

A Lightweight Mobile Remote Collaboration Using Mixed Reality

Jeremy Venerella¹, Lakpa Sherpa¹, Hao Tang^{1,2}, Zhigang Zhu^{2,3}

¹Borough of Manhattan Community College - CUNY

²The City College of New York – CUNY

³The CUNY Graduate Center

jeremy.venerella@stu.bmcc.cuny.edu, lakpa.sherpa5382@gmail.com,
htang@bmcc.cuny.edu, zzhu@ccny.cuny.edu

Abstract

In many complex tasks, a remote expert may assist a local user to guide actions in the local user's environment. Existing solutions allow multiple users to collaborate remotely using high-end Augmented Reality (AR) and Virtual Reality (VR) head-mounted displays (HMD). In this paper, we propose a lightweight remote collaboration approach, with the integration of AR and VR devices, both running on mobile platforms, to tackle the challenges of existing approaches. The AR platform processes the live video and measures 3D geometry of the local environment of a local user. The 3D scene is transited and rendered in the remote side on a mobile VR device, along with a simple and effective user interface, which allows a remote expert to simply manipulate the 3D scene on the VR platform to guide the local user to complete tasks in complex environments.

1. Introduction

Although communication is easy for people in close proximity, we have to spend a large amount of our time apart from each other to avoid long-distance commuting and traveling. Nowadays ubiquitous smartphones play an important role in our communications, yet the interactions are limited to verbal or video content, which are sometimes insufficient. Examples of such tasks may include a customer calling a call center to help troubleshoot a PC, a novice technician calling an expert for guidance in repairing a complex device (e.g., a car engine), or a medical expert guiding a local doctor through a complex procedure.

Augmented Reality (AR) allows users to see virtual objects seamlessly superimposed over the real world and this can allow for co-located users to view shared 3D virtual objects that they interact with, or a remote helper to annotate the live video feed of a local worker, enabling them to collaborate from a distance [4]. Similar AR-base collaboration systems are proposed in [7, 10, 17]; in each of these systems, a remote helper to guide the local user in complex physical tasks. Such an AR system can stick the annotations to the original positions in 3D world.

Recently, HMDs have become increasingly popular as an immersive display device for remote collaboration. AR-based HMDs, including HoloLens and other smart glasses, offer users see-through experience, whereas VR-based HMDs such as Oculus Rift and HTC Vive offer fully immersive VR experience. In [8], video of local workspace is captured and sent to a VR-based HMD worn by the remote user. AR-based smartglasses are used in [9] for both local and remote users in collaboration.

In remote guidance scenarios, the remote user often has to view the local scene from the point of view of the local user. A few studies [5, 13] present collaborative AR systems that allow the remote user to have an independent view into the shared task space. A 3D model is reconstructed by a depth sensor and rendered on a VR HMD worn by the remote user, who can view and annotate the 3D scene from a different perspective, which leads to faster task completion time. More studies follow the idea [5, 13] to provide 3D models of local scenes [6, 11, 15], and use independent depth sensors or HMDs with embedded depth sensor (MS HoloLens) for the local user.

To render smooth and high-quality 3D models on the remote side, high-end VR HMDs are widely used. Although the VR HMDs provide high frame rate, low latency, large field of view (FOV) and high visual quality, they often require wired connections to a high-end workstation, which may not be feasible in many real-world applications. In [14], a mobile remote assistance solution is offered, which streams video while local user stays focused on one area and the remote user's annotation is limited on 2D, making it ineffective in many complex 3D environments. Though high-end AR HMDs provide users with decent AR experiences, they are costly. Therefore, it is desired to develop a portable and affordable remote collaboration solution with effective communication and interaction. However, a few challenges exist:

(1) Only a few mobile devices have depth sensors and offer real-time 3D modeling, thus allowing remote users with independent perspectives and effective referencing object. Existing AR SDKs [1,2,3] don't provide 3D texture modeling out of box. Without 3D modeling, annotations are limited to 2D image/video and may not be effective for remote assistance in many complex 3D environments.

