly provided. Then we describe the system design and implementation details in Sect. 3 and Sect. 4. In Sect. 5, we discuss the findings and provide some implications for future work. Finally, we give a conclusion of this paper in Sect. 6.

2 Related Works

We built our research on several areas of research domains, in terms of MR/AR-based remote collaboration, embodiment in remote collaboration, and collaborative experience using drones.

2.1 MR/AR-Based Remote Collaboration

Remote collaboration is the process of working together to achieve a specific goal from geographically separated locations, usually using digital techniques to support communication and coordination [21]. With the rapid development of Virtual Reality (VR), AR, and MR technologies, there are many compact devices, such as Microsoft's HoloLens[1], Meta Quest[2], Magic Leap[3], HTC Vive[4], Oculus Rift[5], etc., that benefit the collaboration process and attract the attention of academia and industry, yielding positive results. For instance, in terms of perspective sharing and environmental perception, traditional systems utilize 2D video to share the first-person perspective of local users, which heavily relies on the local users and limits the field of view for remote users, resulting in inefficient information acquisition for remote users. To solve the problem, Lee et al. [1] has employed panoramic cameras to share the viewpoint of local users and employed VR technology to visualize it for remote users, so that the remote user can observe the surroundings of local users, allowing perspective independence of local users to a certain extent. In research of Teo et al. [4], the scanned local environmental model is shared with remote users, enabling them to have the imagery of the local environment. However, the scanned models are typically coarse and may even hinder remote users. To address the issue of perspective limitations caused by mobility constraints, researchers have employed Mobile

[1] https://www.microsoft.com/en-us/hololens/.
[2] https://www.meta.com/quest/.
[3] https://www.magicleap.com/en-us/.
[4] https://www.vive.com/us/.
[5] https://www.oculus.com/rift-s/.

Robotic Telepresence (MRP) systems to enable remote users to physically move within the same range as local users. MRP systems integrate video conferencing equipment onto mobile robot devices that can be remotely controlled and navigated [11,12].

2.2 Embodiment in Remote Collaboration

Embodiment refers to the representation of remote users in a way that provides a sense of physical presence in the collaborative environment. It involves creating virtual or augmented substitutes for remote users that facilitate them to interact and engage with others in collaborative work. Embodiment techniques may change how individuals present themselves and have a significant influence on experiencing the physical distance between remote participants. Previous research has shown that the embodiment of users can improve co-presence and social presence in remote collaboration [13].

Many researchers have shed light on embodying remote users with humanoid avatars. Piumsomboon et al. [2] presented Mini-Me, an adaptive avatar that can reflect the remote user's gaze direction and body gestures, to improve the user experience in MR remote collaboration. Yoon et al. [13] conducted user studies to investigate the effect of avatar appearance on social presence and user perception in an AR remote collaboration system. And they found that a realistic whole-body avatar was rated best by the users. In addition to embodying remote users with virtual or AR avatars, some research focused on the physical embodiment. Paulos and Canny proposed Personal Roving Presence [14], which is one of the earliest contributions in this category. Many studies have followed, ranging from tabletop-sized robots [15] to human-scale devices capable of physical movement [16]. These robots have been widely used in home care, education, and working space. Bae [17] introduced "Avatar Drone" to illustrate the potential of a drone as a medium of embodiment. But these systems require a high cognitive load for remote users to control them manually. Recently, Sakashita et al. [18] proposed RemoteCoDe to track a remote user's attention using a depth camera on the smartphone and render it to a local articulated display. However, it is fixed on the desktop, limiting the mobility and navigation of remote users.

2.3 Collaborative Experience Using Drones

How to interact with drones and make use of drones to enhance interaction between humans has garnered the interest of the HCI community. Compared with other telepresence robots, the drone has the potential to provide mobility and navigation beyond the reach of human beings, which enhances collaborative work. Jones et al. [19] investigated the use of semi-autonomous drones for video conferencing, enabling an indoor desktop user to explore the outdoor environment from the drone's perspective. A view that is decoupled and manipulated was provided for the desktop user. Sabet et al. [7] extended Jones et al.'s work by expanding collaborative sharing from one-to-one to multiple users, employing panoramic cameras to capture outdoor environments. Teledrone [20] is another drone-based video conferencing system to foster shared outdoor activities over distance. Although there are existing collaboration systems using drones, the level of interaction and manipulation is relatively low, which is insufficient for complex tasks. In their systems, the local users manipulate the drone, and the video captured by the drone will be transmitted to remote users. The remote users can only view the received video.

