INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA

Motion Capture

Technical Report

Adriana Schulz

Instructor: Luiz Velho

Rio de Janeiro, May 6, 2010

Contents

Chapter 1

Overview

Motion Capture (MoCap) is a technology that allows us to record human motion with sensors and to digitally map the motion to computer-generated creatures[1].

The applications of motion capture go far beyond animation and include biomedical analysis, surveillance, sports performance analysis and input mechanism for human-computer interaction. Each application has its own particular set of singularities and challenges.

In this Technical Report, we will discuss motion capture of full body motion for character animation. We will introduce the basic concepts, detail the technique and describe how it was implemented at the VISGRAF Laboratory.

Chapter 2

Introduction

By definition, to animate is to bring to life, in this case, to make a lifeless object (a graphics model) move. Realistic animation of human motion is a challenging task, firstly, because human movements can be quite complex since joints have many degrees of freedom and, secondly, because people are skilled at perceiving the subtle details of human motion. For example, people are capable of recognizing friends at a distance purely from the way they walk and can perceive personality and mood from body language. This implies that synthetic human motion needs to be very accurate in order to appear real.

It is the opinion of many researchers of the area that synthesis of realistic human motion can only be made possible by an approach that makes extensive use of data from the real world [2]. It this context, it has been quite appealing for animators to used methods of copying the movements of real life objects.

An early example of such a technique is Rotoscoping, in which animators trace over live-action film movement, frame by frame, in order to create realistic motion. Though archives show that most of Snow White's movements were traced from actors motions using this technique, Wald Disney never admired to it. In fact, this was consider "cheating" because it was not produced by the imagination of the artist and also "cheep animation" because animations were supposed to be "bigger than life" not merely a "copy of life" [1].

Rotoscoping was, naturally, a precursor of MoCap and the controversy around it has the same origin. On the one hand, there is the need for creating engaging and expressive characters and on the other hand the need of synthesizing realistic motion efficiently and fast. Therefore, withing the animation community, there is a historical tension between animators and MoCap technicians [3].

The techniques for motion creation are usually divided into three basic categories: manual specification (key framing), motion capture and simulation [4].

Key framing borrows its name from traditional hand animation and is, in fact, very similar to it in the sense that, while computers reduce some of the labor by automatically interpolating between frames, the animator has to specify critical (or key) positions for the objects. This requires great training and talent, since a characters usually have many controls (e.g, each of the main characters of the movie Toy Story, which were animated in this fashion, had more than 700 controls). The advantage of this method is that the artist is able to control subtle details of the motion. Nevertheless, it is very hard to make the characters look real.

Motion Capture, as we have mentioned above, is a process that transfers recorded movement to an animated object. Some of the advantages of motion capture over traditional computer animation are [5]:

- more rapid, even real time results;
- real, naturalistic movement data;
- extremely accurate 3-D data that permits the study of the essence of the movements; and
- data formats that require minimal storage and allow for easy manipulation and processing.

The physically based approach, uses laws of physics to generate motion through simulation and optimization. This technique is largely used, not only for human motion, but also to animate fluids, fire, explosions, face and hair. Simulation techniques supply physical realism, while MoCap allows for natural looking motion. Currently, many applications merge both techniques together in order to create models of human motion that are flexible and realistic [2].

Chapter 3

MoCap Technique

Motion Capture systems can be divided into three different categories [5]: inside-in (sensors and sources located on the body), inside-out (sensors located on the body and sources outside) and outside-in (sources located on the body and sensors outside).

An example of an inside-in system is an electromechanical suit, where sensors are attached to to the performer's body measuring the actual joints inside the body. The advantages of this method are the absence of occlusions (all sensors are always "visible") and the portability of the suits. Nevertheless the actor's movement are constrained by the armature.

In electromagnetic (inside-out) systems, electromagnetic sensors, placed on joints of the moving object, measure their orientation and position with respect to an electromagnetic field generated by a transmitter. This method also directly collects positions and orientations of the sensors and does not have to consider occlusion problems. The drawbacks of this technique relate to the range and accuracy of the magnetic field as well as the constraint of movement by cables.

Optical systems are inside-out systems which use data captured from image sensors to triangulate the 3D position of a subject between one or more calibrated cameras. Data acquisition is traditionally implemented using special retro-reflexive markers attached to an actor and infrared cameras. However, more recent systems are able to generate accurate data by tracking surface features identified dynamically for each particular subject. This is called the marker-less approach.

Optical techniques are widely used since they allow large performance areas (depending on the number of cameras) and performers are not seriously constrained by markers. Nevertheless, orientation information is not directly generated and therefore extensive post-processing needs to be done in order to reconstruct the three dimensional motion.

The three major tasks that need to be undertaken are [6]: markers have to be identified and correlated in the 2D images, 3D locations have to be constructed from the 2D locations, and 3D marker locations need to be constrained to the model being capture (e.g., human skeleton).

The first step is to find the markers at each frame and track them over the video sequence. The latter can be quite difficult because markers often overlap, change position relative to one another and are occluded. In addition, noise can arise from the physical system (markers may move relative to their initial positions) or the sampling process. Hence three major problems occur when tracking the markers and may need user interventions to be resolved: missing data in the presence of occlusions, swapped paths between two markers that pass within a small distance of each other, and noise.

