Immersive Record and Replay for Lively Virtual Environments

Klara Brandstätter*
University College London

Figure 1: (Left): A dancing crowd in VR created by two users and multiple recordings and replays. (Center): Handover of a cupcake RIO between two recorded avatars of the same user. Audio indicators above the avatars' heads indicate when an avatar is speaking. The yellow markers on the audio indicators highlight when the RIO is grabbed by the avatar. (Right): Users performing dialogues in VR.

Abstract

Non-player characters (NPCs) play an important role in virtual environments and make them more lively. I am using recording and replaying techniques in virtual reality which enable a single user to interact with previous recordings of themselves. I conducted a user study that suggested that recordings that are performed by a single user are indistinguishable from multi-user recordings. Encouraged by these results, I would like to use machine learning to improve the existing recording and replaying techniques to create more sophisticated NPCs that can react to the user.

1 Introduction

Background characters or non-player characters are indispensable for many 3D and virtual reality (VR) experiences. They can offer guidance and support for the player during their adventures and make virtual worlds appear full of life. This liveliness is also important for social virtual reality (SVR) experiences where people gather to spend time with others in shared virtual spaces. NPCs in SVR can make the virtual spaces seem more populated and potentially more attractive to users, especially when fewer real people are around. However, animating responsive and realistic NPCs and scripting their behaviour can be a tedious process.

NPC behaviour is very often motion-captured and thus, NPCs look very convincingly human until they get approached by the user who tries to interact with them in a way that was not intended. In conventional 3D experiences, users usually interact with the environment and NPCs via button presses from their input devices (keyboard, joystick, controller). User avatars move in ways specific to the experience, as do the NPCs, and interactions with NPCs are limited to what is possible with the input devices. In VR, the user exists next to the NPCs and the interaction possibilities are much more plentiful. Inappropriate reactions from NPCs are more striking as the user might try to interact with an NPC like they would with a normal human. Especially in SVR experiences where multiple users and NPCs come together in the same virtual space, the difference in motion between avatars - controlled by the users - and animated NPCs is unmistakable.

I propose an approach that enables a single user to create lively and populated environments in VR using advanced recording and replaying techniques. The goals are: (a) to facilitate the creation of lively scenes by letting a single user act different characters in the scene, (b) to present a record and replay tool that supports a single user with creating dialogues and object interaction with past recordings, (c) to use machine learning techniques to modify the behaviour of recorded characters based on the user's actions, and (d) to analyse the plausibility and the realism of social scenes created by only one user.

2 Related Work - Recording and Replaying in VR

In the past, recording and replaying virtual environments has been useful for performance testing, diagnostics, and analysis of user behaviour in VR [1, 8]. Recording and replaying has also been used by Wang et al. [9] to share and relive recorded experiences together with other users in VR. Furthermore, VR is increasingly used for content creation (FlipsideXR, Mindshow), and the possibility to record oneself while embodying different characters makes it easier to produce animated videos all by oneself [2, 4]. Yin et al. [10] proposed the "one-man-crowd" (OMC) paradigm where a single user creates crowd motion data by recording and replaying their movements several times in VR. The OMC paradigm does not support multi-user or audio recordings and replays, therefore, it is possible to record over previous recordings multiple times. Existing content creation tools [2, 4] do support multi-user recordings and replays, however, they cannot record over existing recordings. Interaction with previous recordings of oneself is also difficult, as previous recordings are unable to react to later recordings.

*e-mail: k.brandstatter@ucl.ac.uk

3 Methodology - Interactive Recording and Replaying in VR

In the early stages of my PhD, I developed a record and replay tool for Ubiq, our networking library for research, teaching, and development [3]. It supports networked recording and replaying of multiple users (including their audio) and objects. The tool was tested during a small pilot study where I recorded participants in VR while they had to solve jigsaw puzzles [7]. Afterwards, I replayed the recordings to participants and they had to comment on their actions to figure out if they were aware of certain unconscious behaviours. The developed record and replay tool was an improvement of OMC paradigm as it was suitable for multiple users, and allowed the recording of previous replays.

The record and replay tool was a first step towards creating simple populated environments. In Figure 1 (left), a dancing crowd of cartoony avatars was created from two dancing users that were recorded multiple times in VR. However, these recordings were not interactive. Different recordings could not really interact with each other, especially earlier recordings would never interact with later recordings, as the later recordings did not yet exist. This made verbal interactions or interactions with objects between different recordings very difficult.

