Socially Immersive Avatar-Based Communication

Daniel Roth[1]* Kristoffer Waldow[2,3]† Marc Erich Latoschik[1]‡ Arnulph Fuhrmann[3]§ Gary Bente[2,4]¶

[1]University of Würzburg, Germany
[2]University of Cologne, Germany
[3]TH Köln, Germany
[4]Michigan State University, USA

Abstract

In this paper, we present SIAM-C, an avatar-mediated communication platform to study socially immersive interaction in virtual environments. The proposed system is capable of tracking, transmitting, representing body motion, facial expressions, and voice via virtual avatars and inherits the transmission of human behaviors that are available in real-life social interactions. Users are immersed using active stereoscopic rendering projected onto a life-size projection plane, utilizing the concept of "fish tank" virtual reality (VR). Our prototype connects two separate rooms and allows for socially immersive avatar-mediated communication in VR.

Index Terms: H5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities—; H.4.3 [Communications Applications]: —

1 Introduction

Immersive virtual environments may play an important role as social media of the future. However, multiple technological and social scientific challenges have to be tackled [5]. From a technological perspective, previous research has successfully shown avatar control from vision based systems for the body [7], and the face [9]. In addition, it was shown that (immersive) virtual environments can be created connecting multiple remote users [3, 1, 2]. However, a systematic exploration of the transmission of the full set of realistic human behaviors and their impact on social virtual interaction is still missing. To this regard, the research space of "social immersion" can be structured into four different aspects that may have impact on the user's perception of the social interaction: *Presence*, *Appearance Realism*, *Behavioral Realism*, and *Control Realism*. In contrast to previous approaches, *SIAM-C* (*Socially Immersive Avatar-mediated Communication Platform*) is developed with the requirements to be capable to test the impacts of the realism of behavioral dynamics (facial expression, body motion, gaze - R1), the impact of appearance realism (R2), and the impact of control realism (R3). We set strong focus on the tracking quality and possibilities of recording the performed motions, in order to find implicit differences that may occur due to missing behavioral richness. Therefore, we chose most robust components for our system. Furthermore, the system is designed for the capability of dyadic interactions that should be scalable to smaller groups (R4). The goal of our work was to develop a platform to research the key aspects of social immersion.

*e-mail: daniel.roth@uni-koeln.de
†e-mail: kris_waldow@gmx.de
‡e-mail: marc.latoschik@uni-wuerzburg.de
§e-mail: arnulph.fuhrmann@th-koeln.de
¶e-mail: gabente@msu.edu

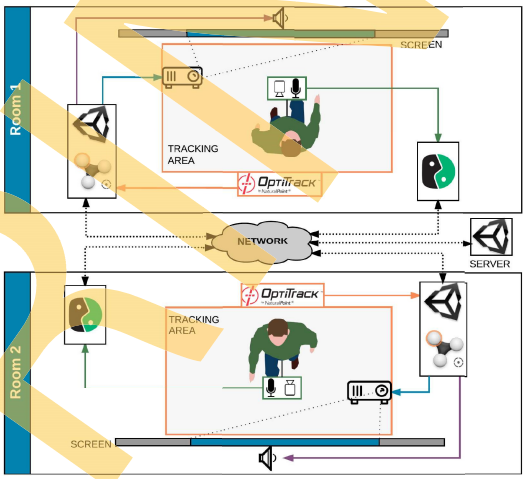

Figure 1: SIAM-C system setup and components. Two remote locations are identically equipped and networked. Two PCs track the users motion and render the simulation. Two laptops track the facial expression and audio. A server controls the system.

2 System and Setup

SIAM-C is designed to be capable of tracking, transmitting and reproducing human behaviors to a large extend to virtual characters. The system's setup is illustrated in figure 1, and includes two remote OptiTrack marker-based body tracking systems (2x8 Flex3 100hz cameras). For the facial (and gaze) tracking, we use two RGB-depth sensors (PrimeSense Carmine 1.09) and facial expression tracking software (Faceshift). We attached the RGB-depth sensor, a LED stripe, and a microphone to a steady shot camera rig worn by the users in front of their body, allowing for freedom of movement. Audio is presented using a Genelec 8020A speaker mounted behind the projection screen or headphones via the Teamspeak. Two active 3D short-throw projectors (Acer H6517ST) mounted in 90° angle render a picture of $2,25m$ x $1,49m$.

The VR projection followed the "fish tank" paradigm [8] using head coupled perspective projection. The avatars can appear in life size on the screen plane. Occlusions first appears about 45cm in front of the projection. We approximated a vertical pixel resolution of 2.86ppi, and 1.85ppi horizontal resolution. We chose to implement a software solution to transmit side-by-side progressive full hd images of 1080x960 pixel per image. The virtual environment is implemented in Unity3D. Figure 2 shows the system's flow for one user. The tracking data is streamed via network analog to [6], and interfaced via Mecanim, driving the avatar behavior. In a combination of centralized/decentralized systems, one machine acted as central server to configure and control both clients with their respective streams (see fig 1). To reduce latency, the actual behavior data

