

# Digital Cloning for an Increased Feeling of Presence in Collaborative Virtual Reality Environments

Sylvain CHAGUE\*, Caecilia CHARBONNIER  
Artanim Foundation, Geneva, Switzerland

DOI: 10.15221/15.251 <http://dx.doi.org/10.15221/15.251>

## Abstract

Embodying an avatar in virtual reality (VR) experiences is very important to reach a high feeling of presence. This is even more important in a collaborative VR scenario where multiple users must interact to achieve a certain task. In this paper, we present a new VR platform combining 3D body scanning, motion capture and head mounted display allowing people to walk freely in a virtual environment, look at their own body, and interact with other users and physical objects. Instead of being represented by generic avatars, users can embody a digital clone of themselves obtained by a 3D scanner.

**Keywords:** virtual reality, body scanning, avatars, character animation

## 1. Introduction

The key point to an effective and natural virtual reality (VR) experience is a strong feeling of presence. According to Jason Jerald [1], this sensation can be divided into four sub-components: 1) the illusion of being in a stable spatial space, 2) the illusion of self-embodiment, 3) the illusion of physical interaction, and 4) the illusion of social communication. The first element is considered the most important and is relatively easy to reach with current consumer VR hardware [2] [3] [4] and game engines. Indeed, the combination of high definition screens, low latency and high framerate with head tracking can easily fool the users' brain into thinking that he is in a physical environment. However, the limited tracking range and the lack of full body tracking currently limit the level of presence that can be achieved in a VR experience.

This latter fact, the illusion of self-embodiment, i.e. the illusion of having a body in the virtual space, requires the use of virtual avatars. If the user can physically move, a tracking system must compute his body motion and transfer it to the avatar so that the motion of the avatar matches the user's one. Our brain is very skilled at detecting and reading human motion. This means that this illusion can be easily broken if the animation or the appearance of the avatar is not plausible. This is even more important when multiple users share the same VR experience in the same physical space and have to interact together.

By combining motion capture with head mounted displays (HMD), the user has the possibility to impersonate an avatar, improving a lot the illusion of self-embodiment [5] [6] [7] [8]. The impact of the realism of the avatar on the feeling of body ownership is still an open question. A previous study by Kilteni et al. [6] showed that this illusion can be intensified by the realism of the avatar (e.g., correct shape, volume or color of the skin). On the other side, another study by Lugin et al. [8] showed that a realistic human appearance was not a critical factor in the illusion of virtual body ownership. People can also rapidly learn how to use a body completely different from their own (e.g., a body with 3 arms) [9].

One of the main difficulties encountered while trying to animate virtual avatars lies in the retargeting of the user's motion on the avatar's animation rig, which can be challenging if the sizes of the user and character greatly mismatch. This could lead to unrealistic motion and thus a lower feeling of presence.

The past years have seen a huge increase in 3D body scanning solutions ranging from affordable solutions [10] [11] [12] [13] [14] [15] to high-end expensive systems [16] [17]. The ability to scan human bodies has been democratized. Solutions to transform raw body scans into game ready digital avatars have also been studied and multiple solutions are now available [10] [12]. We believe that using a digital clone of his self as an avatar in virtual reality can lead to an increased feeling of embodiment and thus an increased feeling of presence.

\* chague@artanim.ch; +41 22 596 45 39; www.artanim.ch

The past years have seen a huge increase in 3D body scanning solutions ranging from affordable solutions [10] [11] [12] [13] [14] [15] to high-end expensive systems [16] [17]. The ability to scan human bodies has been democratized. Solutions to transform raw body scans into game ready digital avatars have also been studied and multiple solutions are now available [10] [12]. We believe that using a digital clone of his self as an avatar in virtual reality can lead to an increased feeling of embodiment and thus an increased feeling of presence.

