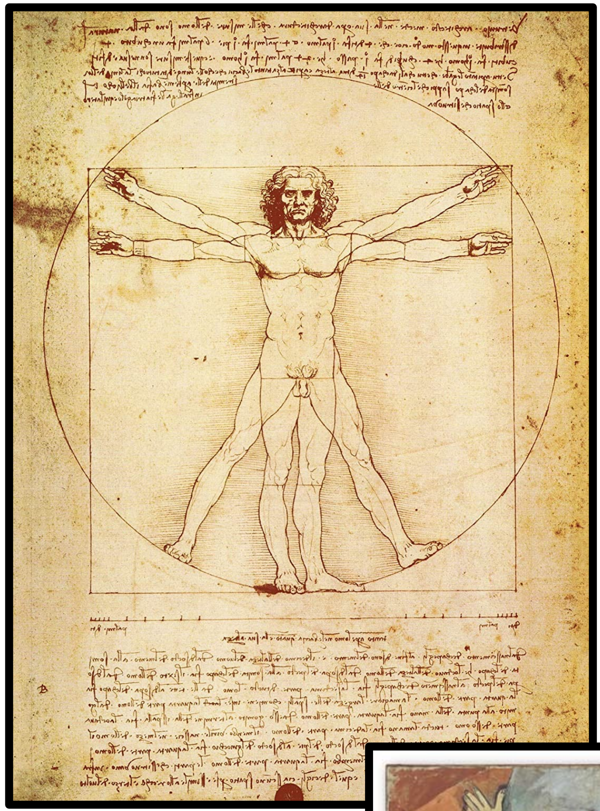
Virtual Humans – Winter 23/24

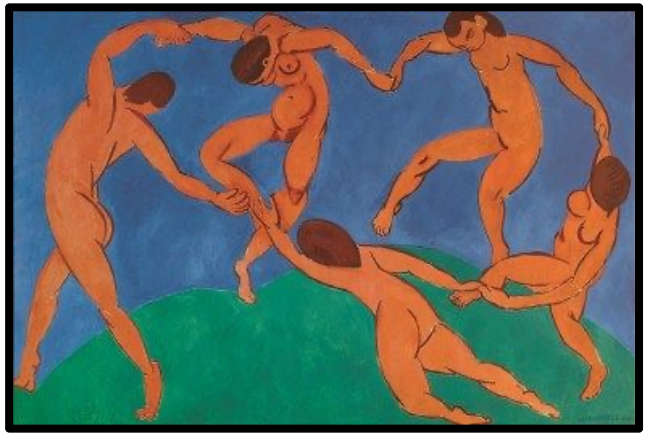
Lecture 1_1 – Introduction to Human Models - History

Prof. Dr.-Ing. Gerard Pons-Moll
University of Tübingen / MPI-Informatics

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN




Images from wikicommons





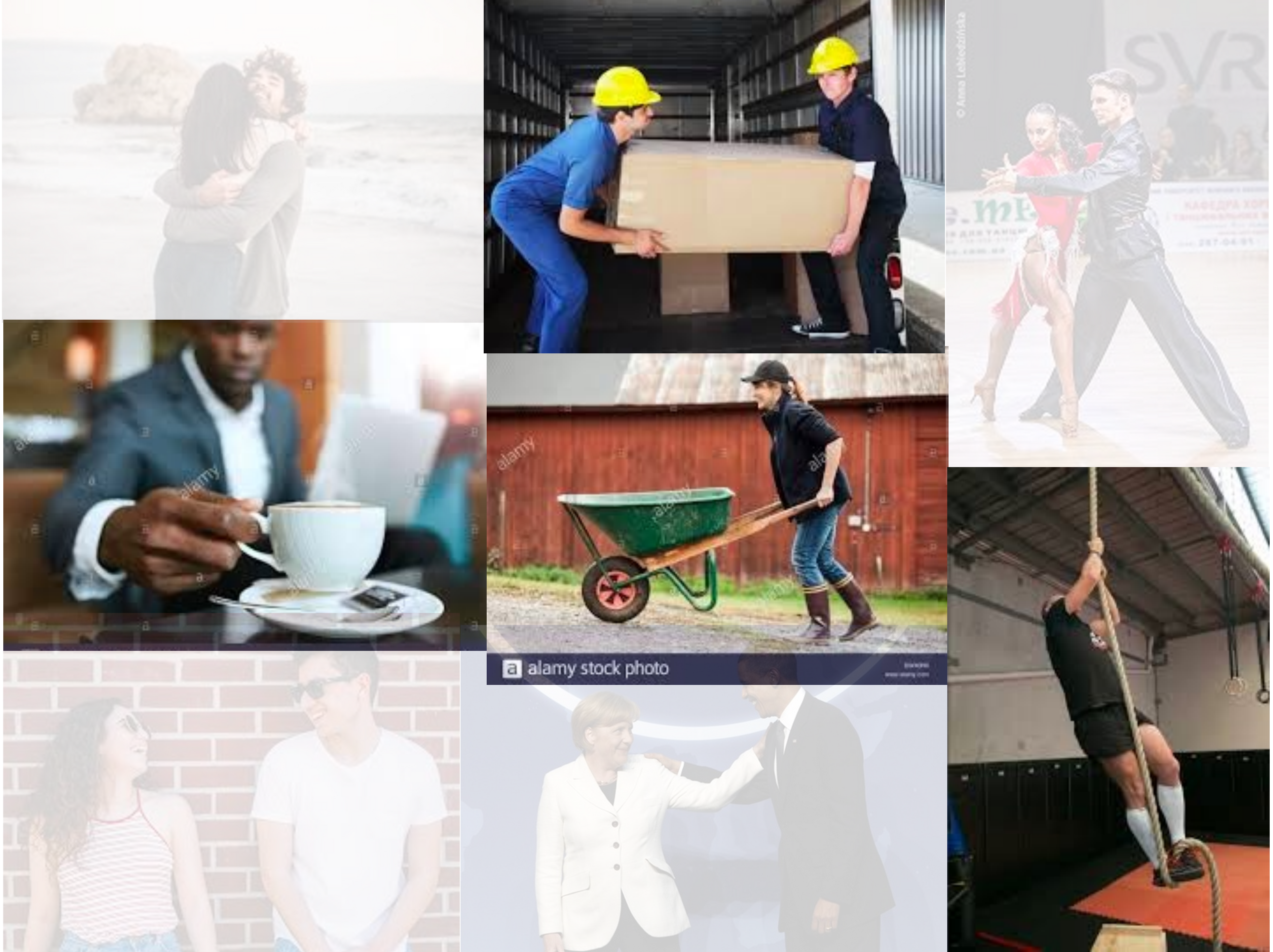











alamy stock photo

Autonomous Driving, Robots, AR/VR

Image source: BDD100K Dataset

Image source: Tilburg University


Image source: Microsoft

People AI at Meta

We are hiring an engineering manager in Zurich, to help us shape the future of human body perception technology for AR and VR. If you are interested feel free to reach out.
#hiring #computervision #machinelearning #augmentedreality #virtualreality #metaverse

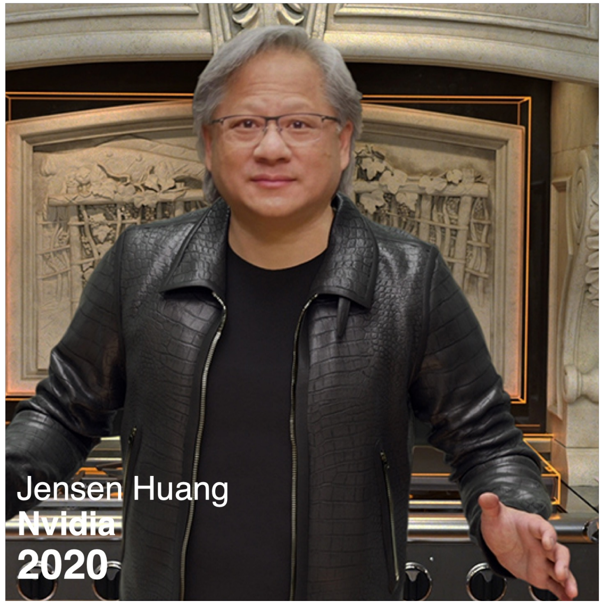
Google

Looking for Research Scientists in Visual Computing and 3D Human Modeling!

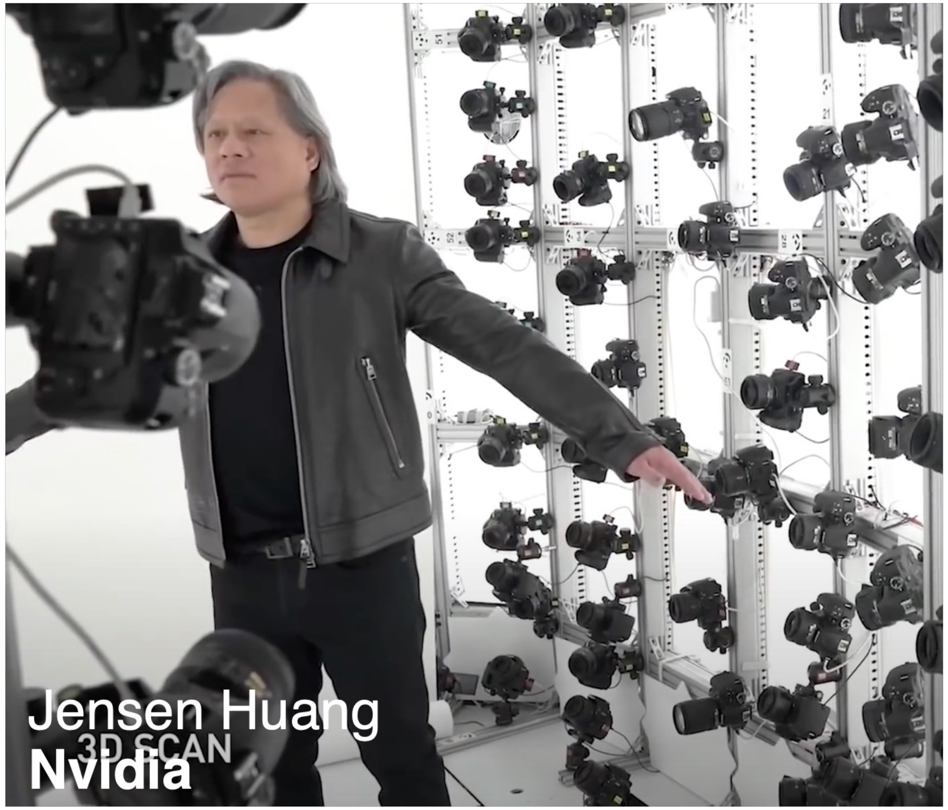
Are you keen on advancing the state-of-the-art research on Human Modeling and at the same time improving the lives of billions of users? Are you passionate about Augmented and Mixed Reality, Visual Computing or Machine Learning? Then our team might be the perfect fit for you!


MPC

Human avatar creation

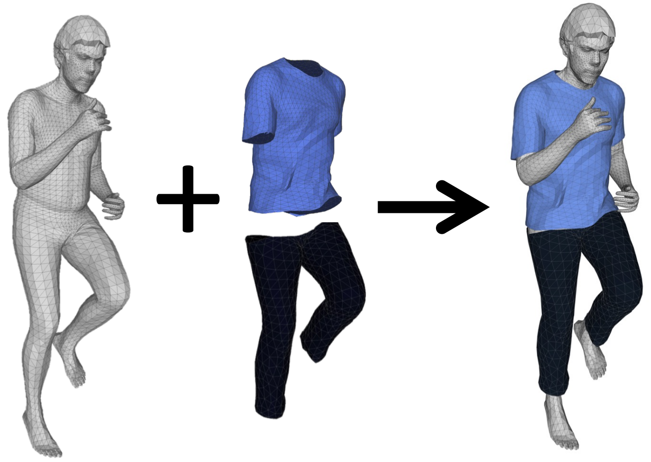


Human avatar creation



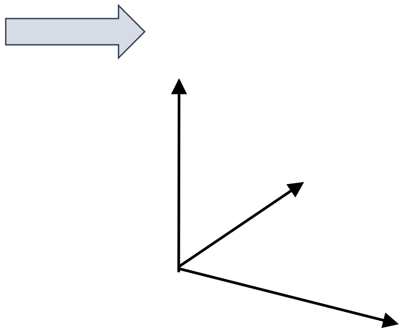
Problem: Time consuming, expensive equipment, specific to one subject, do not scale
Goal: Democratizing human model creation

Goal: **Appearance** Virtual Humans



Generation (+ clothing)