We focus on remote collaborative physical tasks in industrial scenarios. This type of task requires relatively frequent and complex interactions. We aim to explore further the potential of drones in remote collaboration.

3 System Design

In this section, the design of our remote collaboration system using drone and MR technology is described.

3.1 System Overview

The system overview is shown in Fig. 1. The local worker wears an MR head-mounted display (HMD) and performs physical tasks indoors. The remote expert operates a PC to control the drone which is located with the local worker. The drone comes with a camera that can capture video of the local environment, and our system streams real-time video of the local environment to the remote PC. The remote expert can look around the local environment by controlling the navigation of the drone, independent of the local worker's perspective. The remote expert can sketch on the PC to give an illustration. Our system will transfer the sketch to the local worker and visualize the sketch in the real local environment using MR technology. Besides, we propose a new method to embody the remote expert in the local environment to increase the co-presence of both users and enhance collaboration. We propose a drone-driven agent, which combines drone and MR technology. We superimpose a virtual 3D character on the drone using MR technology, embodying the remote expert from both physical and virtual aspects.

Use Case. A local user, usually a novice without much experience, needs to perform some physical tasks, such as common assembly tasks or finding tasks in factories, or visual inspection tasks. Due to a lack of experience, the local user may encounter some obstacles and difficulties that make the task difficult to proceed. At this time, the local user can use our system to connect with a remote user who is usually an experienced expert, and work together to complete the task. The local user wears HoloLens to enter our system. There is a drone embedded with a camera in the local environment. By controlling the drone, the remote user can obtain local environmental information and the work status of the local user. We provide remote users with an interface to operate the drone, and the remote user can adjust the rotation angle of the drone, and make the drone fly up, down, left, and right to visit the surroundings of the local

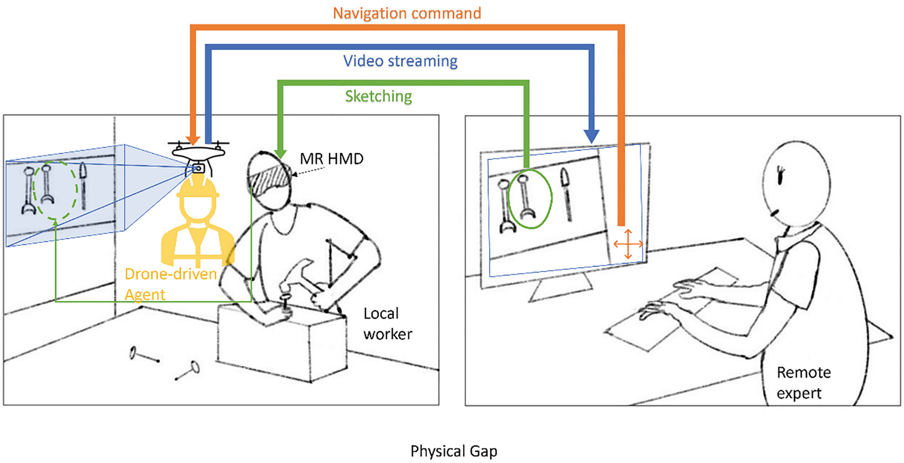


Fig. 1. System overview.

environment. Remote users can also make guidance and instructions to local users by marking and sketching on the PC interface. We will synchronize these guidance and instructions with the local user and use AR technology to display them in 3D in the local environment. Local users can see 3D marks and sketches in HoloLens. We also use a drone as a medium to embody remote users in local environments. The local user can see that at the location of the drone, there is an avatar of the remote user. This is done to enhance social presence.

3.2 Independent View Sharing Using Drone

The drone, equipped with a high-definition camera, provides a unique aerial perspective of the local environment, which is fundamentally different from the first-person viewpoint of the local worker. This capability allows the remote expert to navigate and explore the local environment independently, without being constrained to the local worker's field of view. Our system provides a manipulation interface for the remote user to control the drone in real-time, ensuring a comprehensive and dynamic assessment of the local environment. Figure 2 shows the interface for the remote user. The main image is the view captured by drone. The right-bottom small image is the top view of drone's route. The remote user can make the drone fly forward, backward, left, right, up, and down and make the drone rotate to a degree by pushing buttons on the keyboard. The route of the drone is visualized using a top view with coordinate values. The green point stands for the current position of the drone and the red line stands for the route that the drone has traveled.

