Extending Embodied Interactions
in Mixed Reality Environments

Mohamed Handosa, Hendrik Schulze, Denis Gračanin(✉), Matthew Tucker,
and Mark Manuel

Department of Computer Science, Virginia Tech, Blacksburg, VA 24060, USA
gracanin@vt.edu

Abstract. The recent advances in mixed reality (MR) technologies provide a great opportunity to support deployment and use of MR applications for training and education. Users can interact with virtual objects that can help them be more engaged and acquire more information compared to the more traditional approaches. MR devices, such as the Microsoft HoloLens device, use spatial mapping to place virtual objects in the surrounding space and support embodied interaction with those objects. However, some applications may require an extended range of embodied interactions that are beyond the capabilities of the MR device. For instance, interaction with virtual objects using arms, legs, and body almost the same way we interact with physical objects. We describe an approach to extend the functionality of Microsoft HoloLens to support an extended range of embodied interactions in an MR space by using the Microsoft Kinect V2 sensor device. Based on that approach, we developed a system that maps the captured skeletal data from the Kinect device to the HoloLens device coordinate system. We have measured the overall delay of the developed system to evaluate its effect on application responsiveness. The described system is currently being used for the development of a HoloLens application for nurse aide certification in the Commonwealth of Virginia.

Keywords: Mixed reality · User gestures · Tracking

1 Introduction

Mixed Reality (MR) technologies allow for providing users with an environment that blends the physical surroundings with virtual objects. In order to support user interaction in such environment, there should be means to capture user inputs. Different MR devices tend to support user interaction in a variety of ways, where a user may provide input using physical controls, voice commands, and/or gestures. For instance, the ODG MR-glasses [22] has several input capabilities including on-device buttons and trackpad, a Wireless Finger Controller (WFC) with motion/gesture functionality, and a Wireless Bluetooth Keyboard with multifunction command keys. Another example is the Microsoft HoloLens

device [20], which uses spatial mapping to place virtual objects in the surrounding space and supports interaction with those objects through voice commands or gaze and air-tap. In gaze and air-tap interaction, the user gazes at the object of interest before making the air-tap gesture to trigger an action. Alternatively, the user may trigger the action by pressing a single-buttoned Bluetooth device, namely clicker.

The HoloLens' support for gesture-based input is very limited as it can recognize only two predefined gestures; air-tap and bloom. In several interaction scenarios, users might prefer to interact with virtual objects in a more engaging manner using their limbs almost the same way they interact with their physical counterparts. Unfortunately, this kind of embodied interaction is not supported by the current MR devices. However, support for user interaction in MR environments can be extended with the aid of other devices. Compared to the use of gaze and air-tap or the use of voice commands to interact with virtual objects, the support for embodied interaction can provide a more natural way to interact with the surroundings and allows for developing a rich user interface.

The remainder of the paper is organized as follow. Section 2 provides the related work. Section 3 explores the limitations of MR devices constraining their ability to support user interaction in MR environments. In Sect. 4, we present an approach to extend user interaction in MR environments. Based on that approach, we developed a system that integrates Microsoft HoloLens and Kinect devices. Section 5 demonstrates a case study, where the developed system enables user interaction in an MR space using different body joints. System performance (latency) is discussed in Sect. 6 while Sect. 7 concludes the paper.

2 Related Work

Virtual reality (VR) and Augmented Reality (AR)/MR technologies [5] technologies/applications are becoming more affordable and thus more accessible to general users. While there are over forty years of research in this area, we still need more findings to better understand challenges when it comes to developing MR applications. For example, in the education domain, MR [15]/AR [2], and VR [6] applications have been used mostly in higher education (science, humanities, and art), unlike vocation education. In the medical/health domain, MR has been mostly used for medical treatment, surgery, rehabilitation, education, and training, but not universally to other medical fields [7].

Some of the challenges in creating AR systems were explored by Dunleavy et al. [10]. They studied the limitation and affordances of an MR system. The results showed that although using MR system could significantly increase student engagement, there are still hardware and software issues. The advantages of MR include learning gains, motivation, interaction, and collaboration. Better learning performance, motivation, and engagement demonstrate the effectiveness of MR [2]. However, there is a need for longitudinal studies to study the evolution of knowledge and skills over time and to inform about the suitability of MR for supporting significant learning.

