MIME: A Mixed-Space Collaborative System with Three Immersion Levels and Multiple Users

Inma García-Pereira
Department of
Computer Science,
University of Valencia

Jesús Gimeno*
Department of
Computer Science,
University of Valencia

Manolo Pérez
Department of
Computer Science,
University of Valencia

Cristina Portalés
Institute of Robotics
and ICT (IRTIC),
University of Valencia

Sergio Casas
Department of
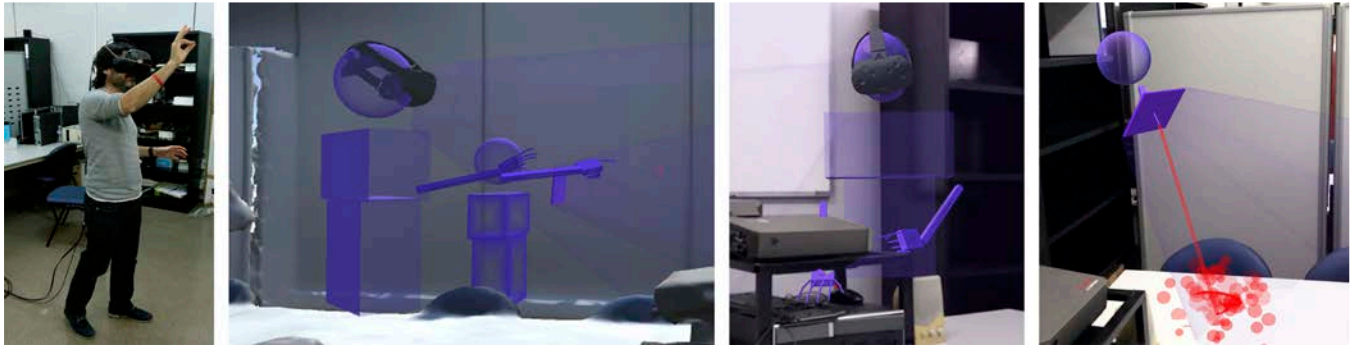Computer Science,
University of Valencia

Figure 1: Images showing the different types of users that can simultaneously collaborate in MIME. From left to right: a user in a remote environment, provided with an HMD; a Mixed Reality space where two virtual users are depicted together with a 3D model of the shared space; the real space with a virtual character, where occlusions are taken into account; a virtual character pointing with his finger to a certain location of the real world and making annotations.

ABSTRACT

Shared spaces for remote collaboration are nowadays possible by considering a variety of users, devices, immersion systems, interaction capabilities, navigation paradigms, etc. There is a substantial amount of research done in this line, proposing different solutions. However, still a more general solution that considers the heterogeneity of the involved actors/items is lacking. In this paper, we present MIME, a mixed-space tri-collaborative system. Differently from other mixed-space systems, MIME considers three different types of users (in different locations) according to the level of immersion in the system, who can interact simultaneously – what we call a tri-collaboration. For the three types, we provide a solution to navigate, point at objects/locations and make annotations, while users are able to see a virtual representation of the rest of users. Additionally, the total number of users that can simultaneously interact with the system is only restricted by the available hardware, i.e., various users of the same type can be simultaneously connected to the system. We have conducted a preliminary study at the laboratory level, showing that MIME is a promising tool that can be used in many real cases for different purposes.

Keywords: Augmented Reality, Virtual Reality, shared spaces, mixed-space, remote collaboration.

Index Terms: K.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities.

