Estimating Human Motion: Past, Present, and Future

Michael J. Black

Max Planck Institute for Intelligent Systems

October 2018

Note

- This is an annotated version of an invited talk I gave at GCPR 2018 addressing the theme “40 years of DAGM”
- It includes a bibliography at the end with links to all papers cited in the talk.
- It is my personal view of the evolution of human motion analysis from video.
- I’ve been working on human motion since 1993 so I only have 25 years of hands-on experience but I look back 40 years.
- I highlight papers that changed how I thought at the time.
- This is not a full review of the literature – it is my personal, and biased, view of it.




SVR

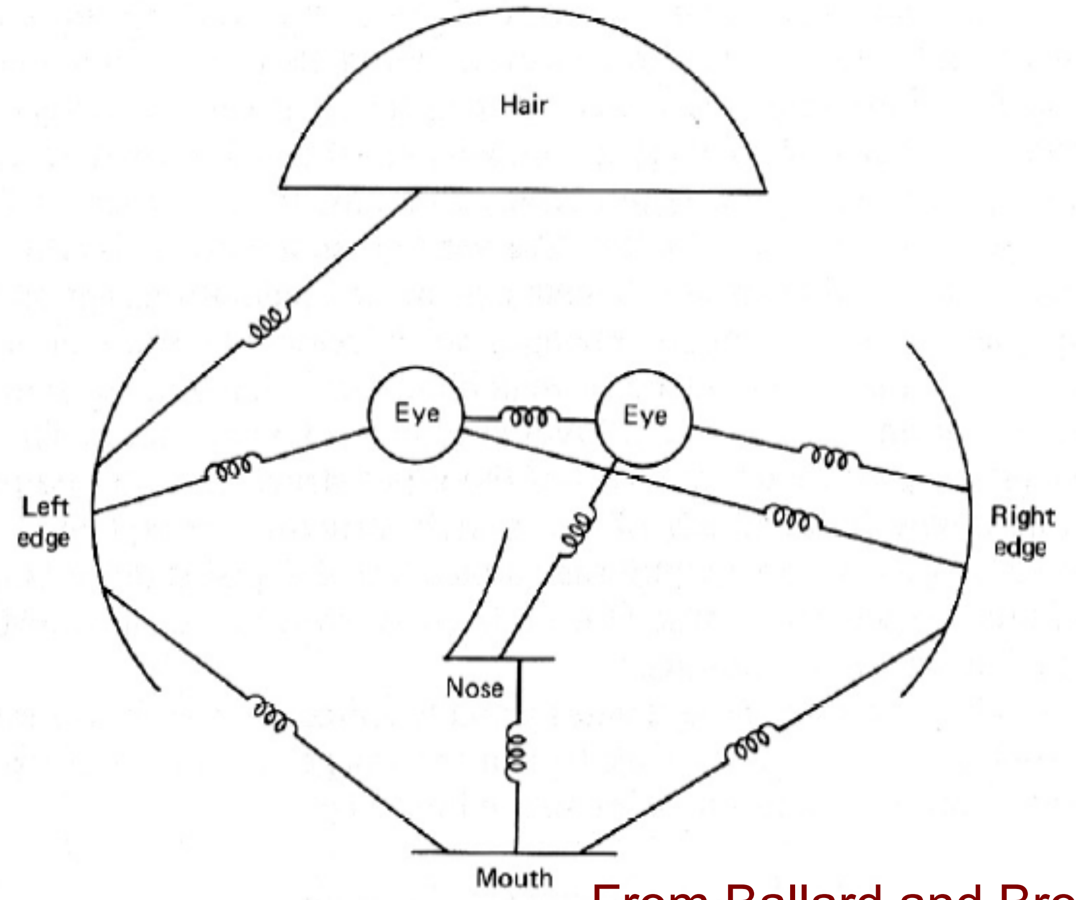









Graph-based models of bodies



From Ballard and Brown

Pictorial structures – Fischler and Elschlager '73

The beginning: 42 years ago

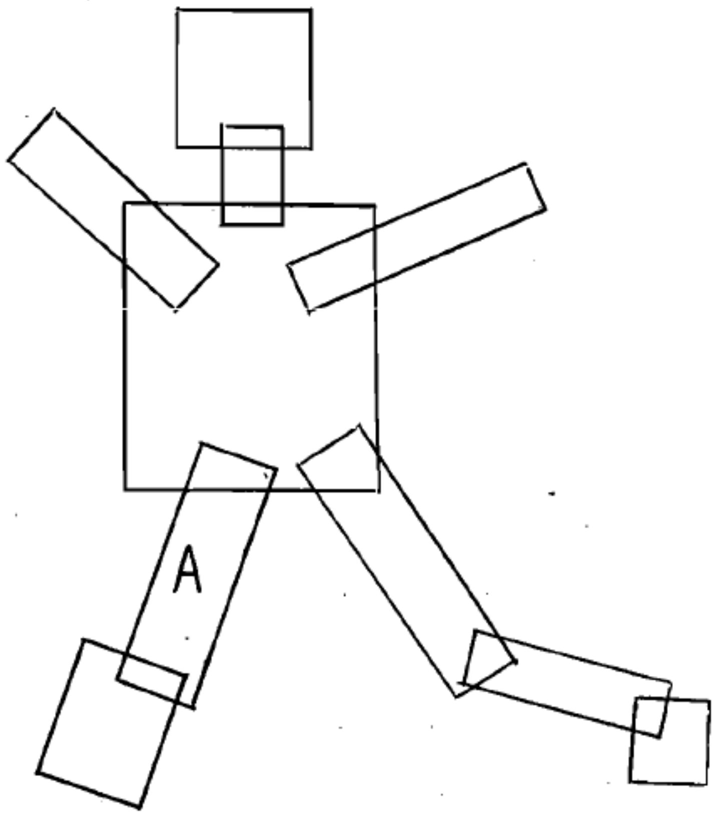


Figure 4. Relaxation picks out the interpretation of A as a thigh even though a calf is a locally better alternative.

G. E. Hinton. Using relaxation to find a puppet. In Proc. of the A.I.S.B. Summer Conference, pages 148–157, July 1976. His first paper!

G.Hinton
Cognitive Studies Program
University of Sussex, Brighton

USING RELAXATION TO FIND A PUPPET

ABSTRACT

The problem of finding a puppet in a configuration of overlapping, transparent rectangles is used to show how a relaxation algorithm can extract the globally best figure from a network of conflicting local interpretations.

INTRODUCTION

The program takes as input the co-ordinates of the corners of some overlapping, transparent rectangles (See figure 1). The problem is to find the best possible instantiation of a model of a puppet. The difficulty is that if we only consider a rectangle and its overlapping neighbours, then each rectangle could be several different puppet parts or none at all, so local ambiguities have to be resolved by finding the best global interpretation. The aim of this paper is to show how a relaxation method can be used instead of the obvious search through the space of all combinations of locally possible interpretations. The relaxation method has several advantages:

1. Using parallel computation the best global interpretation can be found quickly. The time taken is not exponential in the number of local possibilities because combinations are not dealt with explicitly.

2. The computing space required increases only linearly with the number of possibilities, which makes this method better than an exhaustive, breadth-first parallel search, for which there is a combinatorial explosion in space.

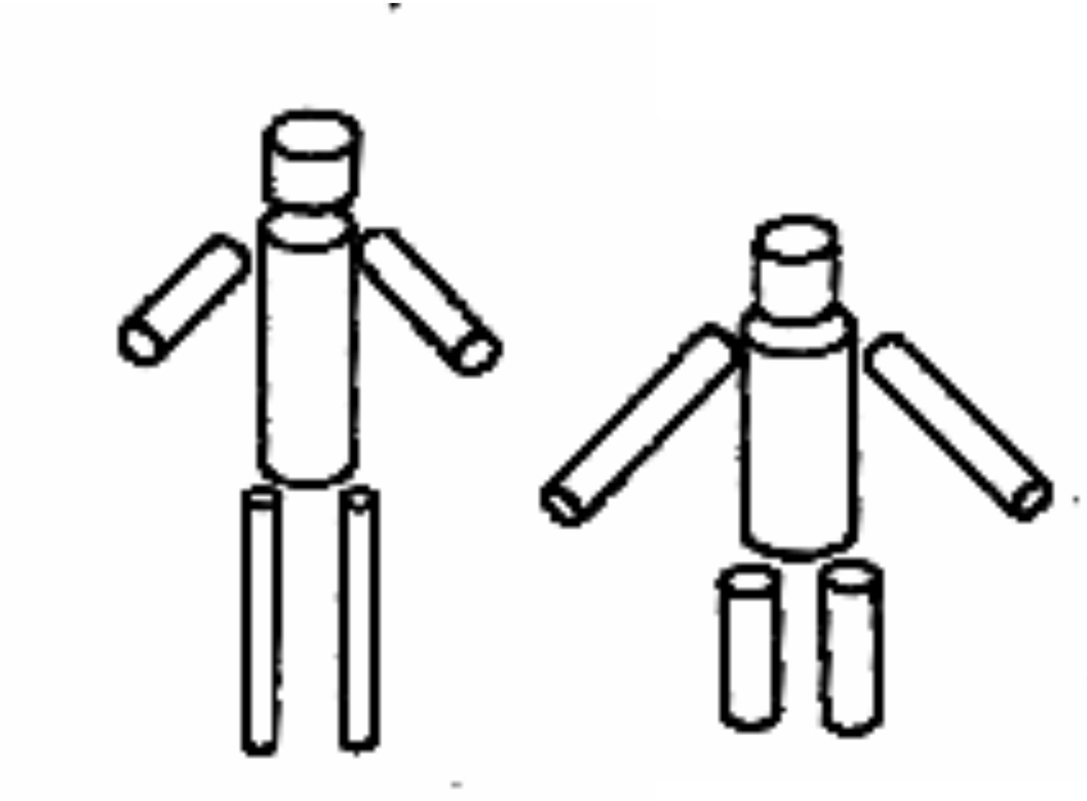
3. It produces the best global interpretation, not just a good one as in heuristic search.

All these reasons make relaxation look good as a model of how the brain resolves conflicting low-level visual hypotheses. A conventional, serial A.I. search would be very slow, given the brain's sluggish hardware (Sutherland 1974).

THE PUPPET MODEL

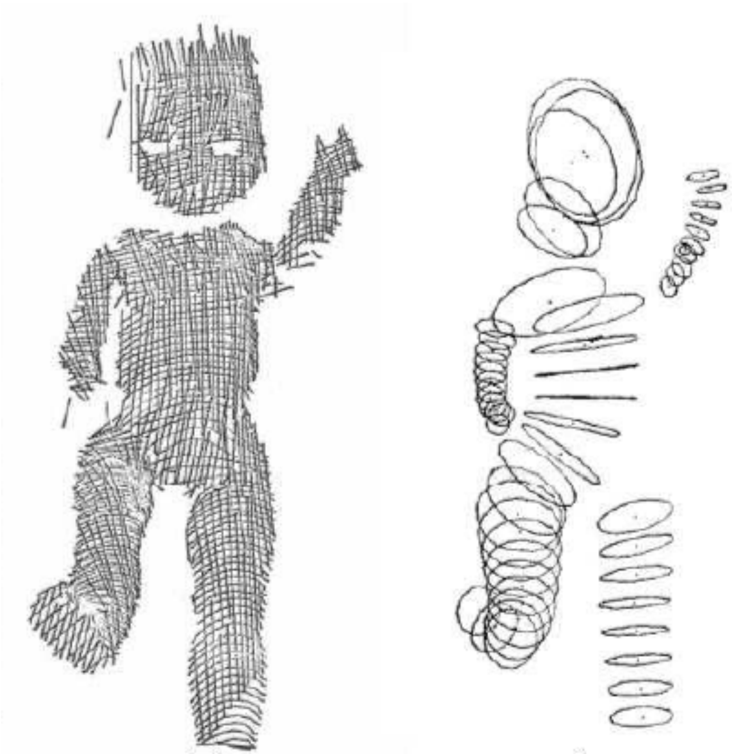
The puppet, which is always depicted in side view, consists of fifteen rectangular parts having the following properties and

The early history was 3D



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

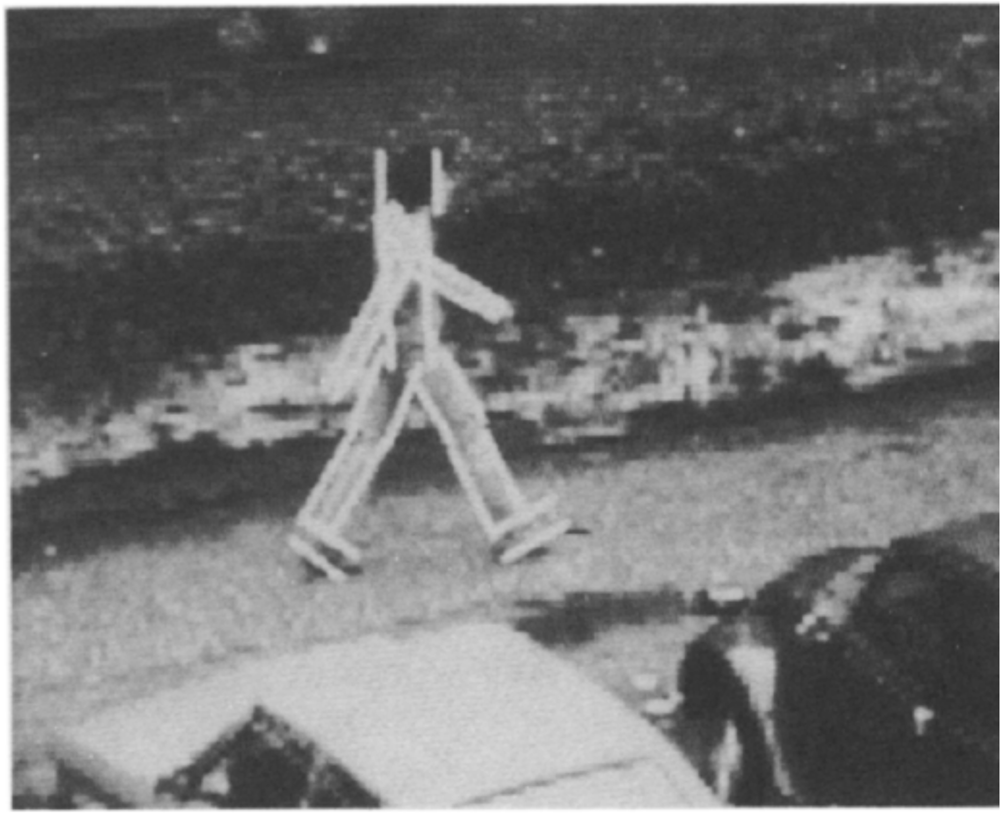


Figure 12. Set of lines which correspond to the imag projections of occluding surfaces. They represent the image in Figure 4

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

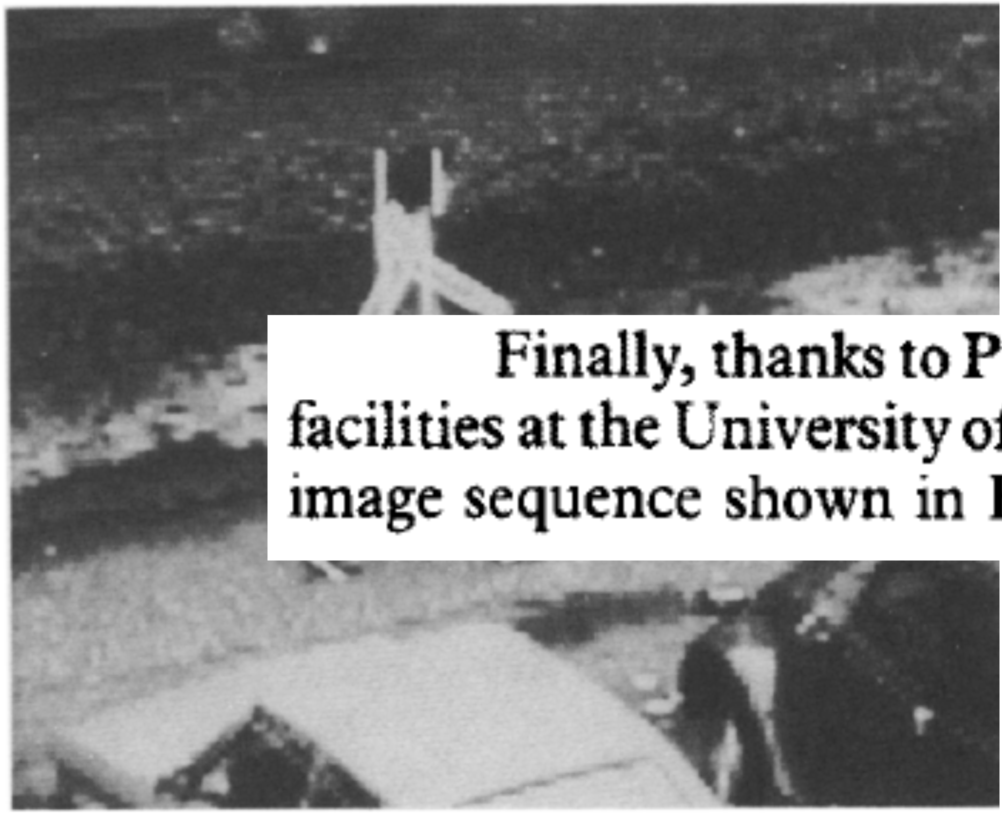


Figure 12. Set of lines which correspond to the imag[e] projections of occluding surfaces. They represent the image in Figure 4

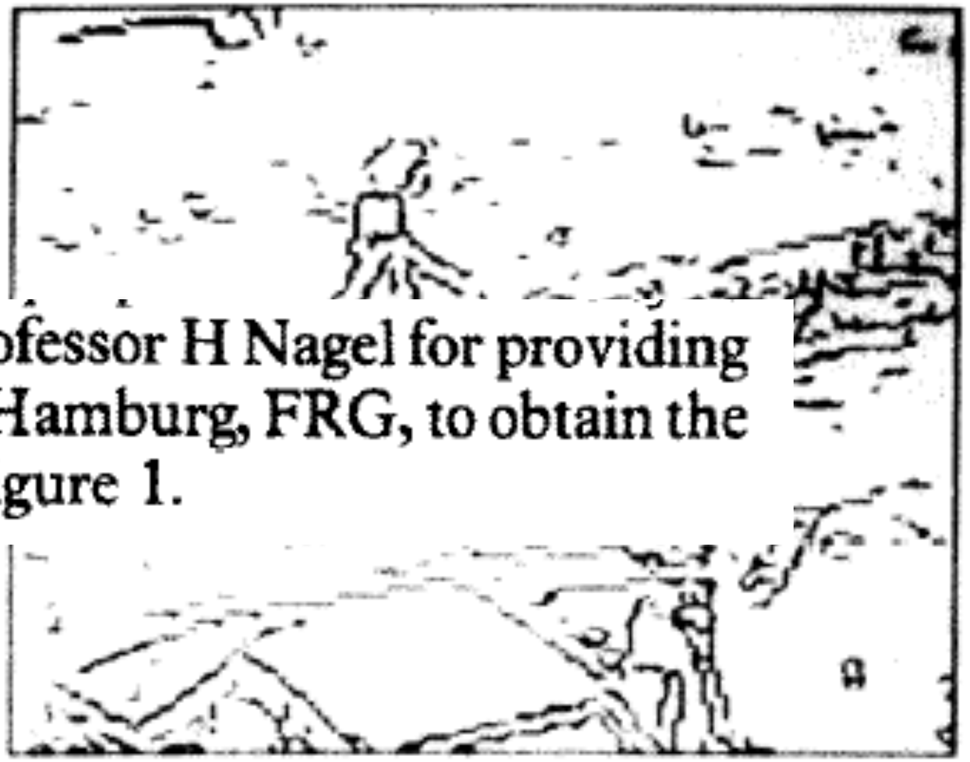
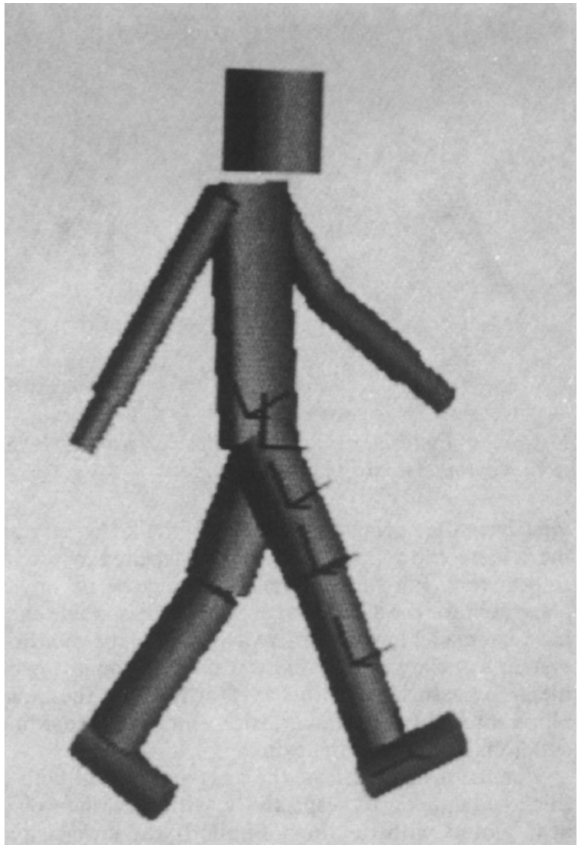


Figure 5. Edge-finding operation applied to the image in Figure 4

Finally, thanks to Professor H Nagel for providing facilities at the University of Hamburg, FRG, to obtain the image sequence shown in Figure 1.