(2) Mobile VR has limitations in its interfaces and

functions, therefore it's difficult for remote users to make effective annotation and interaction.

In this work, we propose a complete mobile solution, MCoAVR (Mobile Collaborative Augmented and Virtual Reality) to tackle the above challenges and provide simple and effective remote collaboration experiences. The collaboration system offers a more flexible yet affordable solution than the high-end A/VR system. The mobile A/VR solution offers a more self-contained experience, easy to use and significantly less expensive and therefore has the potential to act as a means to more accessible educational/training solution.

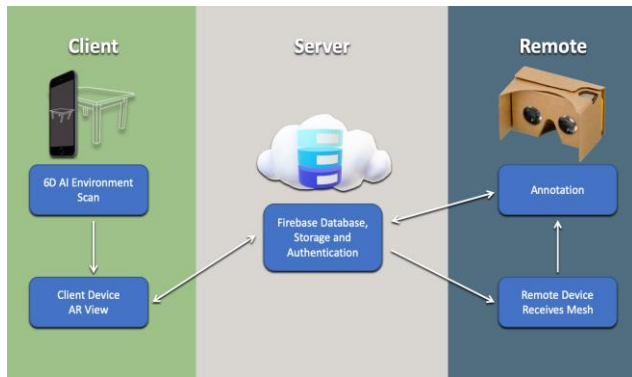


Figure 1. The overview of the proposed MCoAVR system.

2. System Design and Implementation

MCoAVR is a multi-user system with cloud storage. A user on the client side uses a mobile AR system (e.g. a smartphone) to scan local environment and build a 3D model (mesh) of the client's scene. It then uploads the textured 3D mesh onto cloud storage (e.g. Google firebase platform). Users on the remote side use a mobile VR system (e.g. a smartphone with Google Cardboard) to query the client's 3D scene via the textured mesh available on the cloud, allowing remote users to view client's environment from an arbitrary perspective. The interface on the remote side also allows users annotate virtual objects and add virtual landmarks in the client's environment, either to guide the client user or to collaborate with client user to complete a complex task. The system overview can be found in Figure 1.

2.1. Client Side (AR) and 3D Mesh Generation

The mesh generation is provided by 6D.AI [1]. This platform uses AR point cloud data to generate a 3D mesh of the surrounding environment from a mobile device. The near real time networking is built with Google's Firebase platform. The networked mesh allows a remote user to interact with a client's 3D environment. When a client user designates the desired workspace, detected changes in the mesh results (within a user defined region of interest) in an asynchronous upload of the individual generated mesh

chunks in the workspace area into the Firebase storage system. This upload produces a callback triggering a download for the remote user.

The client-side projections are generated during client-side mesh generation. This is done by capturing workspace point of view (POV). Each captured view is automatically sent along with the client device coordinates to the remote user over the cloud server.

2.2. Remote Side (VR) and Its Navigation

Once the remote user receives the captured image and coordinates, a Unity projector is generated at the specified coordinates, projecting the image onto the 3D mesh so remote users can see 3D client environment from any POV.

The remote side runs on a Google cardboard device which comes with very simple interface (one button), hence it's not easy to navigate in the VR world. However, it's necessary to move remote user's perspective to obtain a better POV for more accurate and clearer 3D reference and annotation [16]. We apply the ARKit SDK [3] on the rear camera to obtain user's 6-DOF motion, which allows remote users easily navigate the client 3D scene in the VR view.