* Jesus.Gimeno@uv.es

1 INTRODUCTION

Collaborative systems enable people to interact in the same space, which can be either a real space or a virtual one. When users are co-located in the same real space, Augmented Reality (AR) systems can make possible that users create, manipulate and share virtual information on the real world. If users are in different locations, a Virtual Reality (VR) system can offer a virtual shared space where users can work together. Furthermore, a combination of AR and VR systems can connect co-located and remote users, so they can collaborate. This general classification in three cases (only AR, only VR or a combination of AR/VR), can be widely extended if the different kinds of AR and VR hardware devices are taken into account. In addition to this, each visualization/interaction hardware device offers a different immersion level and distinct interaction capabilities are possible. In conclusion, collaborative systems must be able to communicate heterogeneous users, merging different interaction paradigms in a shared space. In order to deal with this, we propose MIME, a mixed-space collaborative system with three immersion levels and multiple users.

In the academic literature, we can find several collaborative systems. A first classification can be made according to the physical distance between users; when users are collaborating in the same real space it is generally defined as co-located collaboration, whereas remote collaboration is used for distant users. In [1, 2] shared spaces are created using AR where co-located users can interact with virtual models using ARToolKit markers. ShareVR [3] enables co-located asymmetric interaction between users wearing and users not wearing a Head Mounted Display (HMD). Remote collaboration is more often used than co-located collaboration, mainly to create a communication between a user in the real world and a remote expert who may help the user perform some particular task. In [4] a prototype called CoVAR is presented, where two remote users wearing HMDs can collaborate. In this case, the user in the real world wears a HoloLens, an optical see-

through AR visualization device, and the user in the VR wears an HTC Vive. Both users wear also a gaze tracker and a hand tracking device, so different visual cues are used to improve the interaction between them. In this study, both users have independent points of view (PoV), and each one can move with freedom. However, there are many studies where remote users must have the same PoV. For instance, [5-8] are examples where the remote user, which is usually the expert, observes a view of the real world from the other user's PoV in order to try to help him by means of voice explanations, hand gestures, annotations, etc. In our system, co-located users can use an AR interface to collaborate with remote users, who navigate in a virtual scenario, each one of them with his own independent point of view.

A second classification can be made according to the difference between the level of immersion and the interaction ways of the users. Most of the studies deal with similar users, even in the case of remote users using a combination of AR and VR. In [1, 4, 7, 8] both users are wearing HMDs and in [2, 9] all users interact through similar hand-held devices. Collaborative systems with heterogeneous users are more difficult to find, such as the work previously described in [4]. Another example is [10] where a robotic arm placed near the user in the real world is controlled by the remote user to see and project annotations. In [11] two remote users with very different immersion levels share the same virtual scenario. The first one uses a large multiscreen immersive system, with head and hands motion capture, while the second one uses a regular PC. In this case, the PC user has a virtual ray to point at virtual objects, while the first user can point at them with their hands. In our system, we use a very similar pointing interface for users with low level of immersion.

A key aspect of collaborative systems is the way in which users can see the other users in order to create the perception that they share the same space. In co-located systems, users are really in the same space, so this can be reduced to the interactive parts of the user: virtual tools, hands, etc. In [1] the authors emphasize the advantages of real face-to-face collaboration while the AR system can be considered as a collaborative tool to perform operations. This is similar to remote users where both users share the PoV, like [5, 6], where only the virtual hands of the expert are shown to the other user. Nevertheless, this is a key aspect when remote users have independent PoVs. In [4] a user study reveals that showing where the other users are looking to, improves the collaboration between them. In this case, three virtual objects are used: head's 3D model, the field of view's pyramid and a ray with gaze's direction. In [11] not only the user is visually represented but also their hardware device, in order to give a clue to the other users of their interaction possibilities. In our study, we will follow this idea of representing the user's device graphically and we will create fields of view representing the different kinds of users, not only those wearing an HMD.

Finally, in collaborative systems it is possible to make a classification based on the kind of user's interaction. The most common kind of interaction is the "pointing paradigm", which is implemented in most of the collaborative systems, although in different ways. For instance, [5, 6, 9] allow users to point using their hands, whereas for instance [4] uses the gaze of the users as a pointing interface. Another common way of interaction is "object manipulation", where users can handle virtual objects. [12] is a great example of natural interaction where the expert user can grab and move virtual replicas of the real objects to show the other user where to place them. In this same work, an annotation tool makes it possible to create links between virtual and real objects. Annotation tools are a hot topic in research, like in [13, 14], where the ability to create drawings in the space is explored. In our work, we will define how every kind of user can point and create annotations in a natural manner.