3D world

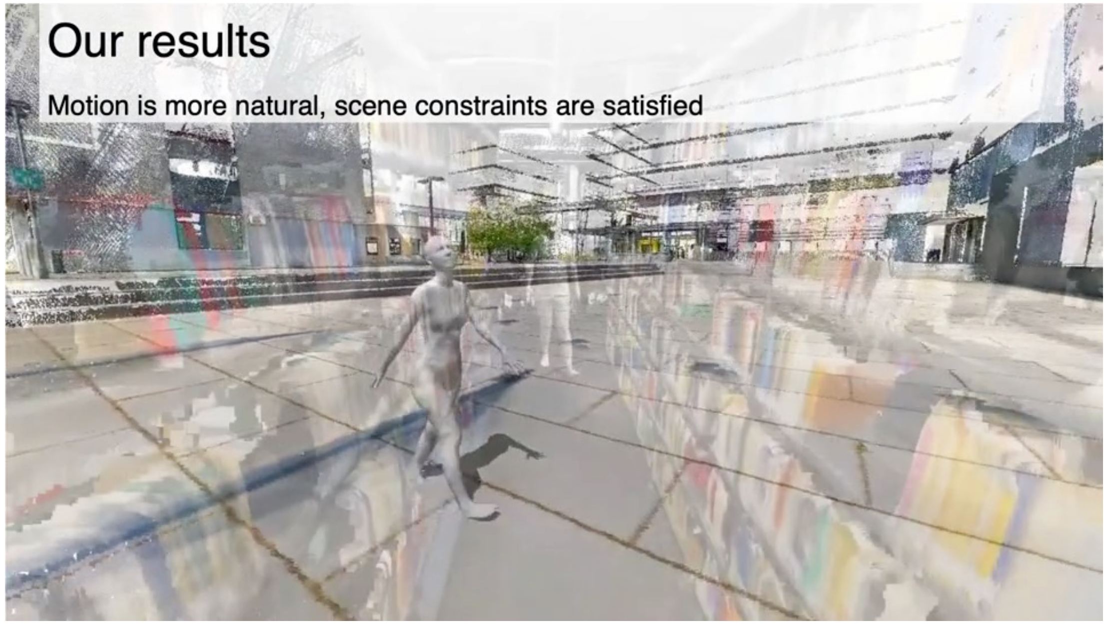


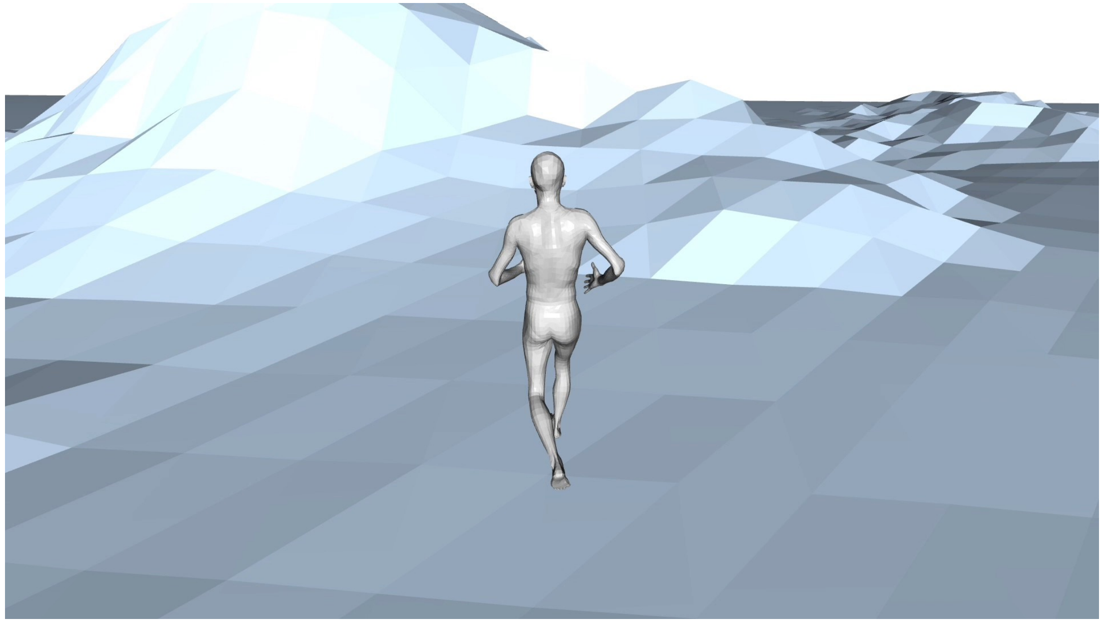
Perception and Analysis

<u>Generate</u> realistic 3D people:
- Move and look like real people
- Easy to control and animate
- Easy to fit to data

<u>Perceive</u> 3D people from images:
- Capture shape, pose, clothing, personal details, illumination, environment …

Goal: **Awaken** Virtual Humans





Perceive: We should be able to reconstruct **real** 3D humans jointly with the objects and the scene they interact with

Generation: **Virtual** humans should be able to move and interact with objects and scenes like real humans

Goals (interrelated)

- Computer Vision: Train computers to “see” us
  - Understand our behaviors, emotions, actions
  - Understand our interactions with each other and the world

- AR/VR/Graphics: Train avatars to mimic us
  - By watching us, learn to behave like us
  - If we can reproduce human-like behavior, then we have understood it at some level

Why is it Difficult?

Loss of 3D in 2D projection

Unusual poses (high D)

Self occlusion

Low contrast



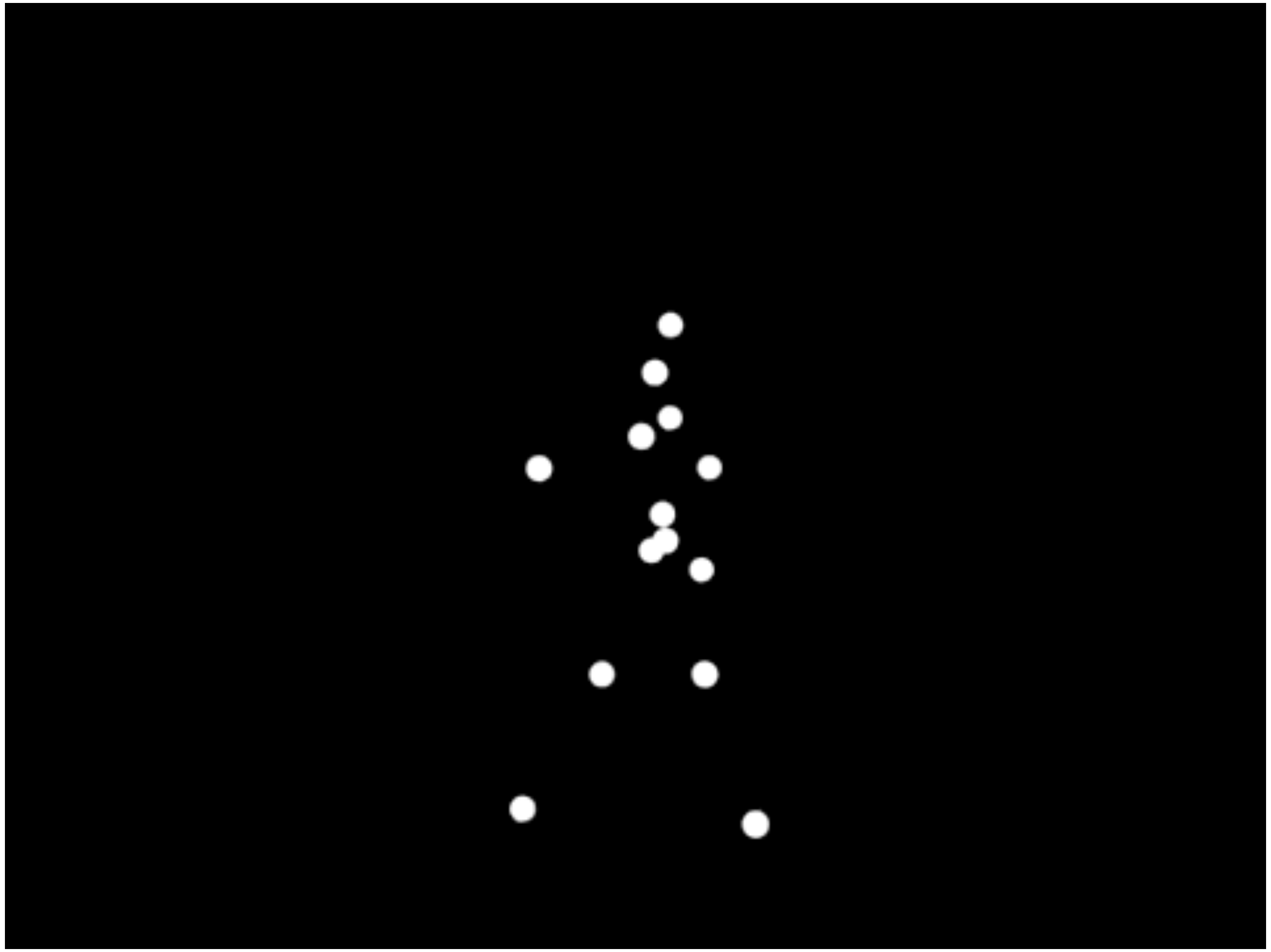



A little history
Early body models

Capturing humans in motion

ETIENNE-JULES MAREY, 1882 chronophotograph.

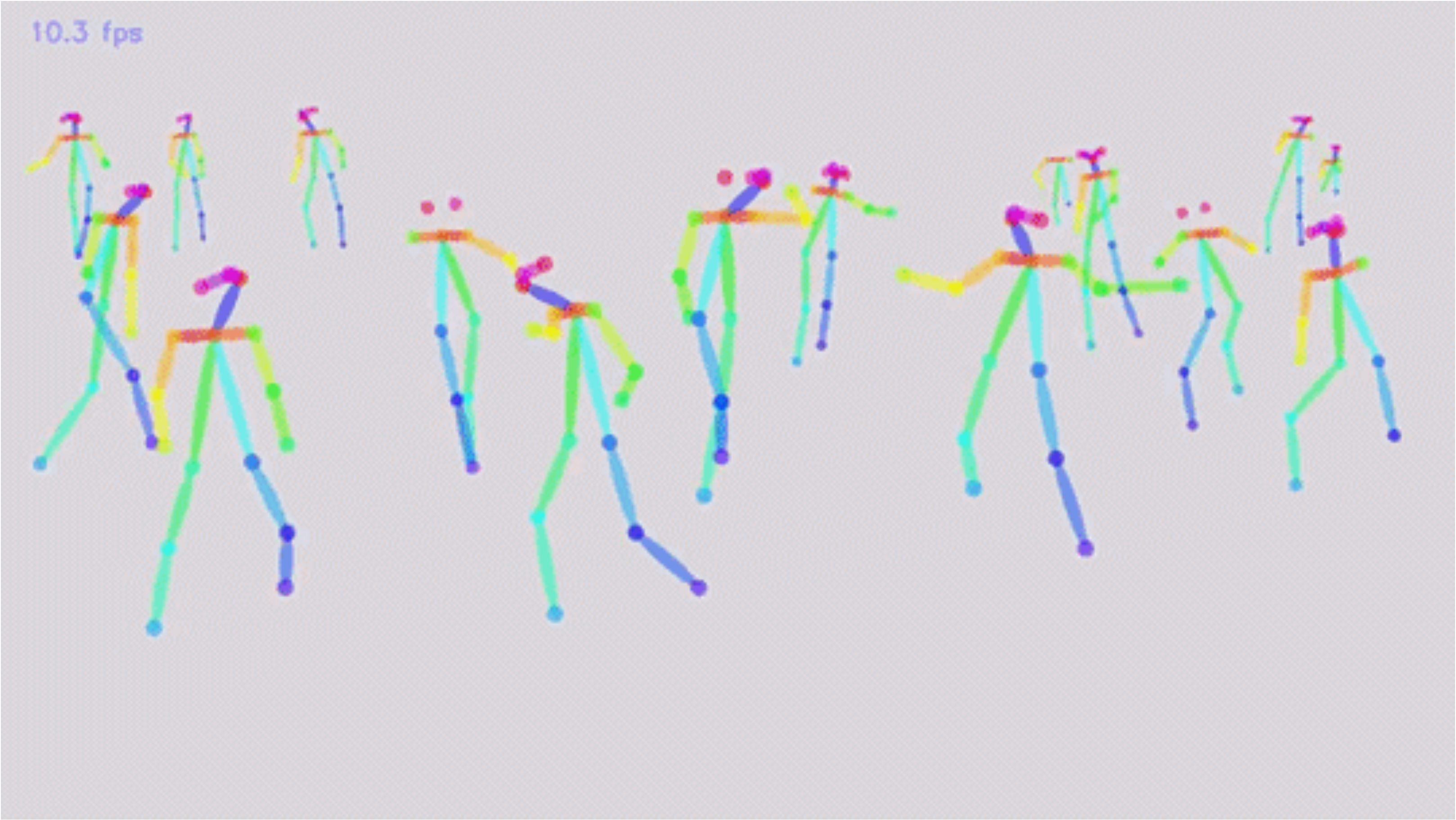
A key influence on the field



"…. the motion of the living body was represented by a few bright spots describing the motions of the main joints…. 10–12 such elements in adequate motion combinations … evoke a compelling impression of human walking, running, dancing, etc."

Gunnar Johansson, Visual perception of biological motion and a model for its analysis, Perception & Psychophysics, 1973.

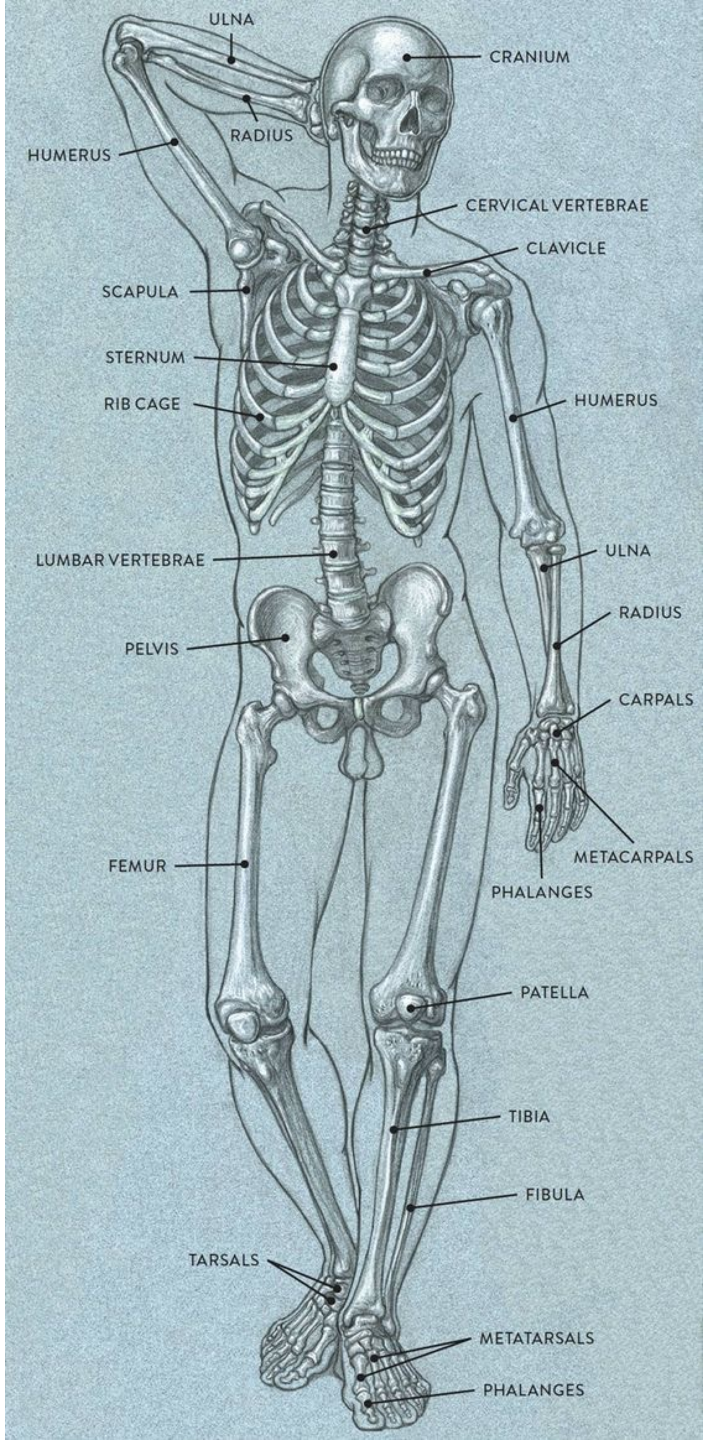
Dominant paradigm: 2D joints



OpenPose, Cao et al., 2017, 2018

Are joints enough?

