CSE 190a Project Proposal:
Converting 2D golf swing sequences into 3D models

Ravi Chugh
rchugh@cs.ucsd.edu

Krystle de Mesa
kdemesa@cs.ucsd.edu

1. Introduction

Innovations in computer graphics and computer vision have been applied in various golf-related applications – motion capture systems to provide detailed swing analysis, video systems to supplement golf instruction, high-precision launch monitors to track club and ball physics, and state-of-the-art graphics techniques to develop golf video games that continue to look more and more realistic. Recently, EA Sports has added a feature to their golf video game to make the user experience especially unique: the ability to upload images of one's own face to map onto 3D characters in the game. We would like to extend this idea by developing a system that takes a 2D video sequence of a golf swing and translates it onto a 3D character model, which, if successful, could reasonably be incorporated into future versions of the game.

Motion capture can be used to create very precise 3D models of a golfer's swing, but it is of course only available to people with access to such studios. It is very accessible and common, however, for golfers to record video of their swings using traditional video cameras or digital cameras. Converting the latter format – a 2D sequence of images – into the former – a 3D model – would approximate the precision of motion capture without the barrier to entry.

We will build a software application that takes as input video of a golf swing from several vantage points and produces a 3D model that approximates the positions of the golfer in the input video. To assist in the conversion process and to achieve more precise results, the user will be asked to calibrate the input swing sequence by identifying key positions in the swing (*e.g.* the locations of the golfer's shoulders, elbows, and hands). These clues will then be used to map the golfer's positions onto a 3D model.

1.1. Questions

There are several key questions we will explore in order to build a successful system.

- What are the key body and joint positions in a golf swing to track?
- What is a set of points to track that is enough to effectively model a swing without maintaining excessive amounts of data?
- Which camera angles capture swing sequences that are the easiest to convert to 3D?
- According to some notion of correctness, how well does the conversion process perform?
- Can additional camera angles significantly enhance the quality of the resulting models?

Time permitting, we will also consider advanced extensions of the system.

- Can the system learn from previous user calibrations to perform better subsequent conversions?
- To what degree can the amount of user calibration be minimized?

2. Project Plan

The system we will build will have three phases of interaction with the user: importing swing sequences, calibrating body parts in these sequences, and displaying the 3D model. Figure 1 diagrams this structure and the processes that will bridge these phases.

In implementing this system, we will employ a variety of applications from animation, graphics, and vision. We will use motion capture to create a base 3D model of a golf swing that will be fitted to the golf swing being imported. Starting with a realistic 3D base will leverage the fact that, despite unique characteristics in a golf swing, a significant amount of the motion and the human body in general is common. Our hope is that making adjustments to an already-realistic model – as opposed to attempting to reconstruct the 3D golf swing motion soley from the images – will give us a better chance of building a successful system.

Given the input swing videos (*i.e.* sequences of 2D images from multiple angles), we will implement various algorithms for tracking and 2D pose estimation to identify the

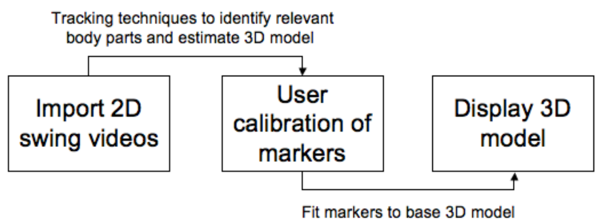

Figure 1. Overview of the system.

critical positions in the golf swing. This will be the most challenging aspect of the system, and we will begin exploring this area by studying the papers discussed in Section 3 to see if any of those techniques can be applied to our application.

This process will produce an estimation of critical points in the images in the form of markers. The user will get a chance to look at these markers and correct them if the estimates do not line up well. The goal will be to have the estimation be as accurate as possible, but the ability for the user to manually correct them will allow the conversion process to be successful even if the vision algorithms we implement do not perform well.

Once the markers have been revised by the user, they will be fitted to the base 3D model by comparing with the corresponding markers of the base 3D model. This will customize the base 3D model to estimate the positions from the input swing sequence, and the motion of the 3D model can then be rendered and animated for display.

2.1. Milestones

We expect that the timeframe for completing different aspects of the system will change as we make progress. As a general guideline for the progress we would like to make, however, we outline several major milestones and the end of the week by which we hope to have them complete.

Week 2	Obtain base 3D model using motion capture
Week 2	Calibrate video cameras and capture golf swing videos
Week 4	Experiment with implementing various tracking and estimation algorithms
Week 6	Prototype of application with only user calibration
Week 8	Prototype of application with automatic estimation and user-assisted corrections
Week 10	Finished application / advanced features

2.2. Advanced features

Depending on how quickly we get the basic system working, we may have time to explore improvements. We discuss two general possibilities.

