FaceBo: Real-Time Face and Body Tracking for Faithful Avatar Synthesis

Jean-Luc Lugrin *
University of Würzburg

David Zilch †
University of Würzburg

Daniel Roth ‡
University of Würzburg
University of Cologne

Gary Bente §
Michigan State University
University of Cologne

Marc Erich Latoschik¶
University of Würzburg

Abstract

This paper introduces a low-cost framework capable of combining both real-time markerless face and body tracking for faithful avatar embodiment in Virtual Reality (VR). We discuss suitable hardware and software solutions and present a first prototype. This work lays the technological basis for further research on the importance of the appearance and behavioral realism of avatars, e.g., for the illusion of virtual body ownership, for social interactions in VR, as well as for VR entertainment applications (immersive games or movies).

Index Terms: H.5.1 [Information Systems]: Artificial—Augmented and Virtual Realities

1 Introduction

Virtual embodiment (or avatar embodiment) is the substitution of a person's body with a virtual replica [6]. The required degree of realism of avatars with respect to both, appearance and behavior [2], is an open research question [3]. Current VR hardware and software systems and avatar interfaces are not capable of reproducing truly realistic replicas. A major challenge is the faithful replication of the variance and subtleties of facial expression and eye gaze in real-time. Both channels convey social signals which can have a critical impact on avatar-mediated communication [4] and possibly on the *illusion of virtual body ownership* (IVBO) [3]. Although the factors and effects of IVBO have been actively researched [6], the importance of faithful avatar appearance and behavioral realism remains unclear [3]. Further research should provide a deeper understanding of self-consciousness mechanisms, and in particular, self-face perception and face ownership [1].

We therefore propose a system for the faithful reproduction of users' behaviors and appearances, providing face scanning and real-time reproduction of body movements, facial expressions, and eye gaze. Our approach targets semi-immersive, (large-)screen-based VR settings where users experience an avatar replicating their behaviors, e.g. by using a magic-mirror metaphor. The technical infrastructure for avatar embodiment in VR system is typically complex, intrusive, expensive, not transportable and requires a considerable amount of time to equip and calibrate the system for each user [6]. To counteract these deficits, we investigated potential hardware and software tools combination and partial solutions given the following requirements: *R1)* low intrusiveness, *R2)* low cost, *R3)* sensors mutual compatibility, *R4)* fast user calibration, *R5)* accurate tracking in a given range of operation, and finally *R6)* Integrated in state-of-the-art graphics or simulation engines (such as *Unity*® or *Unreal*® engines, which are both popular within the VR research and professional communities, and capable of providing high-quality avatar animation and appearance.)

*e-mail:jean-luc.lugrin@uni-wuerzburg.de
†e-mail:david.zilch@stud-mail.uni-wuerzburg.de
‡e-mail:daniel.roth@uni-koeln.de
§e-mail:bente@uni-koeln.de
¶e-mail:marc.latoschik@uni-wuerzburg.de

2 Facial Expression Tracking Software

Facial tracking systems can be divided into i) *user template-based* and ii) *non-template based* applications. They either utilize a) RGB images, b) depth images, or c) both as input [8]. Their output usually is based on virtual markers to drive skeletons or blend shapes. We analyzed four different solutions for facial tracking, namely *Pro Face 2*, *Performer Suite*, *Faceshift Studio* and *FacePlus*. We tested and analyzed critical features to fulfill the requirements for each system using simple demo scenes. Table 1 illustrates our overall comparison criteria and results. We found Faceshift Studio to robustly and adequately track the user's facial expression, due to the user template generation and the utilization of both RGB and depth data. The system can also replicate a faithful 3D model of user's head after a short training phase. This choice led to the selection of Unity® as visualization engine since no plugin exists for Unreal®. Tracking systems requesting markers or only head-mounted camera have not been considered, since they are not satisfying our low invasiveness requirement (*R1*).

Table 1: Comparison of facial tracking software solutions.

	ProFace2	Performer	Faceshift	FacePlus
Robust Tracking	×	✓	✓	✓
Accurate Tracking	×	✓	✓	✓
Approximate Price	$150	$10000	$1500	$1500
Tracking without Training	✓	×	×	✓
Manual Training Process	×	✓	✓	×
Webcam video	×	✓	×	✓
Depth sensing device	✓	×	✓	×
Head Mesh Creation	×	×	✓	×
Integrated in Unity3D	✓	✓	✓	✓
Integrated in Unreal	×	✓	×	×

3 Body Tracking Software

Markerless body tracking systems compatible with Unity® and utilizing the *Kinect v2* can be divided into three types: i) *Unity plugins*(e.g., *Kinect v2 with MS-SDK*), ii) *VR Middleware* (e.g., *MiddleVR*), and iii) *Standalone applications with a Unity plugin* (e.g., *Brekel Pro Body*). The comparison is depicted in Table 2. Brekel Pro Body v2 is our final choice due to its elaborated smoothing techniques and the ready-to-use plugin integration. It allows an extensive control of the applied smoothing and the resulting latency via the user interface and an instant mapping of tracked user movements to the avatars.

Table 2: Comparison of body tracking software solutions.