In this paper, we present a novel multi-user VR platform combining 3D body scanning, optical motion capture and HMD. In this platform, users are represented in the virtual world by a 3D digital clone of themselves created from scan data and whose motion accurately matches their motion. They can also see and physically interact with the other users or with physical objects.

## 2. Methods

### 2.1. 3D body scanning

Before stepping in the VR experience, it is necessary to obtain a realistic digital clone of the users. Users' bodies were hence first 3D scanned using a custom-made photogrammetric 3D scanner (Fig. 1) consisting of 96 Canon Powershot A1400 cameras (12 poles of 8 cameras). Each pole had a dedicated 8 port USB hub connected to the workstation and a 3.3V power supply for the cameras. Two LED strips were positioned on each pole to provide an even illumination of the subject, which is very important to capture a good texture.

The cameras were running the Canon Hack Development Kit [18] (CHDK), an open-source alternate firmware allowing each camera's parameters and tasks to be controlled by a LUA script. In addition, the USB extension chdp-ptp [19] was used to remotely control the cameras from a single computer, trigger synchronized shoot and automatically download the pictures to the computer. This allowed the scanner to be controlled by a simple user interface. The synchronization mechanism was software only which resulted in a maximum sync error of about 100 ms. This can be considered as sufficient for scanning people standing in a specific position. Better synchronization could have been achieved by using a more complicated hardware sync mechanism [20] but this solution was not implemented in our application. Users were asked to stand in an A pose (arms inclined at roughly 45°), which facilitated the post processing of the 3D scan at a later stage of the pipeline (see template fitting below).



Fig. 1. Photogrammetric 3D scanner consisting of 96 Canon Powershot A1400 cameras

The 96 x 16 million pixel colored images were then processed in Agisoft Photoscan Professional [21], a complete photogrammetry suite capable of generating detailed textured meshes from a set of overlapping pictures (see Fig. 3A). However, 3D scans generated by photogrammetric techniques result in meshes of too high density for real time animation. Moreover, since the 3D body scans are often incomplete and exhibit holes, topological errors need to be corrected. Therefore, the next step was to fit a template to the scans in order to tackle these problems.

## 2.2. Template fitting

We first developed two base anthropomorphic human avatars: a male and a female. The avatars were composed of two layers: a texture mesh (skin) used for the rendering of the surface of the avatar and an animation rig (skeleton) – which is a hierarchical tree structure – to deform the mesh. Each mesh was then attached to the virtual skeleton using the standard Skeletal Subspace Deformation technique [22]. Skinning weights were defined for every vertex of the template mesh. The two avatars were modeled in A-Pose – which is the pose users were asked to keep during the 3D scan – with a relatively small number of polygons (15K triangles) in order to be easily rendered in real time.

The 3D scans obtained at the previous step were then used as input to a topology transfer tool [23] which was in charge of fitting the template mesh to the 3D body scan. Two steps were necessary to complete the fitting. First, the template was rigidly aligned to the scan. Second, the template was used as the initial state of an as-rigid-as-possible deformation loop which tried to align the template topology to the input scan while preserving as much as possible the initial shape of the template. The texture from the 3D scan was finally re-projected on the fitted template, yielding a digital clone of the user ready to be integrated in the game engine (Unity 3D [24] for the current application).

Once the template mesh was fitted to the 3D scan, the template skeleton was adapted to the user's dimensions using simple geometric considerations. Skin weights were also copied from the template to the user's avatar.

## 2.3. Motion capture and VR setup

Capturing the motion of an actor using an optical motion capture system usually requires wearing a specific suit with a large number of markers (more than 50) and calibrating a virtual skeleton to the actor's dimensions by capturing and analyzing a complete range of motion. This actor setup is complex and time consuming (around 15 min per user), and the suit itself is not hygienic if multiple persons need to use it during the same day.

