WWoW: World Without Walls Immersive Mixed Reality with Virtual Co-location, Natural Interactions, and Remote Collaboration

Ramesh Guntha(✉), Balaji Hariharan, and P. Venkat Rangan

Amrita Center for Wireless Networks and Applications,
Amrita School of Engineering, Amritapuri Campus,
Amrita Vishwa Vidyapeetham University, Clappana, India
{rameshg, balajih}@am.amrita.edu, venkat@amrita.edu

Abstract. Communicating and sharing knowledge through teleconferencing systems is a common phenomenon now a days. But the traditional remote collaboration systems lack naturalness and are not very immersive. Though existing mixed reality systems support natural interactions, many of them use 3D avatars to represent remote users, hence do not reflect finer movements and emotions of the remote users, and a number of them are quite cumbersome to setup and calibrate. We present our remote collaborative mixed reality environment which provides virtual co-location and gestural interactions using Kinect user image masks and skeletons and is simple to setup. The resulting system is both immersive and natural, gives a feeling to participants that they are in the same physical location, communicating and sharing knowledge objects through natural gestural controls and speech [1].

Keywords: Mixed reality environment · Virtual co-location · Remote collaboration · Kinect user mask streaming · Audio conference · Gestural control

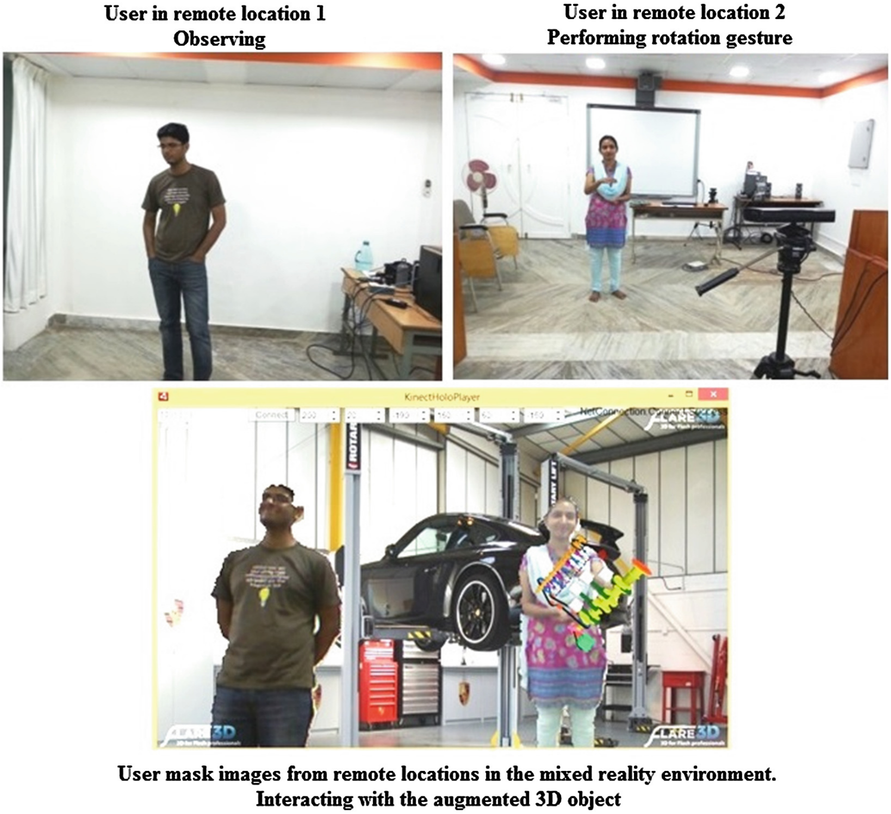
1 Introduction

Communicating and sharing knowledge through teleconferencing systems is a common phenomenon now a days. But the traditional video conferencing and eLearning systems like Skype and Vidyo lack naturalness and are not very immersive. They require users to adopt un-natural interaction mechanisms through keyboard and mouse, resulting in interruptions to the communication flow and thought process. The existing eLearning systems present each knowledge-object such as video, whiteboard, document and 3D models in separate components, hence the users need to focus on multiple parts of the screen(s) simultaneously, resulting in loss of concentration and subsequently loss of interest. In the case of video-conferencing systems the video of each location is presented in a separate window with their native backgrounds. Because of that there is never a feeling of co-location as the remote users are always seen to be separated by virtual walls.