Better automatic estimation. As discussed above, estimating the critical positions in the images of a golf swing will be the most challenging aspect of the project. We anticipate plenty of room for improvement, because the problems of people tracking and pose estimation are hard and open problems. Time permitting, we will continue exploring state-of-the-art techniques to get better results.

Less synchronized inputs. One premise for this system is that the average golfer has access to cameras to record their swings but not motion capture. Our system requires synchronized video from multiple viewpoints, but most people will only have the capacity to record their swing from one angle at a time. To make this system more usable in practice, it would be helpful to relax the requirement of multiple synchronized cameras. This would require allowing the user to input multiple videos corresponding to *different* swings that the user made. The swings would need to be compared and correlated to produce some sort of "average" of the swings, on which the conversion to 3D process could be run.

3. Relevant Work

We will examine some techniques that have been developed to address the problems of people tracking and pose estimation. We list some of the papers with which we will begin this process.

- [5] describes an approach to learning 3D models from small training sets that can then be used to perform people tracking. The model uses a low-dimensinal representation of the pose so that tracking can be done deterministically and online.

- [3] describes an algorithm for automatically detecting and tracking a golf club in a sequence of images. After extracting the club head from the separate images, the trajectory of the club head and the speed at which it travels are estimated.

- A two-level probabilistic approach to extracting 3D figures from 2D images is described in [4]. The first part estimates a 2D pose by running simple body part detectors and then computing a probability distribution for different "loose-limbed" poses comprised by these parts. This 2D pose is then probabilistically mapped to a 3D pose.

- [2] extends the SFS (Shape-From-Silhouette) 3D reconstruction method to dynamic articulated objects over time by using an iterative algorithm to estimate the shape (by segmenting) and motion of an object. Articulated points between pairs of rigid parts are then obtained by using a motion constraint between the connected parts.

- [1] presents a data-driven method that uses a pose deformation model and a separate model of variation (based on body shape) to construct a 3D surface model with realistic muscle deformation for different people in different poses.

4. Resources

Modeling software. For the rendering of the 3D golf swing, we will use either OpenGL or a modeling environment like Maya.

Base 3D model We will need to use the motion capture studio to record a model of a golf swing that our application will use as a base.

Video recording We will need a setup with multiple cameras that record in sync. Having the images in sync will allow us to correlate the markers set in images from different viewpoints. With this we will be able to compute how to fit the motion in the images to the base 3D model.

A. Qualifications

Ravi has taken courses in graphics and animation, and he has some implementation experience from these courses. He has also taken a course in probabalistic reasoning and decision-making, which will be useful in understanding estimation algorithms that employ probabalistic models and belief networks. He has not taken any courses in image processing or vision. In his senior year, he worked on an undergraduate research project for a domain specific string transformation language. For the past several months, he has worked on a concurrent dataflow analysis framework that has been described in a research paper submitted for publication to PLDI.

Krystle has taken graphics, visualization, and animation courses at her undergraduate institution. She has not taken any courses in computer vision, but her involvement in the computer graphics and visualization research group at UCSC has exposed her to some computer vision topics. In the group, she worked on developing real-time 3D scanning algorithms as well as a project involving video conferencing relighting using infrared illumination, which was just recently accepted for publication at Eurographics 2008.

B. Division of Labor

Because neither of has a sufficient background in computer vision, we will work closely together in order to learn the background material and understand the algorithms that may be useful for our system. In terms of building the rest of the system, there will likely be sizable chunks of development that can be done independently. As we start laying the foundation for the system concretely, we will decide how to split the development work so we can maximize the time we spend on grappling with the core vision problems.

References

[1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: Shape completion and animation of people. In *ACM Transactions on Graphics (SIGGRAPH)*, volume 24, 2005. 3

[2] A. Balan, L. Sigal, M. J. Black, J. Davis, and H. Haussecker. Detailed human shape and pose from images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, June 2007. 2

[3] N. Gehrig, V. Lepetit, and P. Fua. Golf club visual tracking for enhanced swing analysis tools. In *British Machine Vision Conference*, September 2003. 2

[4] L. Sigal and M. J. Black. Prediction 3d people from 2d pictures. In *IV Conference on Articulated Motion and Deformable Objects, AMDO*, 2006. 2

[5] R. Urtasun, D. J. Fleet, A. Hertzmann, and P. Fua. Priors for people tracking from small training sets. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, pages 403–410, Washington, DC, USA, 2005. IEEE Computer Society. 2