While being less immersive may be an inherent problem for MR technology, meanwhile it also proposes an interesting question for how we can expand the application scope for full utilization of this technology. One of the main challenges of MR is the limited field of view. As such, how to visualize large chunks of (or big) data is questionable [28]. With technology progressing forward, it is expected that the field of view can be enlarged even beyond of the human field of view in the near future [27,32] so this roadblock can be alleviated.

Another challenge of MR is that while it offers more naturalist interaction experiences [24], it is less immersive compared to the virtual reality technology. Hence, users can be distracted by environmental and other related factors leading to adverse impacts on usability [18]. This implies that MR is not a universal suitable technology for every different application.

MR systems inherently depend on the surrounding space to support user interactions. Our cognitive processes are dependent on how our body interacts with the world (affordances) [16,17] and how we off-load cognitive work onto our physical surrounding (embodied cognition) [30].

Embodied interactions demonstrate the importance of the body's interactions with the physical world. Interaction of our body and the surrounding physical world affect our cognitive processes and embodied cognition [31]. Embodiment cognition leverages the notion of affordances, potential interactions with the environment, to support cognitive processes [8]. Embodied interaction [8] and embodied user interfaces [11] lead towards invisible user interfaces and move the computation from desktop computers to physical space and place [9].

MR can also be used in a collaborative setting. Billinghurst and Kato explored the notion of functional and cognitive seams in collaborative MR systems [4] and reviewed MR techniques for developing collaborative interfaces. These results are reflected in collaborative and standalone MR applications [12]. Sharing gaze, emotions, and physiological cues can enhance collaboration in MR [25] and affect education outcomes [21] and collaboration [26]. The appearance of avatars/virtual agents can affect critical emotional reactions in interpersonal training scenarios as well as users' perceptions of personality and social characteristic [29].

3   Problem Definition

The main goal of MR is to enrich the actual physical environment with digital (virtual) entities. To achieve full immersion, the MR environment should react to user's behavior appropriately and the interaction should be as natural and intuitive as possible. The spatial awareness of an MR device, such as the Microsoft HoloLens device [20], allows a great degree of freedom regarding recognition, movement, and exploration of confined spaces and physical objects, enriched with virtual objects. However, the interaction capabilities of the HoloLens are limited by multiple factors which play an important role regarding the natural feeling to an immersive environment.

The interaction concept is based on voice commands, gaze tracking, hand tracking, and hand gestures. This concept yields the limitations of the device.

We focus on the limitations induced by gaze tracking, hand gestures, and hand tracking. For gaze tracking, the HoloLens uses its orientation to indicate its user's gazing direction. This assumption is not always true as a user may gaze at different directions while maintaining the same head orientation. Therefore, a virtual cursor is usually utilized to help the user perceive the gazing direction assumed by the HoloLens. Adding such extra virtual objects to an MR scene may not be the best way to support natural and intuitive interaction.

For hand gestures, the HoloLens supports only two core gestures, namely bloom and air-tap. The bloom gesture is reserved by the system to perform predefined special actions, such as showing/hiding the start menu and exiting from a running application. This limits the supported gestures when interacting with a HoloLens application to one gesture only, the air-tap, which is a transition between two recognizable hand states, ready and press as shown in Fig. 1.

Ready state   Press State
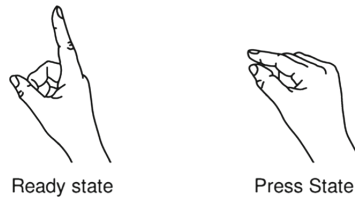
Fig. 1. Air-tap gesture, a switch from the ready to the press state [19].

For hand tracking, the HoloLens can track a hand position only if it is in the ready state (see Fig. 1), a closed fist with the index finger pointing up. Consequently, a user must maintain the ready hand state to enable hand tracking, which might be inconvenient, especially for long interaction scenarios. Moreover, the HoloLens cannot discriminate between left and right hands. In fact, the HoloLens treats a tracked hand as a disjoint object floating in space with no information about its side nor whether it belongs to the user or not. Consequently, the HoloLens may track the hand of a person other than the user and trigger actions accordingly, which can cause interaction conflicts in collaborative spaces with multiple users wearing HoloLens devices.