R. Poppe et al. (Eds.): INTETAIN 2016, LNICST 178, pp. 209–219, 2017.
DOI: 10.1007/978-3-319-49616-0_20

Though there are mixed reality systems which try to solve the above problems of lack of immersion and naturalness, many of them require a lot of equipment, setup and calibration [2–4]. Some of the mixed reality systems try to solve the co-location problem through creation of 3D avatars to represent the remote participants, but the 3D avatars do not reflect finer bodily movements and facial emotions of the remote users.

In this paper we introduce the World Without Walls (WWoW). It is a virtually co-located, naturally interactive, and remote collaborative knowledge sharing and conferencing system which would address the above mentioned issues and it is quite simple to setup and does not require any calibration. WWoW system allows users from remote locations to interact through 3D content as if they are in the same room. The Kinect extracts user mask images and these are streamed into the mixed reality environment, which is shared across all the clients in real-time. As the user masks get assembled against the common background of the mixed reality environment, it appears as though the users are in the same location and interacting with the local 3D objects (Fig. 1).


User in remote location 1
Observing

User in remote location 2
Performing rotation gesture

User mask images from remote locations in the mixed reality environment.
Interacting with the augmented 3D object

Fig. 1. Remote user performing rotation gesture on 3D object.

Apart from extracting user masks, Kinect also extract skeleton joint locations of the tracked users. The relative locations and movements of the joint locations can be used to derive various natural gestures which are used to load, move, rotate, zoom in/out and unload the 3D objects. All the interactions are replicated to all the clients in real-time. The streaming and rendering of image masks against common background, natural gestural interactions with content, replicating it in real-time to all the clients result in immersive and engaging experience.

The rest of the paper contains related work, architecture, testing and data analysis, applications and conclusions.

2 Related Work

Research on mixed reality environments has been going on for decades, as it provides tremendous immersion and user engagement. According to [5] mixed reality is based on the basic principles of immersion, interaction and user involvement. These qualities make it a perfect fit for entertainment games and serious games. Reference [6] developed a collaborated game of ball passing using mixed reality with physics engine. They use Kinect to track user skeletons to identify ball passing and ball catching gestures. The remote users are represented only through skeleton joint frame and hence it does not provide immersive co-location experience.

Reference [7] presents a thorough study of how virtual reality evolved over the period and how it is applied in the fields of education and health. Their study acknowledges the inconvenience of wearing virtual reality helmets and goggles for extended period for ergonomic reasons. In WWoW system users do not have to wear any equipment on them. Reference [8] points out that a player's real world gaming experience consists of physical, social, mental and emotional parts. Our WWoW system provides physical experience through the use of natural gestures to control the 3D objects, best social experience as users feel that they are in the same location and can interact with each other, provides mental experience through immersion and interaction and finally emotional experience through problem solving and learning with 3D objects. Reference [9] states that the future mixed reality system should satisfy the conditions of telepresence, interactivity, connectivity and synthesis. The WWoW system enables telepresence through real-time streaming of audio and video (image masks) of all the users, provides interactivity through gestural control of the 3D objects and connectivity through remote collaboration and synthesis through rendering content and users on the mixed reality environment and sharing it across all the clients in real-time. Reference [10] concludes that a co-located environment which provides freedom to interact with content freely and allows movement of people around the content enables for better collaboration and learning. Reference [3] achieves co-location by developing 3D model of the remote user in real-time using Kinect. This system is quite laborious to setup and needs 6 Kinects to be connected and calibrated precisely, which makes this system not so easy to use.

