Just-in-time, viable, 3D avatars from scans

Andrew Feng
Institute for Creative Technologies
University of Southern California
feng@ict.usc.edu

Evan Suma
Institute for Creative Technologies
University of Southern California
suma@ict.usc.edu

Ari Shapiro
Institute for Creative Technologies
University of Southern California
shapiro@ict.usc.edu

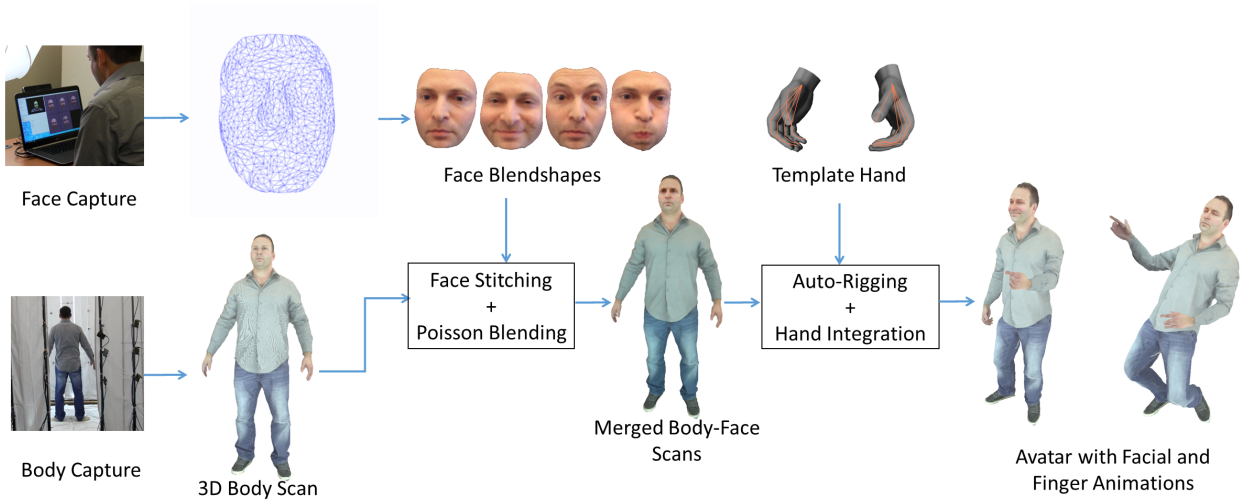

Figure 1: Workflow showing the process of integrating three separate models (body, face and hands) into a controllable, 3D avatar.

CCS CONCEPTS

• **Computing methodologies → Graphics systems and interfaces**; *Procedural animation*; *Rendering*; *Perception*; *Virtual reality*;

KEYWORDS

avatar, 3d, scanning, RGB-D, photogrammetry, rigging, skinning, animation, virtual reality

ACM Reference format:
Andrew Feng, Evan Suma, and Ari Shapiro. 2017. Just-in-time, viable, 3D avatars from scans. In *Proceedings of SIGGRAPH '17 Talks, Los Angeles, CA, USA, July 30 - August 03, 2017,* 2 pages.
https://doi.org/10.1145/3084363.3085045

1 MOTIVATION

We demonstrate a system that can generate a photorealistic, interactive 3D character from a human subject that is capable of movement, emotion, speech and gesture in less than 20 minutes without the need for 3D artist intervention or specialized technical knowledge through a near automatic process. Our method uses mostly commodity- or off-the-shelf hardware. We demonstrate the just-in-time use of generating such 3D models for virtual and augmented reality, games, simulation and communication. We anticipate that the inexpensive generation of such photorealistic models will be useful in many venues where a just-in-time 3D construction of digital avatars that resemble particular human subjects is necessary. Figure 1 shows the overall workflow of our virtual character creation pipeline.

Given the cost and effort needed to generate digital doubles [Perry 2014] for film and video games, it is generally not viable to generate a photorealistic 3D avatar within a short amount of time without expertise. Our goal is to create a photorealistic 3D character using automated methods that require no expertise and can be executed in a reasonable amount of time, while leveraging many existing 3D pipelines and workflows. Thus we seek to answer the following: "Can a viable, 3D photorealistic character be created that for just-in-time use in a game or simulation?"

The uses of such a 3D avatar and by extension the capabilities needed can vary greatly. While some 3D environments might only require a recognizable appearance and little or no movement, others may require the 3D avatar to speak, gesture, move and emote. To this end, we include two aspects that are important for communicative and social 3D avatar environments; facial expression and

SIGGRAPH '17 Talks, July 30 - August 03, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5008-2/17/07.
https://doi.org/10.1145/3084363.3085045

Table 1: Capture and processing times for 3D character construction.

