Markerless Human Motion Transfer

German Cheung  Simon Baker  Jessica Hodgins  Takeo Kanade
{german+, simonb, jkh, tk}@cs.cmu.edu
Robotics Institute, Carnegie Mellon University

Generating animations of real-life characters with realistic motion and appearance is one of the most difficult problems in computer graphics. The ability to create realistic videos of a person performing new motions is essential to games development and special effects in movie production. In this research, we describe a markerless system to transfer articulated motion from one person to another: given video sequences of two people performing two different motions, we generate videos of each person performing the motion of the other person. Our system is based on computer vision techniques and uses no markers in any of the steps.

Figure 1 illustrates the structure of our markerless motion transfer system. Our system uses eight synchronized, calibrated and color balanced cameras. The motion transfer process consists of three steps: (a) human kinematic modeling of both subjects #1 and #2, (b) markerless motion capture of both subjects and (c) image-based rendering of the motion of one subject applied to the other. The first two steps are based on our recent work in the computer vision literature [Cheung et al. 2003]. The main contributions of this research are the development of the image-based rendering algorithm in Step 3 and the demonstration of a complete end-to-end markerless motion transfer system.

The first step is to build kinematic models of the two subjects. Based on a recently proposed shape reconstruction algorithm called Shape-From-Silhouette Across Time [Cheung et al. 2003], we acquired detailed 3D shapes of the two subjects by asking them to stand on an uncalibrated turn-table for 30 seconds. Using a similar algorithm for articulated objects, we also recover the joint location and segmentation of the subjects by asking them to exercise their joints one at a time. Figure 1(a) illustrates the resulting model of subject #1 consisting of a voxel model and a joint skeleton.

Once the kinematic models have been acquired, they are used in the second step to recover the motion data (joint angles) of each subject by aligning the kinematic model with respect to both the silhouette and color images. No markers are required for this motion capture algorithm [Cheung et al. 2003]. Figure 1(b) shows the tracked motion of subject #2 miming a throw motion.

The final step of our system and the novelty of this research is to apply the tracked motion to the kinematic model of the other person and render the results. We have devised an algorithm for performing image-based rendering of articulated objects using an ensemble of spatial and temporal images from the tracked videos of the person. For the example in Figure 1(c), the source images come from the tracked motion videos of subject #1 on the turn-table. Our image-based algorithm produces sharper images than direct rendering of the kinematic model (see Figure 2 for a comparison).

Figure 3 shows selected frames from an input video of subject #2 performing a throw motion and the corresponding frames of the rendered output video. Our system successfully transfers the throw motion from subject #2 to subject #1. The rendered video corresponding to Figure 2(b) is included in the supplementary video clip. Note that our system can also render the subject from novel viewpoints as shown in the supplementary video clip.

References