Mixed reality concepts are applied for learning as well [2, 4, 11, 12]. Reference [11] observes that immersion in a digital environment can enhance education by allowing multiple perspectives, situated learning, and transfer. In the WWoW system, the users can use rotation and movement gestures to interact and see multiple perspectives of 3D content, and in future, solve puzzles and quizzes under the supervision of other participants. Reference [2] proposes an elaborate mixed reality collaboration system around physical artifacts by using virtual reality glasses and pocket computer per participant, 2 Kinects, webcam, computer per location, and a central server, to project 3D avatars of remote participants in the augmented space. We believe this system is too cumbersome and costly to implement and because the remote users are represented as static 3D model avatars their real-time finer emotions and body movements are not represented and another limitation of this system is that only local participants can control the physical artifacts directly, the remote participants can only have indirectly control through requesting the local participants. On the contrary, the WWoW system needs much less hardware and almost no setup, the finer physical moments and emotions of the remote participants are reflected through the user image masks and since the participants control the virtual artifacts, all the participants can control them directly. Reference [4] developed a system that created avatars for remote participants using Second Life. Participants are able to view remote participants through head mounted displays and interact with digital objects. But this system has the same limitation of avatars mentioned above and also requires an identical physical environment to that of virtual environment, which would be cumbersome to achieve for every user. Gestural and speech based controlling of the electronic artifacts is another challenging area. The gestures have to be natural and effective at the same time. Reference [13] studies that direct free-hand manipulation gestures are good for selection, rotation and moving the objects, whereas indirect multimodal gestures perform better for scaling the objects. Reference [14] developed and studied the effectiveness of tracking bare hands to detect natural gestures for picking, moving and releasing objects on the tabletop digital surface and achieved comparable results to that of real world activity. Reference [15] developed a system to interact with and manipulate the objects in virtual world through hand gestures. Users found their system to be very natural and easy to use. In WWoW system, the users use their bare hands to control the 3D objects.

3 Setup and Architecture

WWoW system requires minimum hardware and is very simple to setup. Each client should have a Windows PC with Kinect and the server can run Linux or Windows operating system. The Adobe Media Server (AMS) should be installed on the server (Fig. 2).

The Connection manager on the AMS keeps track of all the connected clients, synchronizes the connection status, and automatically tries to re-establish the lost connections to achieve the fault-tolerance and recovery. The collaboration manager maintains the lifecycle of the shared objects with the help of connection manager to make sure that they get reconnected in the event of connection restoration. Client system also has the corresponding stubs for the connection manager, collaboration manager and shared objects and has the Flare 3D virtual environment to host the image masks and 3D objects and to enable the interactions (Fig. 3).

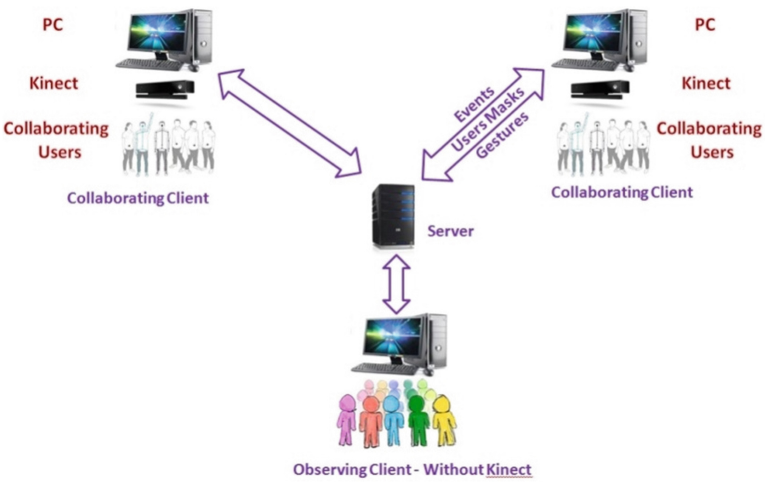


Fig. 2. System setup.