Both gesture recognition and hand tracking require the user's hand to be within the HoloLens' field of view. Additionally, gaze tracking follows the HoloLens' orientation rather than the actual gaze direction of the user, requiring the user to adjust the head orientation towards the object of interest rather than simply gazing at it. Both of these preconditions add a limitation to the possible *space of interaction* and create the need for not necessarily natural behavior patterns in order to interact with objects in a given environment. The lack of custom gestures recognition and full-body tracking (or at least discriminating left and right hands) limits the possible *range of interaction*. Natural interaction patterns such as interaction with both hands at the same time or interaction with multiple objects simultaneously are not (or only to a certain degree) possible.

Furthermore, the requirement of using unnatural hand gesture (ready state) in order to activate hand tracking and interact with the surroundings can greatly affect the *quality of interaction* as it interrupts the immersive experience and poses a difficulty to overcome, especially for users which are inexperienced with the usage of these gestures.

Integrating additional tracking devices, such as the Kinect V2 device, to observe the HoloLens user enables full-body tracking and identification of different body parts. Utilizing such information allows for developing more complex interaction schemes, involving multiple body parts and a higher level of detail for a broader range of recognizable gestures. For example, the recognition of the entire skeleton allows interaction with objects outside of the HoloLens' field of view and interaction with both hands at the same time. Moreover, interaction is not restricted to gestures performed with hands but can be extended to other body parts. The skeletal information, in combination with the spatial awareness of the HoloLens, allows for inferring contextual information from natural body movements.

4   Proposed Approach

We rely on tracking devices to capture the body movements and gestures of the user on behalf of the MR device. Providing such information allows MR applications to overcome the limitations of the MR devices and support rich interaction scenarios. Before MR devices can benefit from user tracking information, this information should be mapped from the tracking device coordinate system to the MR environment coordinate system. Registering two coordinate systems can be achieved by collecting a set of point pairs. Each pair consists of two corresponding points, one from each coordinate system. Once those points are collected, a registration algorithm can be applied to obtain a transformation matrix that maps a point from one world to the other. Several registration algorithms have been proposed such as the algorithm proposed by Besl and McKay [3] and the eight-point algorithm [14].

A system that integrates MR devices with user tracking devices is illustrated in Fig. 2. For each tracking device, there is a server application providing access to the tracking data provided by that device. The MR application should incorporate two modules, a client and a registration module. The client module is responsible for obtaining the tracking data from the server over the network interconnecting them while the registration module is responsible for mapping obtained data to the coordinate system of the MR device. Following this architecture, MR devices can obtain data from several tracking devices and a tracking device can provide data to several MR devices.

Some tracking devices can track several users simultaneously. Consequently, MR devices may receive several tracking data sets for several persons. In that case, the MR device may need to identify which data set belongs to the user. Given that MR devices are head-mounted devices, the current location of a device in the MR coordinate system gives a good indication of the current location of
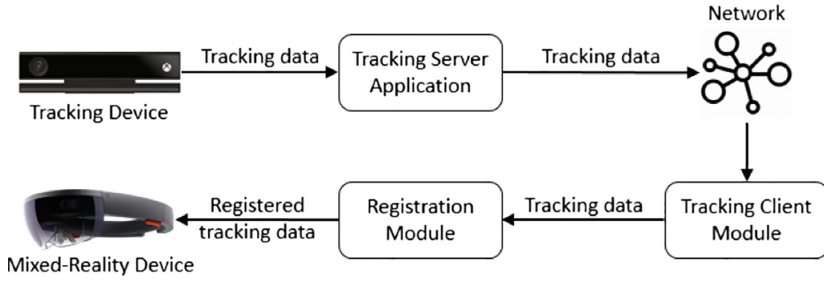
Fig. 2. Integrating MR devices and tracking devices.

the user's head. Comparing the device location with the registered tracking data sets can reveal which data set belongs to the user.