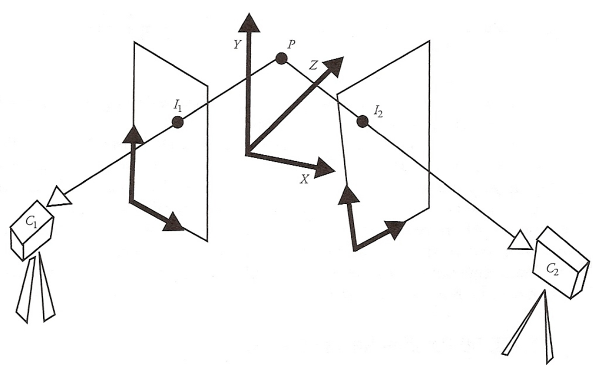


Figure 3.1: Two-camera view of a point. Extracted from [6].

The second step is to reconstruct the 3D points from the 2D trajectories. For this purpose, cameras need to calibrated , i.e., the position, orientation and intrinsic properties of the cameras need to be known. This can be done by recording a number of image points whose locations are known. With calibrated cameras, the three dimensional coordinates of each marker can be recontructed from at least two different views (the more orthogonal the views the better), as shown in Figure 3.1.

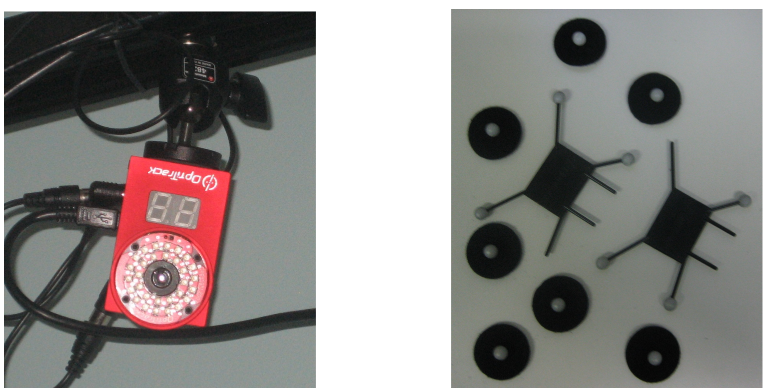
Finally, the 3D marker positions need to be transformed into the motion parameters of a kinematic skeleton model. For this, it is crucial to place the markers in adequate positions. The markers cannot be located exactly on the joint firstly because they are placed on the surface of the skin and second because they have to be set in positions where they will not move according to the performance. This represents a problem because although distances between consecutive joints are always the same, distances between markers may vary. Therefore, in order to locate the joint relative to the marker, we need not only the position, but also the orientation.

To solve these problems, most standard formats require placing three non-linear markers on each rigid body part instead of one in each joint [7]. These locations are then used to determine the position and orientation of each limb and the skeleton configuration is determined while the tracked subject performs an initialization pose (T-pose). Since noise is usually added to the data, directly using the calculated joint positions will probably result in varying bone lengths. To get around this problem, many systems calculate joint rotations and use a skeletal hierarchy to reconstruct the pose of the articulated body.

Chapter 4

MoCap at Visgraph

There are currently several comercial MoCap System available. During the course of this project, motion capture was done in the VISGRAPH Laboratory using OPTITRACK, `NaturalPoint`'s optical MoCap System. The infrared cameras (see Figure 4.1.b) are sensitive to the retroreflexive markers (see Figure 4.1.b) which are placed in the performer's body.



(a) OPTITRACK camera (b) Retroreflexive markers

Figure 4.1: MoCap setup.

As with traditional animation and many other arts, MoCap is actually composed of a number of phases[8]:

- studio set-up,
- calibration of capture area,

- performance and capture of movement,
- clean-up of data, and
- post-processing of data.

The studio was set up with six cameras, as shown in Figure 4.2. Tracking a large number of performers or expanding the capture area is accomplished by the addition of more cameras. Since each marker must be“seen” by at least two cameras, a greater number of cameras diminishes the possibility of occlusions.

Figure 4.2: OPTITRACK cameras set up at the Visgraf Laboratory.

Camera calibration was done with the help of OPTITRACK's software, AREA. This software also specifies the postion of the markers in the performer's body (see Figure 4.3).

The ARENA software (see Figure 4.4) was also used to process the acquired data. The first processing step is to trajectorize the data. This procedure takes the 2D data from each individual camera and changes it to a fully 3D path of each individual marker. After this procedure the software allows for post capture editing (such as filling gaps, fixing swaps and smoothing tracks) and exporting a file in a BVH or C3D format.

In this project, we used the BVH file format, which was originally developed by Biovision, a motion capture services company, as a way to provide motion capture data to their customers. Nowadays, it is widely used, and many applications support importing and exporting files in this format. It consists of two parts, a header section which describes the hierarchy and initial pose of the skeleton, and a data section which contains the motion data.