Stage	Time (min)
3D body capture	1
3D body construction	5
Automatic rigging and skinning	1
3D face capture, 5 expressions	5
blendshape construction	3
Face and body stitching and color correction	2
Hand and body stitching and color correction	1
Total	**18**

finger movements. Facial expressions are important for communication emotion and speech, while finger and hand movements are critical for proper nonverbal communication during gestures, as well as for manipulation of virtual objects. Thus, we demonstrate a semi-automatic process for generating a 3D photorealistic avatar of a human subject within 20 minutes with no requirement of 3D artistic or technical expertise. The avatars generated are capable of emotional expression and speech, as well as manipulation and communication via articulated hands and fingers.

2 CHARACTER CONSTRUCTION

We construct our character through capture and construction of three separate models; body, face and hands. The 3D body model is constructed using a photogrammetric capture cage [Straub and Kerlin 2014] and photogrammetry software (Agisoft Photoscan). The 3D static model is then automatically rigged [Feng et al. 2015] to produce a controllable, 3D character body. Once the rigged 3D avatar is created, it is capable of large scale articulated movements such as walking and jumping, but lacks the capability to perform facial expressions and hand shapes. Thus we perform additional processing to integrate blendshape model and hand rigging with the virtual avatar to enable it for face-to-face expressive communication. The 3D face model is constructed by scanning a set of facial expressions, then constructing a set of blendshapes from the scan data [Casas et al. 2015]. The body and face models are then stitched together by finding matching areas on the face, and then replacing the face geometry and texture of the body model with the blendshape model. A color correction process is then applied to better match the blendshape model against the body model. A controllable hand and finger model is then fashioned onto the 3D body model using the process described in Figure 2

Table 1 shows a summary of all the steps for 3D avatar construction which takes less than 20 minutes.

The ability to construct an expressive, photorealistic 3D avatar near-automatically and in a short amount of time opens the possibility of using such constructs for just-in-time uses. As an example, social virtual reality environments could contain recognizable representations that are immediately identifiable to those familiar with the person, such as in Figure 3. In addition, such constructs could be used in environments where recognition of the individual is important for authenticity.

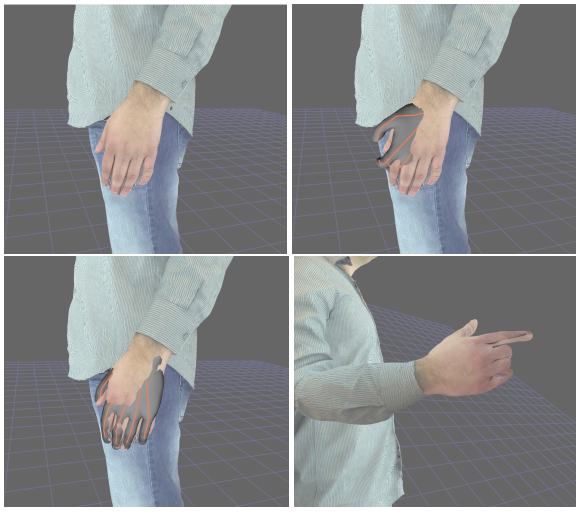

Figure 2: The raw 3D scan does not have enough details to rig for finger control (top, left). Thus a deformable template hand is initialized around the wrist region (top, right). Then the hand rig is fitted to the hand region to replace original hand scan (bottom, left). The resulting articulated hand is capable of various finger poses (bottom, right).

Figure 3: Social VR avatar use.

REFERENCES

Dan Casas, Oleg Alexander, Andrew W. Feng, Graham Fyffe, Ryosuke Ichikari, Paul Debevec, Rhuizhe Wang, Evan Suma, and Ari Shapiro. 2015. Blendshapes from Commodity RGB-D Sensors. In *ACM SIGGRAPH 2015 Talks (SIGGRAPH 15)*. ACM, New York, NY, USA, Article 33, 1 pages. https://doi.org/10.1145/2775280.2792540

Dan Casas, Marco Volino, John Collomosse, and Adrian Hilton. 2014. 4D Video Textures for Interactive Character Appearance. *Computer Graphics Forum* 33, 2 (2014), 371–380.

Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. 2015. High-quality Streamable Free-viewpoint Video. *ACM Trans. Graph.* 34, 4, Article 69 (July 2015), 13 pages. https://doi.org/10.1145/2766945

Andrew Feng, Dan Casas, and Ari Shapiro. 2015. Avatar reshaping and automatic rigging using a deformable model. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*. ACM, 57–64.

Alexandru Eugen Ichim, Sofien Bouaziz, and Mark Pauly. 2015. Dynamic 3D avatar creation from hand-held video input. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 45.

Travis Perry. 2014. Leaving the uncanny valley behind. *Spectrum, IEEE* 51, 6 (2014), 48–53.

Jeremy Straub and Scott Kerlin. 2014. Development of a large, low-cost, instant 3D scanner. *Technologies* 2, 2 (2014), 76–95.