Perspective change in Augmented Reality

Seminar Telecooperation, Summer semester 2023

Jonas Wombacher
Technische Universität Darmstadt

1 INTRODUCTION

Typically, Augmented Reality (AR) is implemented by devices with a transparent display, like the HoloLens 2[1]. Such devices allow you to still see the real world around you, with virtual content overlaid on top. They are not able to fully occlude the real world through virtual content, and the user is therefore restricted to his own point of view (POV) on the environment around him.
Mixed Reality (MR) headsets like the Varjo XR-3[2] or Meta Quest Pro[3] contain fully enclosed, opaque displays. As they implement AR through a pass-through video feed from cameras on the outside shown on the opaque displays on the inside, they are able to occlude the real world completely. This enables them to switch to Virtual Reality (VR) to create the illusion of having changed the user's perspective while still using AR.

Previous research has shown that different contexts can benefit from different POVs.
As the third-person view can improve the user's environment perception [1, 10], it can for example provide an advantage over a first-person view (FPV) in situations requiring the user to assess distances and trajectories [10].
In a different scenario, it is possible to improve the efficiency of exploration, decreasing the physical movement required, by showing the user a manipulated FPV instead of the real one [11].

Partly based on the ability of MR headsets to change the user's POV according to the context, we are going to give an overview of some application scenarios of perspective change in AR, as well as a description of how we applied it in our implementation.

2 RELATED WORK

Different application scenarios for perspective changes in AR have different technical requirements for their implementation. In the following, we are going to cover two groups of exemplary use cases. The first group is based on dedicated POVs provided by cameras worn by users. The second group adds the availability of arbitrary POVs, made possible by switching to VR, showing a virtual model of the environment in order to keep up the illusion of still seeing an AR.

Group 1: Dedicated points of view provided by cameras

Use case: Remote collaboration

This application of perspective change focuses on collaboration, allowing a spectator to experience the remote environment of a second user through that user's FPV.
Such systems were implemented by Kasahara and Rekimoto [4, 3] and by Matsuda et al. [7]. In these related projects, the user to be observed wear cameras, allowing the spectator to perceive his surroundings by watching the cameras' feed on a screen [4] or a head-mounted display (HMD) [3].
The spectator then has the ability to communicate with the observed person through bidirectional voice transmission [7, 3] or through pointing input, which is visualized by an Augmented Reality HMD worn by the observed user [4].
Such a bidirectional flow of information allows the two users to collaborate. This can, for example, be useful in situations, where a remote expert gives instructions to a local worker [4, 3] or where a remote translator helps a person communicate with people speaking unfamiliar languages [7].

Use case: Empathy

Nishida et al. [9] built a system using an HMD and cameras attached to the user's waistline, which gives the user the ability to experience his own surroundings in the same way a smaller person would. This way the user can investigate his environment, looking for potential problems smaller persons would have to face.

Group 2: Arbitrary points of view

The projects presented in the first group relied on predefined perspectives, enabled through cameras worn by the system's user itself or collaborating users. This restricts the possible application scenarios, as the available perspectives are linked to the positions of the available cameras, which are fixed unless the user, or the user's collaborator, respectively, physically moves.
Detaching the perspectives from fixed cameras and allowing arbitrary POVs lets you overcome this limitation. In order to achieve this, the user's environment has to be captured to create a digital twin. Assuming the user is wearing an MR headset, he can then be shown arbitrary POVs of the digital twin in VR instead of the live AR pass-through video feed showing his own POV.

Digital twin through a live capture

One way of creating a digital twin of the environment is to use a live capture, for example by applying the capabilities of depth cameras like Kinect to reconstruct geometry in real-time [8, 2].

Tatzgern et al. [12] built a system with a live capture environment representation, which gives the user the ability to investigate parts of his surroundings that would not be visible from his own POV. The user can zoom in for closer inspection and zoom out, pan around objects or take on automatically computed top-down views for a better overview and the capability to see previously occluded objects.

Komiyama et al. [5] utilized a live capture to improve the type

[1] https://www.microsoft.com/en-us/hololens/hardware
[2] https://varjo.com/products/xr-3/
[3] https://www.meta.com/de/en/quest/quest-pro/

of remote collaboration we presented in the first group. In addition to being able to take on the FPV of other users wearing cameras, the spectator now also has the option to choose arbitrary perspectives in the room, provided by virtual cameras based on the live capture. This can potentially allow the spectator to get a better understanding of the environment and the situation he is watching.

Lindlbauer et al. [6] also enabled the user to freely move or teleport his POV across the environment based on a live capture. However, they also implemented more advanced abilities. By manipulating the captured geometry, they allowed users to modify their surroundings, for example by removing or duplicating objects.
Harnessing the fact, that the live capture continuously recognizes all changes in the environment, both caused by the user and caused by others, Lindlbauer et al. were able to implement time manipulations. Accessing the recorded geometry changes, the user can for example rewind interactions or freeze the surroundings in time in order to inspect a single point in time from different perspectives.

Digital twin through a static model created in advance

Another option for obtaining a digital twin of the environment is to capture a static model in advance, for example through photogrammetry or a LiDAR-based scan.
Unlike a live capture, a static model does not have the ability to depict changes in the real environment, for example other people walking through the room. However, static models have the benefit of being captured in advance and therefore not requiring access to capturing hardware like depth cameras during the application runtime.