4.1   Implementation

Based on the proposed approach, we implemented a system that integrates HoloLens devices with Kinect devices. A Kinect server application tracks the user skeleton using the Kinect device. The HoloLens application obtains tracking data from the server through its Kinect client module before the registration module maps it to the HoloLens coordinate system using a transformation matrix. In order to obtain the transformation matrix, we have developed a four-step process (Fig. 3).
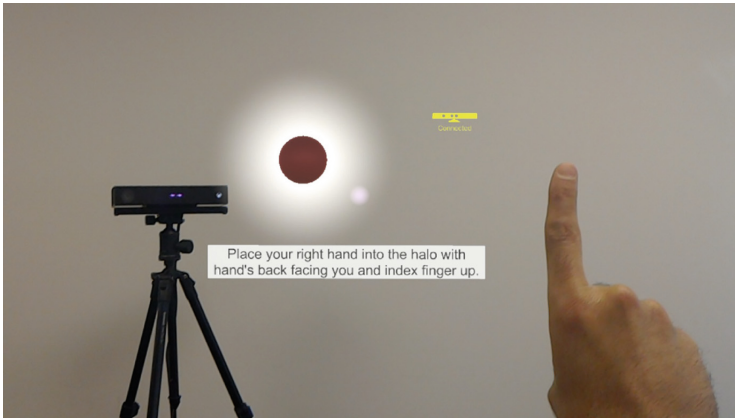
Fig. 3. Calibration process.

   The goal of each step is to collect two corresponding points, one from the Kinect coordinate system and another from the HoloLens coordinate system. The four-point pairs are collected by asking the user to place a hand at four

different positions in space that are indicated by virtual objects. Once the user's hand is in position, hand tracking information is collected from both Kinect and HoloLens to form a point-pair. Afterward, an algorithm is applied to obtain the transformation matrix and save it for later use. Figure 4 shows a HoloLens-rendered virtual skeleton aligned with the corresponding physical body.

Fig. 4. A registered skeleton aligned with the corresponding physical body.

4.2   Network Infrastructure

In order to communicate tracking data from the tracking server to the tracking client, we have tested two communication models, direct and indirect (Fig. 5). For direct communication, we use the User Datagram Protocol (UDP). The server has a predefined listening port to which clients can send subscription requests. The server collects tracking data from the Kinect device before sending it to all subscribing clients. This communication model minimizes the communication delay. However, for multiple-Kinect setup, a HoloLens will need to communicate with multiple servers. Establishing several connections with different servers complicates the networking model and makes network troubleshooting more challenging.

In order to support multiple-Kinect/multiple-HoloLens setups while minimizing the communication model complexity, we use the Message Queue Telemetry Transport (MQTT) protocol [1,13]. MQTT is a publish/subscribe communication protocol that relies on a broker to support indirect communication between publishers and subscribers. Each Kinect server can publish tracking data to a specific topic on the MQTT broker. Unlike direct communication, a client will need to maintain only a single connection with the MQTT broker. The client can subscribe to one topic or more to receive tracking data from one or more Kinect servers. Although, indirect communication may introduce increased delays, it allows for relaxing the complexity of the communication model. Furthermore, additional data sources, including environmental and biometric sensors, can be added.

Fig. 5. Two alternative communication models: (a) UDP-based direct communication and (b) MQTT-based indirect communication.

5 Case Study

In order to evaluate our approach in a non-lab environment, we used it in the development of a HoloLens application for Nurse Aide skills training [23]. The goal of the application is to augment the student's experience in classroom settings and to provide a rich set of educational contents in an MR environment. Students need to learn a set of skills with quite specific steps in a certain order. This requires not only the theoretical knowledge of how a specific skill needs to be done but also practical application in order to manifest the exact workflow required. With limited space and limited availability of hospital equipment in schools, the number of workstations to actually practice the skills are limited as well.