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

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
            c = 0 s = 1

        [liftl]
        posture: [straight]
        position: x = −16 y = 10 z = 0 a = 0 b
            s = 1

class: arm
parts:

    partclass: upper-arm
    weight: 0.5
    position: x = 0 y = −20 z = 0 a = 0 b = 0

    partclass: lower-arm
    weight: 0.5
    position: x = 0 y = −40 z = 0 a = 0 b = [−

class: lower-arm
parts:

    partclass: forearm
    weight: 0.7
    position: x = 0 y = −20 z = 0 a = 0 b = 0

    partclass: hand
    weight: 0.3
    position: x = 0 y = −50 z = 0 a = 0 b = 0

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

The lost decade.

Geometry and optimization: 1994-2004

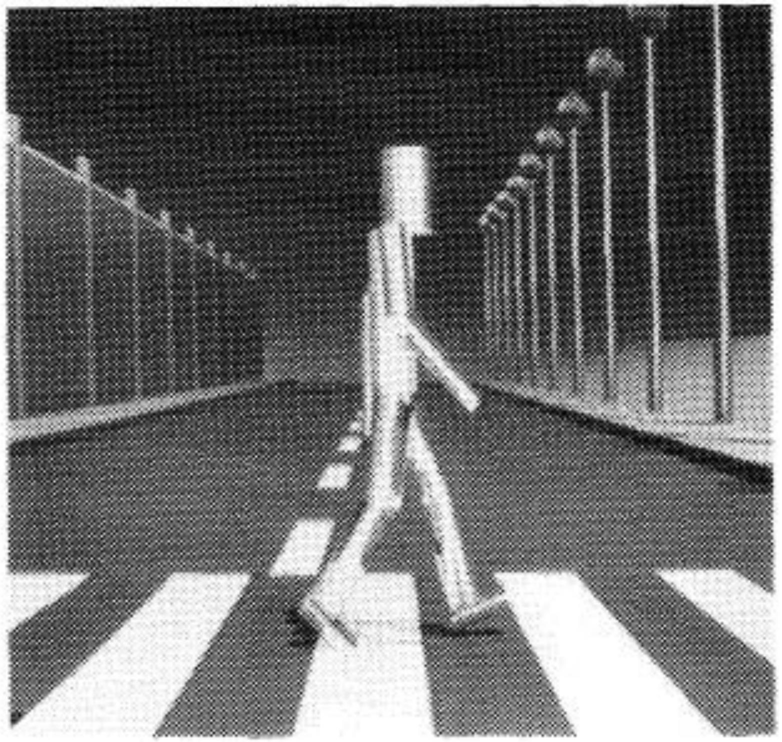


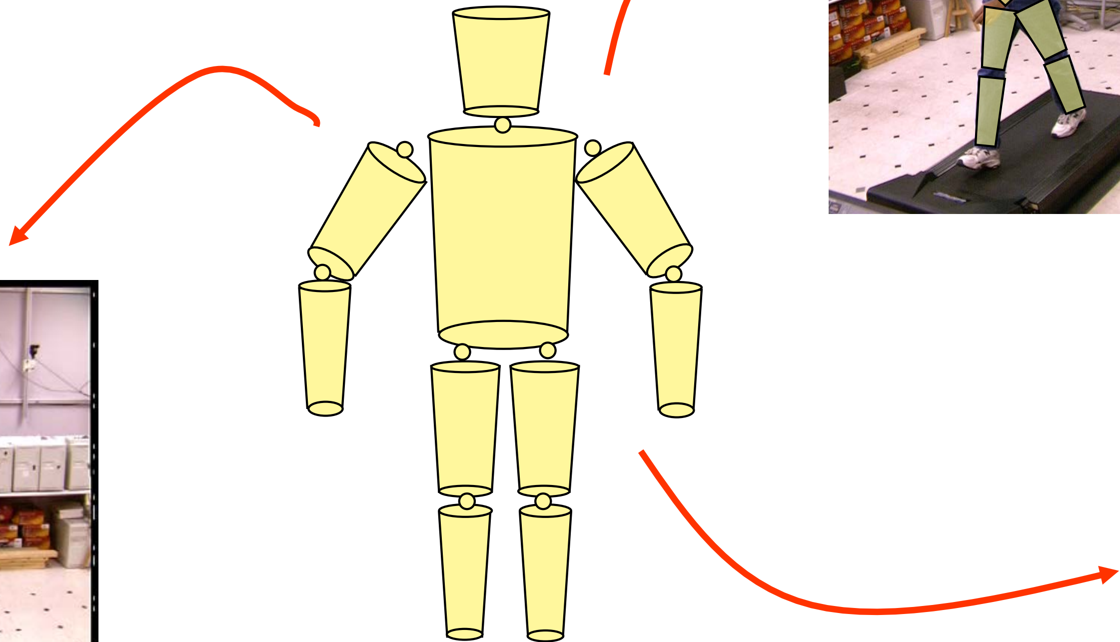
FIG. 4. Model of the human body.

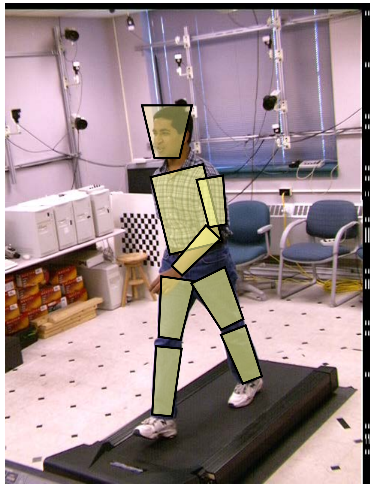
FIG. 20. Determined motion state.

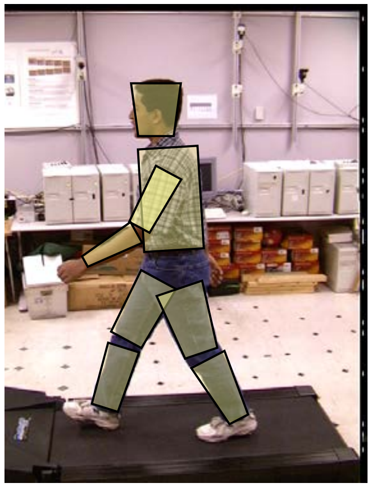
Rohr, Towards Model-Based Recognition of Human Movements in Image Sequences, CVGIP, 1994

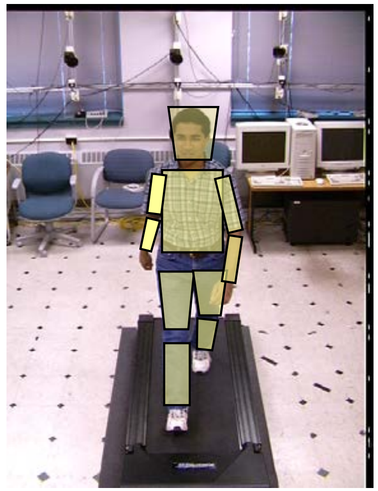
The generative approach

Find the pose θ_t









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

Generative approach

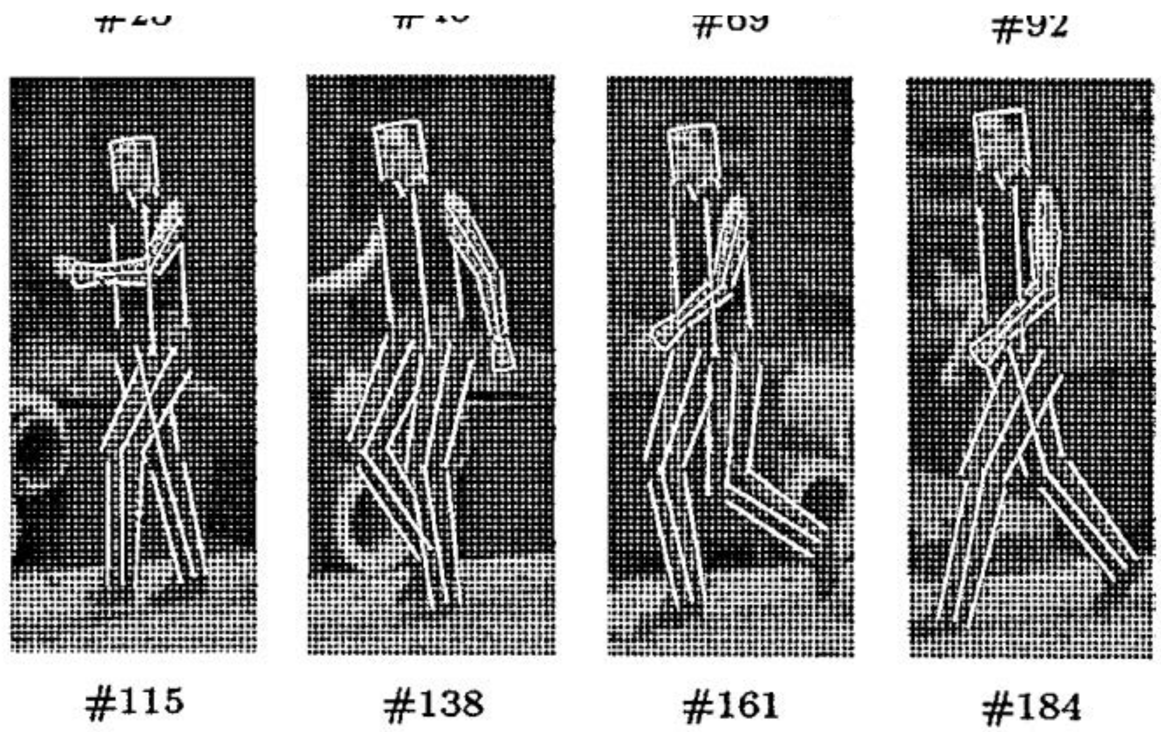
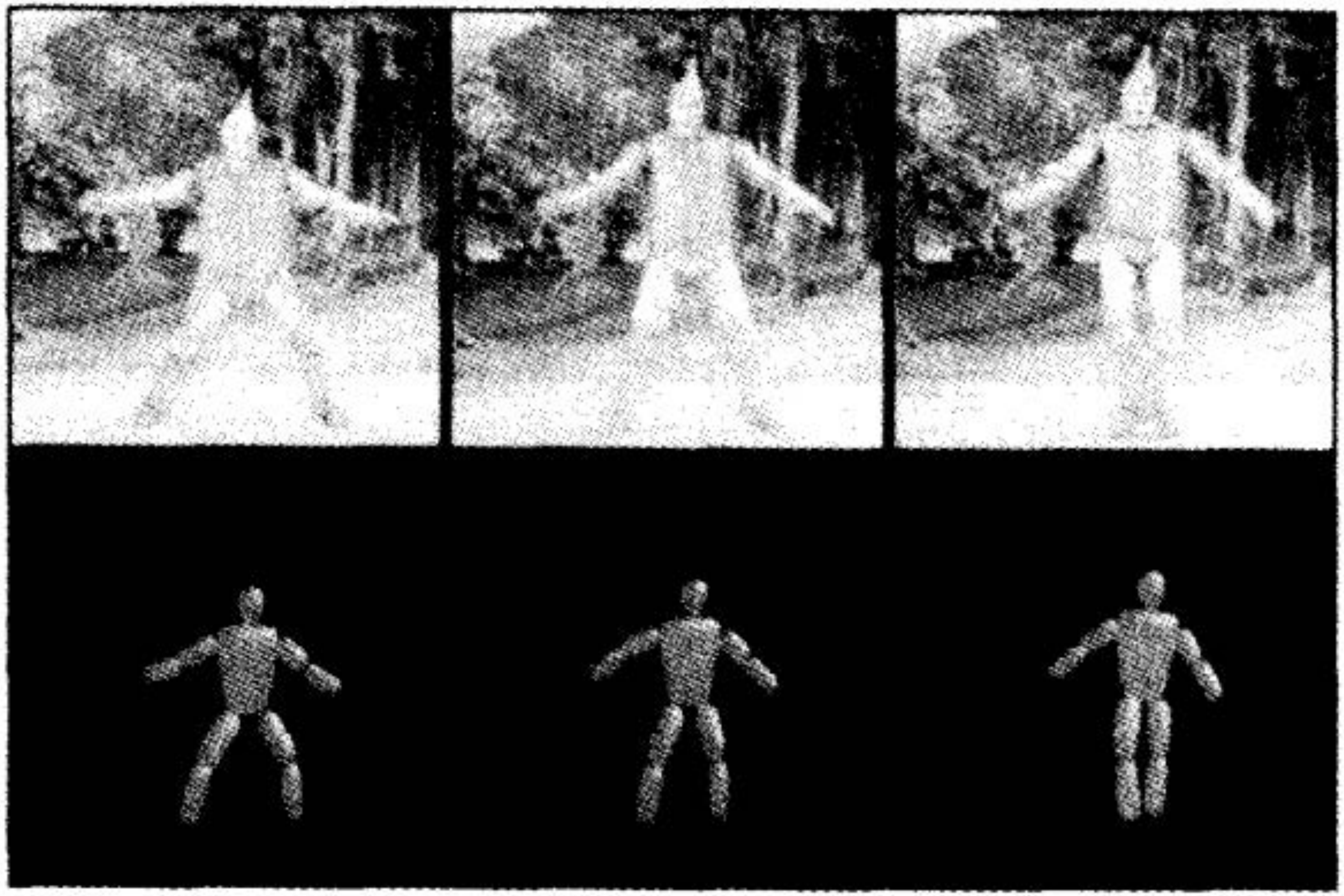


Figure 14. Outdoor walking scene; contours and skeleton are overlaid.