- The joints are **unobserved**.
- **Contact** is key. Joints don't touch the world; the skin does.
- We need to model the surface of the body to reason about contact and expression.
- Our shape is also related to our health and how the world perceives us.



https://doctorlib.info/anatomy/classic-human-anatomy-motion/2.html

Ingredients to infer human models from data

Building a human model

- Kinematic parameterization
  - Rotation Matrices
  - Euler Angles
  - Quaternions
  - Twists and Exponential maps
  - Kinematic chains

- Subject shape model
  - Geometric primitives
  - Detailed Body Scans
  - Human Shape models

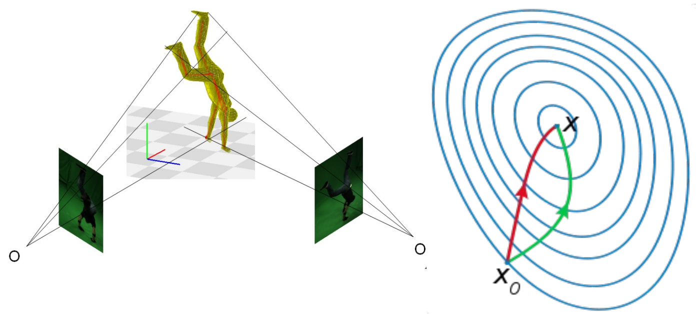
Fitting model to observations

- Inference
  - Observation likelihood
  - Local optimization
  - Particle Based optimization
  - Directly regressing parameters

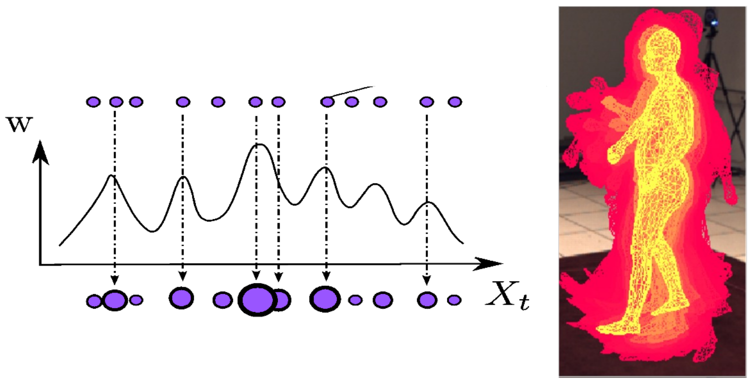
Early works where generative

$$p(\mathbf{x}|\mathbf{I}) \quad \propto \quad p(\mathbf{I}|\mathbf{x}) \quad \times \quad p(\mathbf{x})$$

Posterior Likelihood $\times$ Prior



Optimization

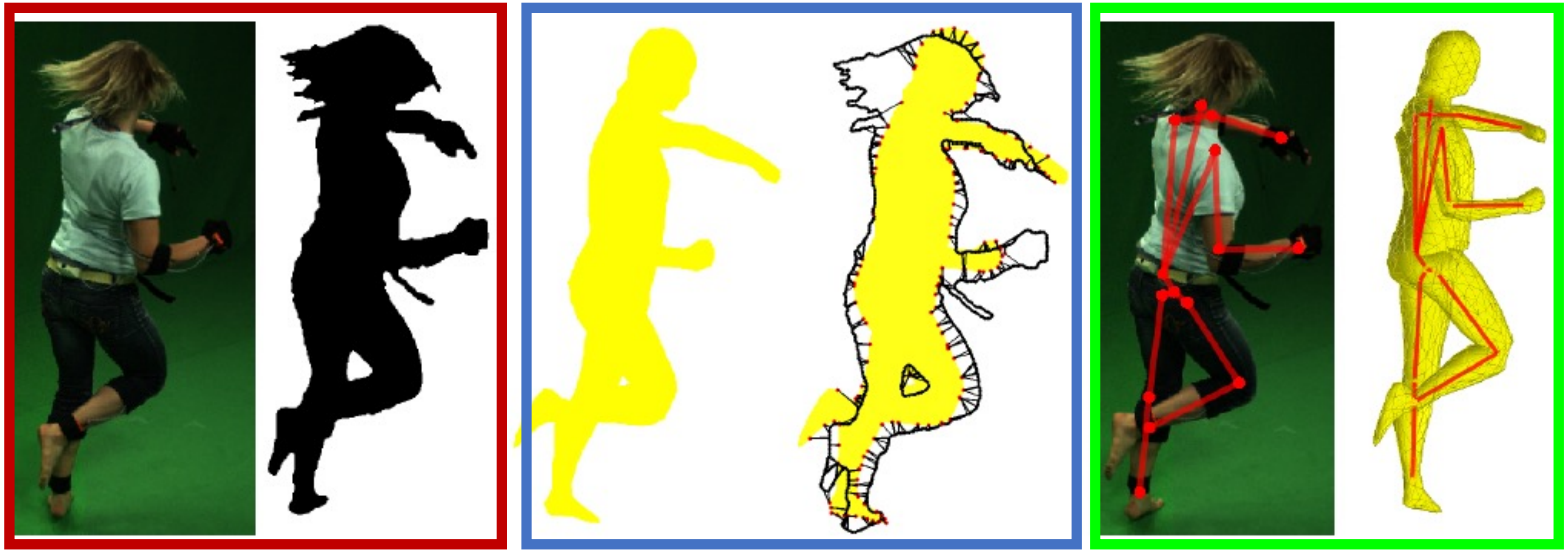


Bayesian models

Map of $p(\mathbf{x}|\mathbf{I})$

Approx. $p(\mathbf{x}|\mathbf{I})$ with weighted samples

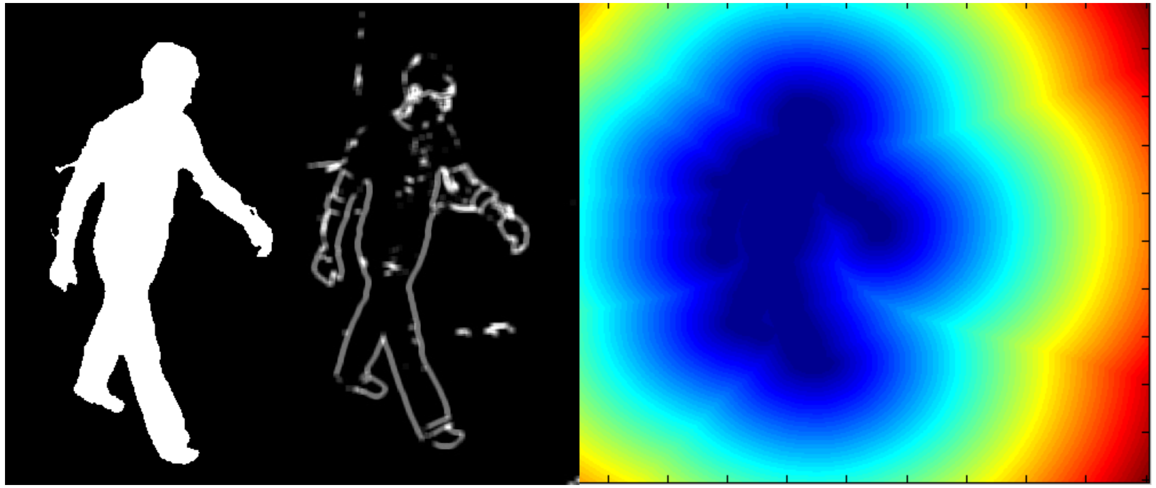
Inferring models from images



Extract features Predict and match Optimize

Matching synthesized features & observation

- Silhouettes
- Edges
- Distance transforms
- SIFT
- Optic flow
- Appearance
- …



Any feature that can be predicted from the model and is fast to compute

The beginning: 45 years ago

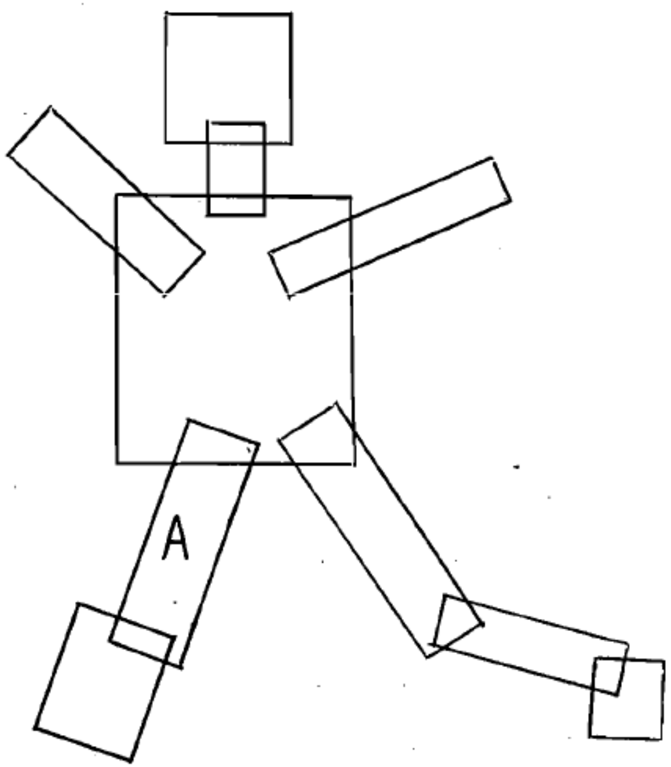
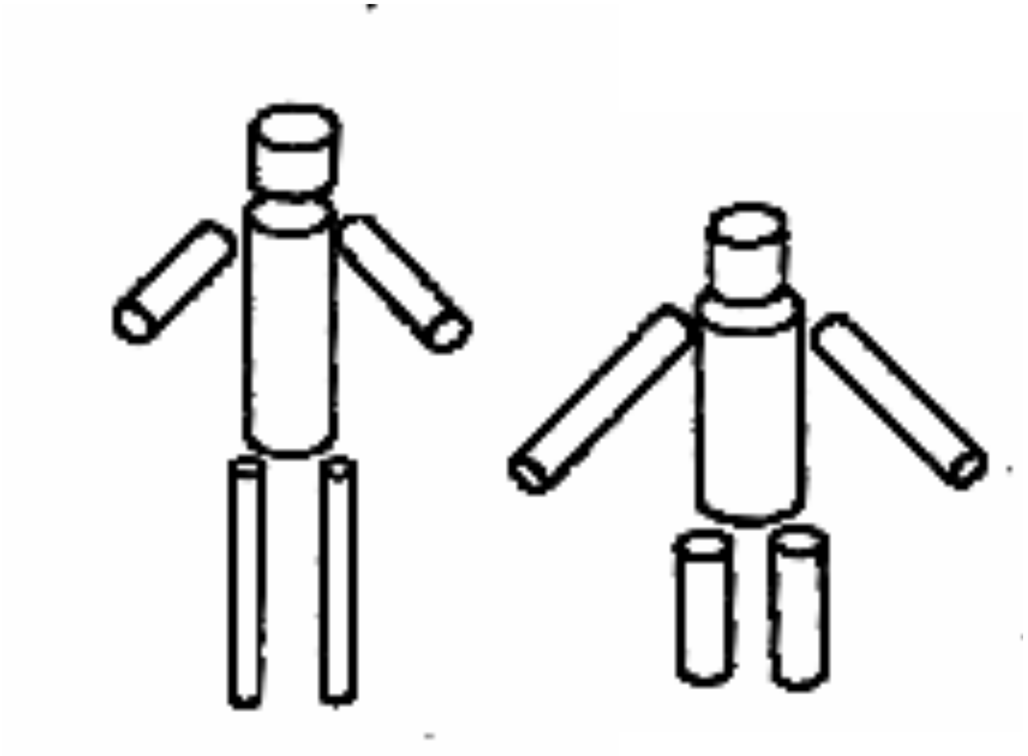


Figure 4. Relaxation picks out the interpretation of A as a thigh even though a calf is a locally better alternative.