We developed a HoloLens application which recreates the scenery of hospital room. Within this virtual hospital room (Fig. 6a) are the required objects and props to perform the skills in a 'close to reality' environment. Figure 6b demonstrates an embodied interaction with digital entities, a denture, and a toothbrush. Our application guides the student through the steps of a skill and requires the student to perform specific and detailed interactions within the MR environment in order to proceed to the next step of a skill.

Fig. 6. (a) The virtual hospital room. (b) Demonstration of using both hands simultaneously for brushing a denture.

Almost all skills require at some point a more detailed user tracking than the HoloLens alone can provide. For example, a crucial part of proper hand washing requires the student to keep the hands and forearms at a downward angle to prevent 'contaminated' water to run down the arms. With the HoloLens alone, there is no possibility to check this condition. Another example is denture brushing where a student should hold a denture in one hand and a toothbrush in the other hand. With HoloLens alone, enabling hand tracking will require the student to maintain both hands in the ready state (Fig. 1) and within the HoloLens field of view, resulting in constrained and unnatural interaction. However, with the additional data about the entire user skeleton provided by the Kinect, we were able to achieve a level of detail and precision to track the user's actions sufficiently.

The application relies on HoloLens-based gesture recognition to support navigation through its menus and to adjust different application settings. However, once a given skill training is started, the application relies on Kinect data to support user interaction. As most of the students had little to no experience with MR devices and MR environments in general, it took some 'warm-up' time for the students to get used to the new experience and to move around and interact with virtual objects comfortably. While the gestures recognized using Kinect data were easy to learn, the original HoloLens gestures (air-tap) required some time to learn. After the initial learning phase, students were able to complete the skill training without or with minimal further guidance. Integrating the Kinect device into the system allowed the students to avoid the difficulties associated with the use of unnatural gestures, which helps reducing the threshold for users to comfortably interact with the virtual objects.

6   Results

Users should receive instant feedback as they interact with objects in an MR space. A noticeable delay in response to user commands can degrade the user experience dramatically. Therefore, the user's skeletal information should be delivered to the HoloLens with minimal latency to ensure the responsiveness of the system. The responsiveness is determined by the delay (latency) between the time at which the user makes a given gesture/move and the time at which the user receives the corresponding feedback through the MR device.

For the purpose of estimating the overall latency, we have captured multiple MR video recordings of user gestures, specifically the closed hand gesture. Exploring the frames of the captured videos revealed that it takes at most four frames for the HoloLens to provide a feedback after the user gesture takes place. Figure 7 shows six consecutive frames from a captured MR video (30 frames per second). The user starts with an open hand and the HoloLens displays a red box indicating that the hand state is open (Fig. 7a). The user starts to close the hand but it is not closed yet (Fig. 7b). The user's hand is closed (Fig. 7c). The user is waiting for the feedback (Fig. 7d, e). The user receives a visual feedback, where a green box is shown (Fig. 7f).
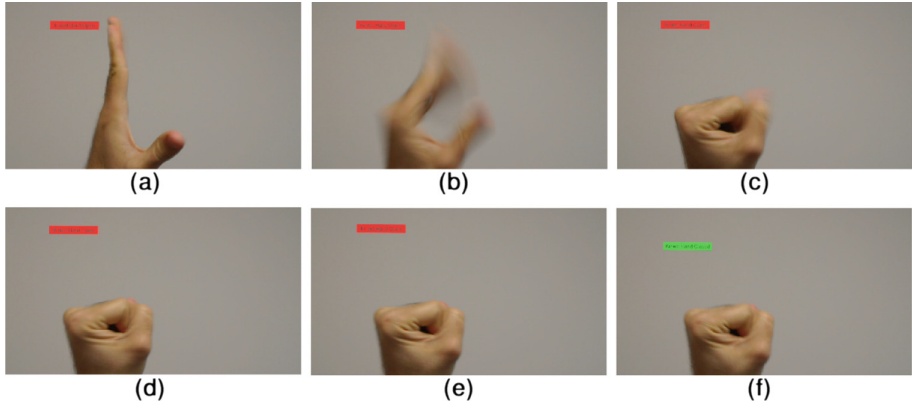
Fig. 7. Video recording consecutive frames for detecting a closed hand gesture using Kinect and providing feedback through HoloLens.

