AvatarMeeting: An Augmented Reality Remote Interaction System With Personalized Avatars

Xuanyu Wang
School of Computer Science and Technology, SPKLSTN Lab
Xi'an Jiaotong University
Xi'an, China

Yang Wang
School of Computer Science and Technology, SPKLSTN Lab
Xi'an Jiaotong University
Xi'an, China

Yan Shi
School of Computer Science and Technology, SPKLSTN Lab
Xi'an Jiaotong University
Xi'an, China

Weizhan Zhang
zhangwzh@xjtu.edu.cn
School of Computer Science and Technology, SPKLSTN Lab
Xi'an Jiaotong University
Xi'an, China

Qinghua Zheng
School of Computer Science and Technology, SPKLSTN Lab
Xi'an Jiaotong University
Xi'an, China

ABSTRACT

To further enhance the immersion perception of remote interaction, avatars can be involved harnessing Head Mounted Display (HMD) based Augmented Reality (AR). In our demonstration, we present an avatar based remote interaction system AvatarMeeting, enabling users to meet with remote peers through interactive personalized avatars just like face to face. Specifically, we propose a novel framework including a consumer-grade set-up, a complete transmission scheme and a processing pipeline, which consists of prescan modeling, pose detection and action reconstruction. And an angle based reconstruction approach is introduced to empower the AR avatars to perform the same actions as each remote real person do in real time smoothly while keeping a good avatar shape.

CCS CONCEPTS

• **Computing methodologies** → **Mixed / augmented reality**.

KEYWORDS

Augmented Reality (AR), Remote Interaction, Personalized Avatar

ACM Reference Format:
Xuanyu Wang, Yang Wang, Yan Shi, Weizhan Zhang, and Qinghua Zheng. 2020. AvatarMeeting: An Augmented Reality Remote Interaction System With Personalized Avatars. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3394171.3414449

1 INTRODUCTION

People have always been chasing a more immersive and realistic remote interaction experience, eager to interact with vivid remote objects or people just as they interact with local ones. With AR HMDs showing us an unprecedented way of displaying and interacting, the future generation of remote interaction can be seen, where remote participators are reconstructed next to the local user with all information in a real face-to-face interaction included such as voices, body movements and facial expressions.

Some AR endeavors have focused on telepresence [4] [10] [9] and avatar appearance and interaction [2] [8] [5]. However, they either do not involve remote human controlled human-avatar interaction or do not provide real-time service in AR environment. And previous remote interaction systems [7] [6] [11] require expensive instrumented areas.

Our demonstration fills the vacancy, proposing a complete implementation framework and introducing an angle based reconstruction method. It enables a user to interact with remote peers with their real-time-moving avatars as the interface. The avatar can also be placed, resized and rotated wherever and however the user wishes.

2 SYSTEM DESIGN

2.1 AR Remote Interaction Framework

The framework of our system is shown in Figure 1.

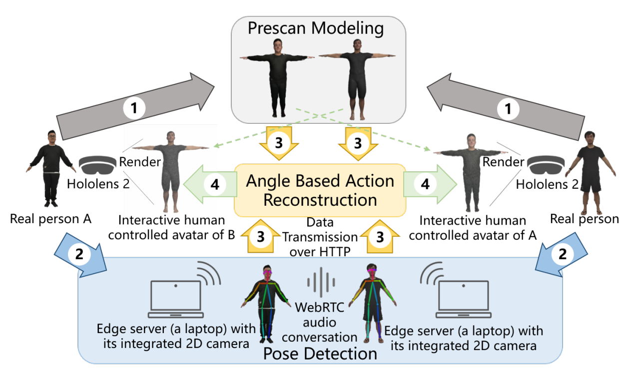

Figure 1: System framework.

ACM ISBN 978-1-4503-7988-5/20/10.
https://doi.org/10.1145/3394171.3414449

The pipeline is composed of three processes. In prescan modeling, from a recorded video, OpenPose [3] is first used to detect 2D joints, based on which a textured 3D model is constructed then referring to the work by Alldieck et al. [1]. Afterwards, in pose detection process, OpenPose is used again to analyze joint positions from the video captured by the 2D camera in real time. And finally with the detected joint data and the prescan models of users, the Hololens 2 renders the human controlled avatar following the action reconstruction approach.

