Rapid Avatar Capture and Simulation Using Commodity Depth Sensors

Andrew Feng*[1], Ari Shapiro†[1], Wang Ruizhe‡[2], Hao Li§[2], Mark Bolas¶[1], Gerard Medioni‖[2], and Evan Suma**[1]

[1]USC Institute for Creative Technologies
[2]University of Southern California

Keywords: animation, image capture, kinect

We demonstrate a process for capturing human subjects and generating digital characters from those models using commodity scanning hardware. Our process is capable of capturing a human subject using still four poses, constructing a 3D model, then registering it and controlling it within an animation system within minutes. The digital representation that our process is able to construct is suitable for use in simulations, games and other applications that use virtual characters. Our technique is able to model many dynamic aspects of human behavior. As shown in Figure 1, our main contribution in this work is a near-fully automated, rapid, low-cost end-to-end system for capture, modeling and simulation of a human figure in a virtual environment that requires no expert intervention.

The subject turns in front of the Kinect sensor in a natural motion, while staying static at 4 key poses, namely front, back and two profiles, for approximately 10 seconds each. For each key pose, a super-resolution range scan is generated as the Kinect device, controlled by a built-in motor, moves up and down. The 4 super-resolution range scans are then aligned in a multi-view piecewise rigid manner, assuming small articulations between them. Traditional registration algorithms (Iterative Closest Point [Chen and Medioni 1992]), which are based on the *shape coherence*, fail in this scenario because the overlap between consecutive frames is very small. Instead, we employ *contour coherence* and develop a contour-based registration method, which iteratively minimizes the distance between the closest 2D contour points. At the final stage, the 4 aligned key poses are processed to generate a water-tight mesh model using the Poisson Surface Reconstruction algorithm [Kazhdan et al. 2006]. The corresponding texture information of the 4 super-resolution range scans are inferred using the Poisson Texture Blending algorithm [Chuang et al. 2009].

The construction of a 3D model take approximately 4 minutes, and the automatic rigging, skinning and skeleton registration [Feng et al. 2013] takes approximately 90 seconds. Models typically contain between 200k and 400k vertices. Simulation and control of the character is performed in real time. The 3D models captured in this way are suitable for use in games where characters need to be recognizable from a distance, but do not require face-to-face or close interactions.

*e-mail:feng@gmail.com
†e-mail:shapiro@ict.usc.edu
‡e-mail:ruizhewa@gmail.com
§e-mail:hao@hao-li.com
¶e-mail:bolas@ict.usc.edu
‖e-mail:medioni@usc.edu
**e-mail:suma@ict.usc.edu

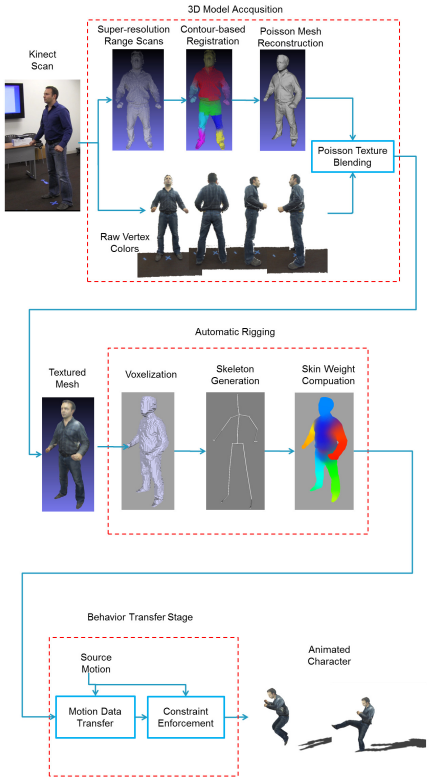

Figure 1: *The overall work flow of our rapid avatar capture and simulation system.*

References

CHEN, Y., AND MEDIONI, G. 1992. Object modelling by registration of multiple range images. *Image and vision computing 10*, 3, 145–155.

CHUANG, M., LUO, L., BROWN, B. J., RUSINKIEWICZ, S., AND KAZHDAN, M. 2009. Estimating the laplace-beltrami operator by restricting 3d functions. In *Computer Graphics Forum*, vol. 28, Wiley Online Library, 1475–1484.

FENG, A., HUANG, Y., XU, Y., AND SHAPIRO, A. 2013. Fast, automatic character animation pipelines. *Computer Animation and Virtual Worlds*, n/a–n/a.

KAZHDAN, M., BOLITHO, M., AND HOPPE, H. 2006. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, Eurographics.