Assuming that frame *(b)* was captured at time 0 and frame *(c)* was captured at time T, then the user gesture takes place at time t_1, $0 < t_1 \leq T$. Similarly, if frame *(e)* was captured at time $3T$ and frame *(f)* was captured at time $4T$, then the feedback has occurred at time t_2, $3T < t_2 \leq 4T$. Consequently, the delay d is $2T < d < 4T$. The video frame rate is 30 frames per second and hence $T = 1/30$ s or 33.33 ms. Therefore, the total system delay ranges between 66 and 134 ms. This estimated latency is caused by the processing and communication steps that take place between a change in user skeleton state and providing the corresponding feedback.

The overall latency $d \geq d_1 + d_2 + d_3 + d_4$, where d_1 is the time it takes the Kinect device to capture a frame and send its data to the workstation; d_2 is the time it takes the workstation to extract skeleton information from the received frame data producing a skeleton information message; d_3 is the time needed to send the skeleton information message from the workstation to the HoloLens; and d_4 is the time it takes the HoloLens to provide a feedback based on the received skeleton information. Delays d_1 and d_4 are device specific and beyond our control. The measured average values of d_2 and d_3 (using UDP-based direct communication) are approximately 0.157 and 0.476 ms, respectively. Compared to the overall latency, both d_2 and d_3 are negligible.

The HoloLens can recognize the press gesture (Fig. 1). On the other hand, the Kinect device can recognize a closed hand gesture. Benefiting from the similarity between the closed hand gesture and the press gesture, we were able to measure the relative latency of the Kinect-based gesture recognition using the HoloLens-based gesture recognition as a reference point. Although the Kinect-based recognition involves several processing and communication steps, results have shown that its performance is comparable to that of the built-in HoloLens recognition. In fact, we noticed that the Kinect-based recognition can often perform faster than the built-in HoloLens gesture recognition. Figure 8 shows a
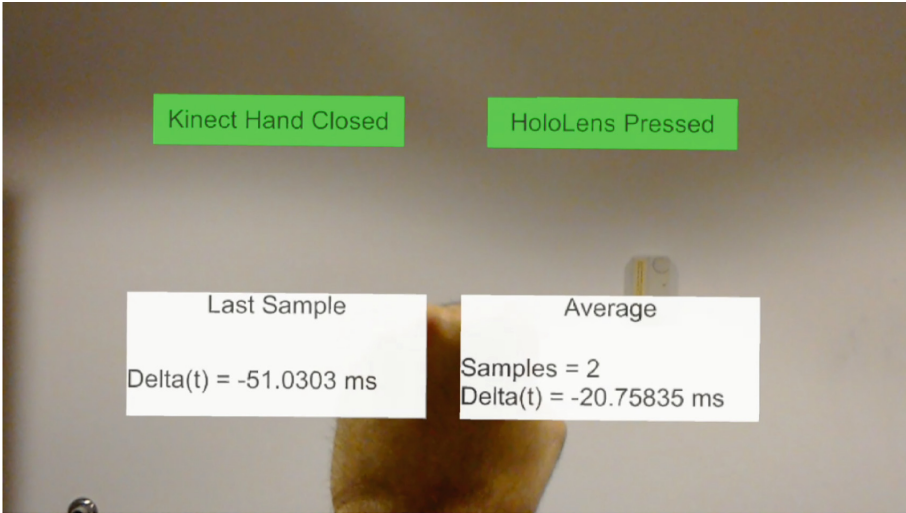
Fig. 8. A closed hand (or press) gesture recognized by both Kinect and HoloLens.

closed hand (or press) gesture, where the Kinect-based recognition outperformed the HoloLens-based recognition by approximately 51 ms.

Using MQTT-based indirect communication can simplify the communication model and make network troubleshooting less challenging, especially for multiple-Kinect/multiple-HoloLens setups. However, a significant increase in communication latency can degrade the user experience. The MQTT protocol supports three Quality of Service (QoS) levels. The message delivery for QoS-0, QoS-1, and QoS2 are at-most-once, at-least-once, and exactly-once, respectively.