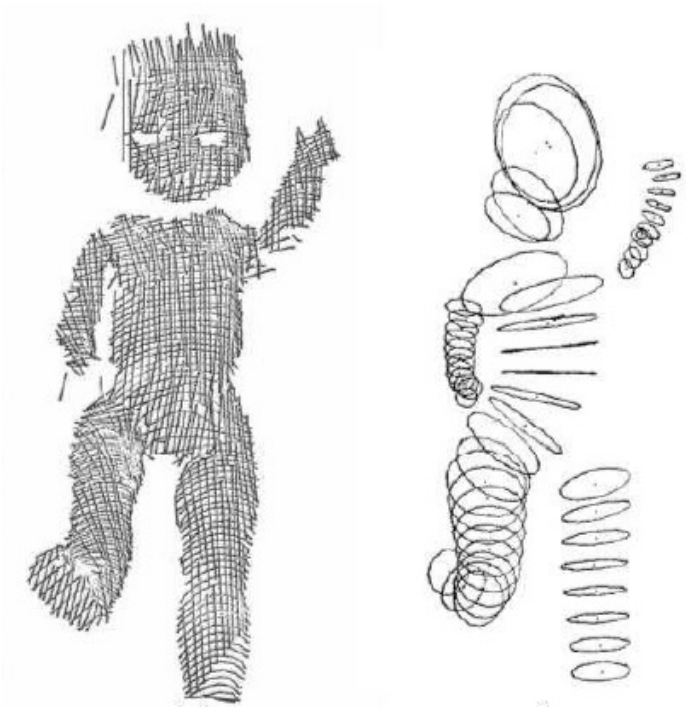
G. E. Hinton. Using relaxation to find a puppet. In Proc. of the A.I.S.B. Summer Conference, July 1976.
His first paper!

The beginning: 3D shape



Marr and Nishihara ’78

Proposal for a general, compositional, 3D shape representation



Nevatia & Binford ’73

Generalized cylinders fit to range data

There were no range scanners!

David Hogg, 1983

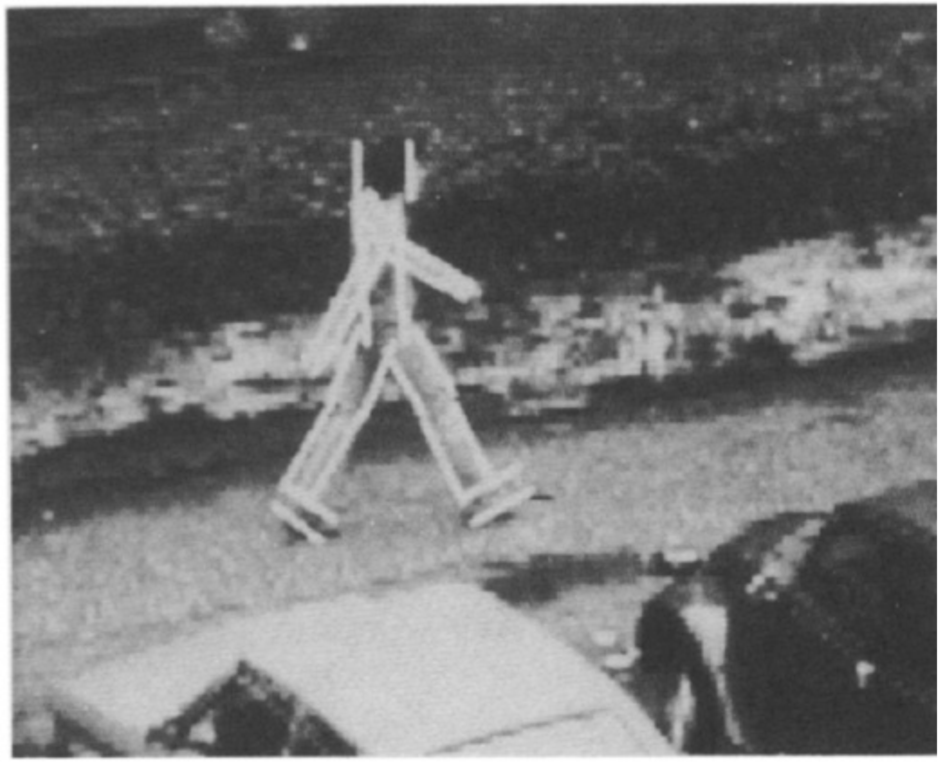


Figure 12. Set of lines which correspond to the imag[e]
projections of occluding surfaces. They represent the image
in Figure 4

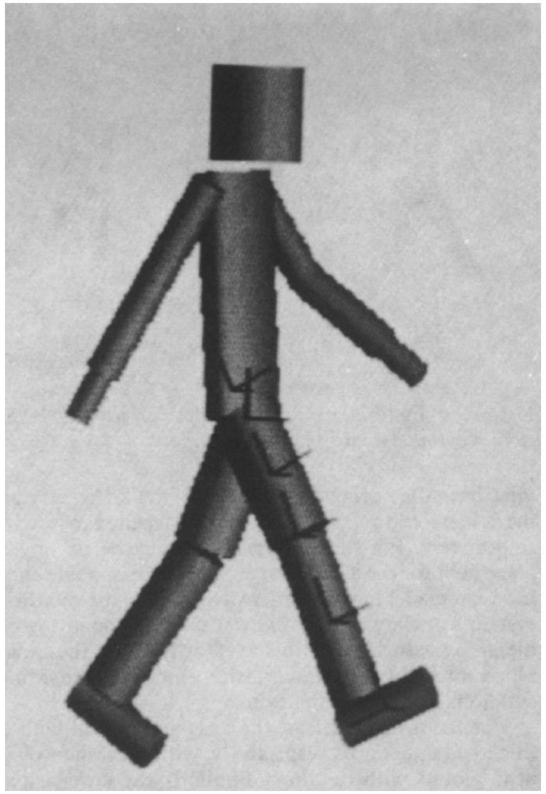
Figure 5. Edge-finding operation applied to the image in
Figure 4

Model-based vision: A program to see a walking person, D Hogg
Image and Vision computing 1 (1), 5-20

David Hogg, 1983

Thanks to Andrew Fitzgibbon for the video.

David Hogg, 1983



```
class: WALKER
parts:

    partclass: person

class: person
postures: [stretchl liftr stretchr liftl]
parts:

    partclass: torso
    weight: 0.05

        [stretchl liftr stretchr liftl]
        position: x = 0 y = 45 z = -5 a = 0 b = -5 c = 0 s = 0.35

    partclass: head
    weight: 0.05

        [stretchl liftr stretchr liftl]
        position: x = 0 y = 112 z = 0 a = 0 b = 0 c = 0 s = 0.14

    partclass: arm
    weight: 0.05

        [stretchl]
        position: x = 26 y = 85 z = -10 a = 0 b = [10 50] c = 0 s = 1
        [liftr]
        position: x = 26 y = 85 z = -10 a = 0 b = [-10 30 -20 0]
            c = 0 s = 1
        [stretchr]
        position: x = 26 y = 85 z = -10 a = 0 b = -50 -10] c = 0 s = 1
        [liftl]
        position: x = 26 y = 85 z = -10 a = 0 b = [-20 40 0 20] c = 0
            s = 1

        [stretchr]
        posture: [straight]
        position: x = -16 y = 10 z = 0 a = 0 b = [-40 -30 -20 20]
            c = 0 s = 1

        [liftl]
        posture: [straight]
        position: x = -16 y = 10 z = 0 a = 0 b = [-30 10 0 15] c = 0
            s = 1

class: arm
parts:

    partclass: upper-arm
    weight: 0.5
    position: x = 0 y = -20 z = 0 a = 0 b = 0 c = 0 s = 0.16
    partclass: lower-arm
    weight: 0.5
    position: x = 0 y = -40 z = 0 a = 0 b = [-40 0 -20 20] c = 0 s = 1

class: lower-arm
parts:

    partclass: forearm
    weight: 0.7
    position: x = 0 y = -20 z = 0 a = 0 b = 0 c = 0 s = 0.16

    partclass: hand
    weight: 0.3
    position: x = 0 y = -50 z = 0 a = 0 b = 0 c = 0 s = 0.08

class: leg
postures: [straight bent]
parts:
```

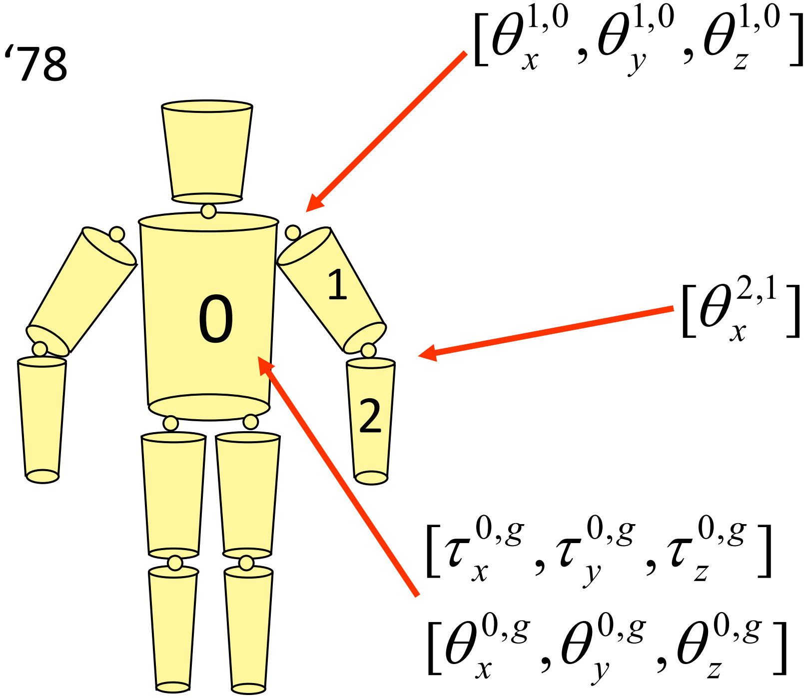
Model-based vision: A program to see a walking person, D Hogg
Image and Vision computing 1 (1), 5-20

1983-1993
The lost decade.

The classical generative approach

Kinematic tree:

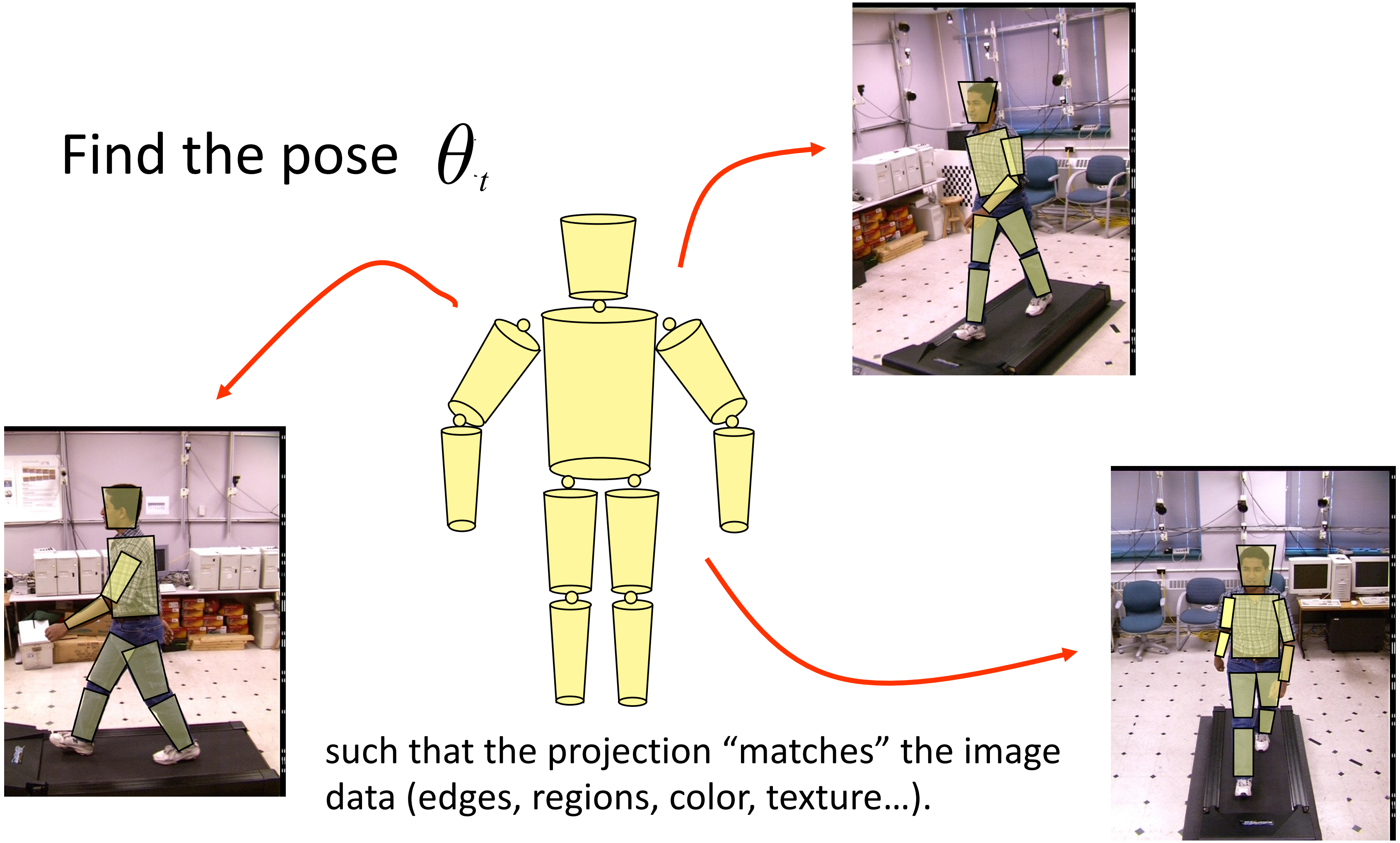
Marr&Nishihara ‘78



Represent a “pose” at time *t* by a vector of parameters: ϕ_t

The classical generative approach

Find the pose θ_t



such that the projection “matches” the image data (edges, regions, color, texture…).