In this work, we present MIME, a mixed-space tri-collaborative system, where the users in the real world can collaborate with other remote users, regardless of how immersive their visualization devices are. The tri-collaboration is achieved with three kind of users that can be classified according to three different levels of immersion: high (users wearing a HMD and with hand tracking); medium (users carrying a handheld device with a touch screen); and low (users interacting with a PC and a mouse). It is worth mentioning that, although other hardware options can be built for the three levels of immersion (e.g. spatial AR or optical see through HMD), we have chosen the given hardware because they are probably the most common setups. However, other hardware devices could be used following the same tri-collaborative space paradigm.

The tri-collaboration in MIME is different to other works, as most of them are based on a bi-collaboration (two kind of users physically located in two different spaces). Additionally, in our system the number of users is only limited by the hardware availability (i.e. number of available hardware devices), unlike most of the systems in the academic literature where the interaction is explored for only two users (one user per space). For instance, there can be h users with high immersion, m with medium immersion and l with low immersion, where h, m and l are natural numbers. Every user will have a self-representation and interaction capabilities to communicate with other users. Our main objective is to support different levels of immersion without reducing the ability to communicate instructions, especially those related to spatial information where the use of hands is the key to signal a place or an object.

2 SYSTEM DESCRIPTION

MIME is a system for remote collaboration in which multiple and heterogeneous users can collaborate in real time for multiples purposes sharing a common space. As explained above, we have considered three types of users according to the level of immersion: high, medium and low. For some of them (those in an AR environment) this space will be the actual space in which they are, and for others (those in a VR environment), it will be a virtual space reproducing and representing the actual shared space. A 3D model of the real space is used to create the visual representation of the space where VR users can move and view the rest of the remote users. The model used in this work was previously scanned with a Tango device and then inserted in Unity 3D. In the following sub-sections we address other characteristics of the system according to (Figure 2): 1) navigation, 2) user representation, 3) hand gesture and 4) annotation.

Level of immersion	Navigation	User represen-tation	Hand gesture	Annotation
High (HMD + hand tracking)	**VR** Walk and look	Head + body PoV	Articulated hands Point with ray gesture	Draw on air
Medium (handheld + touch screen)	**AR / VR** SLAM Image target references	Head + body Device screen PoV	Modeled hand touching screen Point with ray	Draw with ray
Low (PC + mouse)	**VR** FPV keyboard & mouse	Head PC screen PoV	Modeled hand controlled with mouse clicking on screen	Draw with ray

Figure 2: Overview of MIME according to different features and levels of immersion.

2.1 Navigation

In MIME, users can move through the mixed-space individually, so the PoV of each user is different. Navigation differs for the different paradigms (see Figure 3):

- In **highly immersive** setups (with HMDs and hand tracking), the navigation is done within the VR environment (they will be VR users), where users can walk and look to the virtual representation of the shared real environment and to the representation of the rest of users.
- As **medium immersive** setups, we refer to users carrying a handheld device embedding a touch screen. Users can be either in the remote environment (VR users) or in the real shared space (AR users). Navigation in such a setup is done with SLAM (Simultaneous Location And Mapping) techniques. Both the real and the virtual worlds are initially aligned by using reference images; if the alignment is lost, the user can always refer back to any of the reference images, and a re-alignment is automatically performed.
- As **low immersive** setups, we mean users interacting with a desktop PC and a mouse. Users are able to see and navigate through a virtual environment (they will be VR users). They can freely move through the virtual environment, having a FPV (First Person View). For the navigation, rotations are triggered by moving the mouse, whereas walking is possible by pressing the arrow keys in the keyboard.
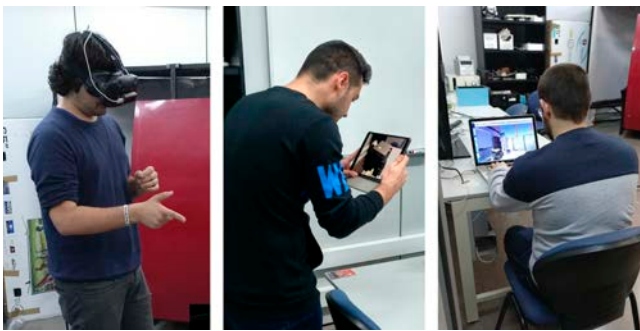