Fig. 2. Third person view of drone and corresponding interface for remote user.

3.3 Embodiment of Remote User Using Drone and Mixed Reality

The drone acts as a physical proxy for the remote expert in the local environment. Through mixed reality, we overlay a half-body 3D avatar onto the drone, representing the remote expert's presence. The head position and rotation of the 3D avatar are aligned with the drone's camera. The 3D avatar can move according to the manipulation of the remote user. This feature enhances the sense of presence and engagement, allowing the local worker to interact with the remote expert as if physically co-located, as Fig. 3 shows. The embodiment through a drone-driven agent provides a more intuitive and natural way for the local and remote users to collaborate, especially in complex tasks requiring precise coordination.

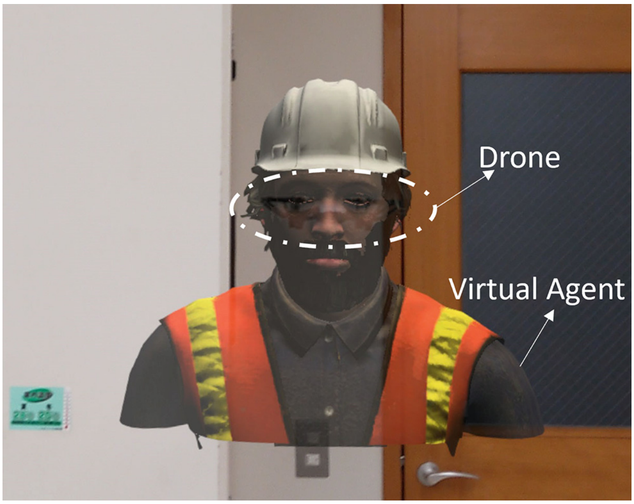


Fig. 3. Drone-driven agent.

3.4 Communication Cues

Apart from traditional audio-visual communication channels, our system incorporates advanced AR-based techniques like 3D sketching and gaze tracking. The drone is a medium to transfer the communication cues between the remote expert and the local worker. For instance, the remote expert can draw sketches or annotate objects in the desktop interface using keyboard, which are then rendered in real-time in the local environment through the MR head-mounted display worn by the local user. This method provides a more interactive and effective way to convey complex instructions or feedback. Additionally, the system includes gaze tracking technology to indicate the remote expert's focus, offering a more intuitive understanding of their intentions and enhancing collaborative efficiency.

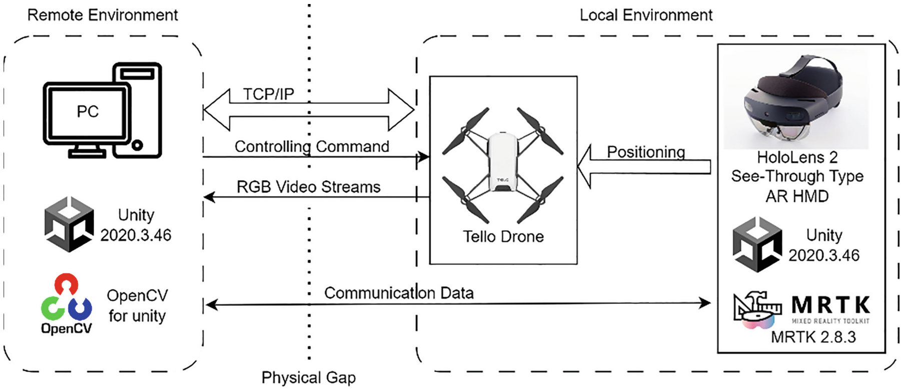


Fig. 4. Implementation Framework

4 Implementation

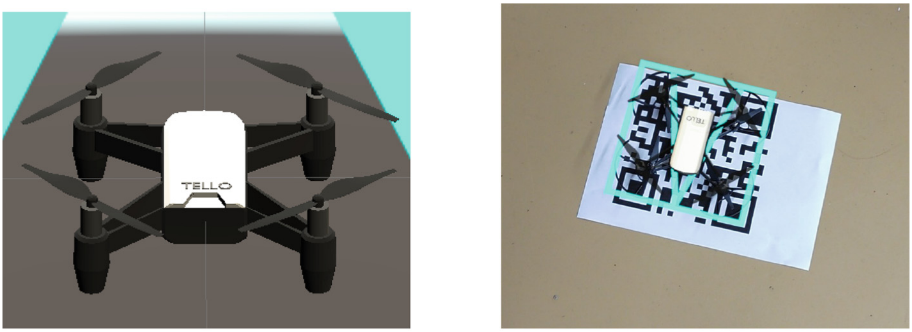
The implementation of our prototype system includes three components:

- A local side where the user wears an MR head-mounted display to receive and follow the guidance to accomplish tasks.
- A drone that is located on the local side but acts as an embodiment of a remote user and can capture the local environment in real time.
- A remote PC side where the user can access the video captured by the drone synchronously and control the navigation of the drone by manipulating the PC.