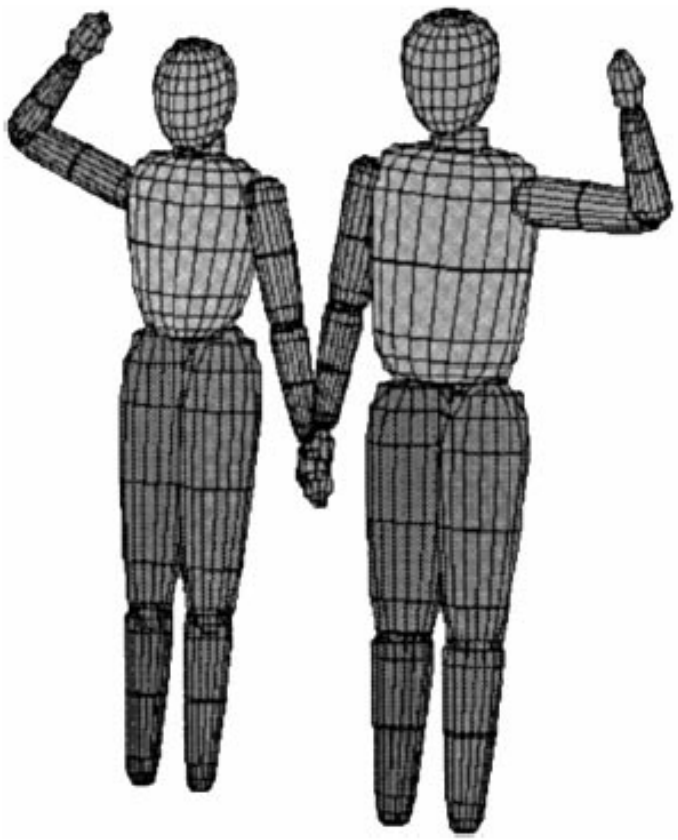
Tracking of persons in monocular image sequences. S. Wachter ; H.-H. Nagel, Proceedings IEEE Nonrigid and Articulated Motion Workshop, 1997

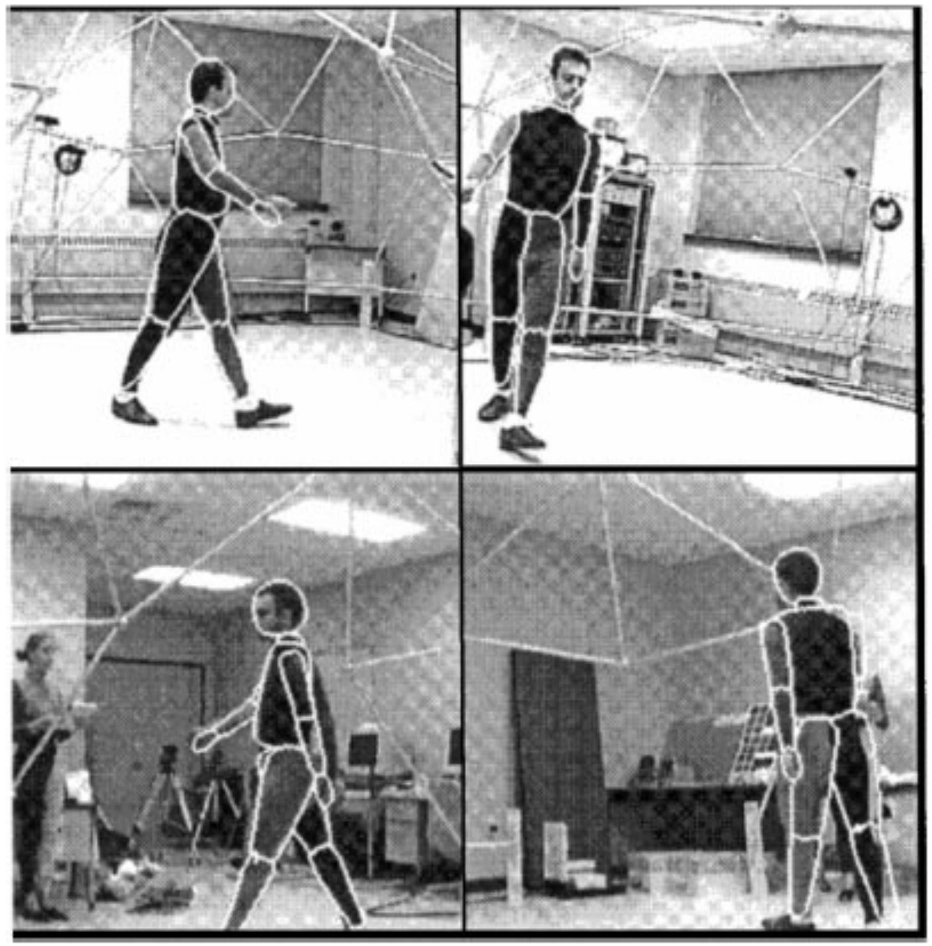
Non-rigid parts



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

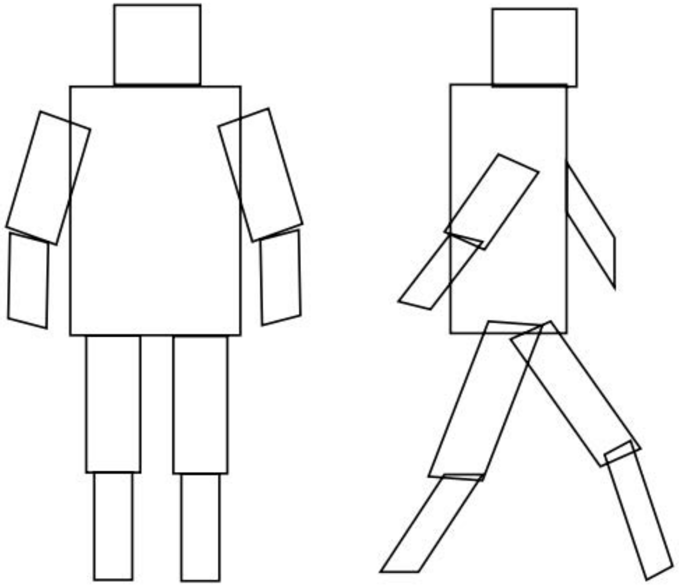
Multi-camera, markerless, mocap

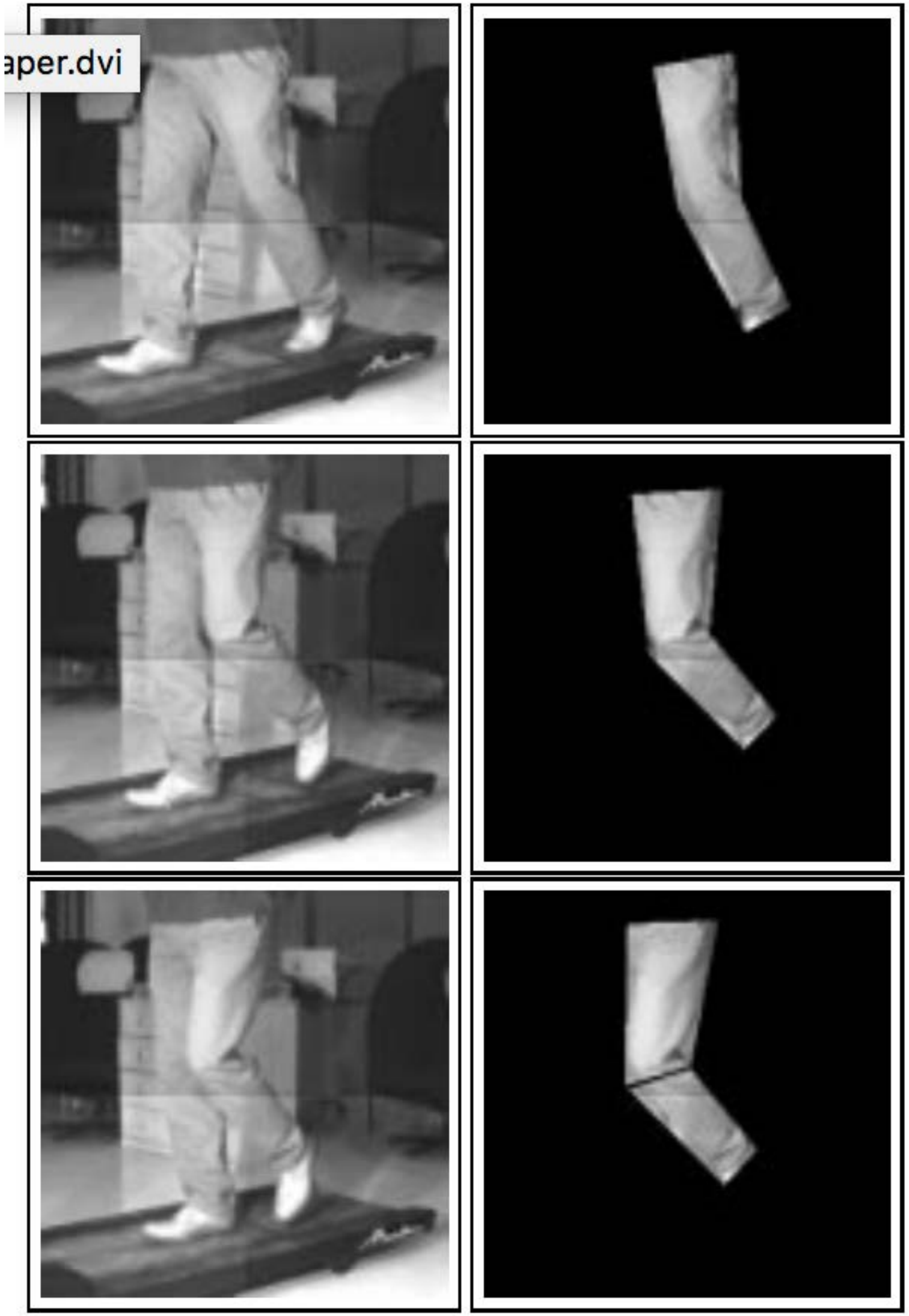




Simple shapes, multi-camera, special clothing.

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



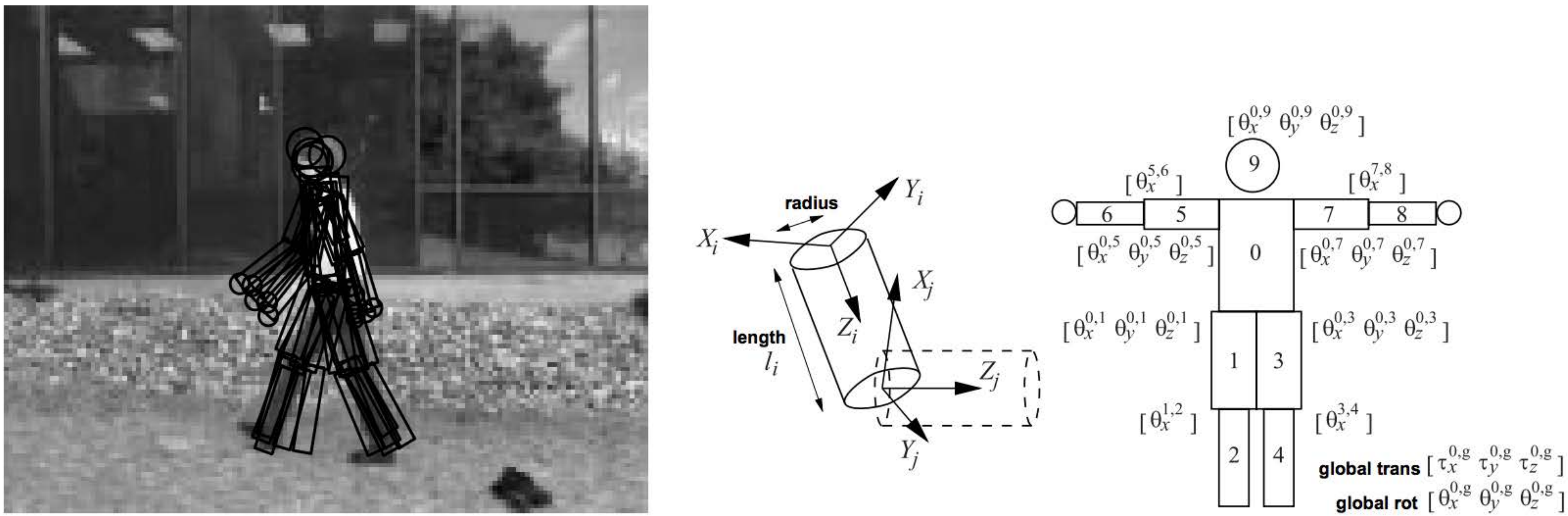


Cardboard people: A parameterized model of articulated motion

Ju, S. X., Black, M. J., Yacoob, Y., Face and Gesture, 1996

Stochastic search
to deal with ambiguity

Represent a distribution over poses

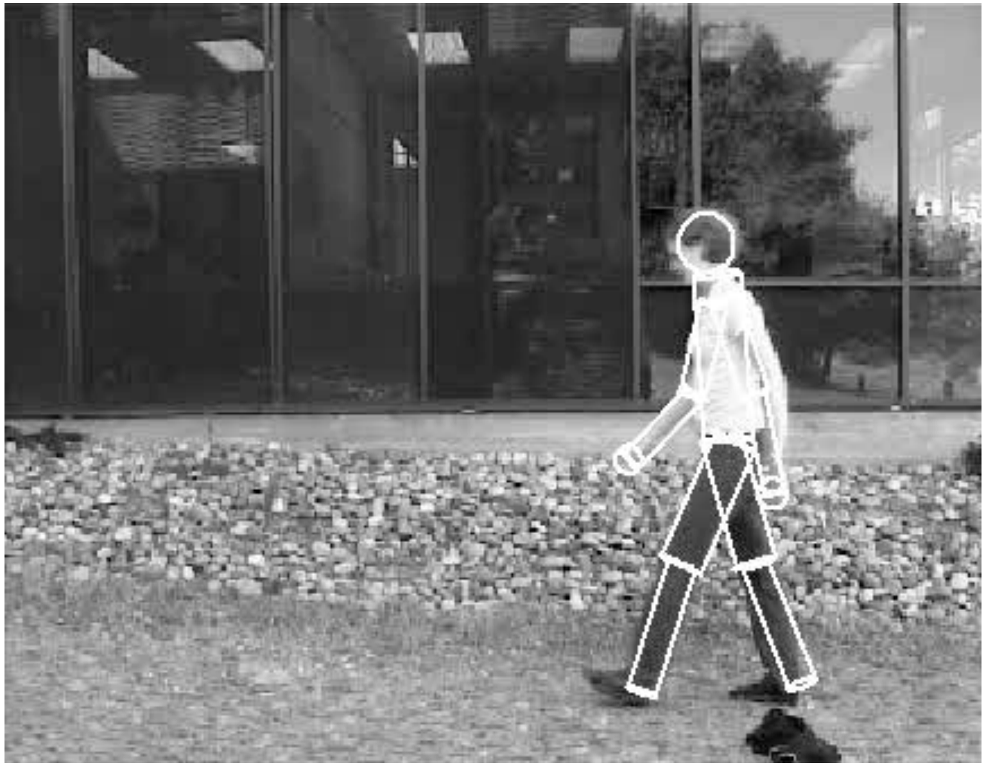
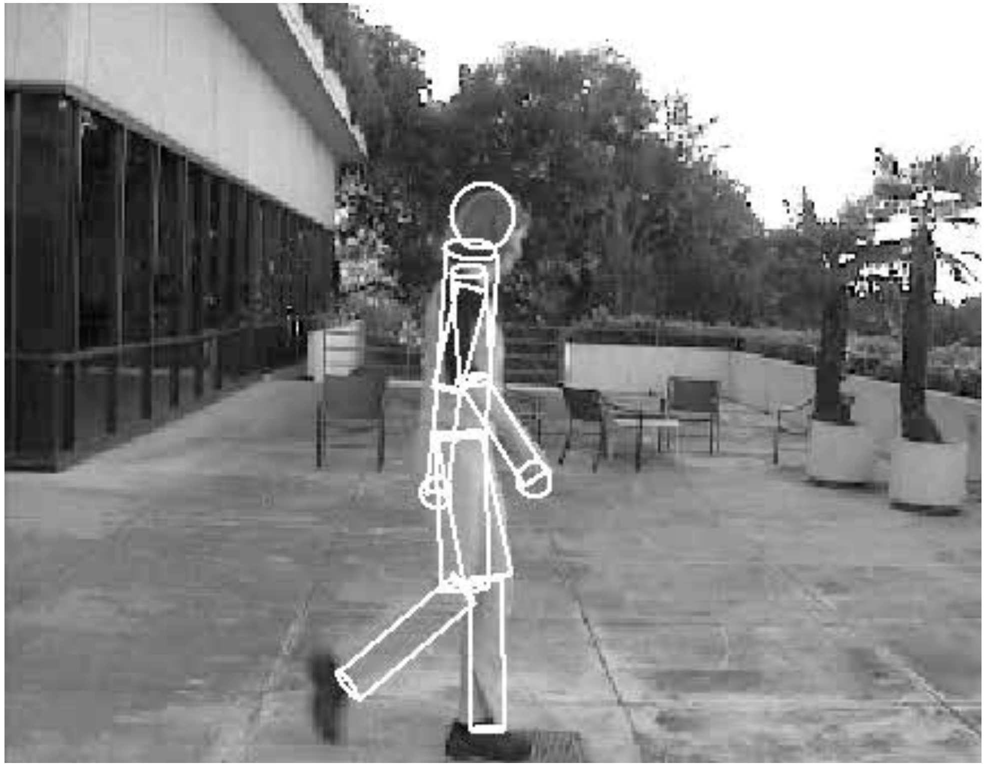


a *b*

- Particle filter to propagate over time

Stochastic tracking of 3D human figures using 2D image motion
Sidenbladh, H., Black, M. J., Fleet, D., ECCV 2000

Represent a distribution over poses

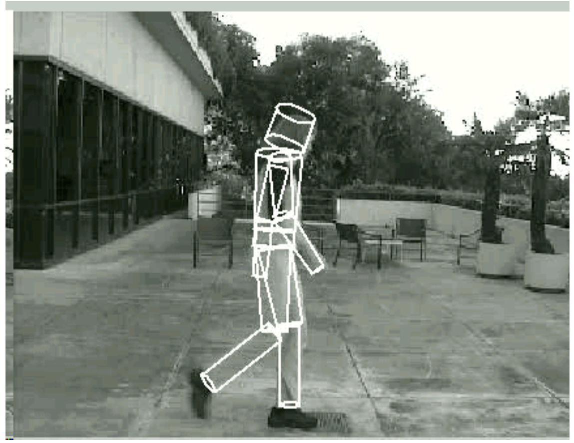


- Particle filter to propagate over time

Stochastic tracking of 3D human figures using 2D image motion
Sidenbladh, H., Black, M. J., Fleet, D., ECCV 2000

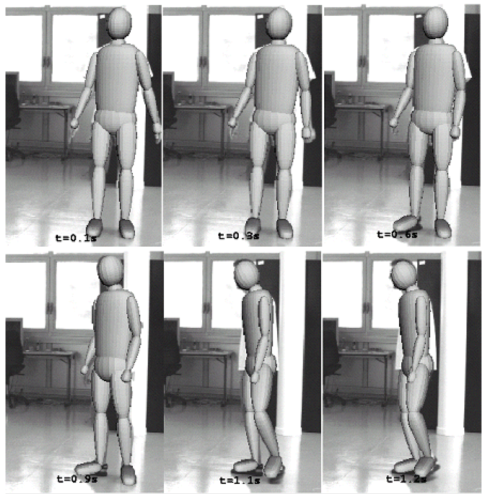
Stochastic search and tracking

Deutscher, North, Bascle, & Blake '99



Sidenbladh, Black and Fleet, '00

Cham and Rehg '99

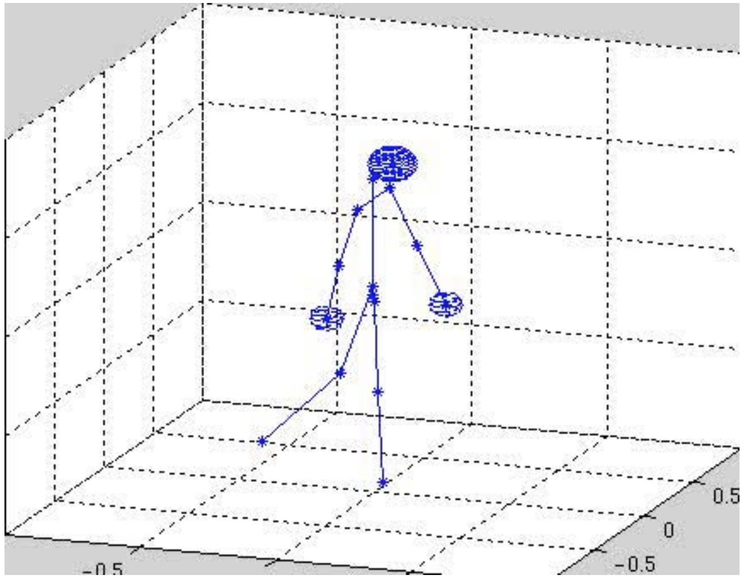


Sminchisescu &Triggs '01

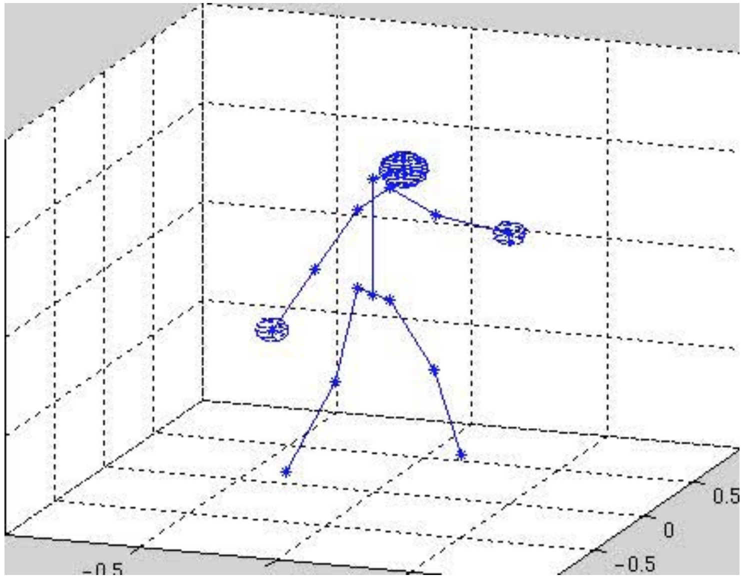
Nothing works.
Add a prior.