Figure 3: Devices used for navigation in the three setups (left to right: high / medium / low).

2.2 User Representation

Each user is able to see a graphical representation on where the rest of users are located (see Figure 4). Furthermore, that representation symbolizes the kind of device used by the rest of users: HMD, handheld device or PC. In this way, we have:

- For the **highly immersive** setup, the user's avatar consists of basic shapes of its head and body. The arms, hands and fingers are also represented, as we are tracking their position with a Leap Motion device. Additionally, attached to its head, we have placed a virtual HMD.
- In the **medium immersive** setup, the user's avatar consists of basic shapes of their head and body and a representation of the tablet, as its position is calculated with SLAM.
- In the **low immersive** setup, the user is represented with their head, close to a representation of the PC screen.

For the three setups, the PoV of users is represented with a pyramidal shape that can be enabled or disabled. If activated, other users can easily see where others are looking at.
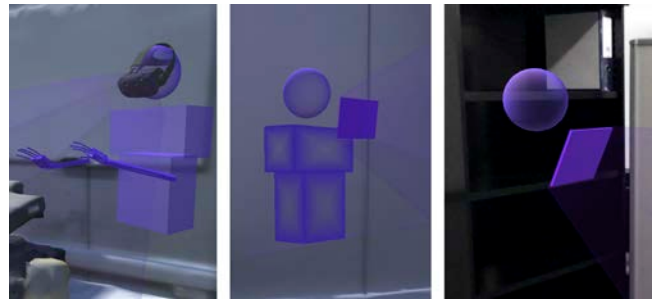

Figure 4: Representation of users in MIME in the three setups (left to right: high / medium / low).

2.3 Hand Interaction

Each user sees a graphical representation of the hands of the rest of users, no matter in which realm they are or the device they are using (see Figure 5). All users can point to a certain object.

- In the **highly immersive** setup, we are using hand tracking with a Leap Motion device attached to the HMD. The users themselves are able to see a representation of their arms, hands and fingers. The user is able to point with their fingers to a certain location. When this happens, we draw a ray emerging from their fingers, which is visible to all users.
- In the **medium immersive** setup, when the user touches the screen of the handheld device, we draw a representation of their hand. This hand has a unique pose, but it is not static. It moves depending on the part of the screen touched. Additionally, they are able to point at objects.
- In a similar way, we draw a ray in the **low immersive setup**, which is triggered by the user with the mouse.
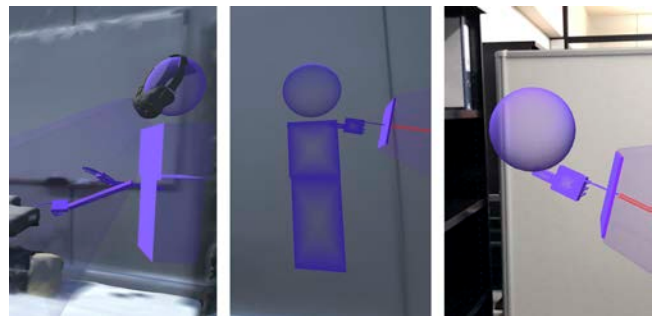

Figure 5: Examples of hand interaction in MIME in the three setups (left to right: high / medium / low).