4 Improving Interactivity between Recordings

To improve the interactivity between recordings, especially single-user recordings, I added audio indicators to the avatars in VR. These can be seen in Figure 1 (center). The audio indicators would show the user when a previous recording would speak. This enabled users to perform pre-scripted dialogues with their past recordings. Additionally, I developed *Recordable Interactable Objects* (RIOs) that could be passed between the user and previous recordings of the user. All the user needs to know is when the RIO will be released or grabbed by the previous recording in order to take over or hand back the RIO respectively. Figure 1 (center) shows two recorded avatars passing a cupcake RIO between each other.

5 Testing for Plausibility

To test if dialogues performed by a single user would be as convincing as dialogues performed normally by two users, I conducted a user study with 50 participants recruited on Prolific. The study consisted of two parts. In the first part, participants had to watch 12 sets of 2 short videos and choose for each set which video they preferred. For each set, one video was performed by only one user while the other was performed by two users. In the second part of the study, participants were told that one video was performed by only one user and their task was to identify the videos that were recorded by one user. They were shown the same 24 videos but in a different order and not in sets of 2. The video dataset was recorded with two additional participants who did not know what the purpose of the dataset creation was and who did not have any acting experience. They both recorded the dialogues together and separately. The dialogues were taken from Amazon's commonsense-dialogues dataset and consisted of 4-6 turns between two people [11]. The voices of the two users were modified using the AI voice transformer *Koe: Recast* [6]. Figure 1 (right) shows a kitchen scene in which the users were performing the dialogues.

6 Preliminary Results

From the 50 results that were gathered from participants, 3 results were excluded as the participants always selected the same answers for each question. To test whether single-user recordings are as good as multi-user recordings, equivalence testing was performed with two one-sided t-tests [5]. The equivalence test was significant, $t(64) = -6.32, p < 0.0005$ (90% CI, -0.015 to 0.033). This indicates that participants could not tell whether a dialogue was performed by one user or two users. The results from the study also suggest that participants preferred single-user videos over multi-user videos.

In a follow-up study, participants will do the same tasks again, but they will be watching the dialogues in VR. This will provide insights into whether participants are better at detecting single-user videos in VR.

7 Conclusion

Using interactive single-user record and replay techniques is a promising way to create multi-user content of equally good quality.

7.1 Future Work and Improvements

I would like to improve the existing record and replay tool and include machine-learning techniques to make recordings more adaptable and reactive to the user, e.g. if the user does not stand in the right position to take over an object from the recording, the recording learns to face the user, or if a user places an object in the wrong position, the recording can adjust and grab the object from the new position. Another idea could be recordings, which are not exact copies of the user but can have different styles or personalities, e.g. gestures could be enhanced or reduced.

Acknowledgments

This work received funding from the EU H2020 project CLIPE (grant number 86076).

References

[1] N. Developer. NVIDIA VR Capture & Replay. `https://developer.nvidia.com/vcr-early-access`, 2022. Accessed: 2022-12-12.

[2] FlipsideXR. Flipsidexr website. `https://www.flipsidexr.com/`, 2022. Accessed: 2022-12-12.

[3] S. J. Friston, B. J. Congdon, D. Swapp, L. Izzouzi, K. Brandstätter, D. Archer, O. Olkkonen, F. J. Thiel, and A. Steed. Ubiq: A system to build flexible social virtual reality experiences. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, VRST '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3489849.3489871

[4] M. Inc. Mindshow website. `https://mindshow.com/`, 2022. Accessed: 2022-12-12.

[5] D. Lakens. Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4):355–362, 2017. PMID: 28736600. doi: 10.1177/1948550617697177

[6] K. Recast. Koe: Recast website. `https://koe.ai/`, 2022. Accessed: 2022-12-12.

[7] A. Steed, L. Izzouzi, K. Brandstätter, S. Friston, B. J. Congdon, O. Olkkonen, D. Giunchi, N. Numan, and D. Swapp. Ubiq-exp: A toolkit to build and run remote and distributed mixed reality experiments. *Frontiers in Virtual Reality*, p. 131.

[8] A. Steed, M. Wang, and J. Drummond. *Recording and Replaying Virtual Environments for Development and Diagnosis*, pp. 349–361. 06 2019. doi: 10.1201/b21598-18

[9] C. Y. Wang, M. Sakashita, U. Ehsan, J. Li, and A. S. Won. Again, together: Socially reliving virtual reality experiences when separated. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376642

[10] T. Yin, L. Hoyet, M. Christie, M.-P. Cani, and J. Pettré. The one-man-crowd: Single user generation of crowd motions using virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2245–2255, 2022. doi: 10.1109/TVCG.2022.3150507

[11] P. Zhou, K. Gopalakrishnan, B. Hedayatnia, S. Kim, J. Pujara, X. Ren, Y. Liu, and D. Hakkani-Tur. Commonsense-focused dialogues for response generation: An empirical study. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, Singapore and Online, 2021.