In order to maintain the setup time under one minute, the users were equipped with a set of rigid bodies positioned on their feet and hands. Each rigid body was a cluster of reflective spherical markers ( $\varnothing$  14 mm) arranged in a unique geometrical pattern allowing the motion capture system to identify it (e.g., left hand player 1) and to compute its absolute position and orientation in the 3D space. The users were also equipped with a mobile and wireless VR system consisting of a small computing unit located in a backpack coupled with a standard HMD (Oculus Rift DK2 [2]) and a pair of headphones. Both the HMD and the user's backpack were covered with a cluster of markers. Additional reflective markers were fixed to physical props and other interactive objects (e.g., a torch or a door that needs to be opened). The complete setup can be seen in Fig. 2.

Motion capture was performed using a MXT40S motion capture system (Vicon system [25], Oxford Metrics, UK), including 16 cameras recording at 120 Hz. A central server running Vicon Tracker 3.0 [25] was used to track all the rigid bodies in the scene and to stream obtained data via Wi-Fi to each computing unit. At the beginning of the VR experience, users were first asked to stay in a calibration pose (T-Pose with arms up and parallel to the floor). This allowed the system to determine the position of the rigid bodies with respect to the user's avatar and to calibrate the avatar to the user's dimensions. Once the calibration was achieved, the position and orientation of the rigid bodies were used to drive the inverse kinematic effectors of the avatar and hence derive full body animation (Fig. 3C). The world position and orientation of the HMD, as well the pose of the avatars, were thus adequately updated at each time instant. The various props were tracked and their digital counterparts updated in the same manner. The rendering of the 3D environment was handled by the Unity3D [24] game engine.

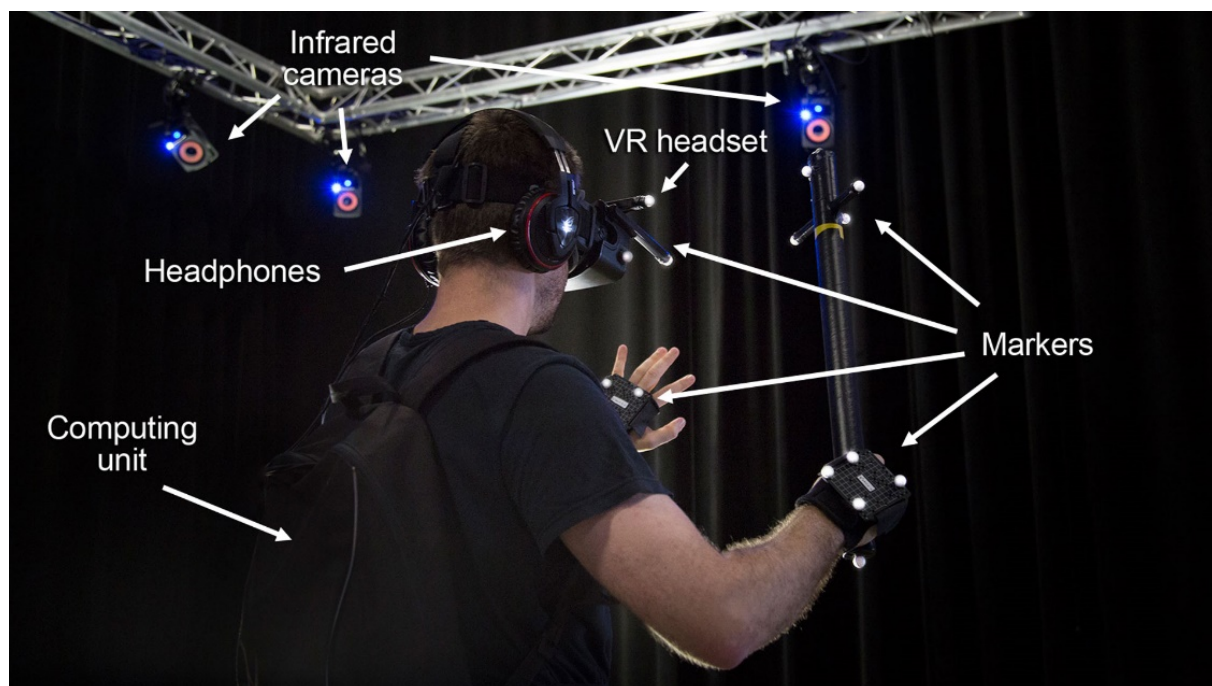


Fig. 2. VR and motion capture setup

### 3. Results

Two demonstration scenarios using this VR platform were developed showcasing different use cases of 3D scans in virtual reality environments. The setup was installed in our motion capture studio allowing the users to navigate in a 6 x 6 meters space.