In the consumer-grade environment set-up, only a laptop and its integrated 2D camera is needed to cooperate with the Hololens 2 HMD. The laptop functions as an edge server, providing computation power and real time updated data.

The transmission scheme contains both pose data and voice transmission. Pose data is transmitted based on the client-server architecture over HTTP. The Hololens 2 makes HTTP requests for each avatar to the corresponding server IP to get the latest data to be rendered. Referring to the traditional video streaming strategy, a buffer is maintained during downloading and rendering. And WebRTC is utilized to provide real time peer audio conversation.

2.2 Angle Based Action Reconstruction

We leverage OpenPose to detect joints from the 2D camera live captured video. A total of 25 joints are pinpointed with their coordinate. As all human actions are driven by joint rotations, we propose an angle based approach to combine joint positions with real human skeleton structure, using only joint rotations to drive the avatar's action. The rotation angles of 10 major joints, the neck, spine, left and right arm, forearm, upper leg and leg are calculated respectively based on the vectorial angle. Take the rotation of the forearm as an example. We vectorize the left arm and left forearm as $\boldsymbol{vector_{arm}}$ and $\boldsymbol{vector_{forearm}}$ by coordinate subtraction as shown in (1). Variables $pos_{shoulder}$,pos_{elbow} and pos_{wrist} are the coordinates of the left shoulder, elbow and wrist respectively.

$$\begin{cases} \boldsymbol{vector_{arm}} = pos_{elbow} - pos_{shoulder} \\ \boldsymbol{vector_{forearm}} = pos_{wrist} - pos_{elbow} \end{cases} \tag{1}$$

The angle is calculated using triangle principles (2) (3). The range of $angle'$ is $[0°, 180°]$ for the use of arccos. But in practice, the joints can rotate both clockwise and anti-clockwise, so we multiple $angle'$ with the sign of the outer product of the two vectors to get $angle$, which ranges from $-180°$ to $180°$, as the input to the avatar.

$$angle' = \arccos\left(\frac{\boldsymbol{vector_{arm}} \cdot \boldsymbol{vector_{forearm}}}{|\boldsymbol{vector_{arm}}| * |\boldsymbol{vector_{forearm}}|}\right) \tag{2}$$

$$angle = angle' * sign(\boldsymbol{vector_{arm}} \times \boldsymbol{vector_{forearm}}) \tag{3}$$

The update rate of Hololens 2 is 60 frames per second, while OpenPose only generates joints coordinate up to 22 frames per second. Therefore, we conduct linear interpolation to fill the vacancy to get a smoother performance. The angle changes from 0° to 180° and then −180° to 0° in clockwise rotation. Especially for the case when the user's limbs move pass 180° to −180°, we interpolate several 180° data in stead of linearly calculate them (in which case the interpolated value would be around 0°) to keep a normal rotation.

3 DEMONSTRATION

The setup of our implementation includes two physical distant users, each wearing a Hololens 2 and facing the integrated 2D camera on an Alienware, which is equipped with an Nvidia GeForce GTX 1080 GPU, 12 Intel(R) Core(TM) i9-8950HK CPU @ 2.90GHz CPUs and a 32GB RAM running Windows 10, respectively.

The users can chat with each other, see each other's avatar moving just as the real person do. They can place the avatar standing in front of them to simulate a face-to-face communication scenario, or resize the avatar and move it anywhere they want (e.g. shrink the avatar and place it on the table to simulate a toy interaction effect). The real scene is shown in Figure 2.