Learning and Tracking Cyclic Human Motion
Sidenbbladh & Black, NIPS 2001

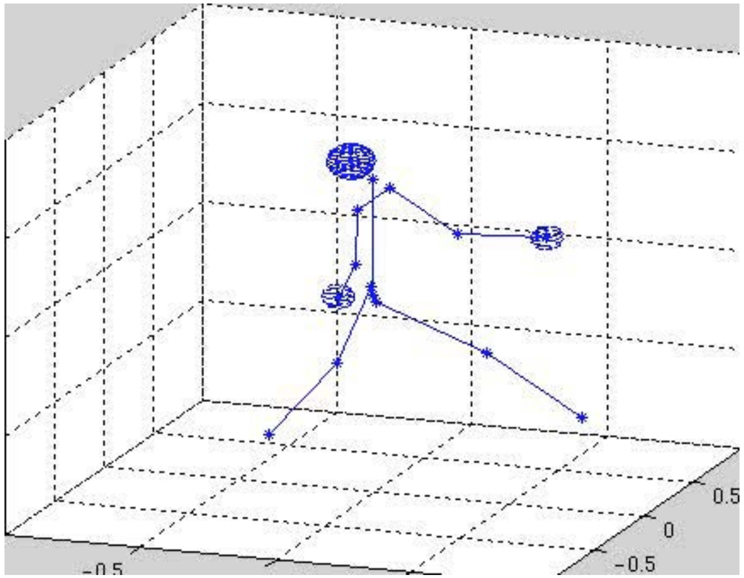
PC1



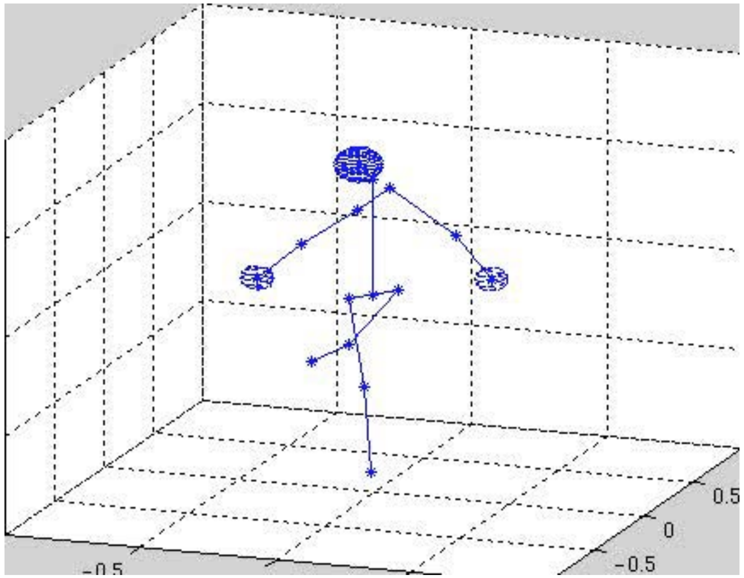
PC2



PC3



PC4



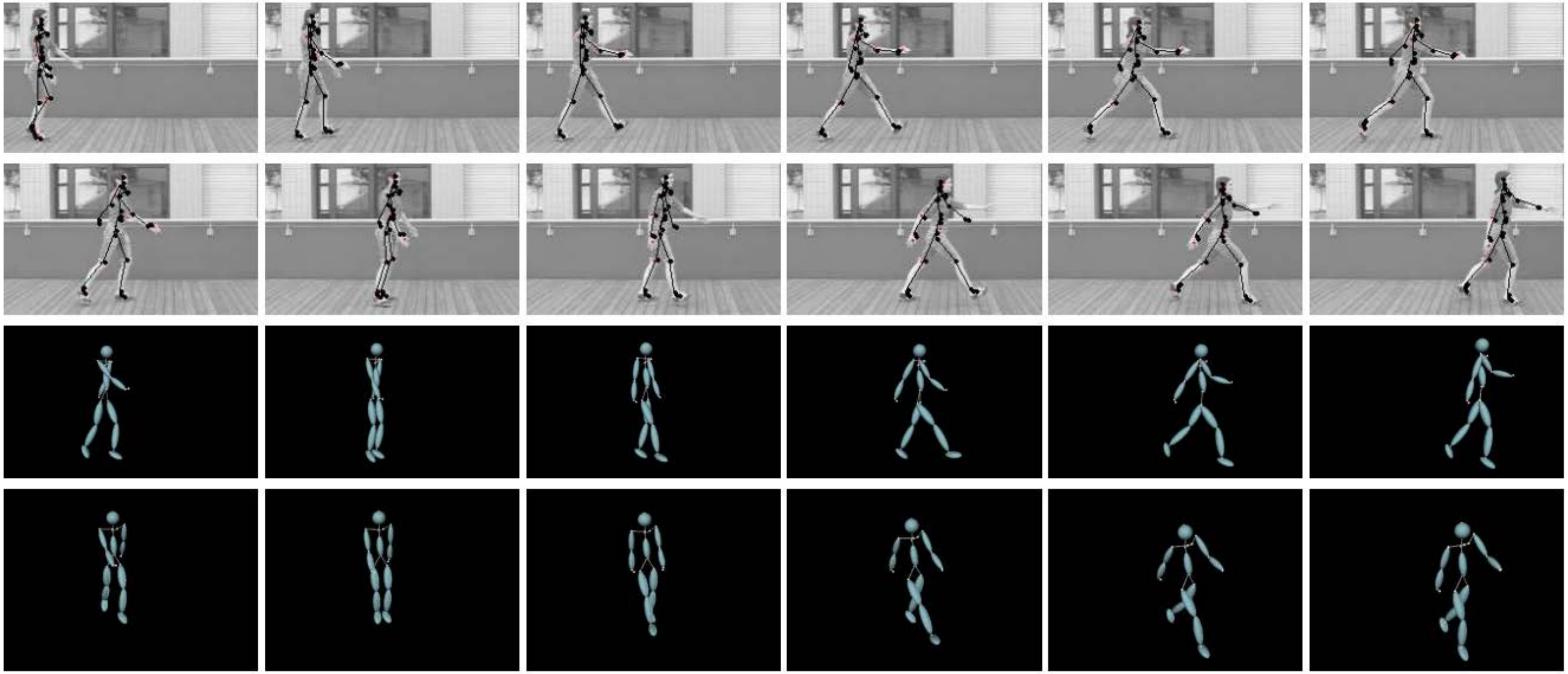
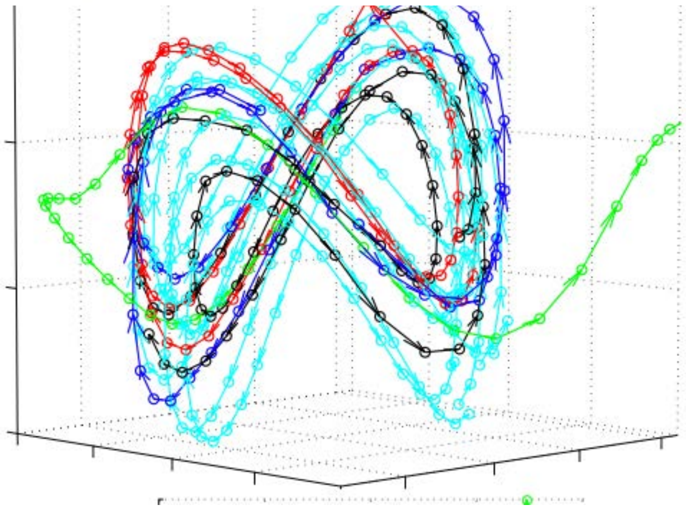


Figure 9. Tracking 37 frames of an exaggerated gait. Note that the results are very accurate even though the style is very different from any of the training motions. The last two rows depict two different views of the 3D inferred poses of the second row.

3D People Tracking with Gaussian Process Dynamical Models, Urtasun, Fleet, Fua, CVPR 2006



Early deep network prior

Restricted Boltzmann machine

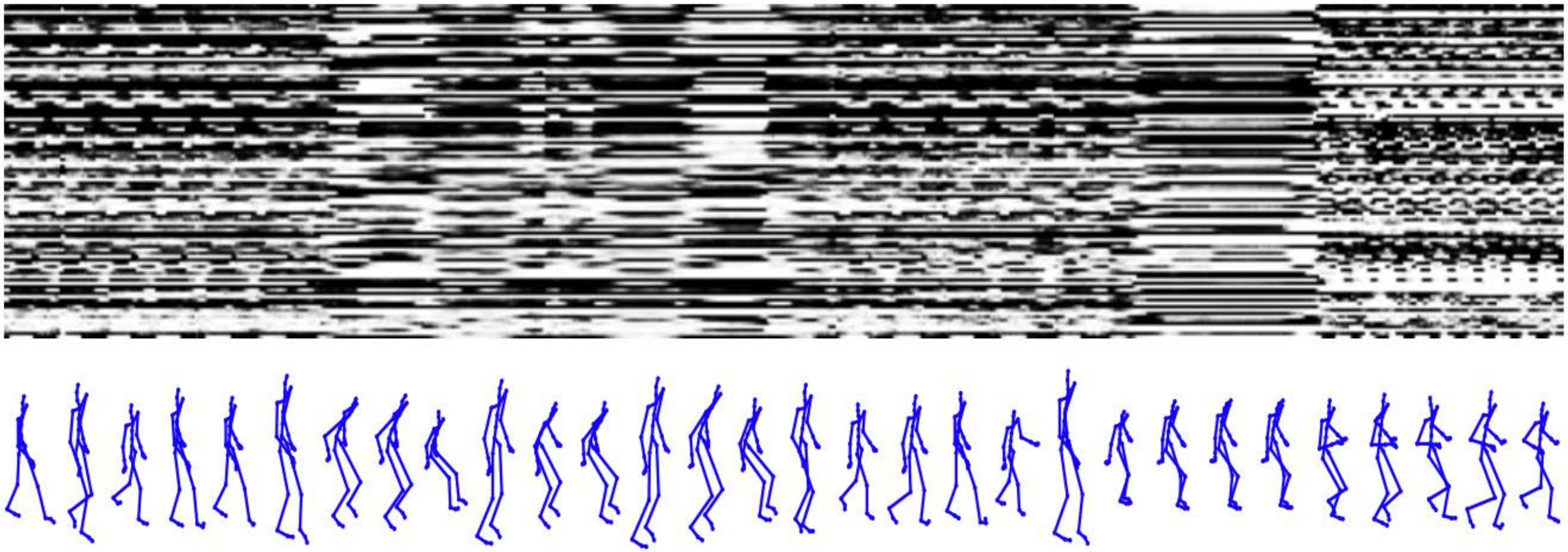


Figure 1: In a trained model, probabilities of each feature being "on" conditional on the data at the visible units. Shown is a 100-hidden unit model, and a sequence which contains (in order) walking, sitting/standing (three times), walking, crouching, and running. Rows represent features, columns represent sequential frames.

Modeling Human Motion Using Binary Latent Variables Graham W. Taylor, Geoffrey E. Hinton and Sam Roweis, NIPS 2007

Priors are crutch for the weak.

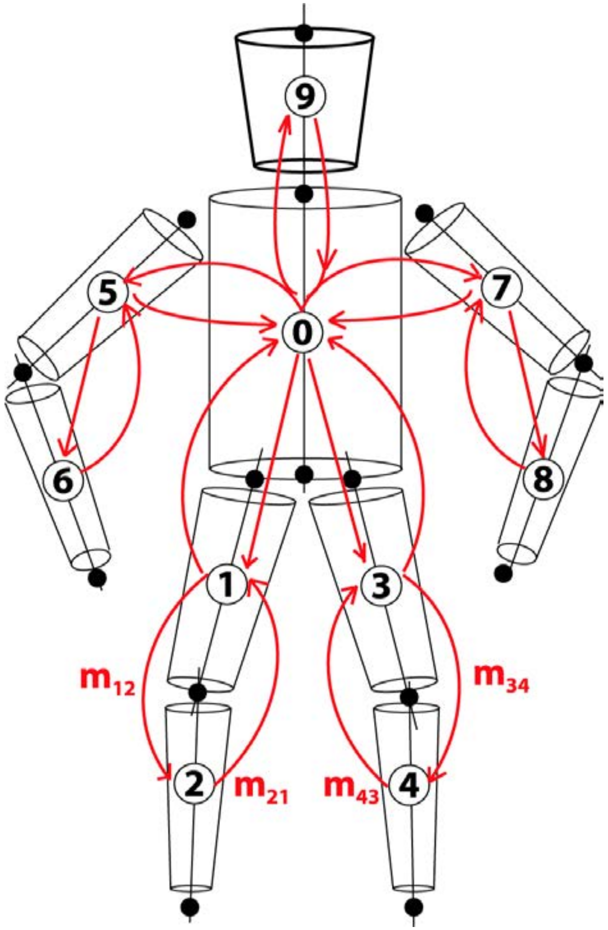
Graphs come back: Belief propagation

Like Hinton but with probabilities

Bottom-up: Find parts. Model inference puts them together.

Felzenswalbb & Huttenlocher, Pictorial Structures for Object Recognition, IJCV 2005,

3D People

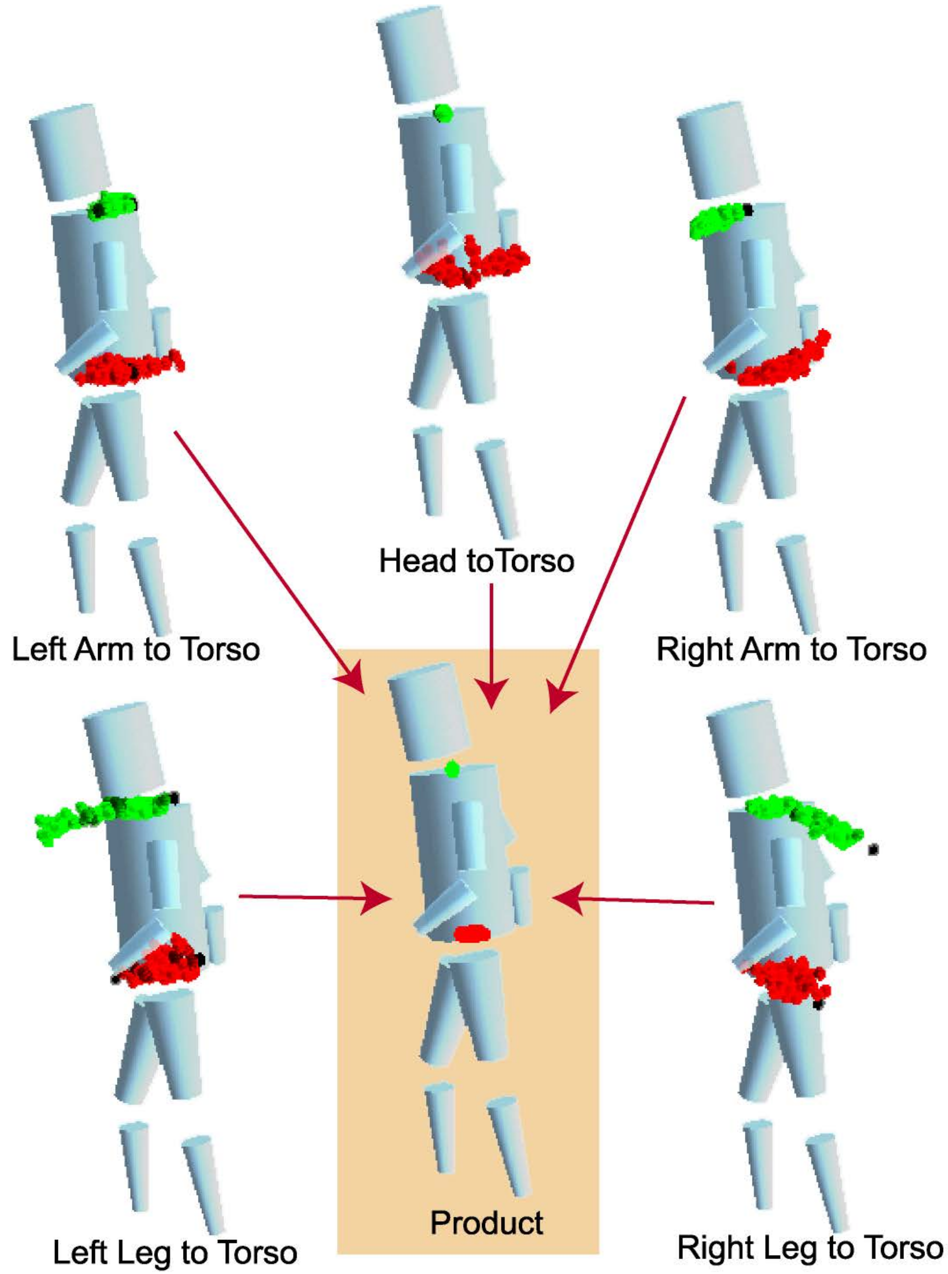


Loose-limbed body
(graphical model)

Attractive people: Assembling loose-limbed models using non-parametric belief propagation Sigal, L., Isard, M. I., Sigelman, B. H., Black, M. J., NIPS 2003