Geometry and optimization: 1994-2004

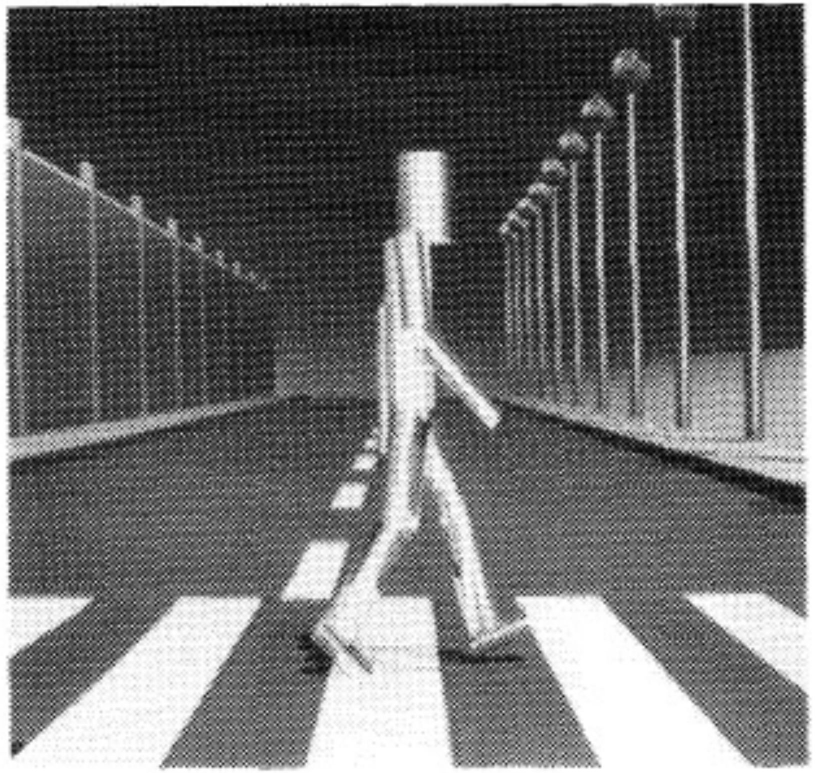
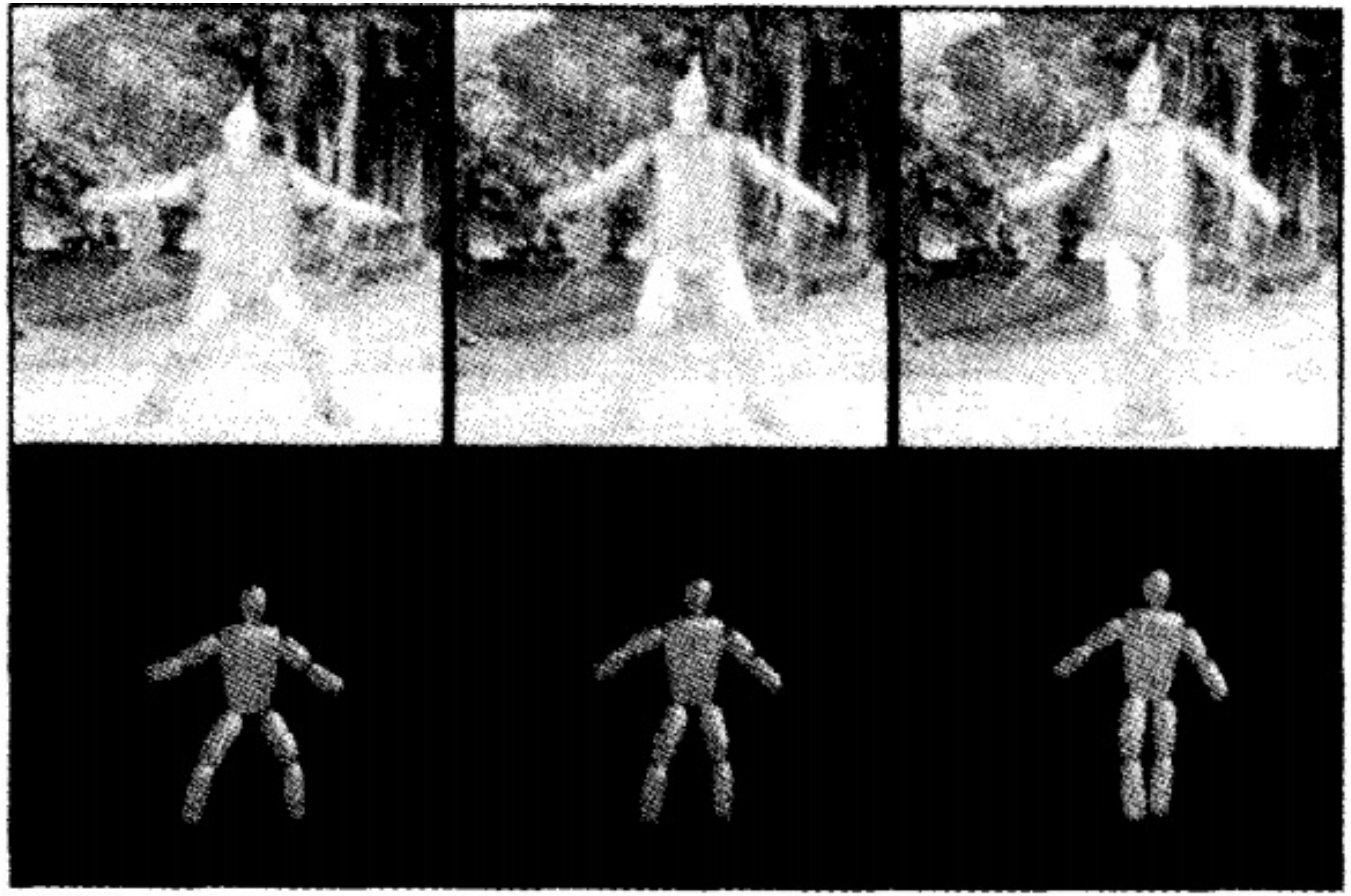


FIG. 4. Model of the human body.

FIG. 20. Determined motion state.

Rohr, Towards Model-Based Recognition of Human Movements in Image Sequences, CVGIP, 1994

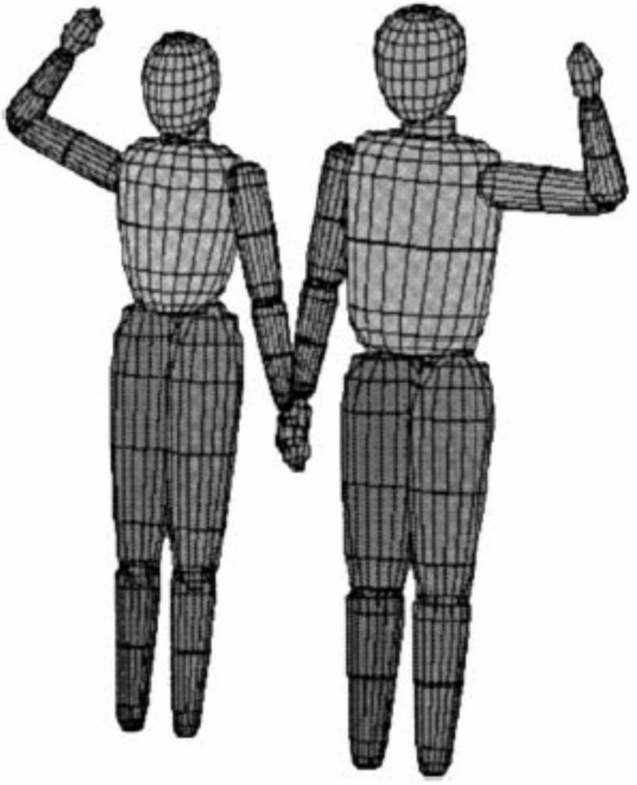
Non-rigid parts

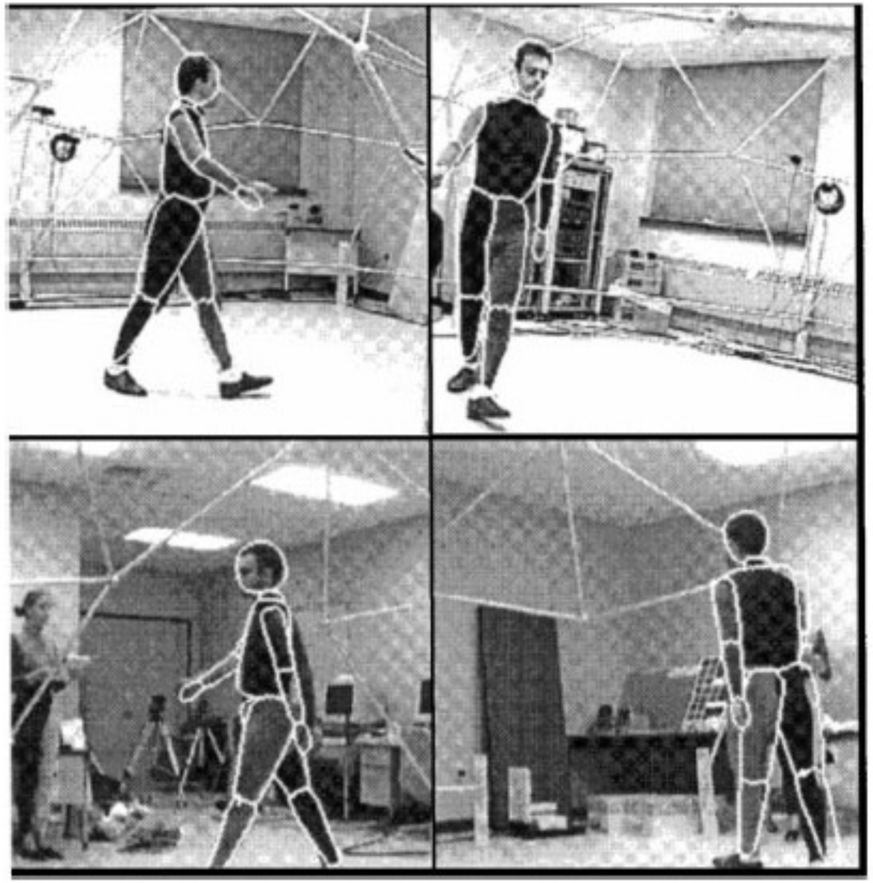


Recovery of Nonrigid Motion and Structure , Alex Pentland and Bradley Horowitz, PAMI 1991

Multi-camera, markerless, mocap

Superquadrics

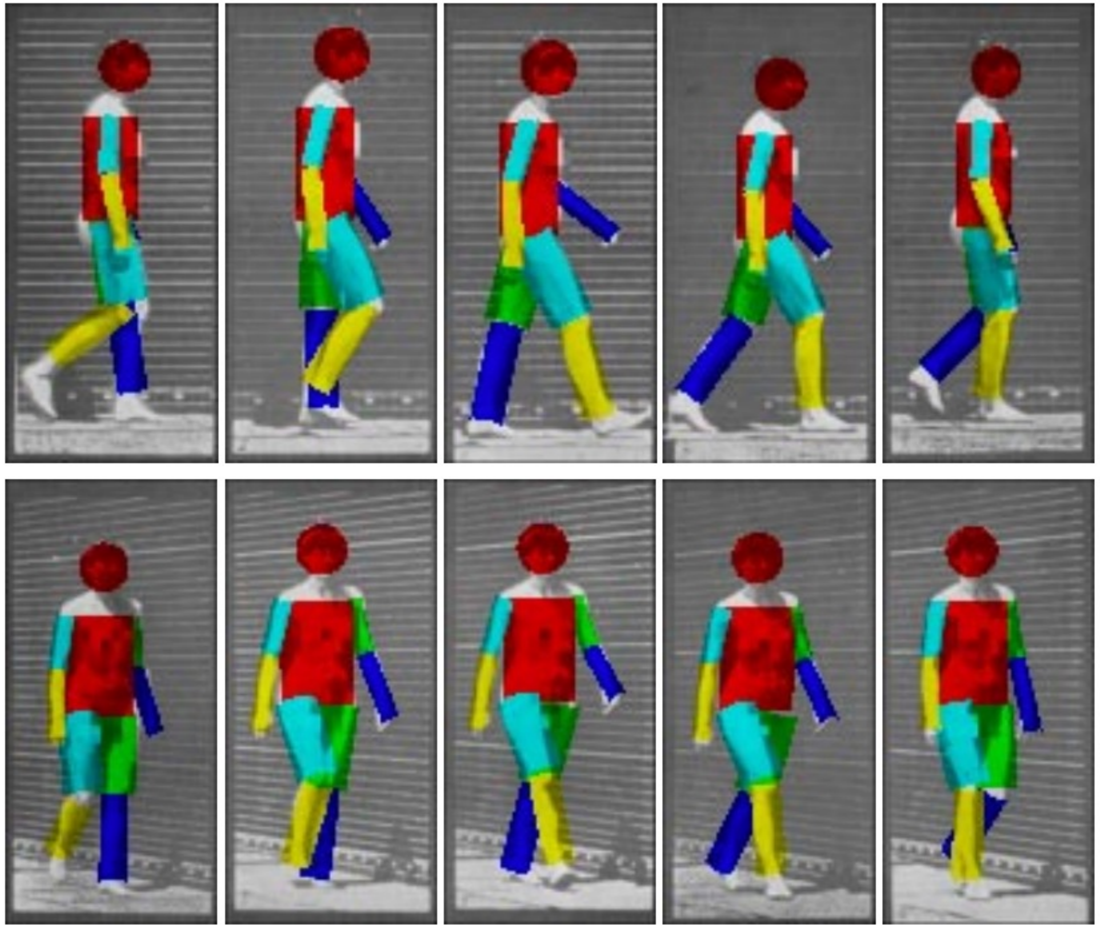




Simple shapes, multi-camera, special clothing.