2.4 Annotation

In MIME, each user can make annotations in the mixed-space by drawing with their hands (see Figure 6).

- In the **highly immersive** setup, the user is able to draw in the air with a pinch gesture. They can make annotations on the virtual objects: the collision of the ray with the virtual objects produces a line, visible for a few seconds. Annotations are triggered with the gesture of pointing with the index finger.
- In the **medium immersive** setup, the user can draw by touching the screen. The collision of the ray with the virtual objects produces a line, which is also visible for a few seconds. In this case, the annotations are activated by default.
- In the **low immersive setup**, the user can make annotations by pressing a button of the mouse.
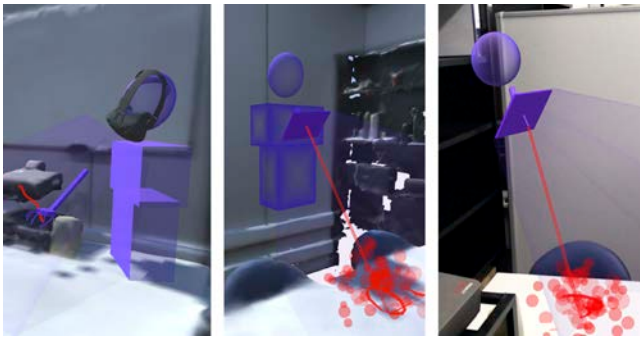
Figure 6: Examples of annotations in MIME in the three setups (left to right: high / medium / low).

3 IMPLEMENTATION

MIME has been implemented in Unity3D. It is based on a server-client architecture. The server is implemented in *Node.js*, using the library *Socket.io* as a basis for the communications. Each client connects to the server and sends the data relative to the user to the server, which is responsible for forwarding it to other clients.

Each client is implemented using Unity 3D. The client of the type "HMD" (high level of immersion) uses a VIVE device for the visualization and capture of the PoV, with a Leap Motion sensor attached to the VIVE device to capture the natural movement of the hands. The client of the type "handheld" (medium level of immersion) uses an iPad Pro 10.5 as a display and the ARKit library for the capture of the PoV. Additionally, the Vuforia library allows to capture the markers that are used as a reference to align the virtual and real worlds in a first step, correcting the initial orientation for ARKit. This user is the only one that is located in the real scenario, previously scanned (into a 3D model). Finally, the client of the type "PC" (low level of immersion) uses traditional interaction devices (keyboard and mouse). Arrow keys control the movement of the PoV in a first-person paradigm, while the mouse rotates the PoV.

An important part of the implementation is that, in the AR environment, occlusions are considered. The scanned 3D model of the real world is used as a phantom object to achieve an occlusion effect. Therefore, users can point at objects and occlusions provide visual cues about the right location of virtual objects (annotations, avatars, virtual hands, etc.). An example of this can be seen in Figure 7, where an avatar is pointing at a computer and their hand is correctly occluded by the projector. The rest of the users will unmistakably recognize this gesture and identify the object the user is referring to.

Figure 7: Occlusions in the AR environment of MIME.

4 EXPLORATORY STUDY

In order to provide a preliminary validation of our system in a laboratory environment, we have conducted a series of experiments with the participation of seven users, with ages ranging from 22 to 40 years old. Each user has tested MIME with three different roles:
- Role 1: the user is in a remote environment, wearing an HMD; she/he is an expert and has to indicate how to do a certain task to the user with the role 2.
- Role 2: this user is physically in the shared space, carrying a tablet; she/he has to follow the indications given by the user with role 1, and perform a task in the real shared environment.
- Role 3: this user is in a remote environment, interacting with a PC; she/he is an expert too, who observes the actions of users with roles 1 and 2 and might help giving indications.