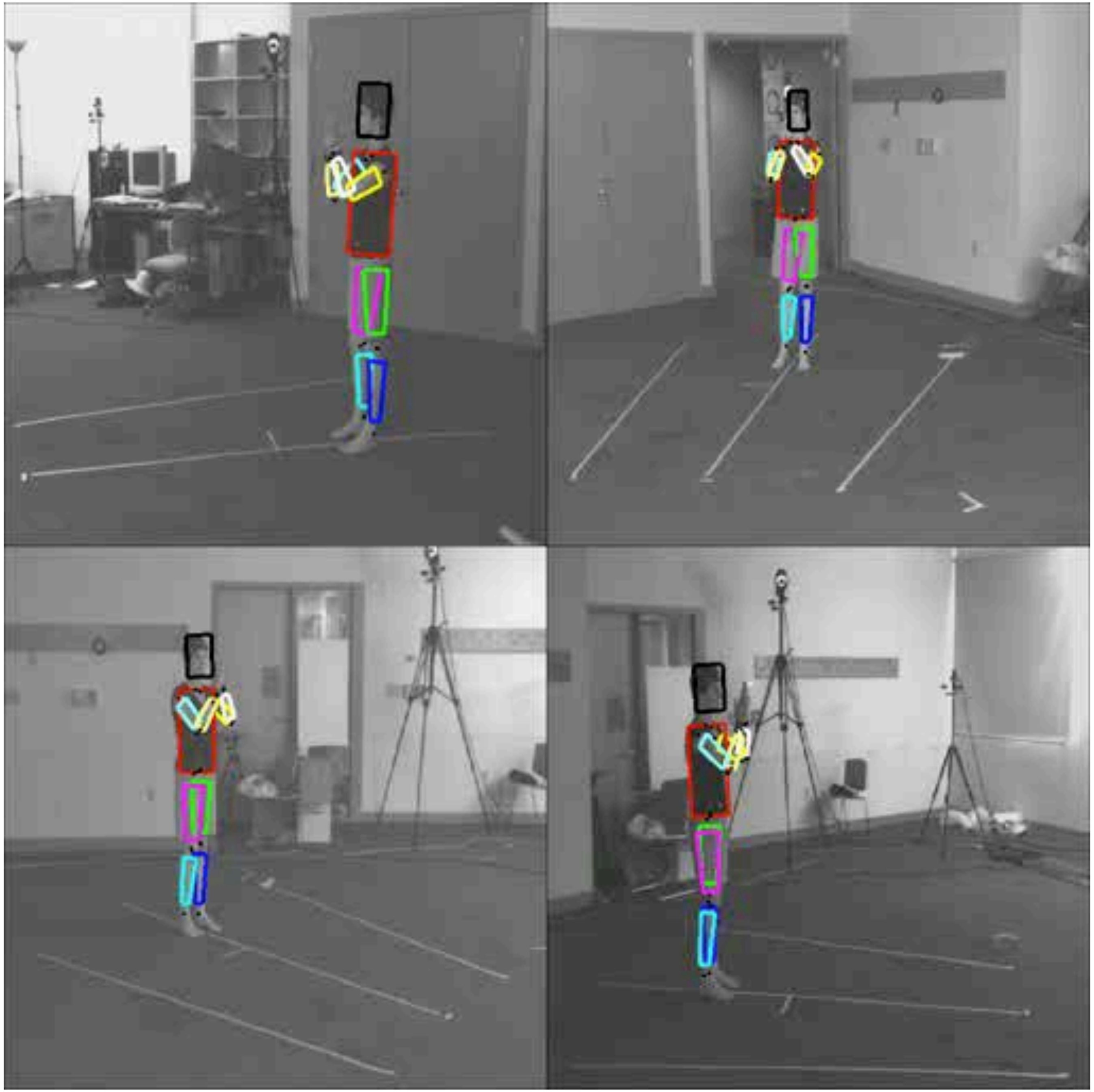
Loose-limbed people, Sigal, L., Isard, M., Haussecker, H., Black, M. J. IJCV 2011.

Illustration of the message product:



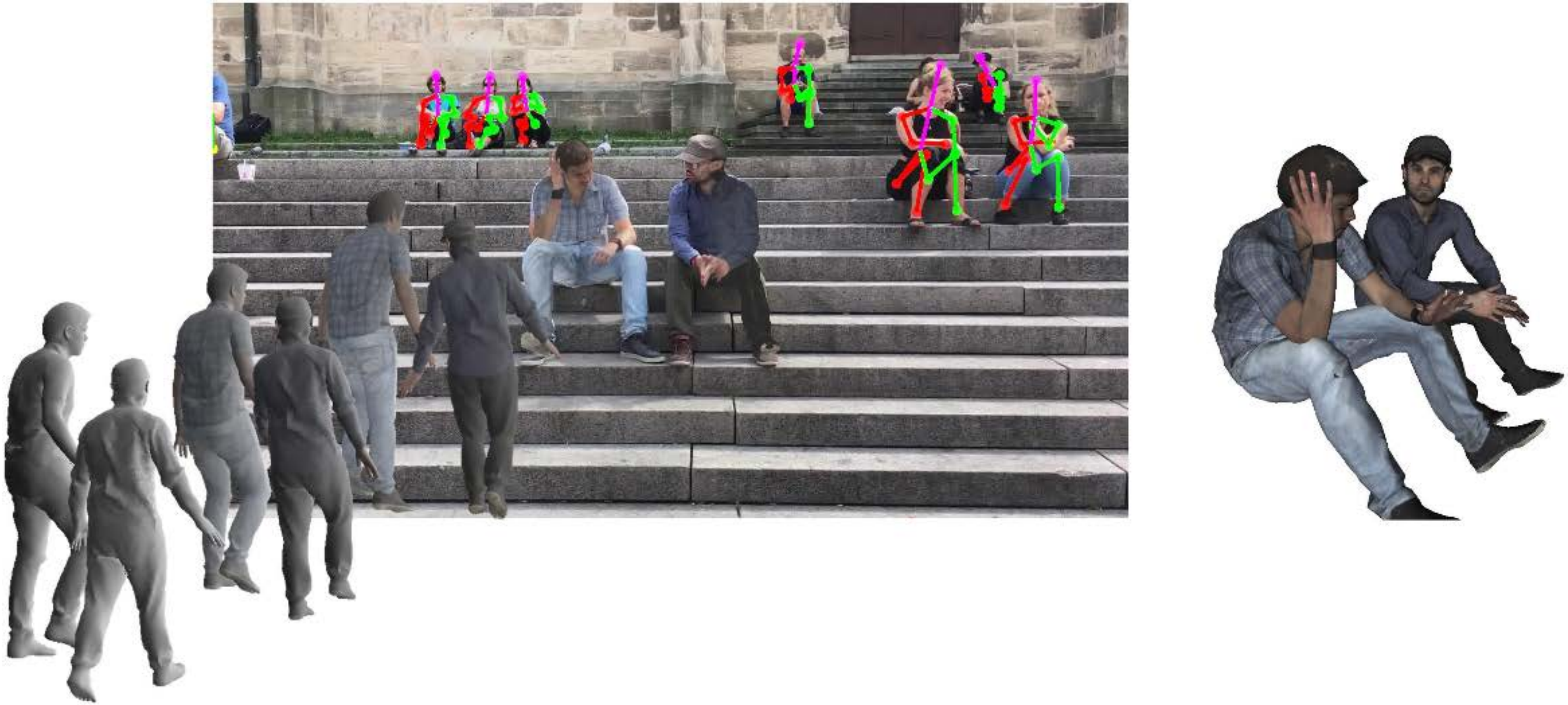
$$m_{ij}(\mathbf{X}_j) = \alpha \int \psi_{ij}(\mathbf{X}_i, \mathbf{X}_j) \lambda(\mathbf{X}_i) \prod_{k \in A_i \setminus j} m_{ki}(\mathbf{X}_i) \, d\mathbf{X}_i$$

Ground truth.
There was none.
Were we making progress?



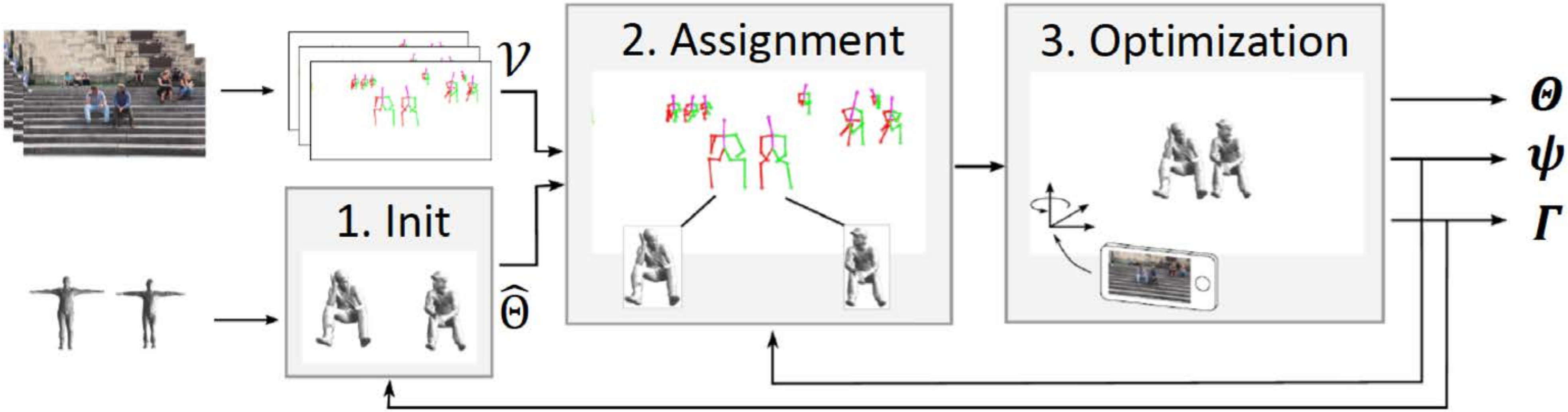
Sigal, Balan, Black, HumanEva, 2004 and IJCV 2010.

3D humans in the wild



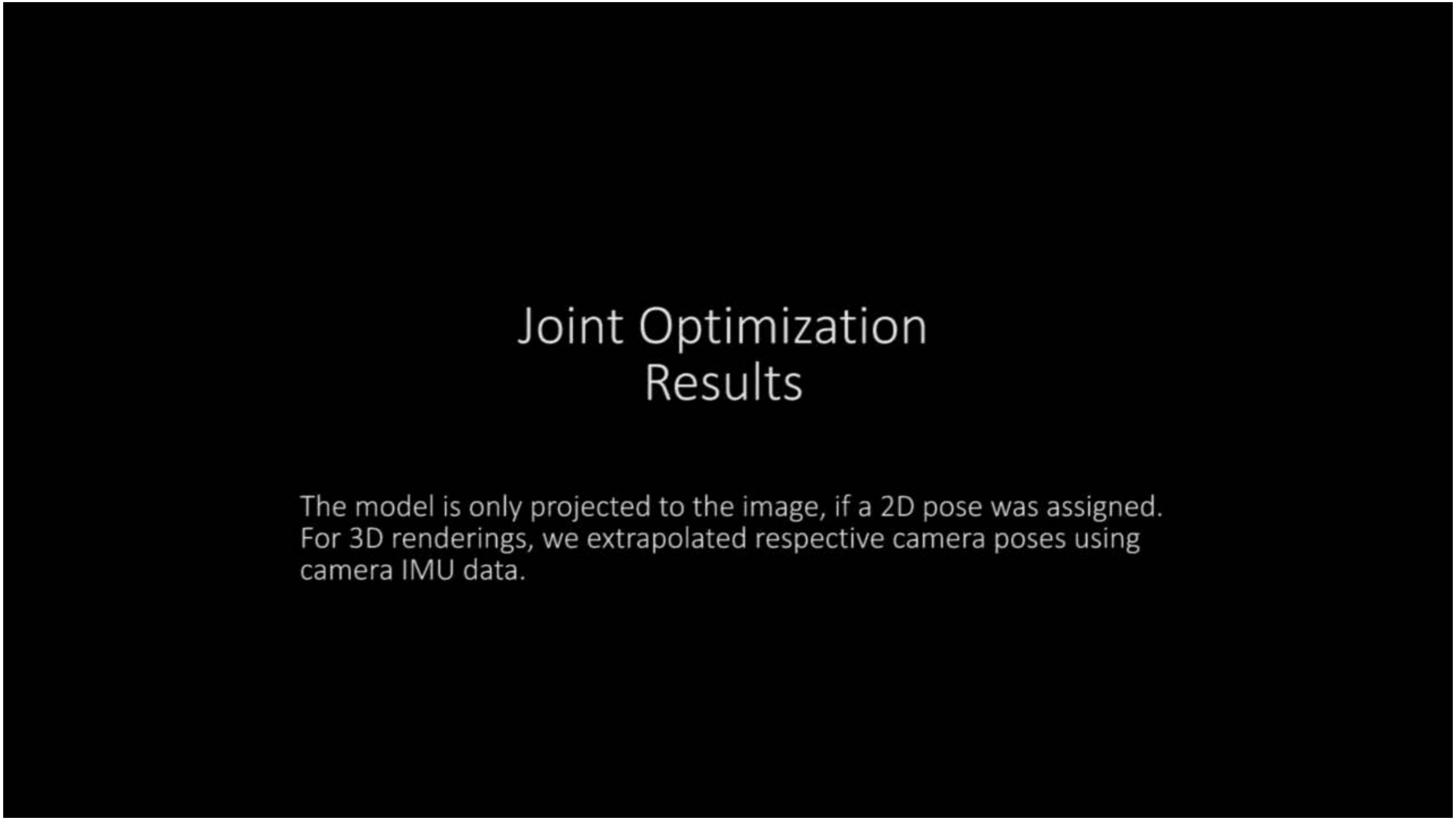
Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera. Marcard, T. V., Henschel, R., Black, M. J., Rosenhahn, B., Pons-Moll, G., ECCV 2018

Video Inertial Poser (VIP)



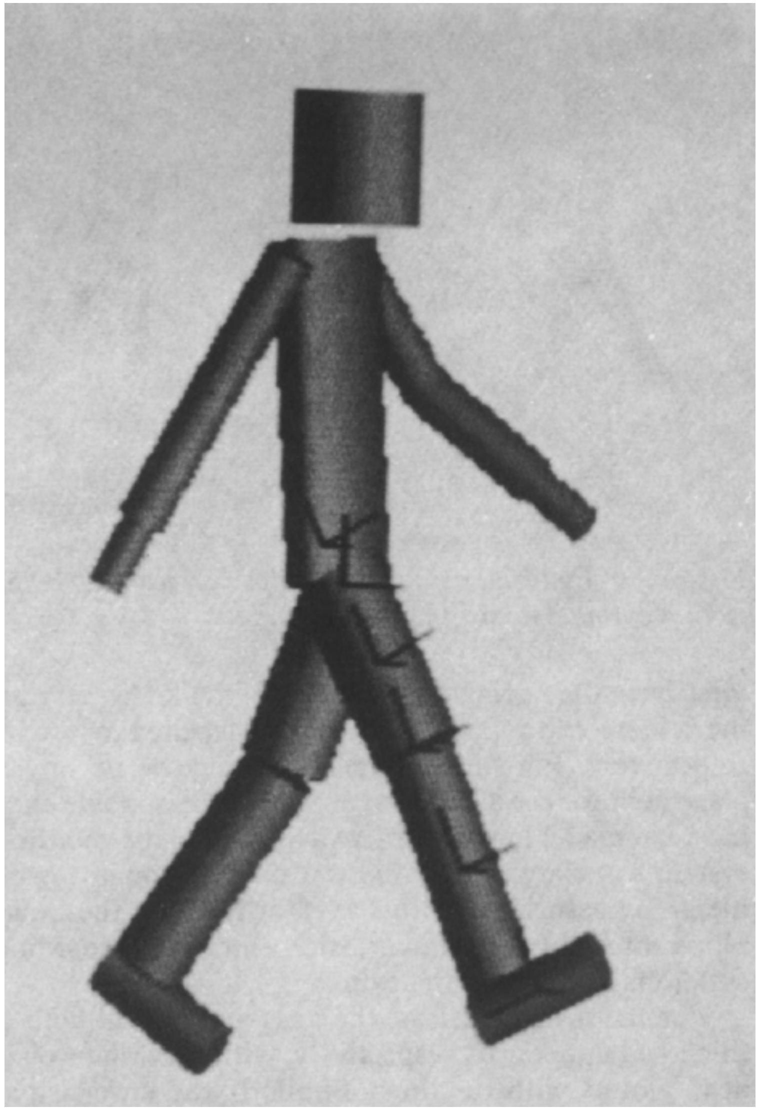
- combines a hand-held camera with body-worn Inertial Measurement Units (IMUs)
- reconstructs accurate 3D poses
- fixes IMU drift problem
- works with multiple, interacting people
- enables 3D Human Motion Capture „in the wild“

3D pose estimation



Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera. Marcard, T. V., Henschel, R., Black, M. J., Rosenhahn, B., Pons-Moll, G., ECCV 2018

The problem



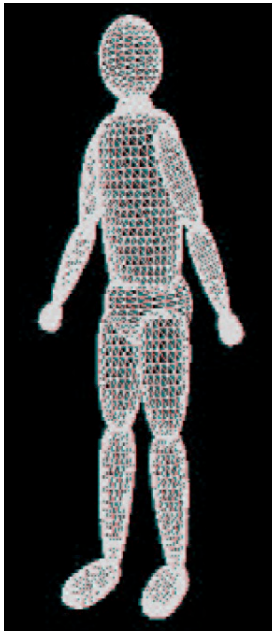
We don’t look like this.

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.