Fig. 3. A) Post-processed 3D scan, B) VR setup, C) View from the 3D game engine.

#### 3.1. Scenario 1

The first scenario was a multi-player virtual visit of an Egyptian tomb. The users were able to walk through the tomb using a virtual torch to light the way and illuminate their surroundings. In this scenario, users had the choice to impersonate a standard avatar or to use their own avatar obtained by 3D scanning. Figure 3 illustrates this scenario. Fig. 3A shows the 3D scan of the user, Fig. 3B shows the user with the equipment and Fig. 3C shows a third person view of the 3D environment.

#### 3.2. Scenario 2

The second scenario was an experimental virtual contemporary dance performance and experience. Two dancers were first 3D scanned (Fig. 4A). A choreography was then captured using state of the art motion capture technique (Fig. 4B). The motion capture data was then applied to the two avatars

created from the 3D scans. A simple 3D environment was then created in which the dancing avatars were integrated (Fig. 4C). As in scenario 1, users could choose to be their own avatar or a generic one. They were then able to freely walk around the virtual dancers and at the same time see their own avatar and the other users. This scenario demonstrated the use of 3D body scan both for the user representation and for the content creation.

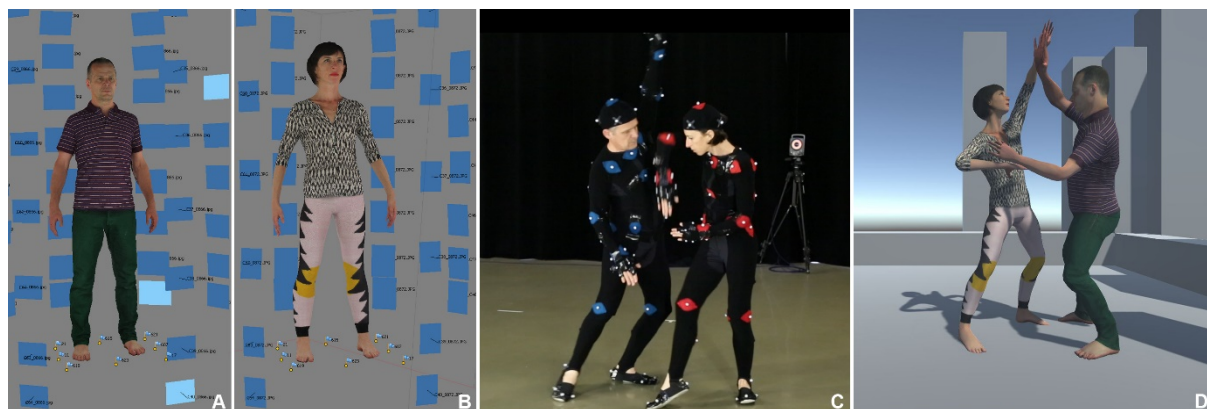


Fig. 4. A) & B) Post-processed 3D scan, B) Motion capture, C) View from the 3D game engine.

## 4. Discussion

### 4.1. Reaching a high level of presence

The setup presented in this paper was an attempt to create a VR experience reaching a high level of presence by answering point by point to the 4 foundations of presence in VR explained in the introductory section.

The combination of a very accurate HMD tracking with the freedom to move in a large space without being wired easily gave the user the illusion of being in a stable spatial space. This worked very well for the two different environments proposed.