2.3. Interaction Between Client and Remote Sides

Simple and effective interaction and communication is the key to a successful remote collaboration system. The proposed system includes different types of interaction features: such as prebuilt/drawn annotations, virtual landmarks with 3D surface alignment, and synchronized or independent POVs with user's head gaze. We will briefly discuss each of them below.

a) *Prebuilt and/or drawn annotations* (circular arrows and red flags in Figure 2b) placed or moved in the remote scene automatically send their coordinates, rotation, annotation type, Globally Unique Identifier (GUID), and color to the client side via database callback. Upon receiving the data, the client application generates a new annotation into the client scene via AR (rendered at the same location/rotation as it's annotated in the remote VR view) or updates an existing based on the GUID.

b) *Virtual landmark* (Figure 2b) has proved effective to reference 3D object in the client's scene. Hence, we implement this feature which allows users on both sides to easily refer to 3D objects. The orientation of virtual landmark is aligned with 3D surface of objects in the local environment (Figure 2b).

c) Remote users can either choose an *independent POV* by default for better perspective to reference 3D objects, or choose a *synchronized POV* with the client user so users on both sides will have the exact same POV if it's more convenient for them to collaborate. Additionally, user on one side can see another user's POV and head gaze (Figure 2a), with boundary of what the users can see through their

display (Figure 2a). It can inform users of what their collaborators can see.

In addition, multi-user networked annotations allow for the ability of multiple remote users to produce annotations for the client-side user. These annotations are used to help guide the client to complete a real-world task.

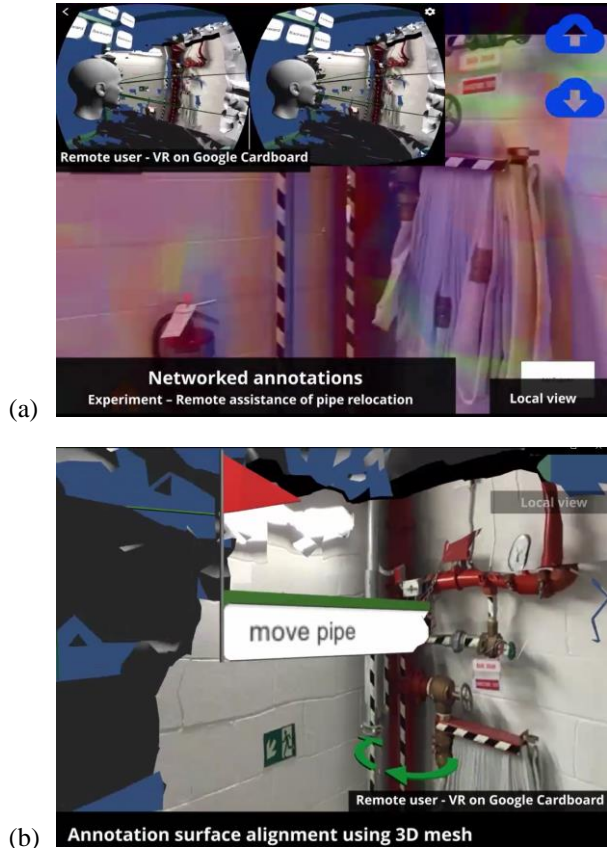


Figure 2. (a) the remote VR view is superimposed to the top left corner of the client AR view, and the head gaze with POV of the client user is rendered in the VR view so the remote users know what their collaborators see; (b) includes a pre-built annotation (circular arrows) and a virtual landmark with text annotation (red flag with “move pipe” annotation) in the remote VR view, note the virtual landmark with annotation is aligned with the 3D wall surface for easy visualization.

3. Exploratory Study

We plan to conduct an exploratory study with 20 participants (10 pairs). We will test if users could use the mobile system to complete a remote help task. To validate the effectiveness of remote collaboration using the proposed mobile system, we will use a similar test in [12] in which a remote user tries to assist a local user to construct a model of wooden blocks. Two participants will be placed in different rooms in each test, one as a client user and another as remote user. The client user follows the remote user’s instructions to build a model. The above test serves

to simulate a local user requiring assistance from a remote user to complete complex tasks.