Test results have shown that the average delay for the QoS-0, QoS-1, and QoS-2 are approximately 2.743, 28.492, and 36.047 ms, respectively. Although QoS-0 provides the smallest average delay, it allows for message dropping. However, this should not be a problem for applications that are interested in receiving the most recent tracking sample rather than receiving every tracking sample. Compared to the UDP-based direct communication average, the MQTT-based indirect communication with QoS-0 does not introduce a significant increase considering the overall delay of the system (Table 1).

Table 1. Average delay (latency) of UDP-based and MQTT-based communication.

	UDP	MQTT		
		QoS-0	QoS-1	QoS-2
Average delay (milliseconds)	0.476	2.743	28.492	36.047

7   Conclusion

MR devices provide an affordable opportunity to develop immersive applications. However, their limited input capabilities constrains the possible interactions. We presented an approach to overcome this limitation by integrating MR devices with tracking devices. An MR application can rely on the user tracking information to extend its ability to capture user inputs and to support interaction scenarios that were not possible before. Based on the proposed approach, we developed a system to integrate Microsoft HoloLens and Kinect devices. We presented a case study, where we utilized the developed system to support user interaction in an MR space using different body joints. The system allows users to interact with virtual objects and receive the corresponding feedback within 66 to 134 ms.

In order to support communication between Kinect devices and HoloLens devices, we have tested two communication models, UDP-based direct communication and MQTT-based indirect communication. Test results have shown that UDP-based communication introduces less average delay (0.476 ms) compared to MQTT-based communication with QoS-0 (2.743 ms). Although the MQTT-based communication introduces a slightly larger delay, it helps relaxing networking and communication complexities, especially for multiple-Kinect/multiple-HoloLens setups.

Current technical constraints and hardware limitations make a comprehensive solution that combines the multitude of required sensors with VR and MR devices difficult to achieve. However, there are many opportunities for combining multiple conventional and unconventional data sources (not only tracking devices) in a comprehensive framework. When the data sources provide overlapping coverage of the MR space, they can be used for internal data alignment, error correction, and increased accuracy.

References

1. Arlitt, M., Marwah, M., Bellala, G., Shah, A., Healey, J., Vandiver, B.: MQTT version 3.1.1 plus errata 01. Standard, OASIS, 10 December 2015
2. Bacca, J., Baldiris, S., Fabregat, R., Graf, S., Kinshuk: Augmented reality trends in education: a systematic review of research and applications. J. Educ. Technol. Soc. **17**(4), 133–149 (2014)
3. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
4. Billinghurst, M., Kato, H.: Collaborative mixed reality. In: Proceedings of the First International Symposium on Mixed Reality (ISMR 1999), pp. 261–284. Springer, Berlin (1999)
5. Bimber, O., Raskar, R.: Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters, Wellesley (2005)

6. Calvo, O., Molina, J.M., Patricio, M.A., Berlanga, A.: A propose architecture for situated multi-agent systems and virtual simulated environments applied to educational immersive experiences. In: Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Toledo Moreo, J., Adeli, H. (eds.) IWINAC 2017. LNCS, vol. 10338, pp. 413–423. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59773-7_42

7. Chen, L., Day, T.W., Tang, W., John, N.W.: Recent developments and future challenges in medical mixed reality. In: Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 123–135, October 2017

8. Dourish, P.: Where the Action is: The Foundations of Embodied Interaction. The MIT Press, Cambridge (2001)

9. Dourish, P.: Re-space-ing place: "place" and "space" ten years on. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW 2006), pp. 299–308. ACM, New York, 4–8 November 2006

10. Dunleavy, M., Dede, C., Mitchell, R.: Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. J. Sci. Educ. Technol. **18**(1), 7–22 (2009)

11. Fishkin, K.P., Moran, T.P., Harrison, B.L.: Embodied user interfaces: towards invisible user interfaces. In: Proceedings of the IFIP TC2/TC13 WG2.7/WG13.4 Seventh Working Conference on Engineering for Human-Computer Interaction, pp. 1–18. Kluwer, B.V., Deventer (1998)