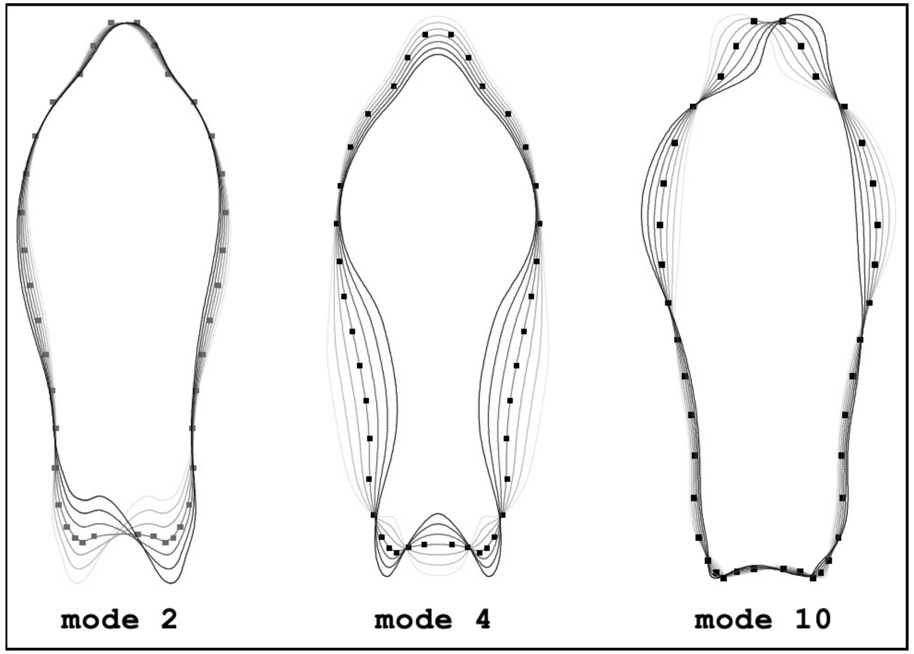
D. Gavrila, Vision-based 3-D Tracking of Humans in Action, Ph.D. thesis, 1996.

Bregler & Malik CVPR 1998

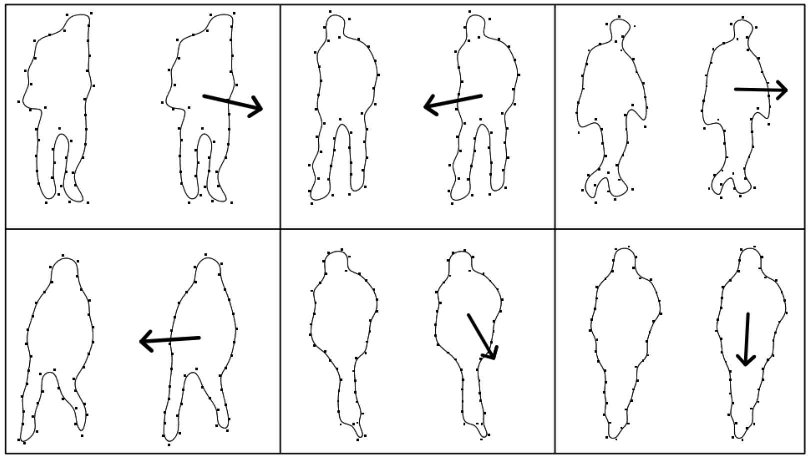


- Tracking People with Twists and Exponential Maps
- 2D motion of a projected 3D model

First learned "body model" was 2D

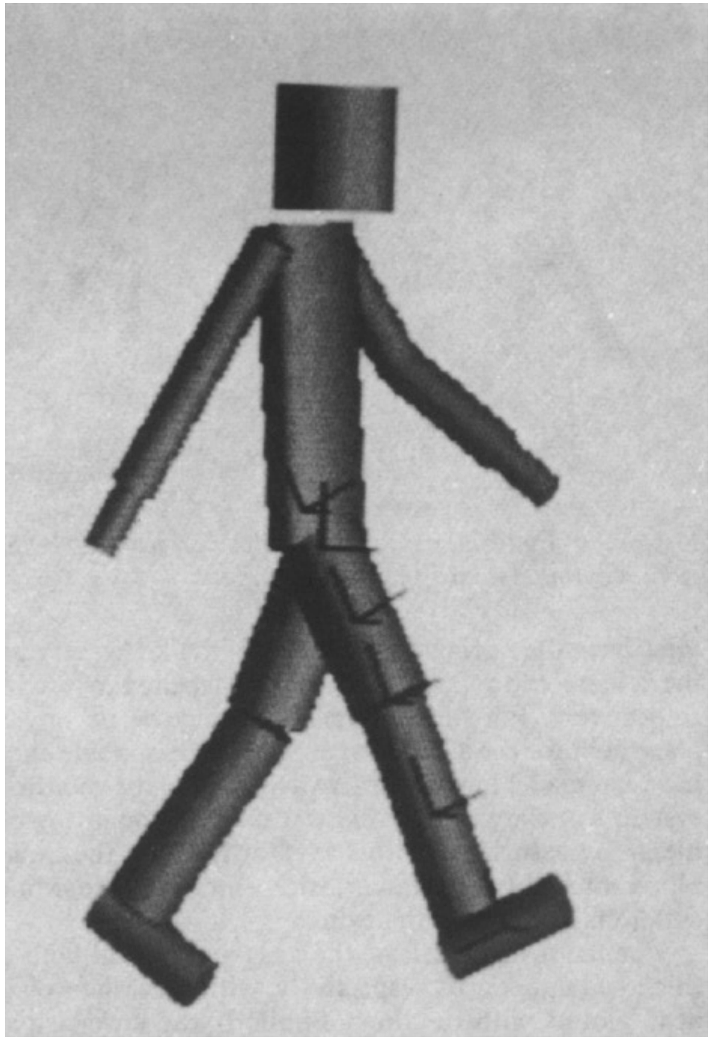


Pedestrian Eigen-shapes



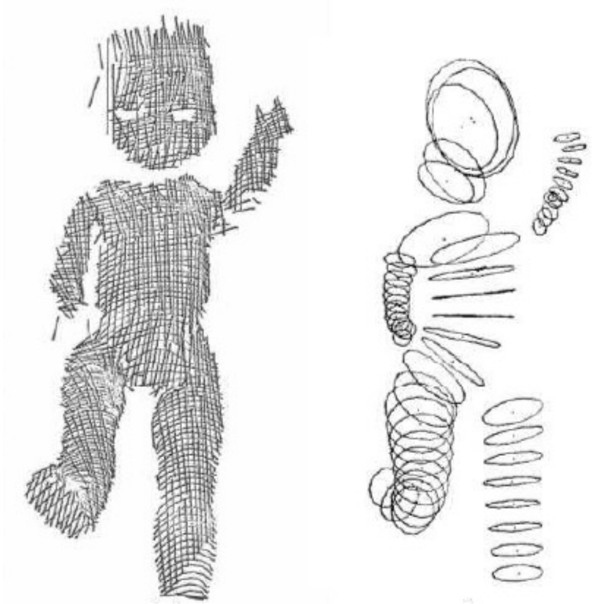
Baumberg and Hogg, Learning Flexible Models from Image Sequences, ECCV '96

The problem…



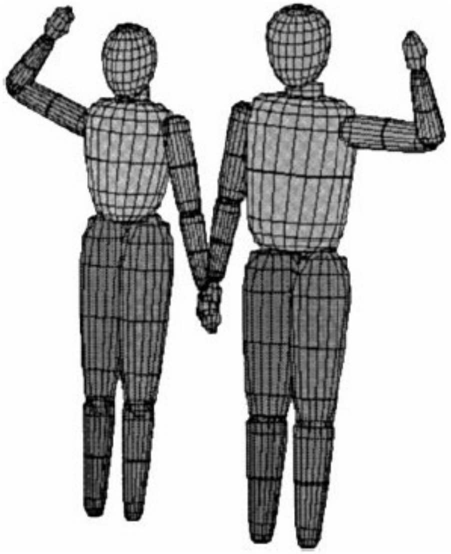
- We don’t look like this.
- Models don’t match the data.
- Systems using such models tend to be brittle.
- We argue that we need a better model of human shape and motion.

