Evaluting Video-Based Motion Capture

Michael Gleicher
Department of Computer Sciences

Nicola Ferrier
Department of Mechanical Engineering
University of Wisconsin-Madison

http://www.cs.wisc.edu/graphics
The Message...

- Motion capture for animation is hard!
- It’s hard in ways that are challenging for computer vision
- Despite advances in computer vision, don’t expect miracles too soon
Outline

- What do we want from motion capture?
- Why is this so hard?
- An experiment
  - What do observations tell you
- Computer vision in this light
Motion Capture

- Observations

- Computationally Tractable Representations

Note:
Motion capture means capturing the motion, not the process of Animation by Observation!
Motion Capture

Observations

How “High Quality”? Needs to be “good enough”!

Computationally Tractable Representations

How cheap, reliable, non-intrusive, …? Can’t be too cheap!
Motion Capture for Animation

Observations

Do whatever it takes!

Computationally Tractable Representations

High Quality “Skeletal” data! (good enough?)
What does animation need?

- Animation doesn’t really need high-precision and accuracy
  - Not concerned about details
  - Not doing measurement

- “Just” need to capture mood, emotion, intent, subtlety, personality, …
  - All those things an actor can do
Two Problems

- Where does X live in the data?
  - Where $X \in \{\text{style, personality, emotion, ...}\}$

- Small artifacts can destroy realism
  - Eye is sensitive to certain details
  - Amazing what you can’t get away with
    - See Kovar, Schreiner and Gleicher, SCA ‘02
How do we handle these problems?

- Don’t know which details are important!
- Must preserve ALL details
  - Since you don’t know what is important
- Need to understand artifacts better

Motion Capture Animators are conservative:
Want excess precision just in case!
Not all Mocap Applications are like this!

- Computer Puppetry
  - Shin, Lee, Gleicher, Shin, Tog ’01.
Dream #1

- Capture “essense” only
- Add details later

- This is equivalent to the vision problem that we’re getting to.
- This motivated our work.
Dream #2

- Cheap capture devices
  - Non-intrusive  Ubiquitous
  - Easy to obtain  Inexpensive
  - Easy to set up

- Single camera, video motion capture!
  - Multiple cameras, might as well be mocap hardware

- How much can you get?
Experiment: Minimal Assumption Mocap

- Pinhole camera model
- Rigid skeleton
- Solve constrained-optimization for locations
Answer: Not a lot!

- See paper for details
  - Surprisingly low precision
  - Surprisingly many ambiguities

- Weak model
  - Few assumptions about motion
  - Distance constraints
  - Assume perfect observations
What’s going on here?

- Limited information -> Limited results
- Not much info in a 2D observation

- 2D observations are a constraint
  - Limit the possible causes
    - But still leave a large space
- How to choose amongst possibilities?
  - Optimization?
  - Probabilities?
Strong Models!
How Computer Vision does better.

- Computer vision human tracking works by using a stronger model
- Use more information about what motions are likely to choose amongst possible interpretations
- Encode what motions are likely
- The hot topic in human tracking
  - Rehg, Black, Forstythe, Reid, Brand, Shin, …
  - Impressive success, varying methods for implementation of “likelihood”
Strong Motion Models

- Encode “likely” or “common” motions
- Observations select from these
  - Extreme: Matt Brand’s work
  - Novice dances, plays motion of expert

- Doesn’t work for animation!
  - Want ununusual, unlikely, specific, …
  - If you had seen the motion before, why not just play it from database?
Better Biomechanical Models

- **Idea**
  - Use more knowledge of human to limit possibilities

- **Problems**
  - Need manipulatable representations for practicality
  - Humans are complex
  - Strong models only good if they are correct
  - Unclear how much more constraint this adds

- Little exploration in vision
Modeling Humans

- Humans are complex!

Human motion can be understood at a very fine level of detail!
Abstractions

206 bones, muscles, fat, organs, clothing, ...

206 bones, complex joints

53 bones Kinematic joints
Abstractions vs. Reality (skeletons vs. humans)

Representation of complex human structure with varying degrees of simplification

- Simple Pin Joint
- Complex tendon and bone system
Prognosis

- Human tracking is improving
  - Primarily through use of strong models
- New approaches may not work for animation
  - Different quality goals
  - Different applications (classification)
- Looks like we’ll have old approaches for a while
  - “engineer” away vision problem
  - Use expensive sensors
Thanks!

- UW Graphics group
- Our friends in the mocap industry
  - For data and ideas

- National Science Foundation
  - Mike: CCR-9984506, IIS-0097456
  - Nicola: IRI-??

- Industrial and University sponsors
  - Microsoft, Intel, Autodesk, Alias/Wavefront, Wisconsin Alumni Research Foundation, IBM, Pixar