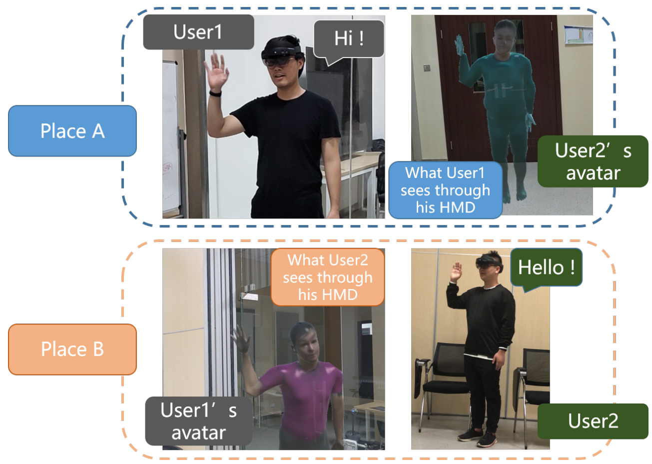

Figure 2: A real scene of the application. Two participators are greeting to each other. The clothes on both avatars are colored (instead of the original color of black) in order to make the holograms more opaque.

4 CONCLUSION

The provision of avatar augmentation aimed ultimately at improving the quality of the interaction experience. In our demonstration, we build AvatarMeeting as a first step into the complete usable avatar based remote interaction system. With the goal to keep the set-up as simple as possible, we use only a 2D camera integrated laptop as the edge server and an AR HMD to present a paradigm of how remote users interact with the augmentation of avatars. A complete data and voice transmission scheme is proposed and implemented. In addition to the whole framework, an angle based action reconstruction approach is introduced to make sure the avatar shape is not distorted and achieve a smoother and more natural moving effect.

ACKNOWLEDGMENTS

This work is supported by the "National Key R&D Program of China" (Grant No. 2018YFB1004500) and the NSFC (Grant Nos. 61772414, 61721002).

REFERENCES

[1] Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, and Gerard Pons-Moll. 2018. Video based reconstruction of 3d people models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8387–8397.

[2] A. Best, S. Narang, and D. Manocha. 2020. SPA: Verbal Interactions between Agents and Avatars in Shared Virtual Environments using Propositional Planning. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 117–126.

[3] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv preprint arXiv:1812.08008* (2018).

[4] Young-Woon Cha, True Price, Zhen Wei, Xinran Lu, Nicholas Rewkowski, Rohan Chabra, Zihe Qin, Hyounghun Kim, Zhaoqi Su, Yebin Liu, et al. 2018. Towards fully mobile 3D face, body, and environment capture using only head-worn cameras. *IEEE transactions on visualization and computer graphics* 24, 11 (2018), 2993–3004.

[5] Youjin Choi, Jeongmi Lee, and Sung-Hee Lee. 2020. Effects of Locomotion Style and Body Visibility of a Telepresence Avatar. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1–9.

[6] Allen J Fairchild, Simon P Campion, Arturo S Garcia, Robin Wolff, Terrence Fernando, and David J Roberts. 2016. A mixed reality telepresence system for collaborative space operation. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 4 (2016), 814–827.

[7] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 741–754.

[8] Tanmay V Randhavane, Aniket Bera, Emily Kubin, Kurt Gray, and Dinesh Manocha. 2019. Modeling data-driven dominance traits for virtual characters using gait analysis. *IEEE transactions on visualization and computer graphics* (2019).

[9] Patrick Stotko, Stefan Krumpen, Matthias B Hullin, Michael Weinmann, and Reinhard Klein. 2019. SLAMCast: Large-Scale, Real-Time 3D Reconstruction and Streaming for Immersive Multi-Client Live Telepresence. *IEEE transactions on visualization and computer graphics* 25, 5 (2019), 2102–2112.

[10] P. Stotko, S. Krumpen, M. Weinmann, and R. Klein. 2019. Efficient 3D Reconstruction and Streaming for Group-Scale Multi-Client Live Telepresence. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 19–25.

[11] Ramanarayan Vasudevan, Gregorij Kurillo, Edgar Lobaton, Tony Bernardin, Oliver Kreylos, Ruzena Bajcsy, and Klara Nahrstedt. 2011. High-quality visualization for geographically distributed 3-d teleimmersive applications. *IEEE Transactions on Multimedia* 13, 3 (2011), 573–584.