12. Gračanin, D., Eck II, T., Silverman, R., Heivilin, A., Meacham, S.: An approach to embodied interactive visual steering: bridging simulated and real worlds. In: Proceedings of the 2014 Winter Simulation Conference (WSC), pp. 4073–4074, December 2014

13. Gračanin, D., Handosa, M., Elmongui, H., Matković, K.: An approach to user interactions with IoT-enabled spaces. In: Proceedings of the 14th International Conference on Telecommunications (ConTEL), pp. 139–146, 28–30 June 2017

14. Hartley, R.I.: In defense of the eight-point algorithm. IEEE Trans. Pattern Anal. Mach. Intell. **19**(6), 580–593 (1997)

15. HoloGroup: Education in mixed reality. http://holo.group. Accessed 6 Jan 2018

16. Kirsh, D.: Embodied cognition and the magical future of interaction design. ACM Trans. Comput.-Hum. Interact. **20**(1), 3:1–3:30 (2013)

17. Klemmer, S.R., Hartmann, B., Takayama, L.: How bodies matter: five themes for interaction design. In: Proceedings of the 6th ACM conference on Designing Interactive systems (DIS 2006), pp. 140–149. ACM Press, New York (2006)

18. Lebeck, K., Kohno, T., Roesner, F.: How to safely augment reality: challenges and directions. In: Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications, pp. 45–50. ACM, New York (2016)

19. Microsoft Corporation: How to perform an air tap. https://developer.microsoft.com/en-us/windows/mixed-reality/gestures. Accessed 2 Mar 2018

20. Microsoft Corporation: Microsoft HoloLens. https://www.microsoft.com/microsoft-hololens/. Accessed 2 Mar 2018

21. Noteborn, G., Carbonell, K.B., Dailey-Hebert, A., Gijselaers, W.: The role of emotions and task significance in virtual education. Internet High. Educ. **15**(3), 176–183 (2012)

22. ODG: Osterhout Design Group. https://www.osterhoutgroup.com/. Accessed 2 Mar 2018

23. Pearson: Virginia Nurse Aide: written (or oral) examination & skills evaluation, March 2017

24. Peters, E., Heijligers, B., de Kievith, J., Razafindrakoto, X., van Oosterhout, R., Santos, C., Mayer, I., Louwerse, M.: Design for collaboration in mixed reality: technical challenges and solutions. In: Proceedings of the 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), pp. 1–7, September 2016
25. Piumsomboon, T., Lee, Y., Lee, G.A., Dey, A., Billinghurst, M.: Empathic mixed reality: sharing what you feel and interacting with what you see. In: Proceedings of the 2017 International Symposium on Ubiquitous Virtual Reality (ISUVR), pp. 38–41, June 2017
26. Quarles, J., Lampotang, S., Fischler, I., Fishwick, P., Lok, B.: Collocated AAR: augmenting after action review with mixed reality. In: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2008), pp. 107–116, September 2008
27. Rakkolainen, I., Turk, M., Höllerer, T.: A compact, wide-FOV optical design for head-mounted displays. In: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, pp. 293–294. ACM, New York (2016)
28. Ren, D., Goldschwendt, T., Chang, Y., Höllerer, T.: Evaluating wide-field-of-view augmented reality with mixed reality simulation. In: Proceedings of the 2016 IEEE Virtual Reality Conference (VR 2016), pp. 93–102. IEEE (2016)
29. Volonte, M., Babu, S.V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., Daily, S.B., Fasolino, T.: Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. IEEE Trans. Vis. Comput. Graph. **22**(4), 1326–1335 (2016)
30. Williams, A., Kabisch, E., Dourish, P.: From interaction to participation: configuring space through embodied interaction. In: Beigl, M., Intille, S., Rekimoto, J., Tokuda, H. (eds.) UbiComp 2005. LNCS, vol. 3660, pp. 287–304. Springer, Heidelberg (2005). https://doi.org/10.1007/11551201_17
31. Wilson, M.: Six views of embodied cognition. Psychono. Bull. Rev. **9**(4), 625–636 (2002)
32. Xiao, R., Benko, H.: Augmenting the field-of-view of head-mounted displays with sparse peripheral displays. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 1221–1232. ACM (2016)