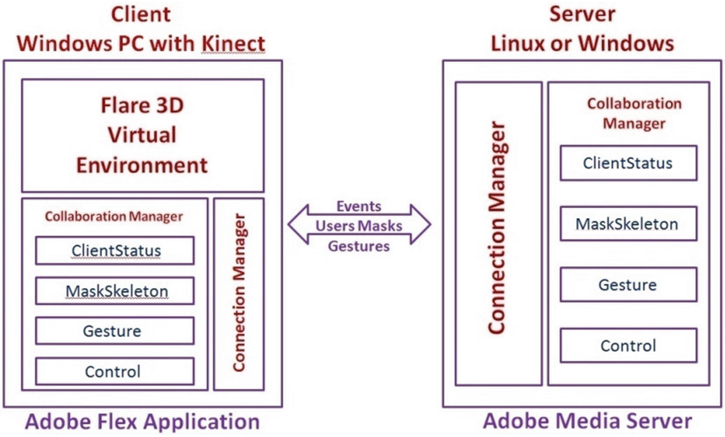


Fig. 3. System architecture.

The shared objects synchronize various data elements across the connected clients to achieve remote collaboration. Each shared object contains data in the form of key-value pairs. When a client modifies the data in a shared object, the changes get propagated to all the shared objects at the connected clients. ClientStatus shared object synchronizes the connection status. The key is clientName and the value is client's status. Control shared object synchronizes the userId of the control user. The key is "ControlUser", the value is the concatenated string of clientName & Kinect userId of the user who has the control over 3D object. MaskSkeleton shared object synchronizes the user mask image and skeleton joint positions of the users. The key is the concatenated string of clientName & Kinect userId and the value contains user mask image and the skeleton object. It is updated with the user mask and skeleton as they are provided by Kinect at 30 frames per second. Gesture shared object synchronizes the current gesture and the related details such as position and rotation angle. The value of the “GestureName” key is the name of the latest gesture performed by the control user and the value of the “GestureDetails” key is the gesture details object (Fig. 3).

4 Testing and Data Analysis

The system is tested for gesture stability, performance, and level of immersion as compared to Skype through a pilot user study. The results are presented below.

4.1 Performance

We have tested the system to analyze how the performance metrics like fps, latency, and bandwidth vary with the number of collaborating users. We used three client nodes and a server, which are connected over 1 GB Ethernet LAN network. Each of the clients have Windows 8.1 PC with Intel Core i7-3770 CPU and 3.4 GHz processor with 8 GB memory and 100 Mbit/s network adapter. The server has Ubuntu 12.0.4 OS with 12 core processor and 16 GB memory.

The FPS start at 22 frames per second, and stay around 15 till 4 collaborating users and comes down to 5 when 8 users are collaborating, similarly the average latency starts at 30 ms from client to client and goes up to 200 ms with 8 users. The average bandwidth consumption starts around 100 Mbps at the server for single user and goes to 250 Mbps for 8 users, the reason it is not growing linearly with the number of users is because fps is reducing as the number of users go up. Similarly the average client bandwidth consumption per transmitted frame is around 5 Mbps for single user and goes up to 33 Mbps for 8 users (Fig. 5).

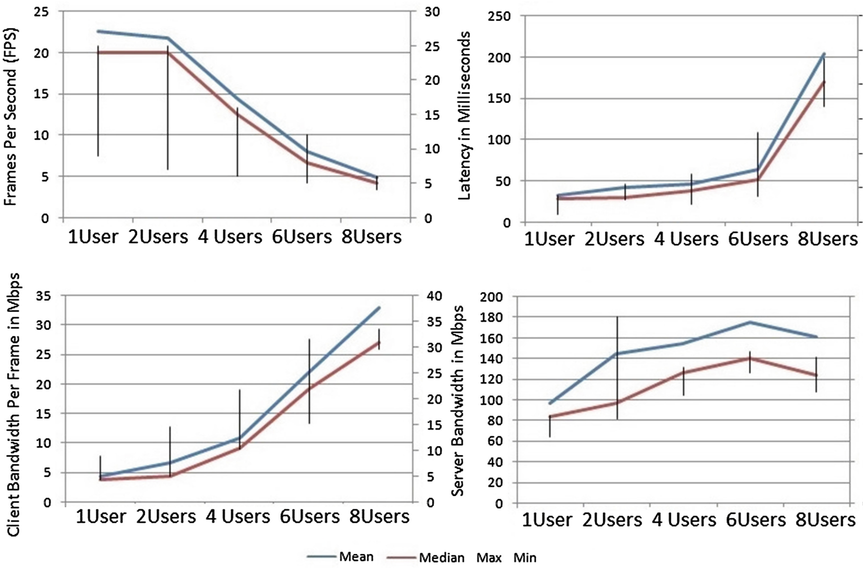


Fig. 5. Performance analysis of WWoW system.