For each experiment, two staff members took two of the roles, and the users testing MIME took the remaining role; after performing the task, they changed the roles until all three roles were covered. In this way, for each experiment there was only one person that had never used the system. In total, we conducted 21 experiments. The task in this exploratory case study was to connect the cables of a projector.

For each experiment, users were asked to fill out a pair of questionnaires related to the usability of the system (Table 1) and the individual satisfaction (Table 2), where the System Usability Scale (SUS) [15] was chosen to measure usability. The results of SUS are depicted in Figure 8. The range 1–5 means: 1: strongly disagree, 5: strongly agree. From these values, the SUS scores are computed [15]. For the role 1, this score reaches 81.43 points; for the role 2, 83.93 points; and for the role 3, 86.79 points. These values can be considered good on the scale of scores provided by the questionnaire, taking into account the fact that a minimum score of 68 would be deemed acceptable for a tool. Therefore, we can state that, for the three roles, users were able to use the tool appropriately.

Table 1. Questions of the SUS questionnaire.

| 1. I think that I would like to use this system frequently |
| --- |
| 2. I found the system unnecessarily complex |
| 3. I thought the system was easy to use |
| 4. I think that I would need the support of a technical person to be able to use this system |
| 5. I found the various functions in this system were well integrated |
| 6. I thought there was too much inconsistency in this system |
| 7. I would imagine that the most people would learn to use this system very quickly |
| 8. I found the system very cumbersome to use |
| 9. I felt very confident using the system |
| 10. I needed to learn a lot of things before I could get going with this system |
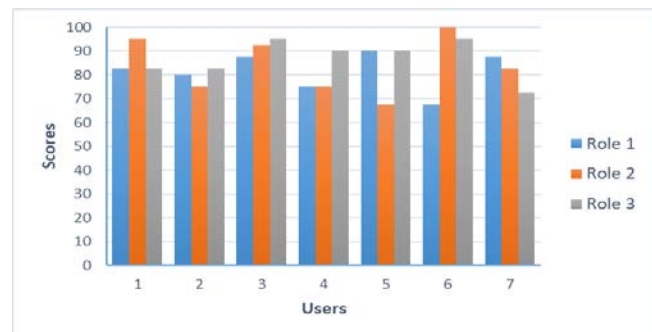
Figure 8: Partial scores obtained for the SUS questionnaire, for each user and for each one of the roles.

Table 2: Questions of the user satisfaction questionnaire.

| |
|---|
| 1. I found it easy to make annotations |
| 2. I have understood the annotations of other users |
| 3. I found it easy to make gestures (only for users with a HMD) |
| 4. I have understood gestures of the rest of users (only for users without a HMD) |
| 5. I found the navigation simple |
| 6. I have recognized the user types by their virtual representation |
| 7. I have recognized the user types by their way of navigating |
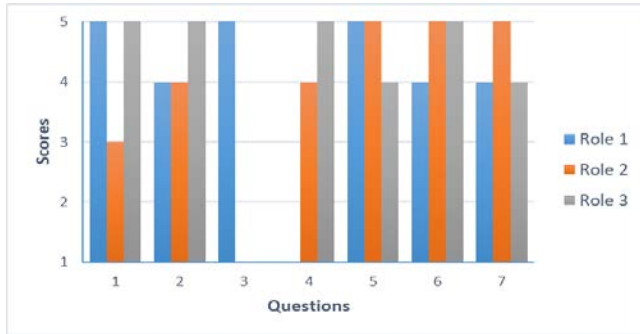
Figure 9: Results of the user satisfaction questionnaire, showing the mode for the three roles and the seven questions.

The results of the individuals' satisfaction questionnaires are given in Figure 9, where the mode is taken as a measure of the score. The meaning of the scores is the same as for the SUS. As it can be seen, only the question 1 for role 2 is below 4 points. The rest get either 5 or 4 points. Therefore, we can state that most users were satisfied or very satisfied with MIME, for the three roles.