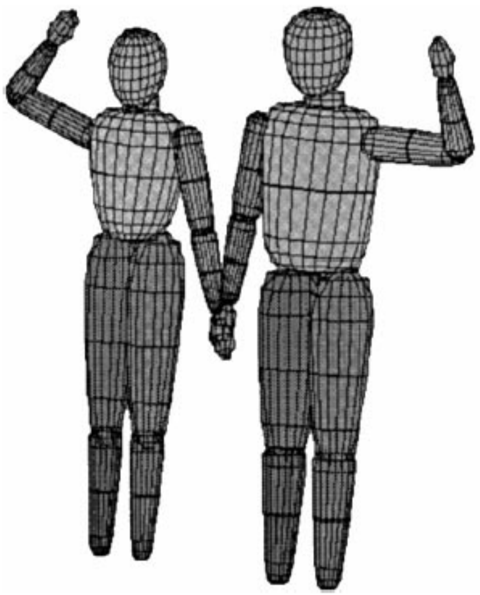
Early body models



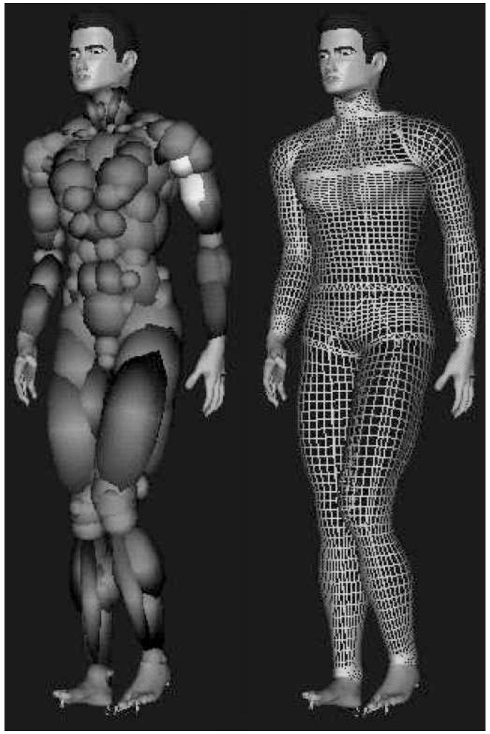
[Sminchisescu
and Triggs ’03]

[Terzopoulos
and Metaxas ’93]

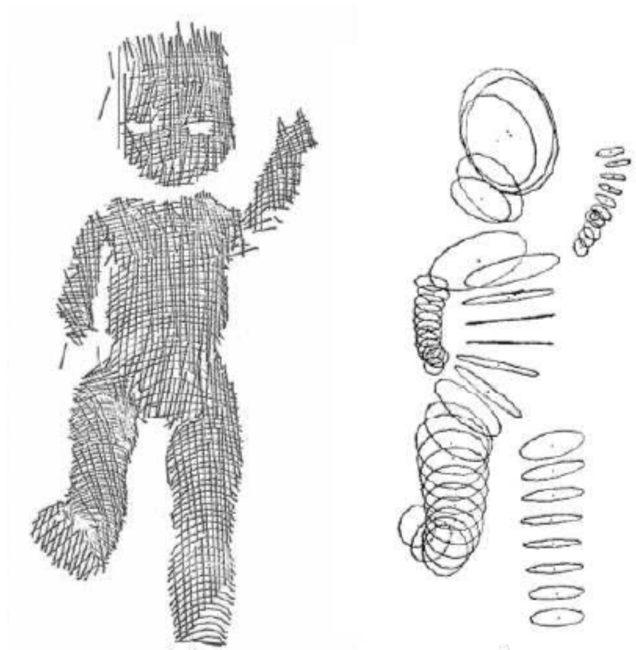
[Kakadiaris and Metaxas ’00]



[Gavrilla, ‘96]

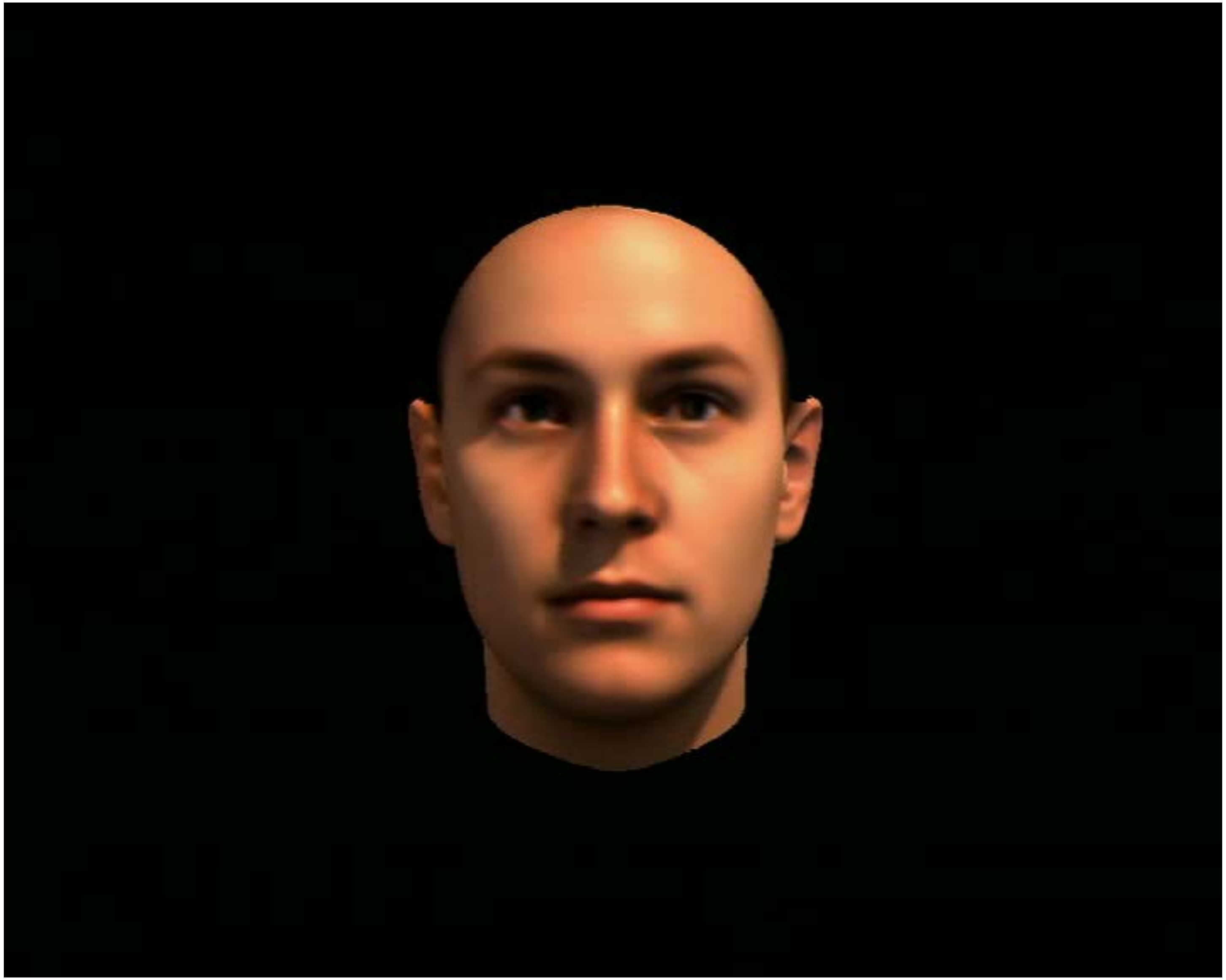


[Plänkers and
Fua ’01]



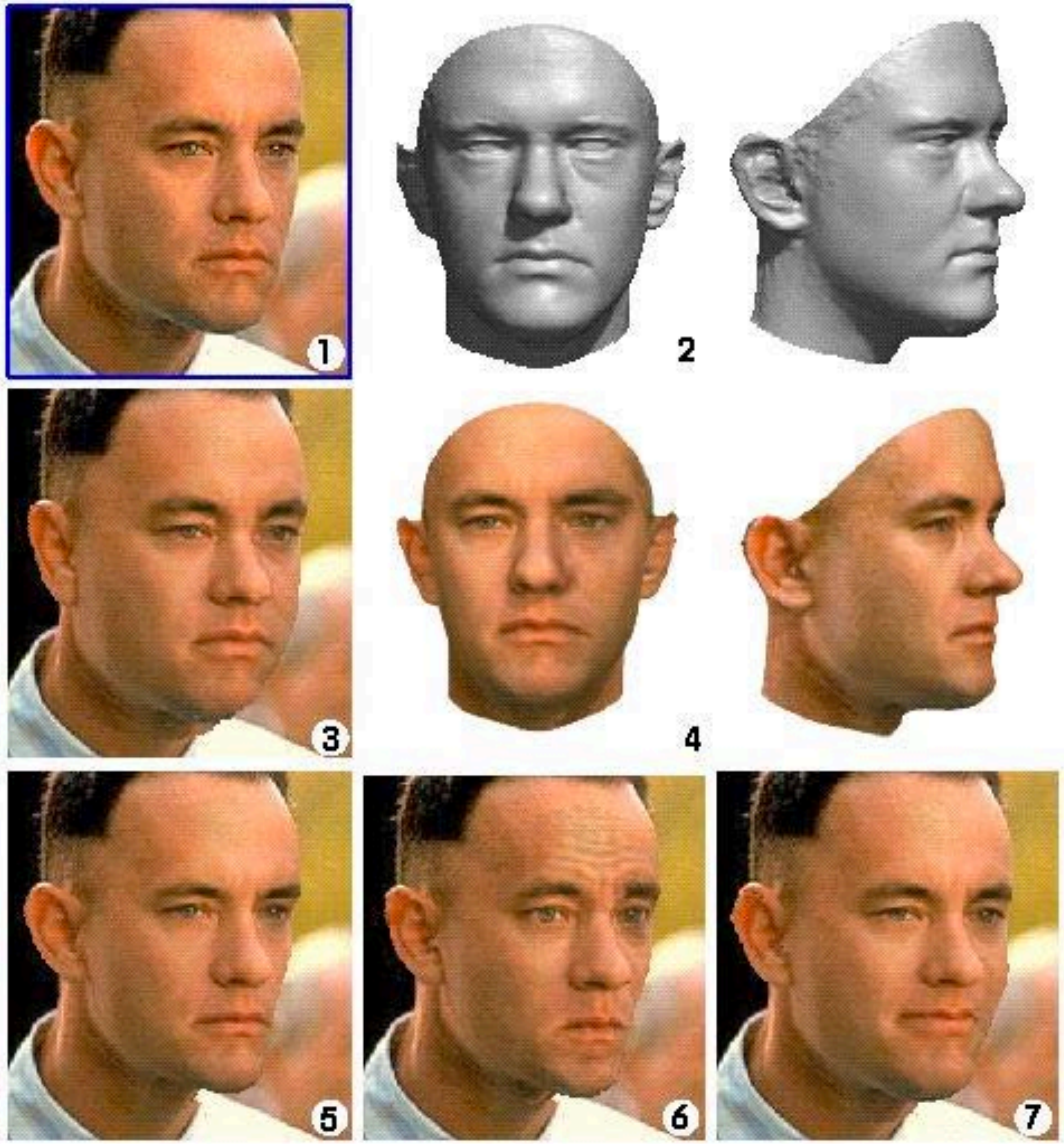
Nevatia & Binford ’73

Learning face shapes



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

Inverse graphics



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

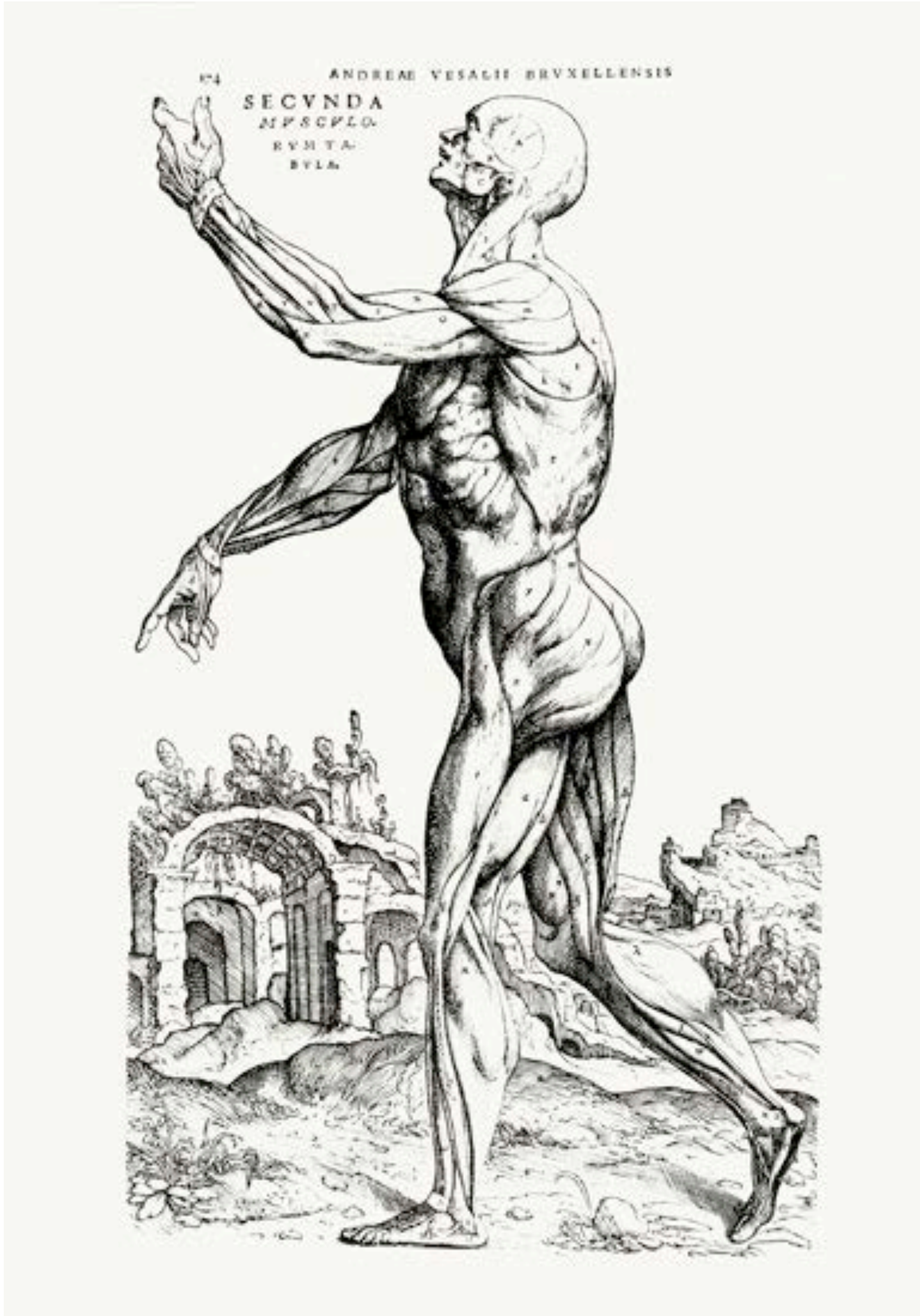
Why is it hard?

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.

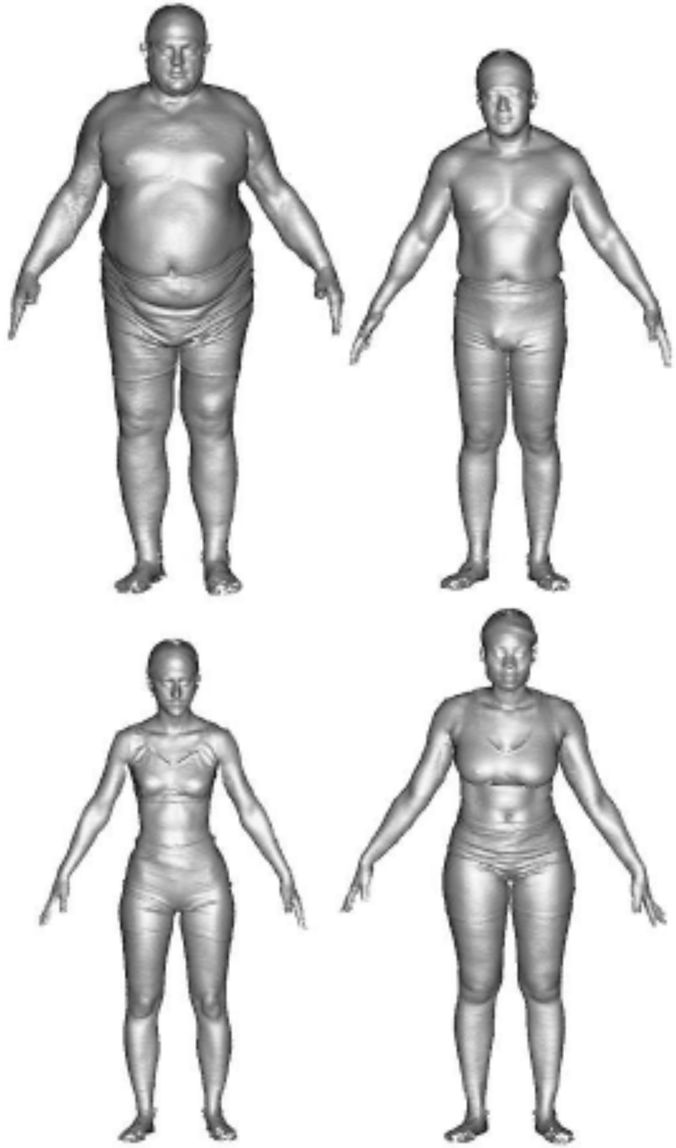
Approach: model only what we can see – the surface.



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

Learning a body model

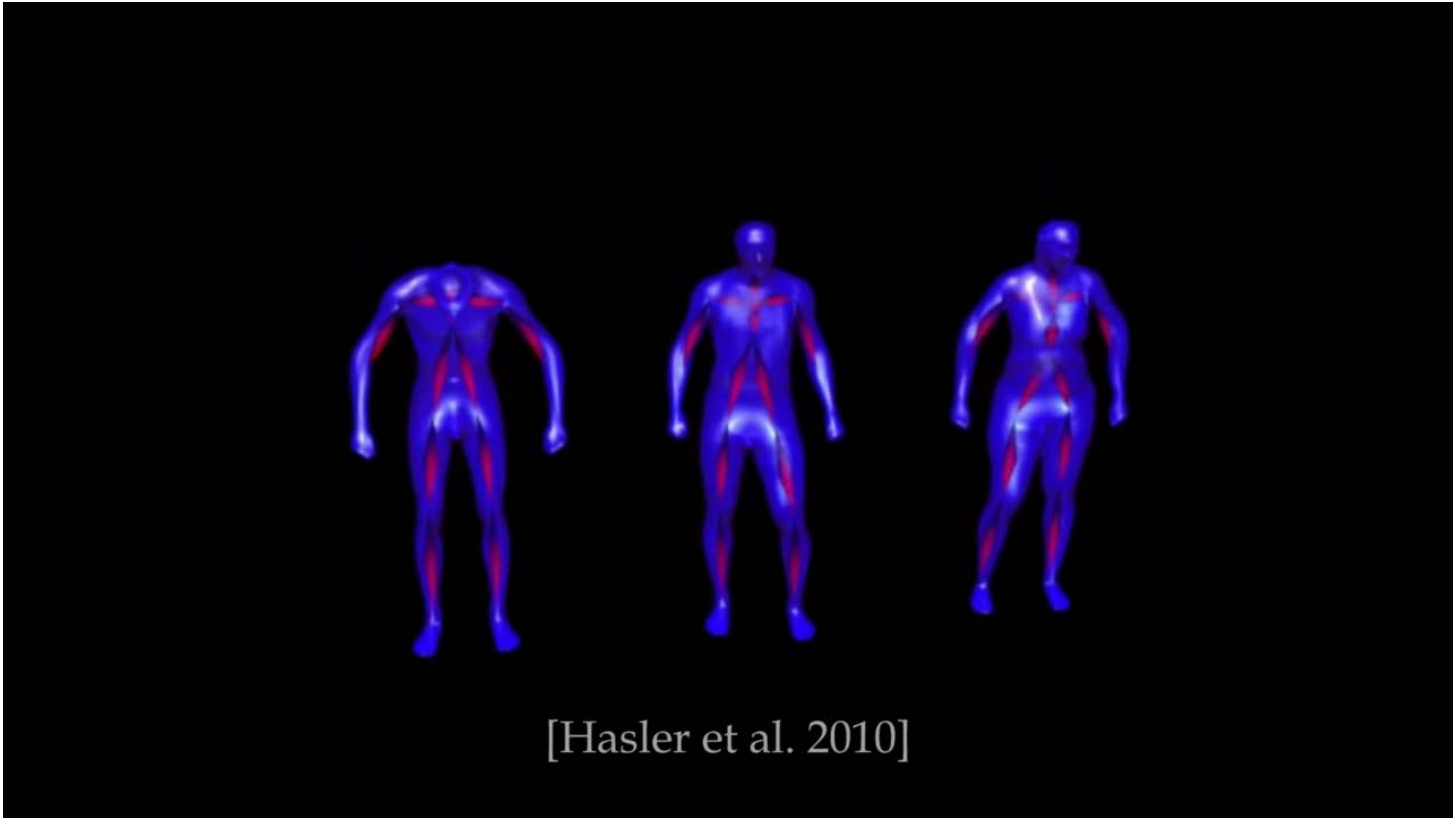
[Cyberware]