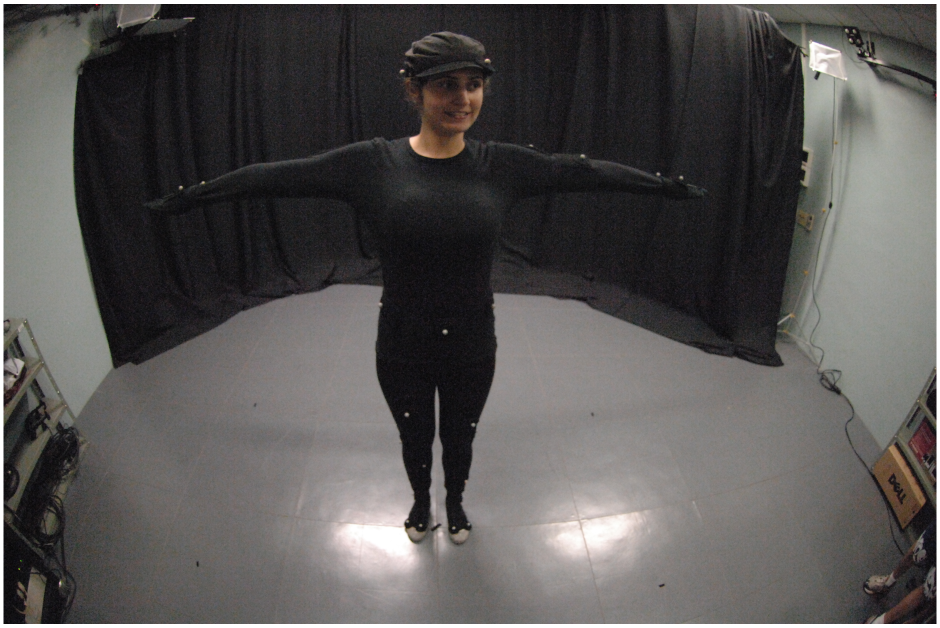


Figure 4.3: Dancer performing at the VISGRAPH Laboratory

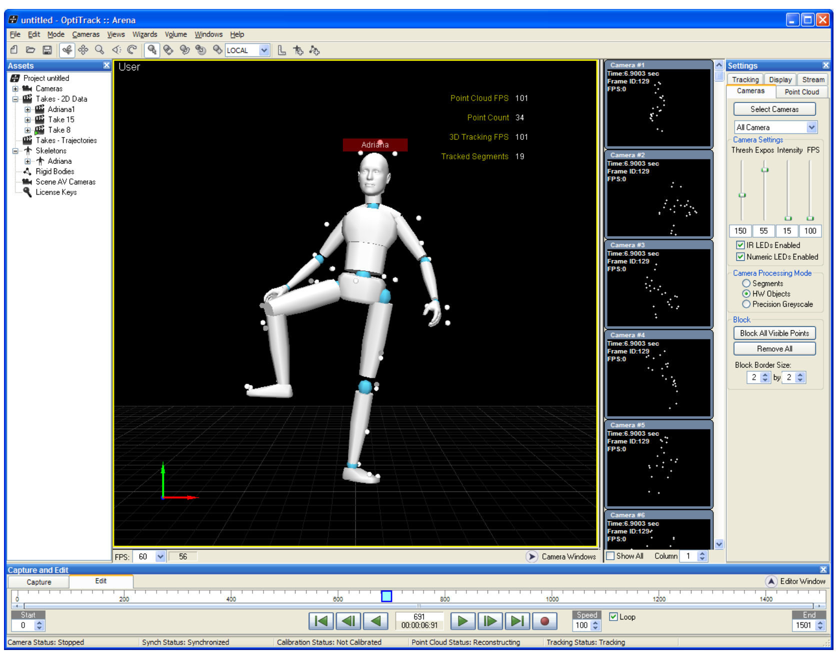


Figure 4.4: The ARENA software.

Bibliography

[1] Chris Bregler. Motion capture technology for entertainment [in the spotlight]. *Signal Processing Magazine, IEEE*, 24(6):160 –158, November 2007.

[2] Jessica Hodgins and Zoran Popovic. *Animating Humans by Combining Simulation and Motion Capture*. SIGGRAPH'00 Course Notes 33, SIGGRAPH-ACM publication, New Orleans, Louisiana, USA, July 2000.

[3] Michael Gleicher. Animation from observation: Motion capture and motion editing. *SIGGRAPH Comput. Graph.*, 33(4):51–54, 2000.

[4] Jessica Hodgins. Animating human motion. *Scientific American*, 278(3), March 1998.

[5] Charlotte Belland, James W. Davis, Michael Gleicher, and Barbara Helfer. *Motion Capture: Pipeline, Applications, and Use*. SIGGRAPH'02 Course Notes 28, SIGGRAPH-ACM publication, San Antonio, Texas, USA, July 2002.

[6] Rick Parent. *Computer animation: algorithms and techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002.

[7] Margaret S. Geroch. Motion capture for the rest of us. *J. Comput. Small Coll.*, 19(3):157–164, 2004.

[8] Maureen Furniss. *Motion Capture*. MIT Communications Forum, http://web.mit.edu/comm-forum/papers/furniss.html.