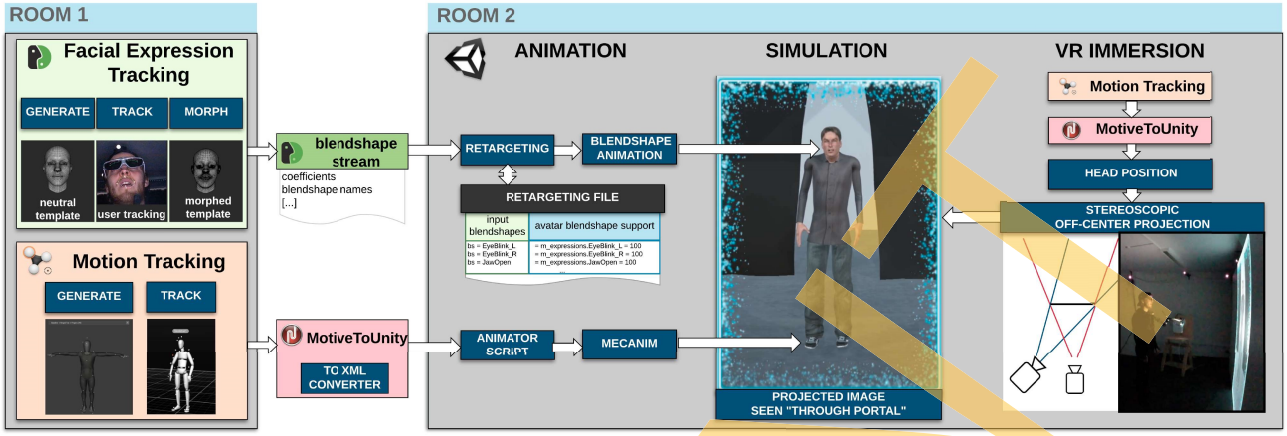

Figure 2: Flow of SIAM-C. Faceshift captures the users facial expression, streaming blendhshape information via network to Unity3D from room 1 to room 2. Using retargeting, blendshapes are mapped to the avatar blendshape targets. The Optitrack motion capture system tracks the users body motion, creates a skeleton and distributes the data via NatNet. The data is then converted to readable XML for its use in Unity3D through the software MotiveToUnity adapted from [3]. They are interfaced with Mecanim and finally mapped on the target avatar. The VR immmersion is a combination of stereoscopy and off-axis projection. The fictional portal enhances the immersion and reduces edge perception.

(i.e. body motion data and blendshape weights) can be sent from client to client in a decentralized way. SIAM-C utilizes a mapping table to retarget facial expressions to the virtual characters.

3 Discussion

With respect our initial requirements we accounted for R1 as a diverse set of avatars are compatible without rescripting the control interface. Our solution also enables a rigid-body based inverse-kinematic approach to reduce the invasiveness of the marker-based tracking for the user. We can control the dimensions of behavioral realism (R2) including the behavioral channels of (lower head) facial expression, voice and body motion in the stereoscopic VR setting, and facial expression, gaze, voice, and body motion in a non-stereoscopic setting. We did not yet test or develop additional control mechanisms requested in R3, although interfaces are present. We accounted and tested for R4 as the system is capable of multiuser interaction. In contrast to [2], our system does not represent real surroundings, but both users can be immersed and 3D tracked. SIAM-C does not have a bezel or occluding objects in the projection except for the steady cam rig, which is set up below the line of vision. The approach by [3] incorporated multiple users. However, the appearance is limited to a small display sized screen. We decided to include a large projection in order to be able to replicate the full range of nonverbal behavior in life size. The approach from [1] also involves full body tracking, however they do not replicate facial expression. Furthermore, SIAM-C is built as full simulation without real-user replications in order to account for R1, R2, and R3. Although the projection is limited, our system allows for the user to be tracked in 360°, which enables CAVE like multi-user scenarios. Using a virtual audio clapperboard, all data can either be recorded synchronously or synchronized in post processing for further analysis.

3.1 Conclusion and Future Work

In this paper, we presented a system capable of exploring the research space of social immersion. Three of the four initial requirements are met. In contrast to other approaches, SIAM-C transmits a larger set of human nonverbal behaviors. Future technological development should investigate and include a solution for the robust integration of eye-tracking sensors into the active shutter glasses and a more stable configuration for facial expression tracking in the stereoscopic VR setting. A user study aims at investigating the effects of behavioral realism on communication, systematically varying behavioral channels.

Acknowledgements

We thank Dmitri Galakhov, Arvid Hofman, Felix Stetter, Carola Bloch, Sebastian Lammers, Bastian Jarczewsky, Dominik Gall and NaturalPoint for their support in this project.

References

[1] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):616–625, 2013.

[2] A. Maimone and H. Fuchs. A first look at a telepresence system with room-sized real-time 3d capture and life-sized tracked display wall. *Proceedings of ICAT 2011*, pages 4–9, 2011.

[3] K. Otsuka. MMSpace: Kinetically-augmented telepresence for small group-to-group conversations. In *Virtual Reality (VR), 2016 IEEE*, pages 19–28. IEEE, 2016.

[4] D. Roberts, R. Wolff, J. Rae, A. Steed, R. Aspin, M. McIntyre, A. Pena, O. Oyekoya, and W. Steptoe. Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. In *2009 IEEE Virtual Reality Conference*, pages 135–142. IEEE, 2009.

[5] D. Roth, M. E. Latoschik, K. Vogeley, and G. Bente. Hybrid Avatar-Agent Technology A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges. *i-com*, 14(2), Jan. 2015.

[6] D. Roth, J.-L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. Avatar Realism and Social Interaction Quality in Virtual Reality. In *Proceedings of the IEEE Virtual Reality (IEEE VR) Conference 2016*, Greenville, 2016.

[7] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[8] C. Ware, K. Arthur, and K. S. Booth. Fish tank virtual reality. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 37–42. ACM, 1993. 00343.

[9] T. Weise, S. Bouaziz, H. Li, and M. Pauly. Realtime Performance-based Facial Animation. In *ACM SIGGRAPH 2011 Papers*, SIGGRAPH '11, pages 77:1–77:10, New York, NY, USA, 2011. ACM.