Fig. 4 shows the implementation framework of our prototype system. Overall, the implementation is divided into remote side and local side with hardware and software.

The hardware used in the prototype includes Microsoft HoloLens 2 which is a see-through-type MR head-mounted display, a PC with Intel i9 CPU and 32 GB RAM, and a Tello drone which is embedded with a micro camera[6]. The camera features a 5-megapixel photo resolution and an 82.6-degree field of view, capable of recording HD720P30 video in MP4 format. Additionally, it supports electronic image stabilization to ensure steady footage. We chose Tello because it is small, its size is only $98 \times 92.5 \times 41$ mm, and its weight is only about 80 g, which is very suitable for indoor flying such as industrial physical tasks.

The prototype system was developed based on the Windows 10 operating system using the Unity 3D[7] 2020.4 game engine. The MixedRealityToolkit[8] was used to develop the local user's MR application of HoloLens and the remote user's manipulation application. The OpenCV for unity was used to visualize the sketching communication cues. The communication between remote and local environment was done by TCP/IP protocol, including controlling command from remote side to local drone, RGB video streams from drone to remote PC, and communication data between the remote and local sides.



(a) 3D model of Tello drone. (b) Tracked QR code.

Fig. 5. Calibration of coordination using QR code.

The coordination of the drone, local MR environment, and remote manipulation need to be calibrated. To calibrate the drone and local MR environment, we use a QR code that is placed on the floor of the local physical world. By scanning the QR code, the MR device knows the position of the QR code, and a virtual 3D model of the Tello drone is shown on the QR code. Then we place the drone onto the QR code to make the QR code the origin of coordination (Fig. 5).

[6] https://www.ryzerobotics.com/.
[7] https://unity.com.
[8] https://github.com/microsoft/MixedRealityToolkit.

5 Discussion and Future Work

In this paper, we present a remote collaboration system using drone and MR technology. We propose the drone-driven agent to embody the remote user in the local environment and share the local environment through a drone-embedded camera to provide a fully independent view sharing. The communication cues are provided to facilitate instruction and communication between collaborators. By combining drone and MR technology, we intend to explore a new direction of remote collaboration.

The view sharing using drone provides a more independent way of sharing compared with previous camera based sharing. However, the manipulation of drone from remote space is sometime not stable because of the delay of network. In addition, the drone used in this paper is a small-size drone, and its flight direction is easily affected by airflow and other factors in the environment. Future work needs to consider the above issues and overcome them when designing remote collaboration systems with drones. For example, the drone-involved system can work well in the indoor workspace, but may not be as good outdoors.

Augmenting a 3D virtual character to the drone provides a new embodiment method of remote user. However, the alignment of virtual character and real drone is depend on the situation of drone. In our system, the position of drone and virtual character is calculated by command from the remote user. Future system can consider to detect the drone directly so that the alignment will be more precise.

Our current system primarily focuses on one-to-one collaboration scenarios. Future work should enable multiple users to collaborate simultaneously. This could involve multiple drones or advanced MR techniques to represent multiple remote users in the local environment.

For the safety concerns, as drones become more integrated to the collaborative workspace, there are issues that drones may collide with physical objects. Future systems should consider intelligent security mechanisms to protect users from hurt.

Formal user studies are needed to evaluate drone-driven agents in remote collaboration systems. To measure the effectiveness and user acceptance of drone-driven agents in remote collaboration environments, comprehensive user research is needed. User research should evaluate the technical performance of the system and should also collect user feedback on usability, engagement, and overall satisfaction to guide further improvements to the system to make it more user-friendly and efficient in practical applications.

6 Conclusion

In summary, this paper proposes a new remote collaboration system integrating drones and mixed reality technology. Our system provides remote users with greater independence and enhanced presence in the local environment through augmented drone using a 3D virtual avatar. Drone-driven agents and independent view sharing via drone-embedded cameras help remote collaborative work a lot, especially for physical tasks.