	Pro Body v2	MiddleVR	Unity Plugins
Tracking without Training	✓	✓	✓
Training Process available	✓	×	×
Smoothened Tracking Data	✓	×	×
Network Capability	✓	✓	×

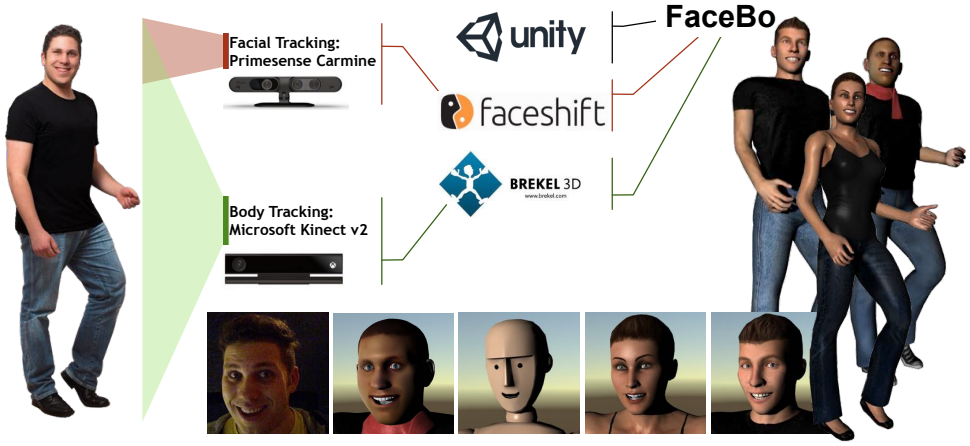

Figure 1: FaceBo - A low cost and markerless framework integrating both real-time face and full-body avatar embodiment in VR.

4 Prototype: The FaceBo System

Following our initial requirements and comparative tests, we selected Pro Body and Faceshift Studio as the target software solutions. Both deliver state-of-the-art tracking quality (*R5*), fast user calibration (*R4*) and model integration into Unity® (*R3, R6*). A fully-working prototype has been developed. The system is called ***FaceBo***, and its overall architecture is depicted in Figure 1. The system drives Faceshift Studio with the Carmine 1.09 and Pro Body with the *Kinect v2* (both software and hardware sensors fulfill *R1*, *R2*, *R3*, *R5* and *R6*). Both sensor devices work well in combination: the Kinect v2 utilizes time-of-flight to measure depth [5] and the Carmine 1.09 uses structured light [7]. As depicted in Figure 1, FaceBo is capable of projecting both user's facial expression and body movement to a multitude of different avatars in real-time with a very short calibration and equipment phase (*R4*).

The FaceBo system includes a Unity® module providing a generic avatar, animated by the combined input from both tracking systems. It also provides a high-level API to customise and calibrate avatar's dimension to individual body sizes and proportions. In addition, our generic avatar permits to rapidly import character models produced by popular 3D modelling softwares, such as *Autodesk Character Generator* or *Poser*. One of the most interesting features is truly the large spectrum of avatar types possible, and their interchangeability; from low realistic (e.g., mannequin avatar) to moderately realistic (e.g., human male or female avatar) to highly realistic (e.g., custom head mesh and texture) avatars. FaceBo allows developers to quickly import new avatars as well as to replace their head by the scanned user's head model from Faceshift Studio, after a short modelling phase of approximately 15-20 minutes. The optimal operating range of the system is represented by a volume of approximately $1.3m$ length $\times 1.3m$ width $\times 2m$ height *(R5)*, which suitable for a large variety of VR applications and experiments settings, especially for magic-mirror configurations.

5 Conclusion

We have analyzed hardware and software tools and partial solutions for a real-time markerless combined body and face tracking. The process was guided by 6 initial requirements constraining potential solutions with respect to general criteria like overall cost (of the equipment as well as of the final usage), tracking quality, and mutual compatibility. A suitable combination of partial solutions could be identified and a first prototype has been developed. The preliminary results of this prototype are promising. Our future work will focus on a formal evaluation of the system performance in terms of end-to-end latency, perceived embodiment quality, as well as API and workflow usability. We will investigate alternatives to our selected tracking systems (e.g., *Intel RealSense*, *Faceware Live*,) in order to both, respond to novel technologies, and/or possible distribution or licensing variations of current ones (e.g., Faceshift).

The presented prototype therefore lays the ground for further investigations of avatar's appearance, behavioral realism and avatar-mediated communication in VR. For instance, it is currently applied to research on the impact of avatar realism on social interaction in collaborative virtual environments, as well as on the exploration of a potential uncanny valley effect towards realistic avatars [3]. The system also offers novel perspective to VR-based entertainment applications (e.g., immersive games or movies) to easily integrate full user avatar embodiment in current or future applications. We hope that the replicability and planned open-source distribution will foster future applications and research on avatar embodiment.

References

[1] O. Blanke, M. Slater, and A. Serino. Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*, 88(1):145 – 166, 2015.

[2] D. Bombari, M. Schmid Mast, E. Cañadas, and M. Bachmann. Studying social interactions through immersive virtual environment technology: Virtues, pitfalls, and future challenges. *Frontiers in Psychology*, 6(869), 2015.

[3] J.-L. Lugrin, J. Latt, and M. E. Latoschik. Anthropomorphism and illusion of virtual body ownership. In *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*, 2015.

[4] 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 Journal of Interactive Media*, 14(2), 2015.

[5] H. Sarbolandi, D. Lefloch, and A. Kolb. Kinect range sensing: Structured-light versus time-of-flight kinect. *Computer Vision and Image Understanding*, 139:1 – 20, 2015.

[6] B. Spanlang, J. Normand, D. Borland, K. Kilteni, E. Giannopoulos, A. Pomés, M. González-Franco, D. Pérez-Marcos, J. Arroyo-Palacios, X. Navarro, and M. Slater. How to build an embodiment lab: Achieving body representation illusions in virtual reality. *Front. Robotics and AI*, 2014, 2014.

[7] S. Vongbunyong and W. H. Chen. *Disassembly Automation: Automated Systems with Cognitive Abilities*. Springer, 1 edition, 2015.

[8] 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.