By embodying their own virtual avatar and seeing its motion matching their own, users reached a strong feeling of embodiment. Furthermore, having a digital clone of the user limited possible issues arising when retargeting motion capture data with inverse kinematics to an avatar of different size, thus allowing a more faithful reproduction of the user's motion on his avatar. As previously discussed in the introduction, the quality of the motion reproduction is very important and errors at this stage can easily break the illusion.

The illusion of physical interaction was achieved by tracking various physical props and allowing each user to interact with them. For instance, the torch in the Egyptian tomb proved to be a very important element of the experience. The sense of touch and the reactivity of the virtual lighting and shadowing allowed a high feeling of presence.

Lastly, being able to see the other players' avatars fully animated and being able to physically interact with them (shake hands, exchange physical elements like the torch) provided a strong feeling of social communication. Users could easily communicate by gestures or by voice while collaborating on specific tasks.

### 4.2. Limitations & future work

Several limitations were observed while testing this setup. The lack of finger tracking was sometimes breaking the illusion. Users – while looking at a digital version of their hands – were expecting their virtual fingers to move as their real ones. Future work will study the integration of finger tracking technologies in this setup.

For the face, the feeling was similar. In particular the view of a static face, while discussing with another user, was problematic. Incorporating eye [26] and facial [27] tracking in the setup could be interesting but the hardware setup and the necessity to rig the avatar's face makes these elements complicated to integrate and implement. This would also increase the risk of falling into the uncanny valley [28]. Indeed, this is a common issue arising when trying to render realistic humans in VR.

Humans have a life time experience looking at other humans so they detect very quickly when something is wrong with the digital character. This is even more important with facial animation.

Finally, we were not able to provide a definitive answer whether users reached a higher level of presence when they embodied their digital clone compared to a generic humanoid avatar. Further tests including a measure of the feeling of presence (via questionnaire [29]) could be very interesting in order to understand the importance of the embodied avatar during a VR experience.

## 5. Conclusion

We presented a new VR platform combining 3D body scanning, motion capture and head mounted display allowing people to walk freely in a virtual environment, look at their own body, and interact with other users and physical objects. 3D body scanning can provide effective tools to increase the feeling of virtual body ownership as well as facilitate the creation of realistic digital characters in VR environments.

This framework opens up new digital experiences. For instance, users can jointly explore long lost historical sites (e.g., a Pharaoh's tomb before it was looted, Maya temples), participate in collaborative games or interactive story-telling experiences. Future work should focus on improving and simplifying the avatar creation and on adding finger animation and simple facial animation on the avatars to increase the feeling of self-awareness and multi-users presence.

## Acknowledgment

We would like to thank Kenzan SA for the creation of the 3D content of the Egyptian tomb and Cie Gilles Jobin, Gilles Jobin and Susana Panades Diaz for the dance performance.