Overall the system seems to be stable and within the limits of acceptable fps and latency for a smooth collaborative and interactive session.

4.2 Gesture Stability

While positioning the 3D Object based on the hand coordinates it is noticed that the 3D object is shaky, even though the user is keeping the hands as still as possible. The xyz coordinates of the left and right hand joints are analyzed to assess the variations (Fig. 4). The measurements are taken for the duration of 10 s after the user is in

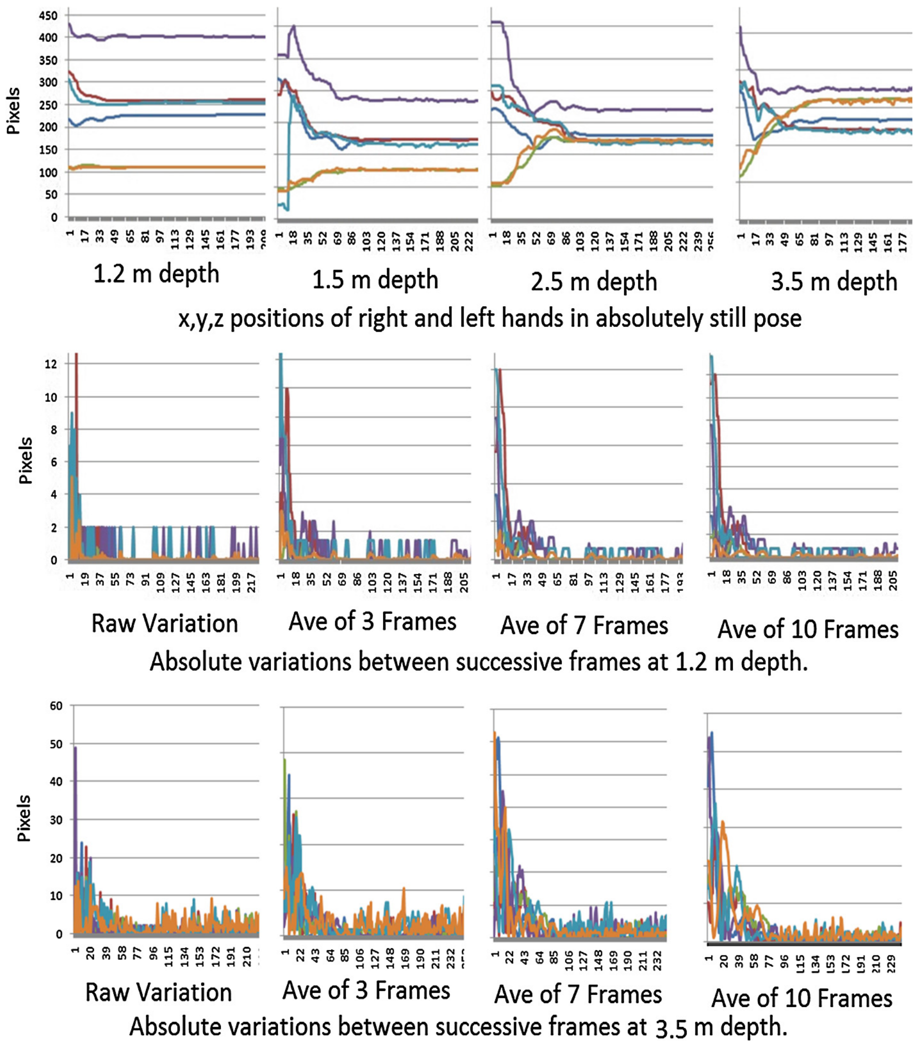


Fig. 4. Analysis of Kinect joint position variations.

absolutely still position. The top row presents the variations of xyz coordinates of both hands; while it takes few seconds stabilize in the beginning, they become quite stabilize with some jitter. The jitter seems to increase as the user moves away from Kinect. The second and third rows show the plot of the moving average of variations between subsequent frames at the depth 1.2 and 3.5 m respectively with. It is noted that both 7 and 10 frame moving averages show much less variations compared to raw and 3 frame ones. So we chose 7 frame moving average as a balance between stability and responsiveness.