We will evaluate the system with a baseline which only uses Skype video chat for collaboration, as well as a commercial software: Vuforia chalk (free evaluation) [14]. A few participants will be designated as remote users, who will be trained to be familiar with the manipulation tool and 2D annotation on the proposed mobile VR system. Others will be assigned as local users and use the AR system. Each local user will perform three trials with one remote user, using the above three systems. The three wood block models will be carefully designed for different trails so all models have comparable complexity. The completion time will be recorded for all trails for performance analysis. Additionally, each user will complete a survey, including various questions such as ease of use, fatigue, awareness of remote actions, collaborativeness, ease of annotation and overall satisfaction rate.

4. Conclusion

In this paper, we propose a lightweight remote collaboration system, with the integration of mobile AR and VR devices, which are more affordable and accessible. A local user uses an AR device to reconstruct simple 3D mesh models of the local environment. The 3D model is transited and rendered in the remote side on a mobile VR device, along with a simple and effective user interface. The proposed mobile A/VR solution allows a remote expert to simply manipulate the 3D scene on a mobile VR platform to guide the local user to complete a complex task.

5. Acknowledgments

This research is partially supported by IC CAE Critical Technology Studies Program at Rutgers University, U.S. Department of Education: Minority Science Engineering Improvement (MSEIP) Grant, PSC-CUNY Research Grant, and a CCNY CiPASS Grant.

References

- [1] 6D.AI <http://6D.AI>, last access 4/25/2019
- [2] ARCore <https://developers.google.com/ar/> last access 4/25/2019
- [3] ARKit <https://developer.apple.com/arkit/> last access 4/25/2019
- [4] Ronald Azuma, (1997). A Survey of Augmented Reality. In Presence: Teleoperators and Virtual Environments, vol. 6, no. 4, pp. 355-385.
- [5] Matthew Tait and Mark Billingham: The Effect of View Independence in a Collaborative AR System Computer Supported Cooperative Work 24(6) · August 2015
- [6] F. Tecchia, L. Alem and W. Huang. 2012. 3D Helping Hands: A Gesture Based MR System for Remote Collaboration. In Proc. ACM VRCAI. 323-328.

- [7] Matt Adcock and Chris Gunn. 2015. Using Projected Light for Mobile Remote Guidance. *Computer Supported Coop. Work* 24, 6 (December 2015), 591-611.
- [8] Judith Amores et al. ShowMe: A Remote Collaboration System that Supports Immersive Gestural Communication, CHI 2015
- [9] Kevin Wong, HandsOn: A Portable System for Collaboration on Virtual 3D Objects Using Binocular Optical Head-Mounted Display, Master thesis, MIT 2015
- [10] Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. (2012). TeleAdvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New-York: ACM Press. pp. 619–622.
- [11] Thammathip Piumsomboon, Lee, Y., Lee, G. and Billinghamurst, M, 2017. CoVAR: a collaborative virtual and augmented reality system for remote collaboration SIGGRAPH Asia 2017 Emerging Technologies, ACM, Bangkok, Thailand, 1-2.
- [12] Ranjan, A., Birnholtz, J. P., and Balakrishnan, R. Dynamic Shared Visual Spaces: Experimenting with Automatic Camera Control in a Remote Repair Task. In *Proc. of CHI (2007)*, 1177-1186.
- [13] R.S. Sodhi, Jones, B. R., Forsyth, D., Bailey, B.P. and Maciocci, G. 2013. BeThere: 3D mobile collaboration with spatial input *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Paris, France, 179-188.
- [14] Vuforia chalk, <https://chalk.vuforia.com/> last access 4/25/2019
- [15] Wang et al. Augmented Reality as a Telemedicine Platform for Remote Procedural Training, *Sensors* 2017, 17, 2294
- [16] Benjamin Nuernberger et al., Multi-view gesture annotations in image-based 3D reconstructed scenes, 22nd ACM Conference on Virtual Reality Software and Technology, 129-138.
- [17] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, Tobias Höllerer, World-Stabilized Annotations and Virtual Scene Navigation for Remote Collaboration, the 27th annual ACM symposium on User interface software and technology, Honolulu, Hawaii, October, 2014, 449-459.