Although combining MR and drone technologies has great potential, our system still faces some challenges, such as network latency, drone environment limitations, and user interface. Future research is needed to refine the system and expand its applicability to these challenges. For future work, we plan to conduct a formal user study to evaluate the proposed system.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Lee, G.A., Teo, T., Kim, S., Billinghurst, M.: Mixed reality collaboration through sharing a live panorama. In: SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications, New York, pp. 1–4. ACM (2017)
2. Piumsomboon, T., et al.: Mini-me: an adaptive avatar for mixed reality remote collaboration. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, New York, pp. 1–13. ACM (2018)
3. Bai, H., Sasikumar, P., Yang, J., Billinghurst, M.: A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In: Proceedings of the 2020 CHI Conference on Human Factors in computing systems, New York, pp. 1–13. ACM (2020)
4. Teo, T., Lawrence, L., Lee, G. A., Billinghurst, M., Adcock, M.: Mixed reality remote collaboration combining 360 video and 3d reconstruction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, New York, pp. 1–14. ACM (2019)
5. Kim, S., Lee, G., Huang, W., Kim, H., Woo, W., Billinghurst, M.: Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, New York, pp. 1–13. ACM (2019)
6. Hassanalian, M., Abdelkefi, A.: Classifications, applications, and design challenges of drones: a review. Prog. Aerosp. Sci. **91**, 99–131 (2016)
7. Sabet, M., Orand, M., McDonald, D.W.: Designing telepresence drones to support synchronous, mid-air remote collaboration: an exploratory study. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, New York, pp. 1–17. ACM (2021)
8. Amazon Prime Air prepares for drone deliveries. https://www.aboutamazon.com/news/transportation/amazon-prime-air-prepares-for-drone-deliveries. Accessed 24 Dec 2023
9. Drones Survey Services. https://dronesurveyservices.com/do-drones-deliver-packages/. Accessed 24 Dec 2023
10. DJI drones helped track and stop the Notre Dame fire. https://www.theverge.com/2019/4/16/18410723/notre-dame-fire-dji-drones-tracking-stopped-thermal-cameras. Accessed 24 Dec 2023

11. Biehl, J.T., Avrahami, D., Dunnigan, A.: Not really there: understanding embodied communication affordances in team perception and participation. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, New York, pp. 1567–1575. ACM (2015)
12. Kristoffersson, A., Coradeschi, S., Loutfi, A.: A review of mobile robotic telepresence. In: Advances in Human-Computer Interaction 2013, p. 3 (2013)
13. Yoon, B., Kim, H., Lee, G.A., Billinghurst, M., Woo, W.: The effect of avatar appearance on social presence in an augmented reality remote collaboration. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), New York, pp. 547–556. IEEE (2019)
14. Paulos, E., Canny, J.: PRoP: personal roving presence. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, pp. 296–303. ACM(1998)
15. Neto, I., Johal, W., Couto, M., Nicolau, H., Paiva, A., Guneysu, A.: Using tabletop robots to promote inclusive classroom experiences. In: Proceedings of the Interaction Design and Children Conference, New York, pp. 281–292. ACM (2020)
16. Gealy, D.V., et al.: Quasi-direct drive for low-cost compliant robotic manipulation. In: 2019 International Conference on Robotics and Automation (ICRA), New York, pp. 437–443. IEEE (2019)
17. Bae, I. H.: Avatar drone: drone as telepresence platform with 3D mobility. In: 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), New York, pp. 452–453. IEEE (2016)
18. Sakashita, M., Ricci, E. A., Arora, J., Guimbretière, F.: RemoteCoDe: robotic embodiment for enhancing peripheral awareness in remote collaboration tasks. In: Proceedings of the ACM on Human-Computer Interaction, New York, vol. 6, no. CSCW1, pp. 1–22. ACM (2022)
19. Jones, B., et al.: Elevating communication, collaboration, and shared experiences in mobile video through drones. In: Proceedings of the 2016 ACM Conference on Designing Interactive Systems, New York, pp. 1123–1135. ACM (2016)
20. Shakeri, H., Neustaedter, C.: Teledrone: shared outdoor exploration using telepresence drones. In: Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, New York, pp. 367–371. ACM (2019)
21. Lee, G., Kang, H., Lee, J., Han, J.: A user study on view-sharing techniques for one-to-many mixed reality collaborations. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), New York, pp. 343–352. IEEE (2020)