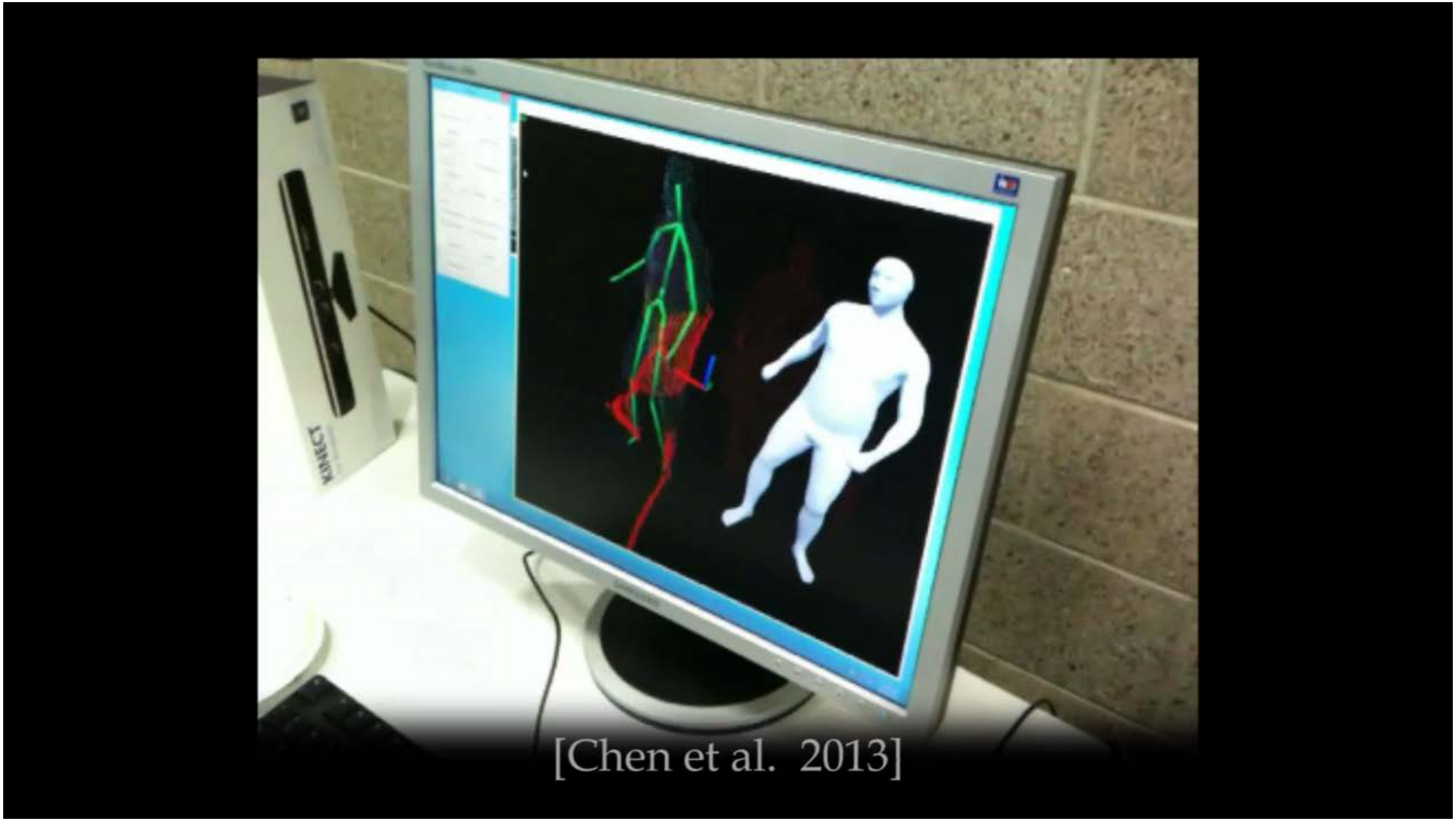
CAESAR dataset – 2001.

Learning body models (2003-2013)

Learning body models (2003-2013)



Learning body models (2003-2013)

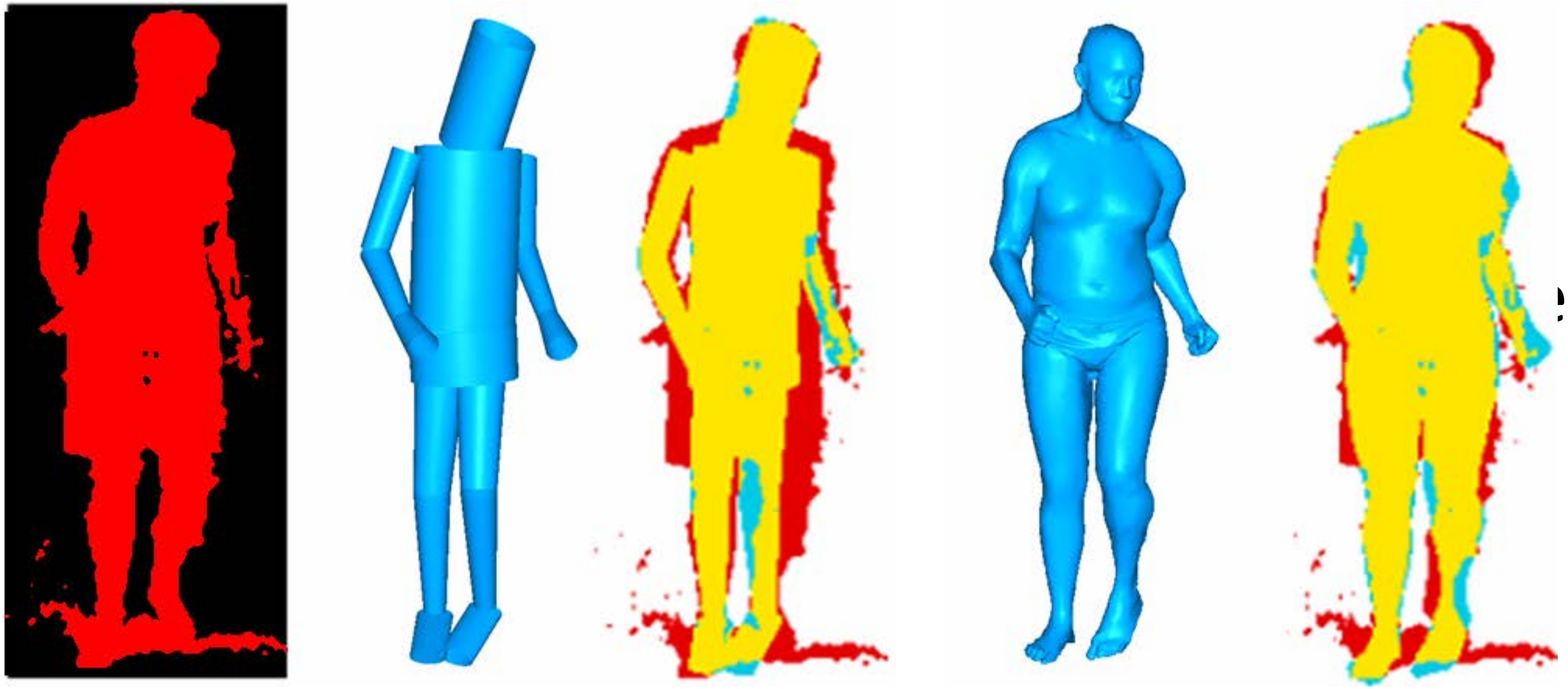


[Chen et al. 2013]

Learning body models (2003-2013)

Anguelov et al., SCAPE, 2005

Generative models of bodies



Traditional model — Proposed model

Detailed human shape and pose from images
Balan, A., Sigal, L., Black, M. J., Davis, J., Haussecker, H., CVPR 2007

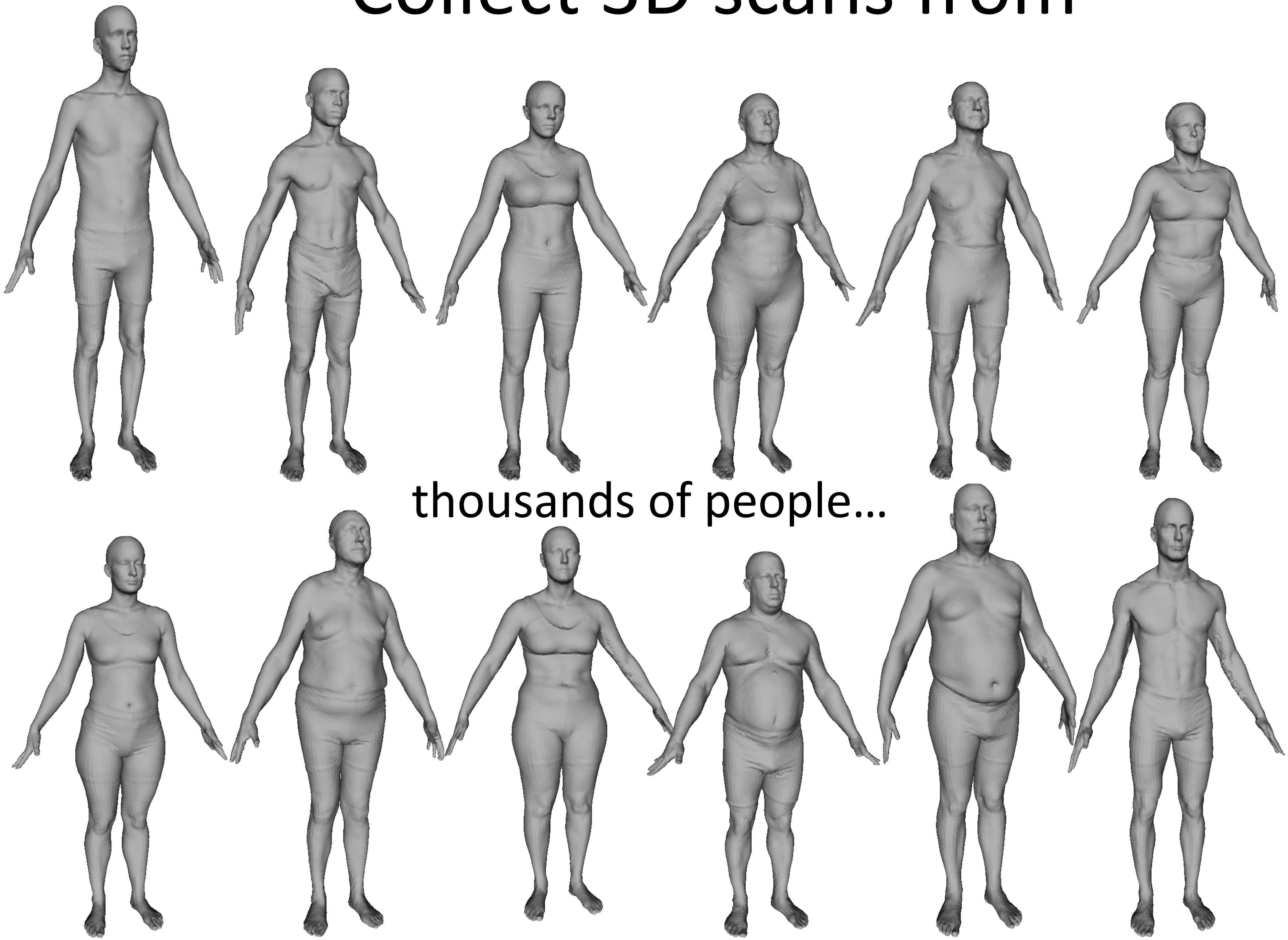
Goal: Virtual humans

Define a simple **mathematical model** of body shape.
It should **look** like real people.
It should **move** like real people.
It should be low-D, differentiable, have joints, and be easy to animate and fit to data.

