Automatic Acquisition and Animation of Virtual Avatars

Ari Shapiro* Andrew Feng* Ruizhe Wang† Gerard Medioni† Mark Bolas* Evan A. Suma*

*USC Institute for Creative Technologies †University of Southern California

Abstract

The USC Institute for Creative Technologies will demonstrate a pipline for automatic reconstruction and animation of lifelike 3D avatars acquired by rotating the user's body in front of a single Microsoft Kinect sensor. Based on a fusion of state-of-the-art techniques in computer vision, graphics, and animation, this approach can produce a fully rigged character model suitable for real-time virtual environments in less than four minutes.

Keywords: Avatars, depth sensors, reconstruction, animation

Index Terms: H.5.1 [[Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality

1 Introduction

Recent advances in low-cost scanning have enabled the capture and modeling of real-world objects into a virtual environment in 3D. There is great value to the ability to quickly and inexpensively capture real-world objects and create their 3D counterparts. For example, a table, a room, or work of art can be quickly scanned, modeled and displayed within a virtual world with a handheld, consumer scanner. Automatic scanning and reconstruction of human models is potentially even more impactful than static objects. However, the organic, dynamic nature of the human body introduces a number of non-trivial challenges, most notably: (1) acquisition of stable 3D measurements from a single noisy, low-cost depth sensor, (2) reconstruction of a recognizable 3D model from a non-rigid body that may move during scanning, and (3) animation of the static reconstructed model to produce lifelike human motion.

We propose to demonstrate an approach for rapid and automatic generation of lifelike animated human avatars using a single Microsoft Kinect, a low-cost consumer-level depth sensor with integrated RGB camera. With a total acquisition and processing time of less than four minutes, attendees of the IEEE Virtual Reality conference may opt to be scanned and see their avatar brought to life immediately. In addition to the real-time demonstration, we will also provide a copy of the generated avatar to each scanned participant via email, in a file format suitable for rendering in real-time virtual environment platforms such as Unity.

2 Avatar Generation Pipeline

Cheap, fast and accurate 3D body scanning is of great interest to the computer graphics. Unlike the traditional method which uses multiple 3D sensors and requires the subject to stay as static as possible on a turn table [5], we use only a single Kinect sensor and ask the subject to turn. The subject turns and pauses at four key poses: front, back and two side profiles (see Figure 1). At each pose, the subject stays static for approximately 10 seconds while the

*e-mail: {shapiro, feng, bolas, suma}@ict.usc.edu
†e-mail: {ruizhewa, medioni}@usc.edu

IEEE Virtual Reality 2014
29 March - 2 April, Minneapolis, Minnesota, USA
978-1-4799-2871-2/14/$31.00 ©2014 IEEE

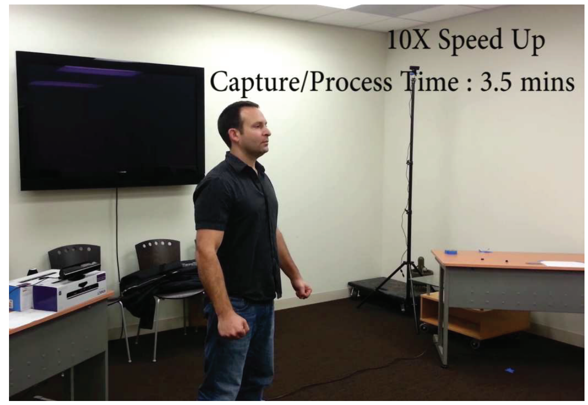

Figure 1: Participants are scanned by rotating their bodies in front of a single Kinect sensor with pauses at four key poses (front, back, and sides). The total scanning time is under a minute.

Figure 2: The overview of autorigging process. Our system first converts the model into the voxel representation to ensure a clean watertight mesh. The control skeleton and skinning weights can then be robustly computed from voxels.

Kinect sensor, controlled by a motor, moves up and down to scan the whole body. For each key pose, roughly 200 frames are merged by KinectFusion [4] and reprojected to the image plane to generate a super-resolution range scan. These 4 super-resolution range scans are further aligned in an articulated manner using a contour-based method [6]. At the final stage, the Poisson Surface Reconstruction Algorithm [3] is used to generate the watertight mesh model.