4.3 Level of Immersion - Comparison with Skype: A Pilot User Study

The level of immersion provided by WWoW system is tested by comparing it with Skype. We have setup a two location conference with Skype and WWoW system in two separate rooms of our lab. For the Skype session, the topic of Windmill is taught with the help of 3D model through screen sharing and video conference. For the WWoW session the topic of 4-Stroke 4-Cylinder Internal Combustion engine is taught with 3D model augmentation. We tested the two sessions with the same set of users consisting of men and women of ages between 25 and 30, none of are familiar with the topics taught. Same method of teaching and testing is used for both the sessions; The remote instructor taught the subject, then gave opportunity for the students to ask questions and then left the students for themselves to explore the subject by interacting with the system, later the students are called to the teacher's room one by one and are interviewed. The interview questions are both qualitative such as level of immersion, learning experience and difficulty of the subject etc., and also specific questions such as locate various parts, name various parts, explain the principle, explain the working of the system etc. (Fig. 6).

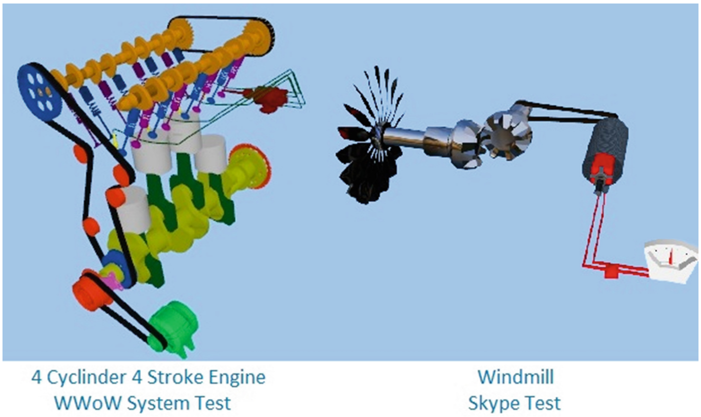


Fig. 6. Models used for immersion test.

Even though the topic in WWoW system is much more complex and took much more time to teach, the students exhibited lot of interest and asked a lot of questions to the teacher and also spent much more time in interacting and exploring with the system and among themselves. As the results suggests that WWoW system performed better than Skype in qualitative terms by providing much more engaging and immersive experience with lot of student-to-student and student-to-teacher interaction (Table 1).

The students have done equally well in answering subject specific questions in both the methods of teachings, but they preferred learning though WWoW system as they found it much more interesting to see the 3D object augmentation on the remote teacher's image, and to interact with the 3D model through natural hand gestures.

During the post-test discussion the students mentioned that there is a lot of potential to the WWoW concept if we can improve it by implementing more natural and smooth gestures and improve the video quality and gaze alignment.

Table 1. Level of immersion test results.

Test criteria	Skype	WWoW
Lecture duration	3 min	15 min
Number of questions by students	0	7
Student's exploration duration	2 min	18 min
Complexity of the subject	Medium	High
Level of immersion	Medium	High
How other students spent time during interview	On smart phone	Discussing and playing with system
Students explanation of principles	Excellent	Excellent
Number of subject questions asked in the interview	9	12
Average percentage of questions answered	100 %	100 %
Level of immersion	Medium	High
Preference to learn complex topics through the system	Medium	High
Overall experience	Medium	High

5 Applications

WWoW system has many applications in education, meetings and panel discussions, trainings, demos, and presentations.

In education it can be used for teaching of complex engineering, medical and science subjects with engaging and interactive multi-media content, it can be used for quizzes, to test the assembly of various engineering components, it can be used for virtual labs, e.g., students can learn to operate various machinery in mechanical engineering labs, students can learn to make various circuits or electronic boards in electrical and electronics labs, in medical labs, students can experiment various medical equipment to examine body parts.