Early body models

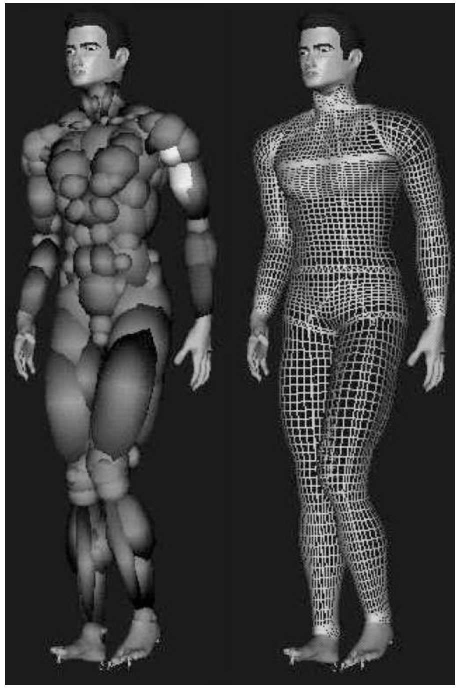


Nevatia & Binford '73

Terzopoulos
and Metaxas '93

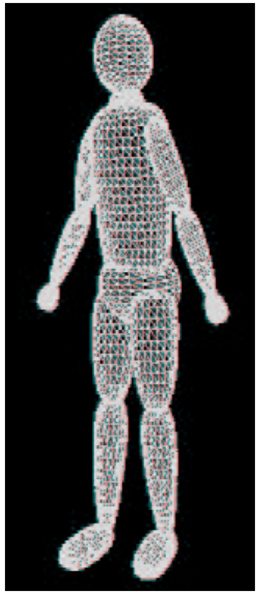


Gavrilla, '96



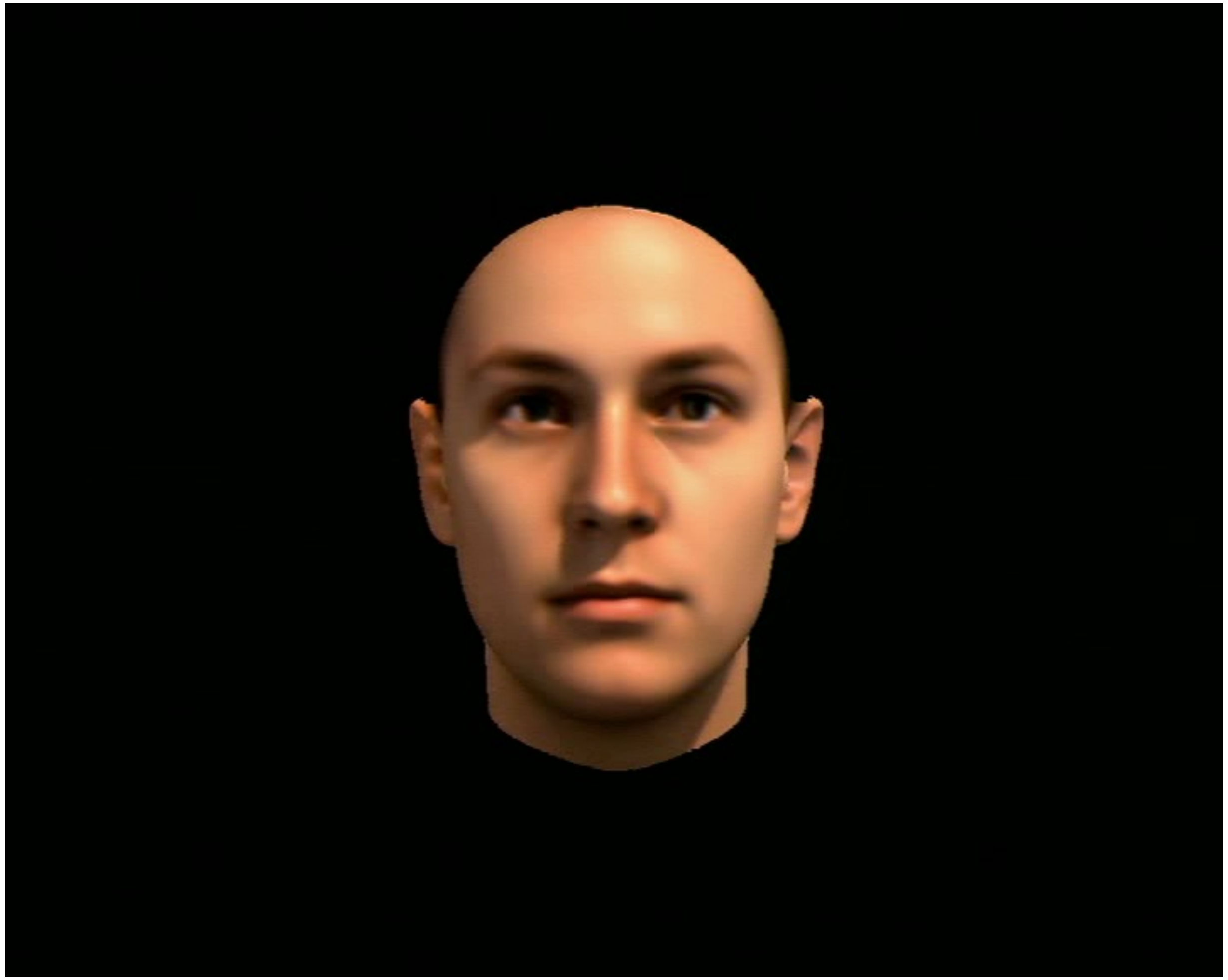
Plänkers and Fua '01

Kakadiaris and Metaxas '00



Sminchisescu
and Triggs '03

The breakthrough started with the face



Blanz & Vetter, A Morphable Model for the Synthesis of 3D Faces, SIGGRAPH 1999

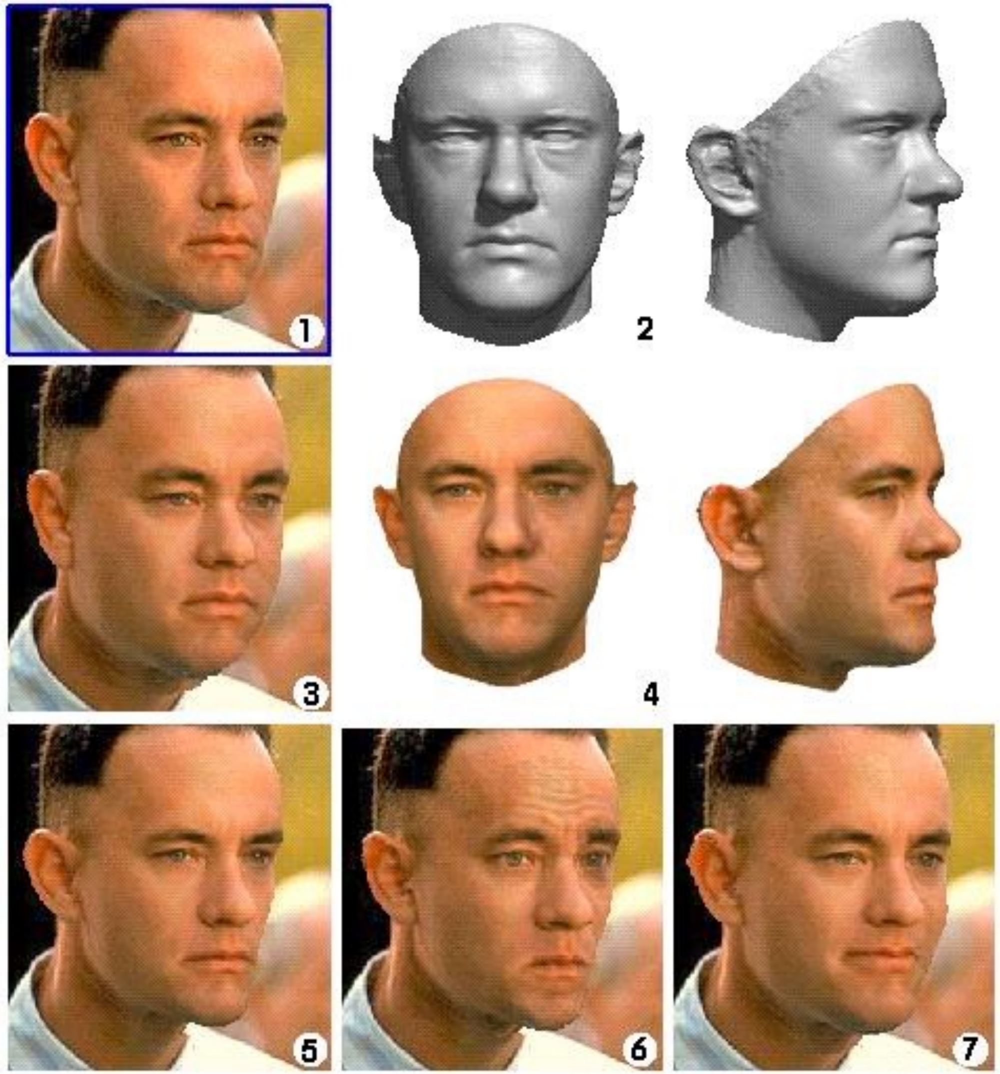
Face scanner

Cyberware scanner

Idea:
Scan faces and learn a statistical model of shape and appearance.

1989 – first 3D body scan

Inverse graphics



Blanz & Vetter, A Morphable Model for the Synthesis of 3D Faces, SIGGRAPH 1999

Let’s do that for bodies!

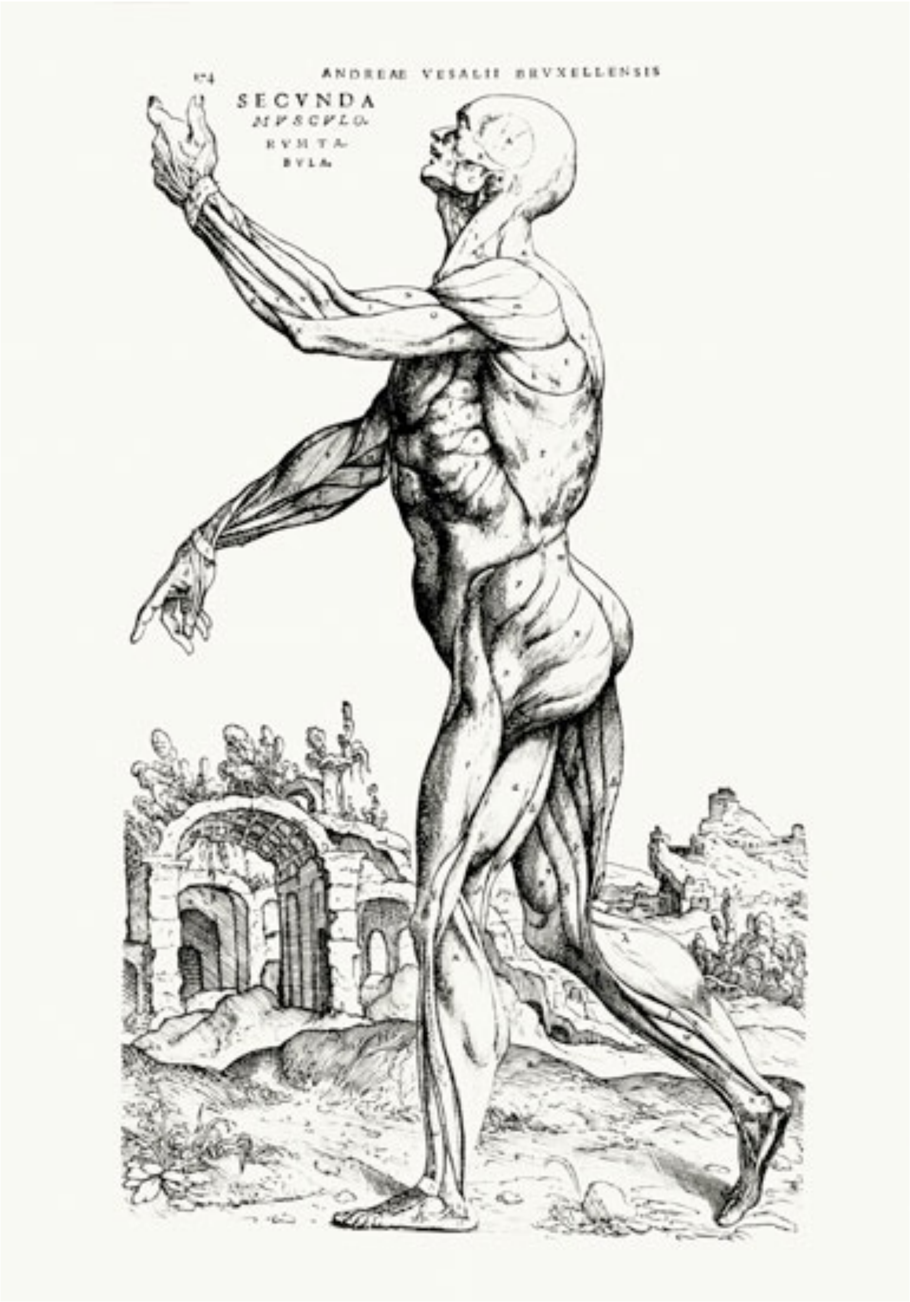
Why is it hard?

The body has about
600 muscles,
200 bones,
200 joints, and
many types of joints.

We also bulge, breath, flex, and jiggle.

Our shape changes with our age, our fitness level, and what we had for lunch.

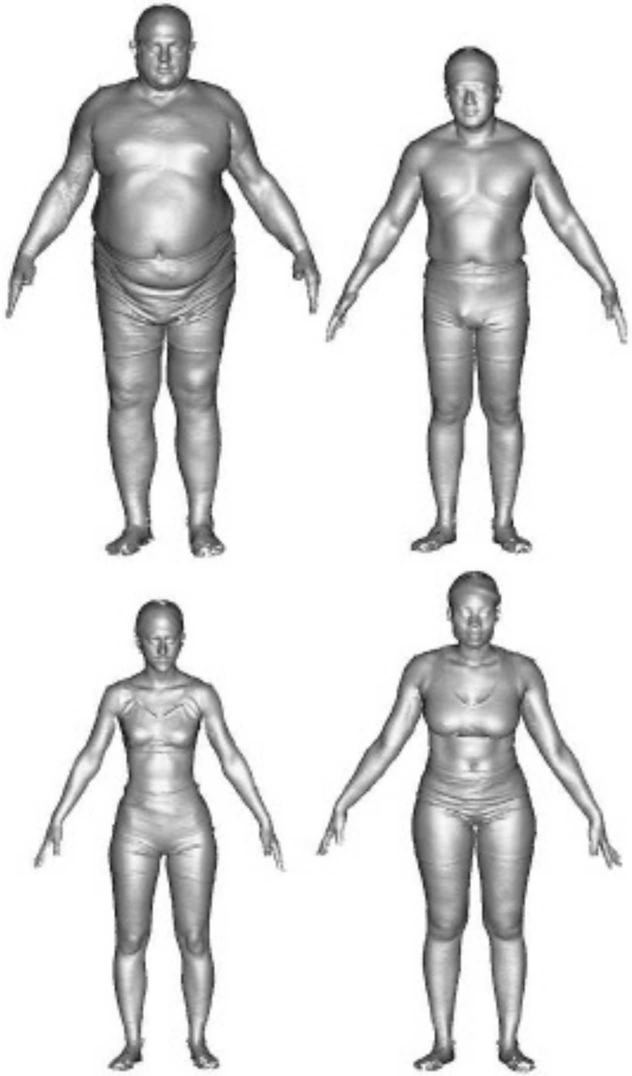
Approach: model only what we can see – the surface.



ANDREAS VESALIUS, Musculature Structure of a Man, c. 1543.

Learning a body model

Cyberware



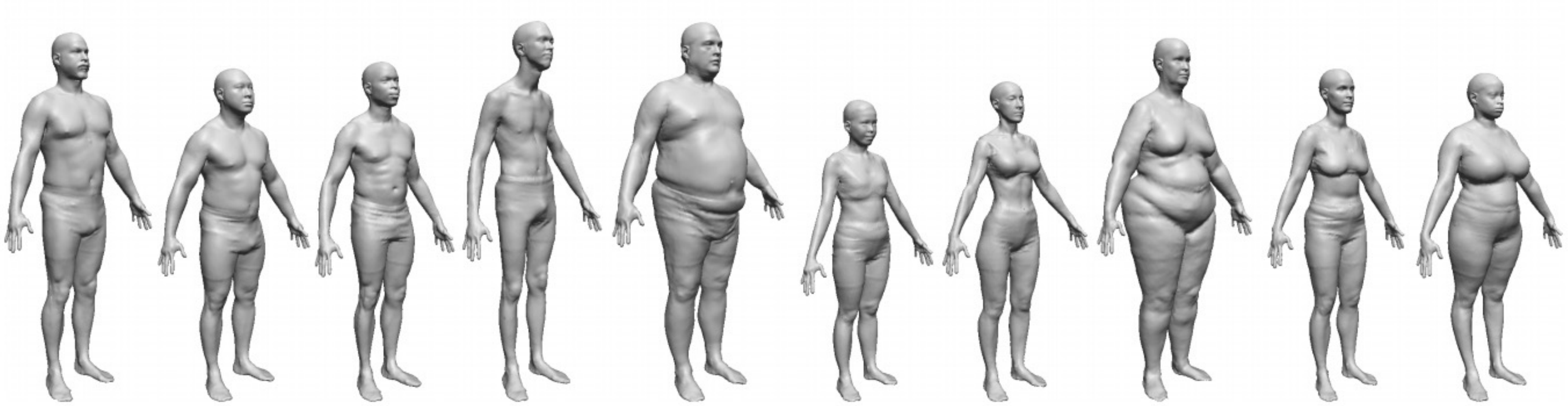
CAESAR dataset – 1999-2001.

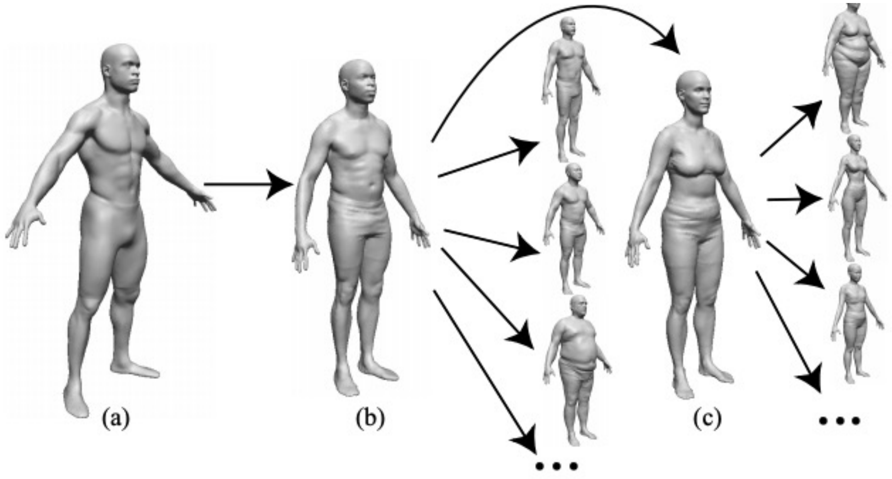
Based on 1990 US census data.

2000 men and 2000 women from the US and Europe.

Pioneers: Allen et al.