5 CONCLUSION

In this paper, we have presented MIME, a mixed-space tri-collaborative system, which allows remote collaboration of heterogeneous users with shared spaces. Differently from other research works, MIME entails a heterogeneous solution for remote collaboration in mixed-spaces with multiple users. By heterogeneous we mean that we consider different navigation options, different representation of users in the MR world, hand gestures and annotations with different kind of devices.

Through a preliminary validation of MIME at the laboratory level, we have showed its great capabilities with both a usability and a satisfaction test for three different roles. Although we have considered a small group of users (seven in this case), the results obtained are promising. These results show a high SUS score, which is surprisingly higher for the least immersive setup. Nevertheless, all three setups/roles are positively assessed and the differences are fairly small.

As further work, the system could be extended to other types of immersive and non-immersive devices, such as AR glasses or smartphones. The representation and the annotation capabilities of the users could be also improved with new features, so that users are able to communicate more information in the shared collaborative space, such as voice or new gestures. In addition, further evaluation experiments should be conducted in order to know the opinion of the prospected users and improve the system. In any case, we believe MIME is already a step forward in the development of heterogeneous collaborative shared spaces with multiples users.

REFERENCES

[1] M. Billinghurst, et al., "Mixing realities in shared space: An augmented reality interface for collaborative computing," Proc. Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, IEEE, 2000, pp. 1641-1644.

[2] M. Sugimoto, et al., "Caretta: a system for supporting face-to-face collaboration by integrating personal and shared spaces," Proc. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 2004, pp. 41-48.

[3] J. Gugenheimer, et al., "ShareVR: Enabling co-located experiences for virtual reality between hmd and non-hmd users," Proc. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, 2017, pp. 4021-4033.

[4] T. Piumsomboon, et al., "Exploring enhancements for remote mixed reality collaboration," Proc. SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications, ACM, 2017, pp. 1-5.

[5] W. Huang, et al., "HandsIn3D: augmenting the shared 3d visual space with unmediated hand gestures," Proc. SIGGRAPH Asia 2013 Emerging Technologies, ACM, 2013, pp. 10.

[6] K. Robert, et al., "MobileHelper: remote guiding using smart mobile devices, hand gestures and augmented reality," Proc. SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications, ACM, 2013, pp. 39.

[7] M. Le Chénéchal, et al., "Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality," Proc. Collaborative Virtual Environments (3DCVE), 2016 IEEE Third VR International Workshop on, IEEE, 2016, pp. 9-12.

[8] G.A. Lee, et al., "Mixed reality collaboration through sharing a live panorama," Proc. SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications, ACM, 2017, pp. 14.

[9] R.S. Sodhi, et al., "BeThere: 3D mobile collaboration with spatial input," Proc. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2013, pp. 179-188.

[10] P. Gurevich, et al., "TeleAdvisor: a versatile augmented reality tool for remote assistance," Proc. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2012, pp. 619-622.

[11] T. Duval, et al., "Improving awareness for 3D virtual collaboration by embedding the features of users' physical environments and by augmenting interaction tools with cognitive feedback cues," Journal on Multimodal User Interfaces, vol. 8, no. 2, 2014, pp. 187-197.

[12] O. Oda, et al., "Virtual replicas for remote assistance in virtual and augmented reality," Proc. Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, ACM, 2015, pp. 405-415.

[13] Y.S. Chang, et al., "Evaluating gesture-based augmented reality annotation," Proc. 3D User Interfaces (3DUI), 2017 IEEE Symposium on, IEEE, 2017, pp. 182-185.

[14] B. Nuernberger, et al., "Interpreting 2D gesture annotations in 3d augmented reality," Proc. 3D User Interfaces (3DUI), 2016 IEEE Symposium on, IEEE, 2016, pp. 149-158.

[15] J. Brooke, "SUS-A quick and dirty usability scale," Usability evaluation in industry, vol. 189, no. 194, 1996, pp. 4-7.