The meetings in WWoW can be refreshing to see the distant participants in the same virtual room, interacting with power point presentations or presenting latest design models by controlling rotations and zoom, all with through using only natural gestures.

Engineers and marketing personnel in the industry can present and interact with various 3D models of components or products and see how they look by altering various physical properties such as color and sizes of these models in run time.

6 Conclusions

This work needs to expand to include many more gestures to control variety of multimedia content. The resolution of the image masks can be improved greatly if we use KinectV2 as it provides HD resolution image masks. User masks can be replaced with 3D textures of users built from point clouds in real-time, such 3D representations of the users can be used to interact with mixed reality environment much more intimately. Much more work needs to go in to the positioning of user's masks in the mixed reality environment, so that there is proper gaze alignment to bring even more naturalness in conversations, as the current system shows only the frontal view of all the participants to each other, which is ideal for instruction and demo scenarios, but it is not quite suited for a discussions and round table meetings, where each participant should be presented with different angular perspectives of the remote users.

References

1. https://www.youtube.com/watch?v=1Lv9A2pnnEE&feature=youtu.be
2. Weigel, J., Viller, S., Schulz, M.: Designing support for collaboration around physical artefacts: using mixed reality in learning environments. In: 2014 IEEE International Symposium on Mixed and Mixed Reality (ISMAR), pp. 405–408. IEEE, September 2014
3. Maimone, A., Bidwell, J., Peng, K., Fuchs, H.: Enhanced personal autostereoscopic telepresence system using commodity depth cameras. Comput. Graph. **36**(7), 791–807 (2012)
4. Kantonen, T., Woodward, C., Katz, N.: Mixed reality in virtual world teleconferencing. In: 2010 IEEE Virtual Reality Conference (VR), pp. 179–182 (2010)
5. Pinho, M.S.: Realidade Virtual. PUC, Rio de Janeiro (2004)
6. Tang, T.Y., Winoto, P., Wang, Y.F.: Alone together: a multiplayer mixed reality online ball passing game. In: Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing, pp. 37–40. ACM, February 2015
7. Carvalho, B., Soares, M., Neves, A., Soares, G., Lins, A.: The state of the art in virtual reality applied to digital games: a literature review. In: 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014, July 2014
8. Nilsen, T., Linton, S., Looser, J.: Motivations for mixed reality gaming. Proc. FUSE **4**, 86–93 (2004)
9. Lau, H.F., Lau, K.W., Kan, C.W.: The future of virtual environments: the development of virtual technology. Comput. Sci. Inf. Technol. **1**, 41–50 (2013)

10. Church, T., Hazelwood, W.R., Rogers, Y.: Around the table: studies in co-located collaboration. In: Adjunct Proceedings of the 4th International Conference on Pervasive Computing (2006)
11. Dede, C.: Immersive interfaces for engagement and learning. Science **323**(5910), 66–69 (2009)
12. Marzouk, D., Attia, G., Abdelbaki, N.: Biology learning using mixed reality and gaming techniques. Environment **2**, 3 (2013)
13. Piumsomboon, T., Altimira, D., Kim, H., Clark, A., Lee, G., Billinghurst, M.: Grasp-Shell vs gesture-speech: a comparison of direct and indirect natural interaction techniques in mixed reality. In: 2014 IEEE International Symposium on Mixed and Mixed Reality (ISMAR), pp. 73–82. IEEE, September 2014
14. Figueiredo, L., Dos Anjos, R., Lindoso, J., Neto, E., Roberto, R., Silva, M., Teichrieb, V.: Bare hand natural interaction with augmented objects. In: 2013 IEEE International Symposium on Mixed and Mixed Reality (ISMAR), pp. 1–6. IEEE, October 2013
15. Tecchia, F., Avveduto, G., Carrozzino, M., Brondi, R., Bergamasco, M., Alem, L.: Interacting with your own hands in a fully immersive MR system. In: 2014 IEEE International Symposium on Mixed and Mixed Reality (ISMAR), pp. 313–314. IEEE, September 2014