We used a similar autorigging method as the one proposed in [1] to builds a skeleton from the 3D scanned character. The original method requires a watertight and single component mesh to work correctly, which poses a big restriction on the type of 3D models the method can be applied to. While the specific reconstruction approach described above generates a watertight mesh, many other scanned meshes may often contain holes. Additionally, non-manifold geometry or other topological artifacts may require additional clean-up. To alleviate this limitation, we proposed a modified auto-rigging method with voxelization. It is done by converting the mesh into voxels using depth buffer carving in all positive and negative x,y, and z directions. The resulting voxels are naturally free from many topological artifacts and can be easily pro-

Figure 3: With the help of retargeting, the acquired 3D skinned character can perform a wide range of common human-like behaviors such as kicking, running, jumping.

cessed to produce a clean, watertight mesh for autorigging. Figure 2 demonstrates the overall autorigging process.

The motion retargeting process works by transferring an example motion set from our canonical skeleton to the custom skeleton generated from automatic rigging. Here we use the method from [2] to perform motion retargeting. The retargeting process can be separated into two stages. The first stage is to convert the joint angles encoded in a motion from our canonical skeleton to the custom skeleton. The second stage is to enforce various positional constraints such as foot positions to remove motion artifacts such as foot sliding. After the retargeting stage, as shown in Figure 3, the acquired 3D skinned character can be incorporated into the animation simulation system to execute a wide range of common human-like behaviors such as kicking, running, jumping, and so forth.

3 Research Lab

This research demo is a collaboration between the MxR Lab and the Character Animation and Simulation group at the Institute for Creative Technologies (ICT), as well as the Computer Vision Laboratory at the University of Southern California. Established in 1999 by the University of Southern California with funding from the US Army, the ICT is a University Affiliated Research Center (UARC) that was formed to leverage academic research, the Hollywood movie industry, and the game industry in order to advance the state of the art in learning, simulation, and training. The mission of the ICT is to create synthetic experiences so compelling that participants reach as if the synthetic experiences are real.

The ICT MxR Lab (http://www.mxrlab.com) explores techniques and technologies to improve the fluency of human-computer interactions and create visceral synthetic experiences. One focus area for the lab is the research and development of immersive virtual reality technologies. The lab is led by Mark Bolas, the Associate Director for Immersion at ICT and an Associate Professor of Interactive Media at the USC School of Cinematic Arts, along with co-directors David Krum and Evan Suma. The lab maintains a philosophy of understanding through building and prototyping. To this end, the labs facilities include two motion tracking and capture stages, and an optics/electronics workroom, and a workshop containing a CNC mill, laser cutter, and a variety of power tools. Valuing a breadth of approaches and skills, MxR Lab personnel have a wide variety of backgrounds including perception, electronics, optical design, film production, mechanical engineering, and computer science. The lab has also hosted students for research and internships from various universities and various programs, including human-computer interaction, computer science, and interactive media.

The ICT Character Animation and Simulation group seeks to discover and develop methods for synthesis of motion on virtual characters. The group investigates methods for movement, manipulation and emotional expression. The group is headed by Ari Shapiro and includes Andrew Feng and Yuyu Xu.

The Computer Vision Laboratory at the University of Southern California is one of the major centers of computer vision research for thirty years. They conduct research in a number of basic and applied areas. Research topics include image and scene segmentation, stereo, motion and video analysis, range data analysis, perceptual grouping, shape analyis and object recognition. They have worked on applications to robotics, manufacturing, mobile robots, surveillance and aerial photo interpretation. Their approach emphasizes the use of segmented (part-based) symbolic descriptions of objects. Such representations have long been believed to have significant advantages over other alternatives but have been difficult to infer from the sensed data. The group has been successful in making significant progress on this topic and is pursuing the use of these techniques for practical applications.

References

[1] I. Baran and J. Popović. Automatic rigging and animation of 3d characters. *ACM Transactions on Graphics*, 26(3):72, 2007.

[2] A. Feng, Y. Huang, Y. Xu, and A. Shapiro. Fast, automatic character animation pipelines. *Computer Animation and Virtual Worlds*, pages n/a–n/a, 2013.

[3] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, 2006.

[4] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.

[5] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan. Scanning 3d full human bodies using kinects. *Visualization and Computer Graphics, IEEE Transactions on*, 18(4):643–650, 2012.

[6] R. Wang, J. Choi, and G. Medioni. Accurate full body scanning from a single fixed 3d camera. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 432–439. IEEE, 2012.