Shin et al. [11] created a digital twin by scanning the environment with the LiDAR of an iPhone 13 Pro. By showing virtual POVs of the digital twin, which were offset and rotated relative to the user's real orientation in the room, they enabled amplified movement and amplified rotation. This reduced the amount of physical movement required for a user to explore the environment.

3 IMPLEMENTATION

Our goal was to explore the concept of applying XR to enable embodiment in the context of human motion simulation. To achieve this, we implemented a Unity application which we executed on a Meta Quest Pro. For the digital twin of the environment, we produced a static model in advance using an iPhone's LiDAR sensor with the app Polycam.
This allowed us to let the user leave his own FPV to take on the POVs of virtual avatars simulating human behavior. By taking the perspectives of avatars with different body heights and different kinds of visual impairments, users can investigate their environment in order to find problems they were unable to notice through their own POV.

The application included two main exemplary scenarios allowing you to experience the application of embodiment in different contexts, both safety-critical and mundane.
The objective of the first scenario was to make sure that all simulated persons present in the room could escape unhindered in the event of fire. By embodying the virtual avatars, one could notice that the fire alarm was placed too high up to be reachable by some of the smaller persons and that the emergency exit sign was indistinguishable for color-blind people. After fixing these problems, examining the simulated escape routes revealed potentially unnecessary detours around some heavy boxes lying on the ground. By removing them virtually, the user could investigate the unhindered escape routes without having to actually move the heavy boxes first.
As a contrast to the safety-critical first scenario, the second scenario covered a mundane context, namely practicing for a presentation. First, the user had to choose an appropriate projector screen height, embodying the simulated audience members to make sure everybody had an unobstructed view without having to stand up. As some of those audience members were affected by myopia, the user also had to adjust the presentation's font size. Finally, the user could practice his body language, guided by the feedback of the simulated audience, which for example would complain when the presenter would put their hands in their pockets.

REFERENCES

[1] Geoffrey Gorisse, Olivier Christmann, Etienne Armand Amato, and Simon Richir. 2017. First- and third-person perspectives in immersive virtual environments: presence and performance analysis of embodied users. *Frontiers in Robotics and AI*, 4. DOI: 10.3389/frobt.2017.00033.

[2] Shahram Izadi et al. 2011. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (UIST '11). Association for Computing Machinery, Santa Barbara, California, USA, 559–568. ISBN: 9781450307161. DOI: 10.1145/2047196.2047270.

[3] Shunichi Kasahara and Jun Rekimoto. 2015. Jackin head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology* (VRST '15). Association for Computing Machinery, Beijing, China, 217–225. ISBN: 9781450339902. DOI: 10.1145/2821592.2821608.

[4] Shunichi Kasahara and Jun Rekimoto. 2014. Jackin: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference* (AH '14) Article 46. Association for Computing Machinery, Kobe, Japan, 8 pages. ISBN: 9781450327619. DOI: 10.1145/2582051.2582097.

[5] Ryohei Komiyama, Takashi Miyaki, and Jun Rekimoto. 2017. Jackin space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference* (AH '17) Article 14. Association for Computing Machinery, Silicon Valley, California, USA, 9 pages. ISBN: 9781450348355. DOI: 10.1145/3041164.3041183.

[6] David Lindlbauer and Andy D. Wilson. 2018. Remixed reality: manipulating space and time in augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, Montreal QC, Canada, 1–13. ISBN: 9781450356206. DOI: 10.1145/3173574.3173703.

[7] Akira Matsuda, Kazunori Nozawa, and Jun Rekimoto. 2018. Jackin neck: a neckband wearable telepresence system designed for high comfortability. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (ISS '18). Association for Computing Machinery, Tokyo, Japan, 415–418. ISBN: 9781450356947. DOI: 10.1145/3279778.3279917.

[8] Richard A. Newcombe et al. 2011. Kinectfusion: real-time dense surface mapping and tracking. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, 127–136. DOI: 10.1109/ISMAR.2011.6092378.

[9] Jun Nishida, Soichiro Matsuda, Mika Oki, Hikaru Takatori, Kosuke Sato, and Kenji Suzuki. 2019. Egocentric smaller-person experience through a change in visual perspective. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). Association for Computing Machinery, Glasgow, Scotland Uk, 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300926.

[10] Patrick Salamin, Daniel Thalmann, and Frédéric Vexo. 2006. The benefits of third-person perspective in virtual and augmented reality? In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (VRST '06). Association for Computing Machinery, Limassol, Cyprus, 27–30. ISBN: 1595933212. DOI: 10.1145/1180495.1180502.

[11] Yerin Shin and Gerard Jounghyun Kim. 2022. Leveraging vr techniques for efficient exploration and interaction in large and complex ar space with clipped and small fov ar display. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology* (VRST '22) Article 3. Association for Computing Machinery, Tsukuba, Japan, 10 pages. ISBN: 9781450398893. DOI: 10.1145/3562939.3565613.
[12] Markus Tatzgern, Raphael Grasset, Denis Kalkofen, and Dieter Schmalstieg. 2014. Transitional augmented reality navigation for live captured scenes. In *2014 IEEE Virtual Reality (VR)*, 21–26. DOI: 10.1109/VR.2014.6802045.