KinÊtre: Animating the World with the Human Body

Jiawen Chen Shahram Izadi Andrew Fitzgibbon
Microsoft Research Cambridge

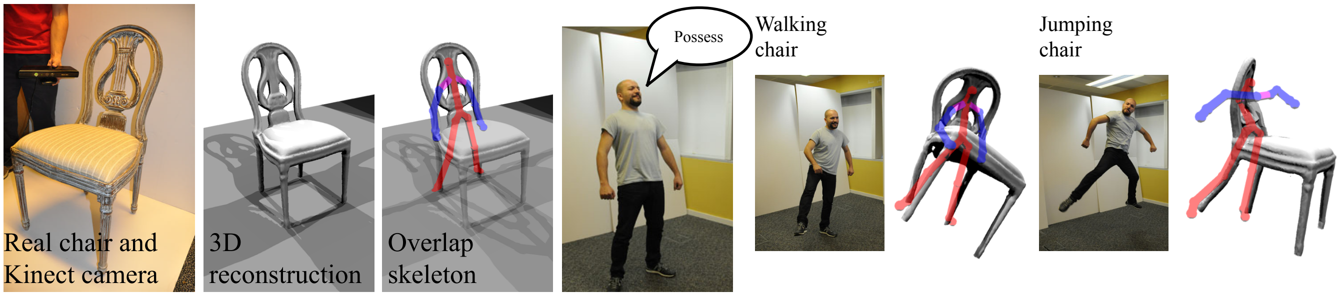

Figure 1: *KinÊtre in action. The user rapidly scans a real chair using a Kinect camera. He then sets the camera down, which tracks his motion as he moves into position. After "possessing" the chair, it walks and jumps with him.*

1 Introduction

Imagine you are asked to produce a 3D animation of a demonic armchair terrorizing an innocent desk lamp. You may think about model rigging, skeleton deformation, and keyframing. Depending on your experience, you might imagine hours to days at the controls of Maya or Blender. But even if you have absolutely no computer graphics experience, it can be so much easier: grab a nearby chair and desk lamp, scan them using a consumer depth camera, and use the same camera to track your body, aligning your virtual limbs to the chair's geometry. At one spoken command, your limbs are attached to the chair model, which follows your movements in an intuitive and natural manner. KinÊtre is such a system. Rather than targeting professional animators, it brings animation to a new audience of users with little or no CG experience. It allows realistic deformations of arbitrary static meshes, runs in real time on consumer hardware, and uses the human body for input in conjunction with simple voice commands. KinÊtre lets anyone create playful 3D animations.

2 Approach

The user begins by scanning a physical artifact using KinectFusion [Izadi et al. 2011]. Using a depth camera, we can *acquire* and *segment* a human-scale object in minutes, producing a 3D mesh.

KinÊtre lets a user scan an arbitrary object and make it move by moving his body. While we can track his motion with the same camera, using an off-the-shelf system, given that the object can be anything, how should this mapping be done? Rather than trying to discover properties of the object or guess the user's intentions, we opt for a direct manipulation approach. After scanning the object, he moves it out of the way, which automatically segments the mesh from the background. Then he embeds his body to position vacated by the object. On a voice command, "possess", we *attach* his limbs to the parts of the mesh he overlaps. These attachments serve as constraints in our deformation model.

Our mesh deformation model is inspired by the Embedded Deformation (ED) method of Sumneret al. [Sumner et al. 2007], a system designed for interactive shape manipulation using the mouse and keyboard. Compared to other mesh deformation techniques, ED is also able to handle unstructured data such as polygon soups and particle systems. The ED model operates on a set of *point constraints*: when the user drags a point, a nonlinear optimization finds the smoothest, non-stretching set of transformations that satisfies the constraints. Qualitatively, the deformation feels rigid when constraints do not conflict, and stretches like rubber when they do. For efficiency, the set of transformations are computed on a deformation graph—a subsampled version of the input mesh. KinÊtre uses largely the same energy model, with the user's limbs serving as the constraints. Our contributions are:

- A new method for generating the deformation graph and an improved energy function that makes our system robust to the kinds of surfaces acquired by 3D scanning—in particular, partial scans.
- A fast nonlinear optimizer that can handle multiple simultaneous constraints that come from a tracked human skeleton.

In our scenario, users rapidly scan objects, which can create incomplete 3D scans with numerous holes and islands. With traditional ED, such inputs lead to an ill-posed optimization problem. We guarantee a well-posed problem by sampling the mesh with an orientation-aware metric, connecting disconnected components at critical points using a spanning-tree algorithm, and introducing an additional energy term that enforces rigid connections between islands. These extensions ensure that KinÊtre is robust on a wide variety of scenarios, described below.

KinÊtre lets users scan *arbitrary* objects, we envision children using the system to animate household objects for storytelling. Our accompanying video and supplementary materials show a number of examples, including a chair, bookcase, stepladder, and lamp. Another application area for our system is in creating avatars for gaming or teleconferencing scenarios. Users can scan themselves to drop into a game, or each other and switch bodies. A key advantage of our system is its robustness to low-quality partial scans. Finally, meshes possessed by KinÊtre can be efficiently incorporated into physics-based interactions by approximating the surface with a set of spheres. Imagine taking over a friend's avatar and rampaging through a virtual world, or making a chair play dodgeball.

References

IZADI, S., KIM, D., HILLIGES, O., MOLYNEAUX, D., NEWCOMBE, R., KOHLI, P., SHOTTON, J., HODGES, S., FREEMAN, D., DAVISON, A., AND FITZGIBBON, A. 2011. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *UIST 2011*.

SUMNER, R. W., SCHMID, J., AND PAULY, M. 2007. Embedded deformation for shape manipulation. *SIGGRAPH 2007*.

SIGGRAPH 2012, Los Angeles, California, August 5 – 9, 2012.
ISBN 978-1-4503-1435-0/12/0008