## References

- [1] J. Jerald, "The VR Book: Perception and Interaction Design for Virtual Reality", ACM & Morgan & Claypool, DOI 10.1145/2792790
- [2] Oculus VR, <https://www.oculus.com/en-us/dk2/>
- [3] Google cardboard, <https://www.google.com/get/cardboard/>
- [4] HTC VIVE, <http://www.htcvr.com/>
- [5] Y. Yuan and A. Steed, "Is the rubber hand illusion induced by immersive virtual reality", in *IEEE Virtual Reality Conference (VR)*, 2010
- [6] K. Kilteni, R. Groten and M. Slater, "The Sense of Embodiment in Virtual Reality", in *Presence*, 2012, 21(4), 373-387. DOI: 10.1162/PRES\_a\_00124
- [7] A. Maselli, M. Slater, "The building blocks of the full body ownership illusion", in *Frontiers in human neuroscience*, 2013, DOI: 10.3389/fnhum.2013.00083
- [8] J. Lugrin, J. Latt and M. Latorschik, "Avatar anthropomorphism and illusion of body ownership in VR", in *IEEE Virtual Reality (VR)*, 2015, 229, 230, 2015, DOI: 10.1109/VR.2015.7223379
- [9] A. Stevenson Won, J. Bailenson, J. Lee and J. Lanier, "Homuncular Flexibility in Virtual Reality", in *Journal of Computer-Mediated Communication*, 20(3), 241-259, 2015
- [10] B. O'Farrell, "Bodyhub.com: A Cloud-Based Service for Automatically Creating Highly Accurate Articulated 3D Models from Body Scans", 3D Body Scanning Conference, 2013
- [11] H. Li, E. Vouga, A. Gudym, L. Luo, J. T. Barron and G. Gusev, "3D self-Portraits", ACM Transactions on Graphics, Proceedings of the 6th ACM SIGGRAPH Conference and Exhibition in Asia 2013, 11/2013
- [12] A. Shapiro, A. Feng, R. Wang, H. Li, M. Bolas, G. Medioni and E. Suma, "Rapid avatar capture and simulation using commodity depth sensors", in *Comp. Anim. Virtual Worlds 14*, 2014, 25, 201-211, DOI: 10.1002/cav.1579
- [13] R. Garsthagen, "An Open Source, Low-Cost, Multi Camera Full-Body 3D Scanner", in *Proc. of 5th Int. Conf. on 3D Body Scanning Technologies*, Lugano, Switzerland, 2014, pp. 174-183, [doi:10.15221/14.174](https://doi.org/10.15221/14.174).
- [14] C. Kopf et al., "ReconstructMe - Towards a Full Autonomous Bust Generator", in *Proc. of 5th Int. Conf. on 3D Body Scanning Technologies*, Lugano, Switzerland, 2014, pp. 184-190, [doi:10.15221/14.184](https://doi.org/10.15221/14.184).
- [15] J. Tong, J. Zhou, L. Liu, Z. Pan and H. Yan, "Scanning 3D Full Human Bodies Using Kinects", in *IEEE Transactions on Visualization and Computer Graphics*, 18(4), 2012

- [16] P. Debevec, "The Light Stages and Their Applications to Photoreal Digital Actors", in *SIGGRAPH Asia*, 2012
- [17] Infinite Realities, <http://ir-ltd.net/>
- [18] Canon Hack Development Kit, <http://chdk.wikia.com/wiki/CHDK>
- [19] chdkptp, <https://www.assembla.com/spaces/chdkptp/wiki>
- [20] CHDK USB Remote, [http://chdk.wikia.com/wiki/USB\\_Remote](http://chdk.wikia.com/wiki/USB_Remote)
- [21] Agisoft Photoscan, <http://www.agisoft.com/>
- [22] N. Magnenat-Thalmann, R. Laperriere and D. Thalmann, "Joint-dependent Local Deformations for Hand Animation and Object Grasping", In *Proc. Of Graphic Interface '88*, 1988, pp26-33
- [23] R3DS WrapX 1.3.4, <http://www.russian3dscanner.com/>
- [24] Unity3D, <http://unity3d.com/>
- [25] Vicon system, <http://vicon.com/>
- [26] S. Deng, J. Kirkby, J. Chang, J. Zhang, "Multimodality with Eye tracking and Haptics: A New Horizon for Serious Games?", in *International Journal of Serious Games*, 1(4), 2014
- [27] H. Li, L. Trutoiu, K. Olszewski, L. Wei, T. Trutna, P. Hsieh, A. Nicholls and C. Ma, "Facial Performance Sensing Head-Mounted Display", *ACM Transactions on Graphics, Proceedings of the 42nd ACM SIGGRAPH Conference and Exhibition*, 2015
- [28] M. Mori, "The Uncanny Valley", in *Energy*, 1970, 7(4), 33–35. DOI: 10.1162/pres.16.4.337. 50
- [29] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire", in *Presence: Teleoper. Virtual Environ.* 7(3), 225-240, 1998