Motion Tracking Avatars in Desktop Multiplayer Online Games

CAELEN WANG, San Francisco State University, USA

Piloting a virtual avatar using physical motion allows multiplayer online games to foster vibrant worlds where digital personas feel alive and connected to their corporeal counterparts. However, this affordance has been exclusive to virtual reality and console games that require specialized hardware. This paper proposes a system to seamlessly integrate real-time motion puppetry into a traditional third-person desktop game using webcam video. The player retains the full set of controls and conveniences offered by the mouse and keyboard, while having their hand gestures and facial expressions reflected on a humanoid avatar without it being intrusive to a stationary, seated playing experience. This feature enhances virtual embodiment and presence, which in turn increases intrinsic motivation towards game objectives, and engenders engaging social interactions with other players.

CCS Concepts: • **Human-centered computing** → **Gestural input**; *Mixed / augmented reality*.

Additional Key Words and Phrases: virtual avatars, social interactions, multiplayer online games, pose estimation

ACM Reference Format:
Caelen Wang. 2022. Motion Tracking Avatars in Desktop Multiplayer Online Games. 1, 1 (November 2022), 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Virtual reality allows the player to step into immersive worlds through an avatar that reflects physical motion as if they are in the character's shoes. This level of virtual embodiment present in VR is lost in desktop games due to the fact that the avatar is operated through an intermediate non-diegetic input scheme, such as mouse/keyboard or controller. The system presented in this paper aims to bridge the gap between desktop and virtual reality by establishing a corporeal correspondence between player and avatar bodies. While playing seated in front of a webcam, the player's hand gestures and facial expressions are mapped onto their avatar as they are detected in real-time.

The player-avatar relationship sits at the forefront of any gaming experience, since it is how the player interacts with the game world [2]. One value that players often look for in their avatars is self-expression that produces the illusion of a digital self [6]. The motion tracking feature enables the player to easily express their physical state through their avatar, which strengthens affinity towards the in-game character. Self-identification with one's avatar can ultimately lead to intrinsic motivation for playing the game [1]. Through gesture and expression mirroring, the player gains emotional incentive to play the game beyond those laid out explicitly in the game design.

These embodiment effects become compounded as they are experienced with other players. In traditional multiplayer games, social interactions are conducted through emotes with preset animations. There is high latency in such exchanges with no immediate indication of when an interaction evokes emotional response. Emotes alone are not responsive enough to form attentive affective signaling found in virtual reality settings [4]. The motion tracking feature mitigates

Author's address: Caelen Wang, ywang36@sfsu.edu, San Francisco State University, 1600 Holloway Ave, San Francisco, California, USA, 94132-1722.

Manuscript submitted to ACM

this shortcoming by enabling genuine, fluid interactions as they occur. Through the reflection of player emotions onto their avatars, social interactions in multiplayer games become more personal and engaging.

Fig. 1. Character Select

Fig. 2. Multiplayer Hub

2 FEATURE DESIGN

During the initial character selection stage, the player is met with a closeup, mirror view of the avatar alongside webcam pose estimation results. This is an opportunity for the player to acclimate to the motion tracking feature as well as adjust external factors such as camera position and lighting to ensure the best experience. As the player sees their gestures and expressions in the avatar mirror, an enfacement effect will conduce self-identification with the avatar [3]. To best illustrate the embodiment effects, the feature is built on top of a third-person game with a positional tracking, free rotating camera. The WASD keys navigate the avatar and the camera within the game, while the mouse rotates the camera so that the avatar can be viewed from any direction. The avatar performs preset animations such as idle, walk, run, and jump following traditional game logic. The motion tracking system consists of three modular, independent components: face, arms, and hands. Pose estimation is performed each frame and results are mapped onto the avatar as they become available, overriding the preset animation of the target body part. The design follows a zero downtime principle, where the avatar is always active whether through motion tracking or preset animation. The game demo is set in a mountainside village, reminiscent of an MMORPG hub town. Within the hub world, players can greet one another with physical gestures and expressions. This behavioral realism augments social presence beyond that of a traditional multiplayer game [5].

3 TECHNICAL IMPLEMENTATION

The game demo is built using Unity, with multiplayer functionality provided by Photon. The avatars are created through ReadyPlayerMe, which provides detailed personalization. Body pose and facial feature estimation is performed using Google MediaPipe [7]. In particular, MediaPipe Holistic is used for simultaneously face, body, and hand landmarks inference. A hand-tuned blendshape solver uses the face landmarks to animate avatar facial expressions. A forward kinematics solver uses the body and hand landmarks to set avatar arm and finger joint rotations.

4 CONCLUSION

This paper proposes a novel avatar motion control system for desktop games using webcam video. The system does not require specialized hardware besides a webcam, and seamlessly integrates into a seated gaming experience. Mapping

the player's body and face motions onto their avatar creates a level of virtual embodiment unseen in previous desktop games. This feature allows players to identify with their character and gain intrinsic motivation towards the game. The increased sense of social presence encourages engaging and personal virtual interactions.

REFERENCES

[1] Max V. Birk, Cheralyn Atkins, Jason T. Bowey, and Regan L. Mandryk. 2016. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2982–2995. https://doi.org/10.1145/2858036.2858062

[2] Alison Gazzard. 2009. The Avatar and the Player: Understanding the Relationship beyond the Screen. In *2009 Conference in Games and Virtual Worlds for Serious Applications*. 190–193. https://doi.org/10.1109/VS-GAMES.2009.11

[3] Mar Gonzalez Franco, Anthony Steed, Steve Hoogendyk, and Eyal Ofek. 2020. Using Facial Animation to Increase the Enfacement Illusion and Avatar Self-Identification. *IEEE Transactions on Visualization and Computer Graphics* (March 2020). https://www.microsoft.com/en-us/research/publication/using-facial-animation-to-increase-the-enfacement-illusion-and-avatar-self-identification/

[4] Heidi S. Kane, Cade McCall, Nancy L. Collins, and Jim Blascovich. 2012. Mere presence is not enough: Responsive support in a virtual world. *Journal of Experimental Social Psychology* 48, 1 (2012), 37–44. https://doi.org/10.1016/j.jesp.2011.07.001

[5] Simon Kimmel and Wilko Heuten. 2022. Elevating Social Presence in Multi-User VR by Increasing Behavioral Realism.

[6] Ian J. Livingston, Carl Gutwin, Regan L. Mandryk, and Max Birk. 2014. How Players Value Their Characters in World of Warcraft. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work amp; Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1333–1343. https://doi.org/10.1145/2531602.2531661

[7] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. *CoRR* abs/1906.08172 (2019). arXiv:1906.08172 http://arxiv.org/abs/1906.08172