CAESAR scans

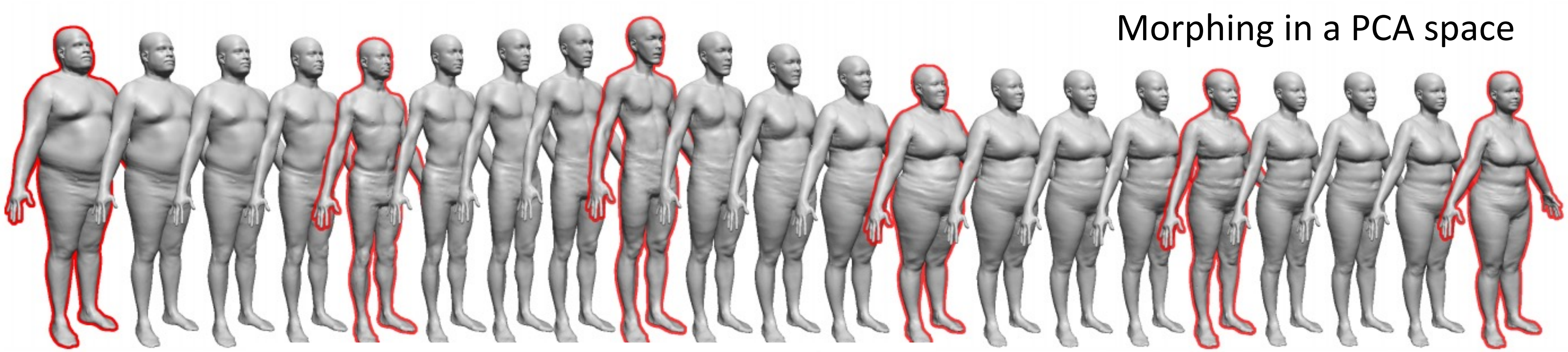




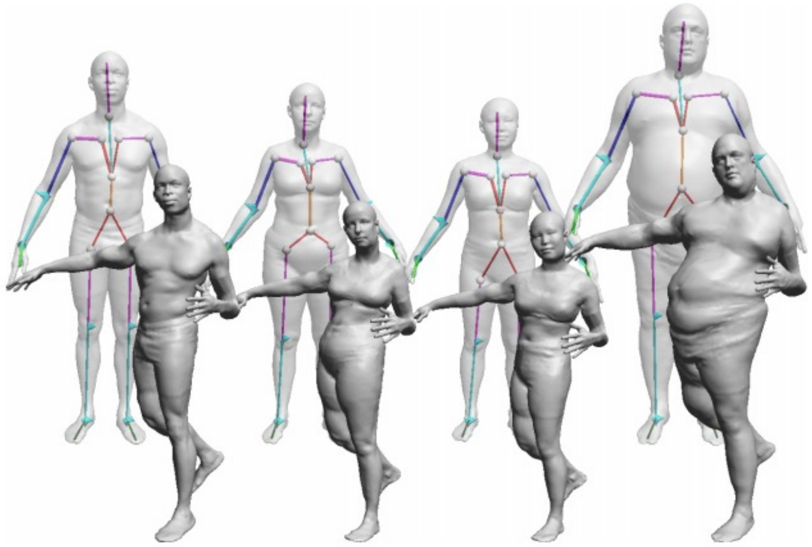
Register a common mesh template to the scans. This was hard and they didn't register many.

The space of human body shapes: reconstruction and parameterization from range scans, Allen, Curless, and Popovic, SIGGRAPH, 2003.

Pioneers: Allen et al.



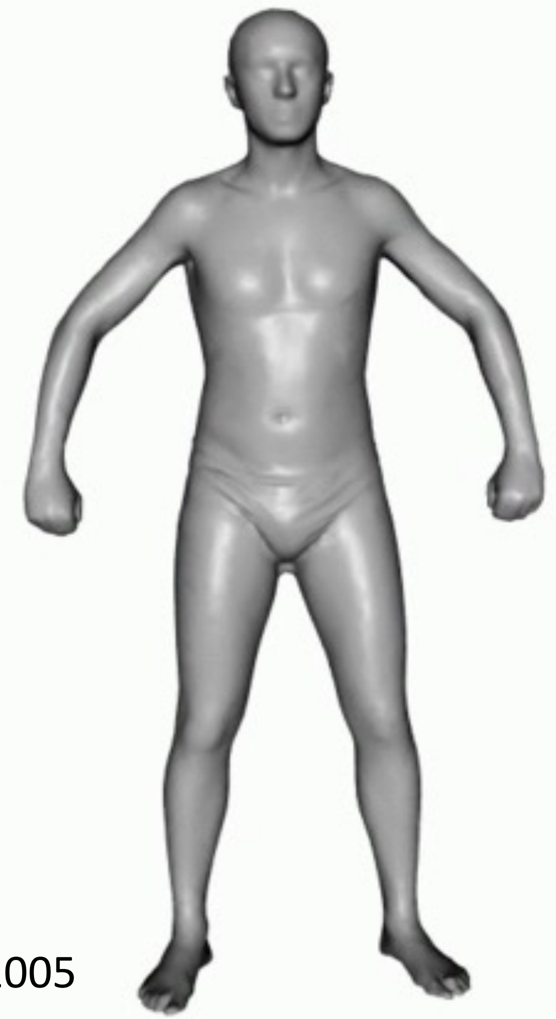
Morphing in a PCA space



Rigging and animating. Lacked realism because there were no pose-dependent deformations.

The space of human body shapes: reconstruction and parameterization from range scans, Allen, Curless, and Popovic, SIGGRAPH, 2003.

Learning body models (2003-2013)



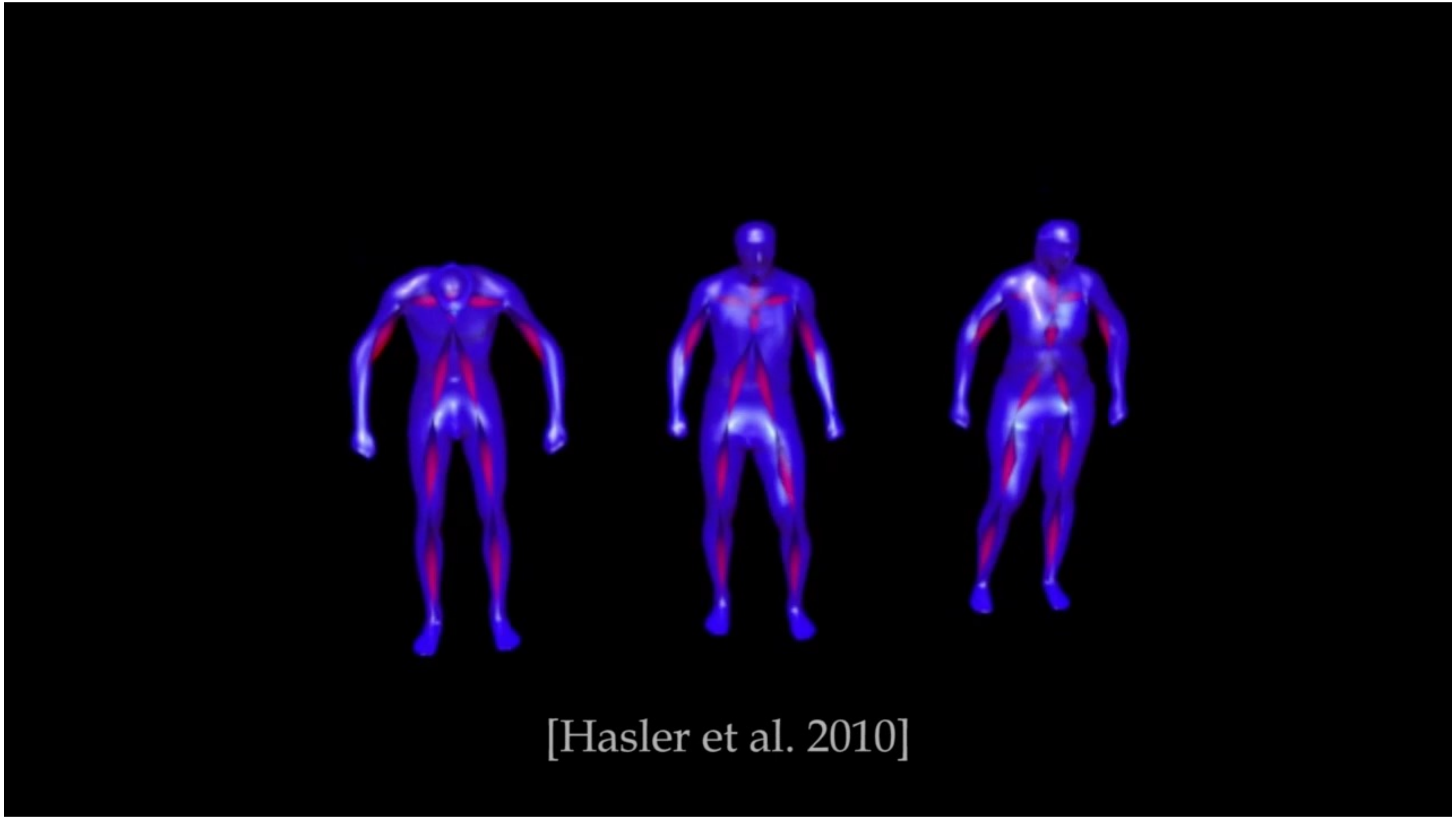
First to combine static scans of several people with scans of one person in many poses.

Based on *triangle deformations*.

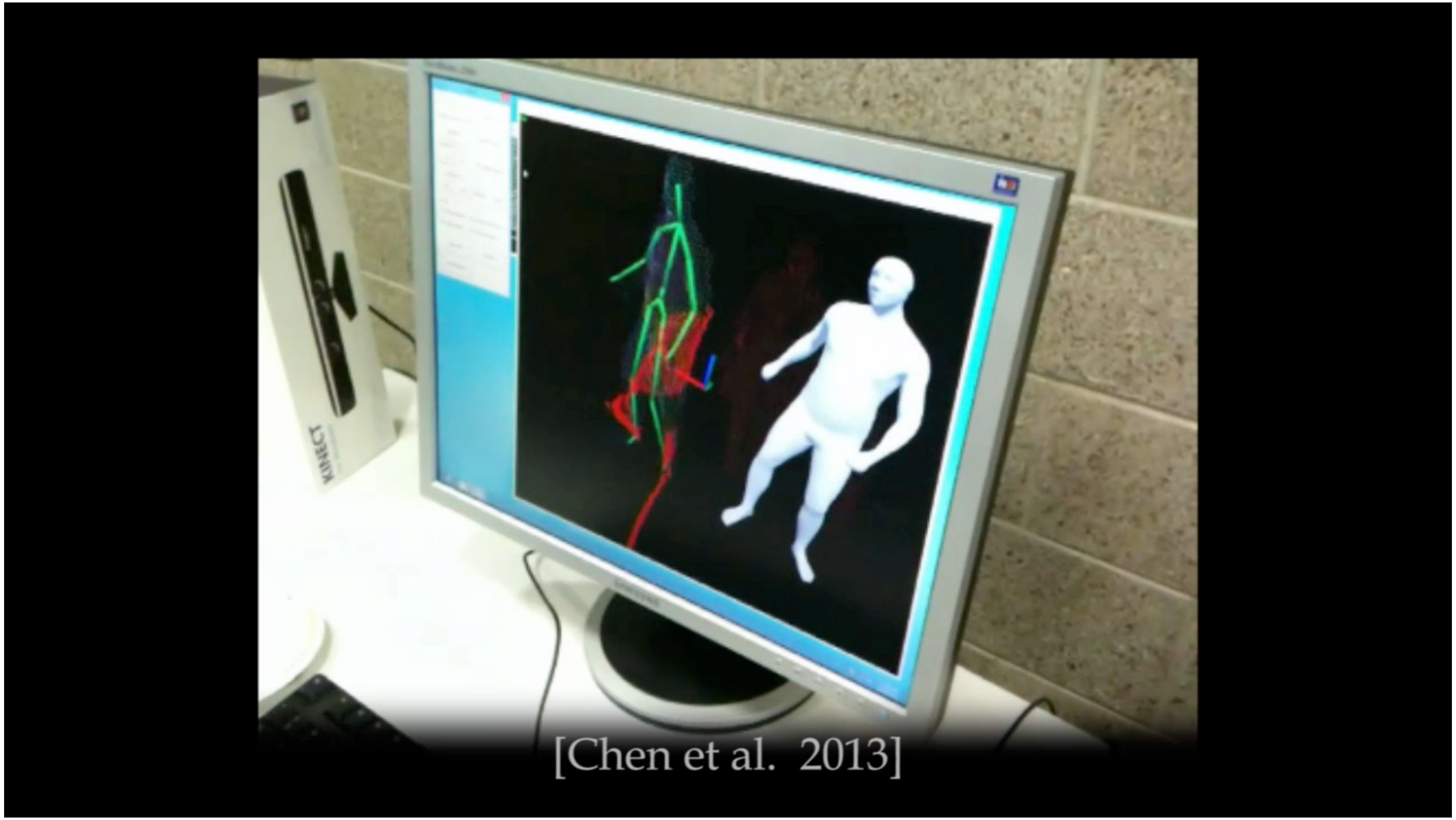
Anguelov et al., SCAPE, 2005

Learning body models (2003-2013)

Learning body models (2003-2013)



Learning body models (2003-2013)



“Tenbo”

Subject specific body models (~2010)

Rigged Subject Scan

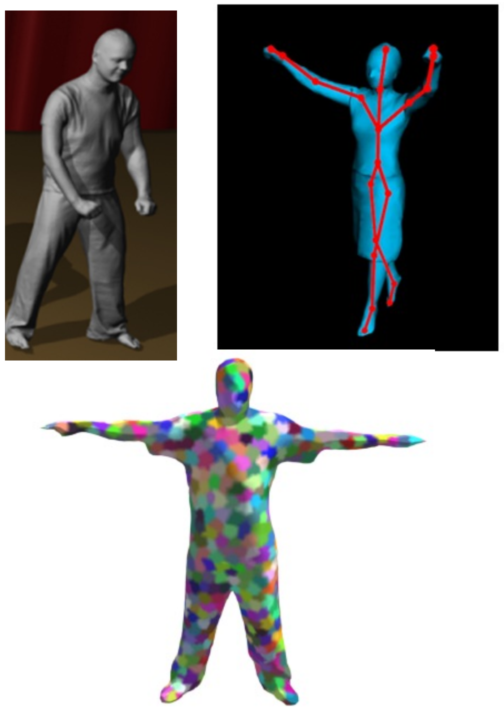
~ 30 DoF

- Kinematic model

Pons-Moll et.al.
Rosehnahn et.al.
Hasler et.al.

Free form Surface

- > 1000 DoF
- with ++ constrains



Aguiar et.al.
Gall et.al
Cagniart et.al.