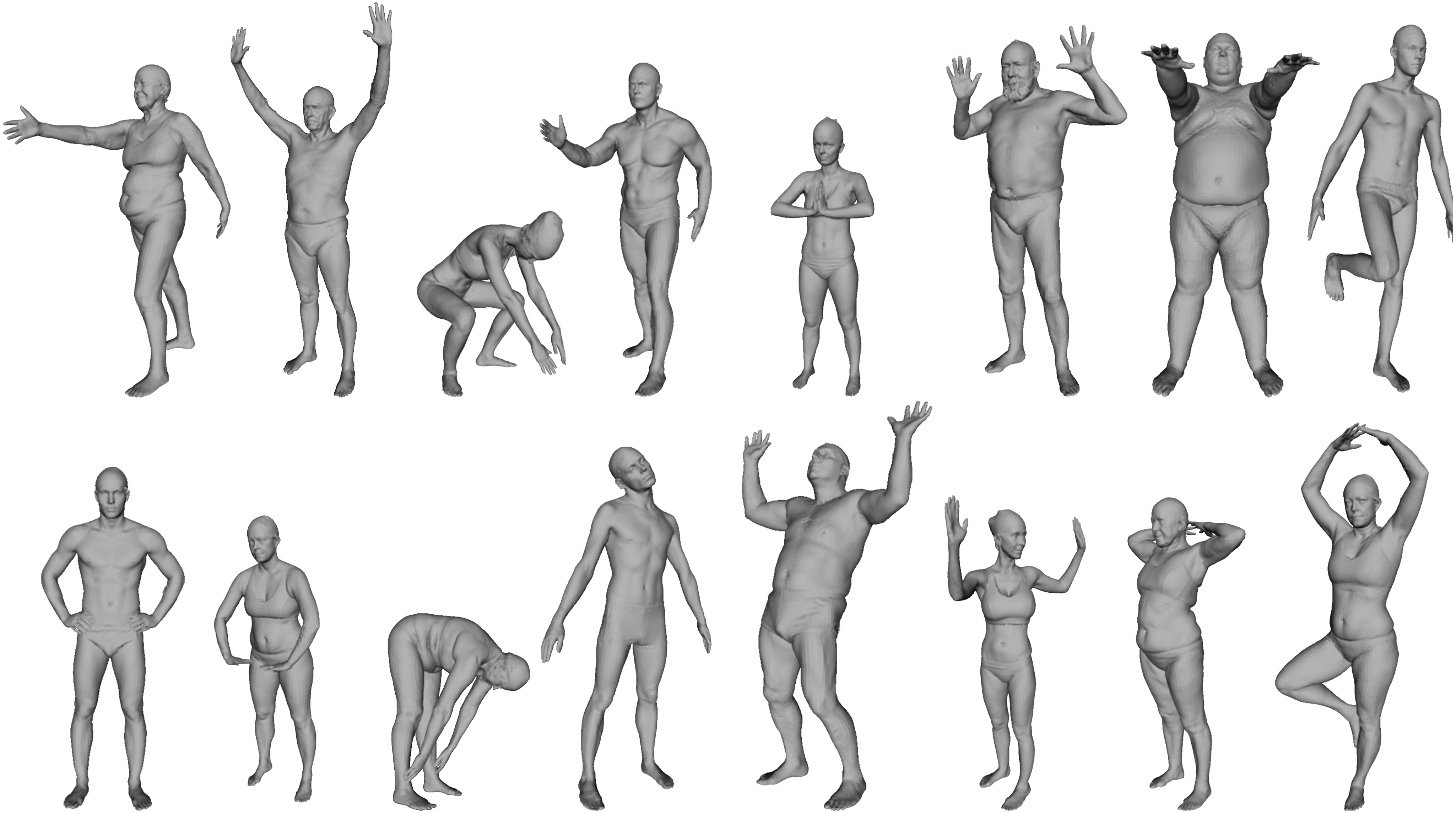
4D scanner:3D at 60 fps

Collect 3D scans from

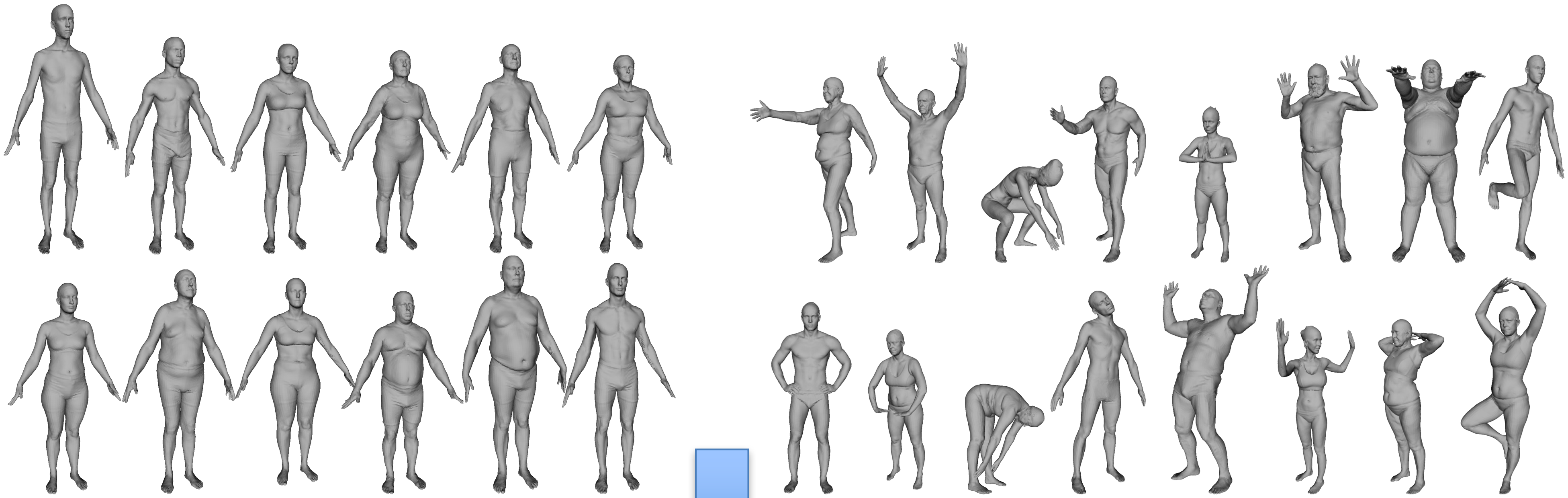
thousands of people…



and thousands of poses



1000’s of high-resolution scans of different shapes and poses



$$M(\theta, \beta, \delta, A)$$

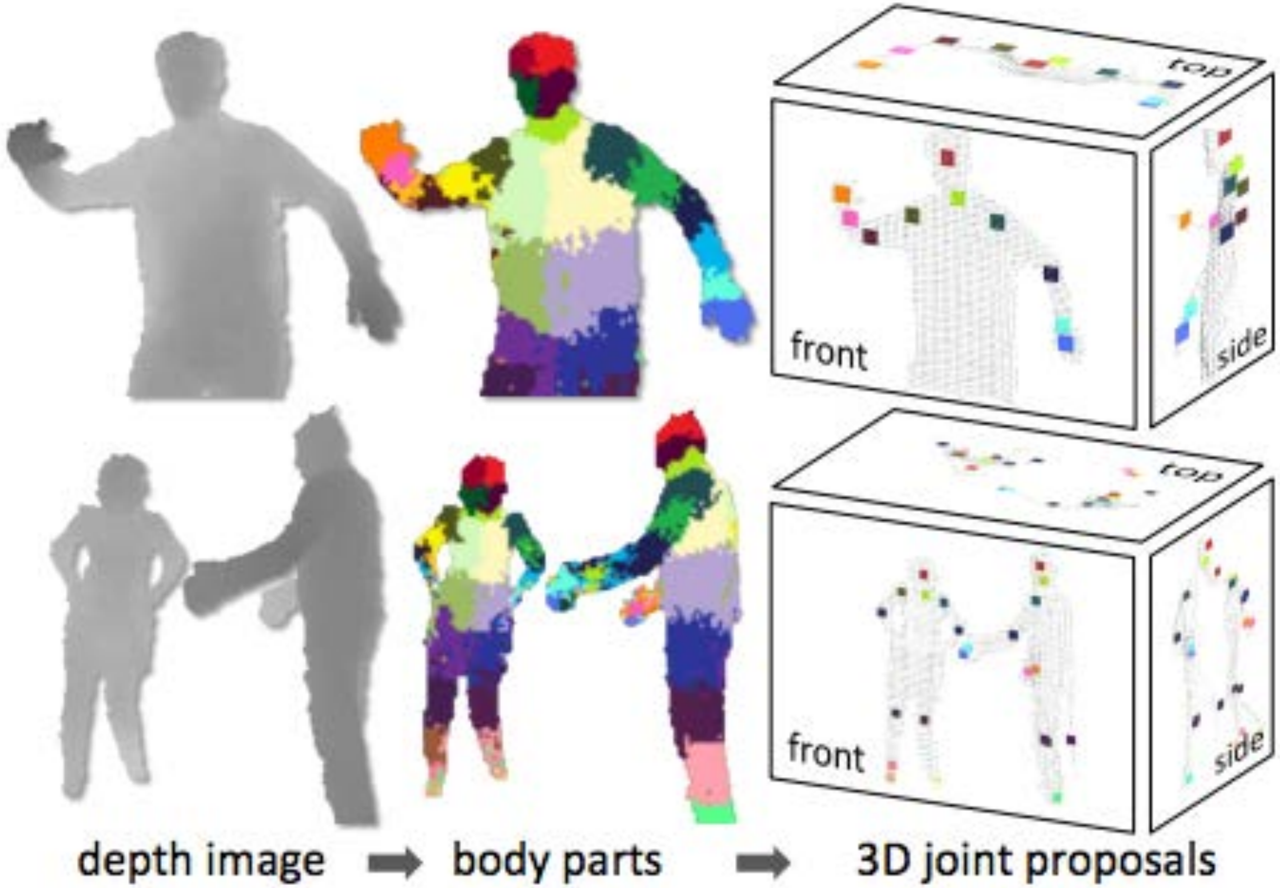
A body model M takes a small number of pose, shape, and other parameters and returns a 3D mesh.

Key idea: Everything is learned from registered data to minimize surface-to-surface error.

SMPL: A Skinned Multi-Person Linear Model, Loper et al., SIGGRAPH 2015

RGB-D

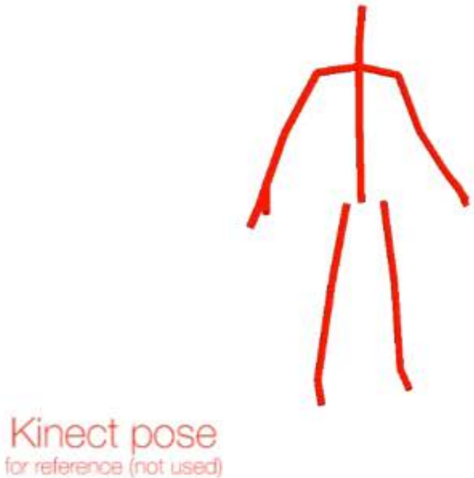
Kinect

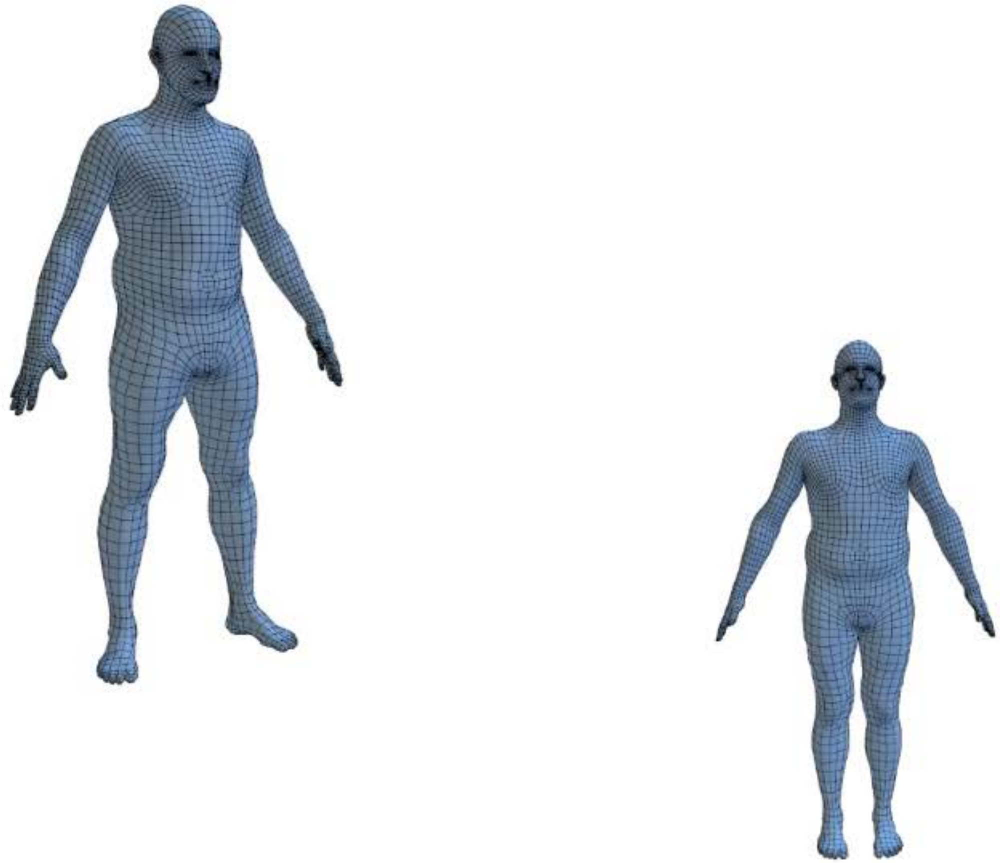


Synthetic data.
ML approach.
Bottom up.
Fast, reliable.

Real-Time Human Pose Recognition in Parts from Single Depth Images, Shotton et al., CVPR 2011

For reference. Not used.



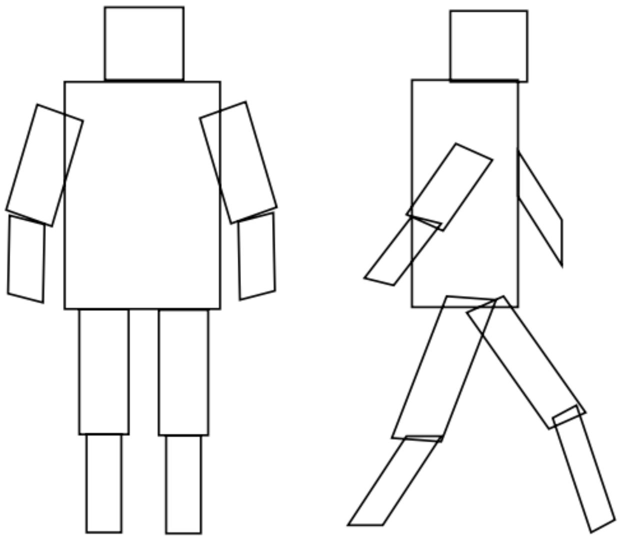


Average Euclidean surface-to-surface error over 7 subjects: 2.4mm

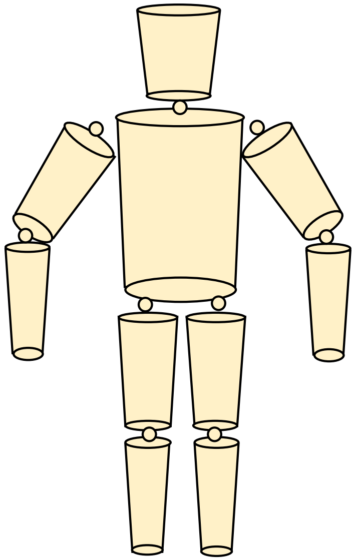
Bogo et al., ICCV 2015.

The evolution of body models

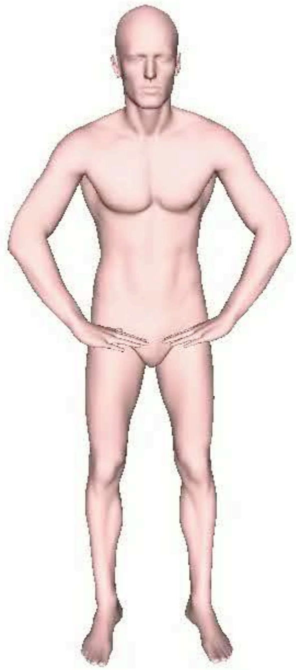
1996



2006



2016

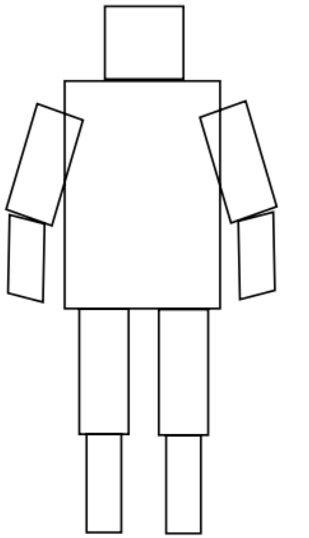


Learned 3D model of body shape and pose from 3D scans.

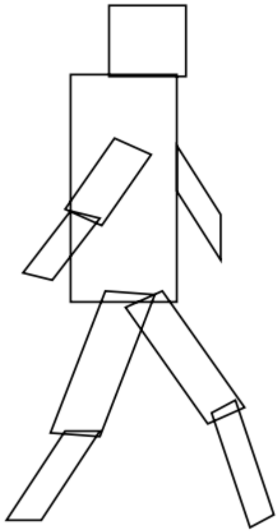
Loper et al., SMPL, SIGGRAPH Asia 2015

The evolution of body models

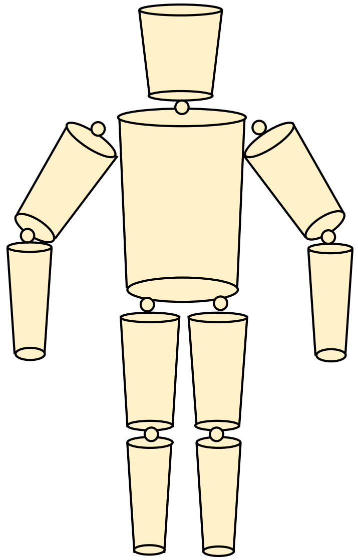
1996

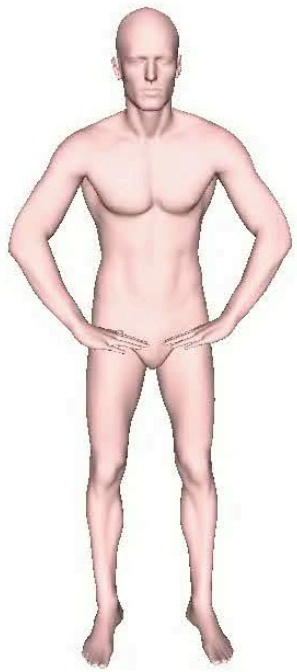


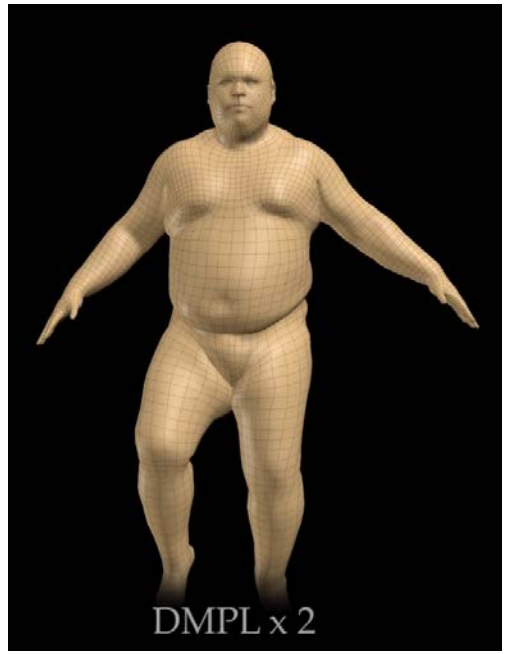
2006



2016

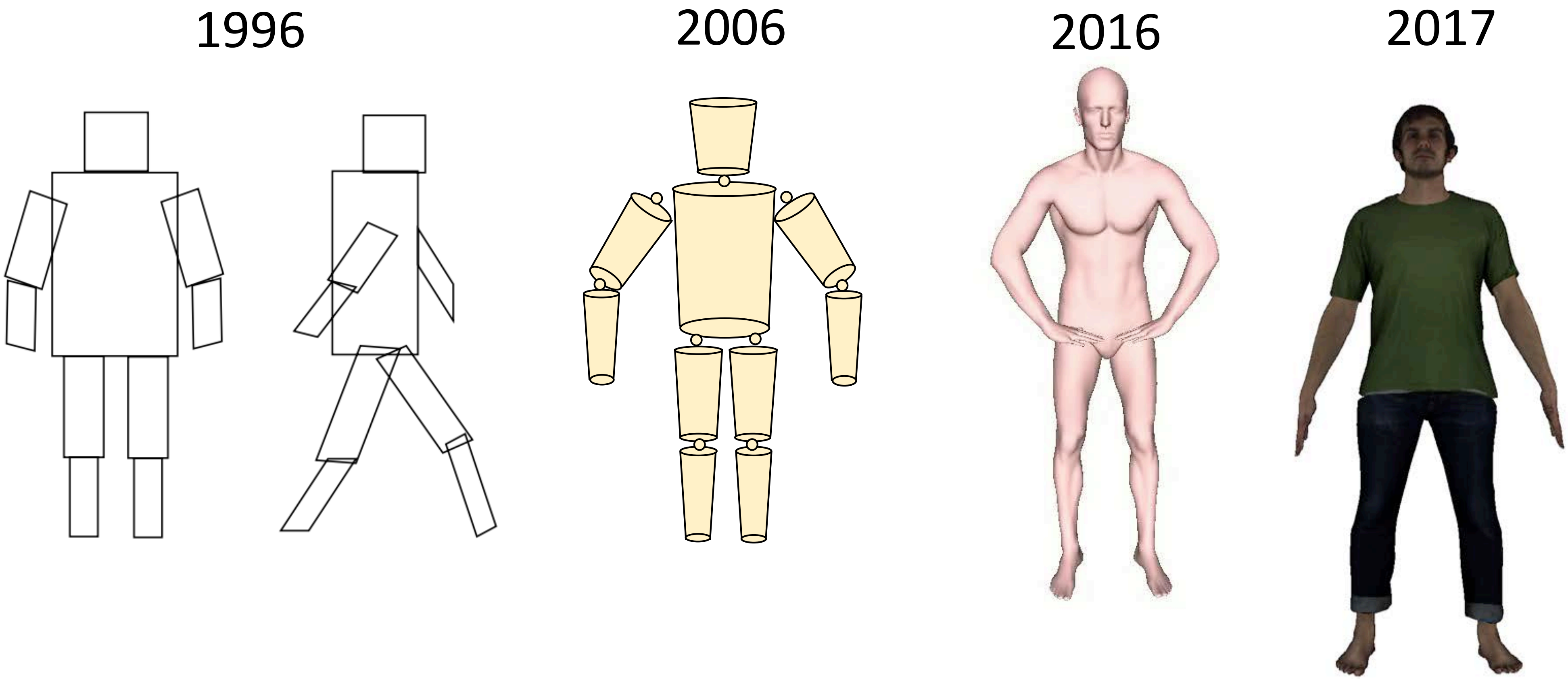






Dyna: A Model of Dynamic Human Shape in Motion,
Pons-Moll et al, SIGGRAPH 2015

The evolution of body models

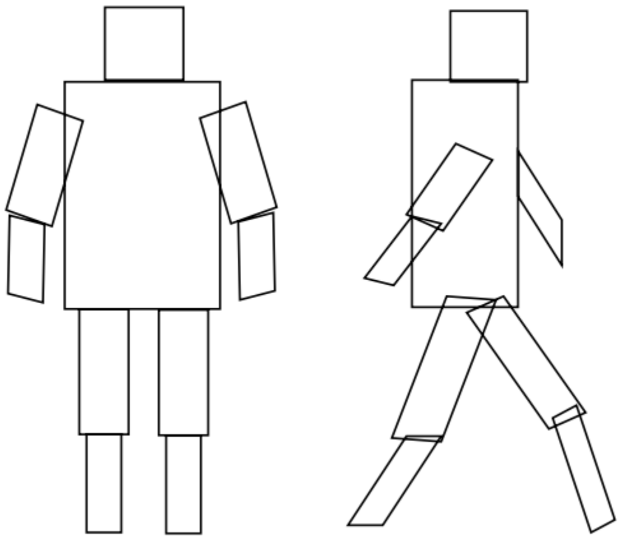


"ClothCap: Seamless 4D Clothing Capture and Retargeting,"
Pons-Moll, G., Pujades, S., Hu, S., Black, M.J.,.
ACM Transactions on Graphics (SIGGRAPH), 2017.

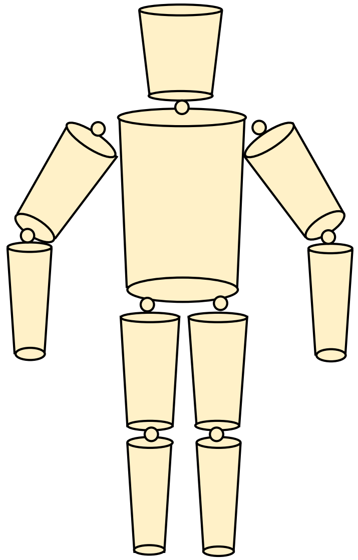
Capture and model clothing

The evolution of body models

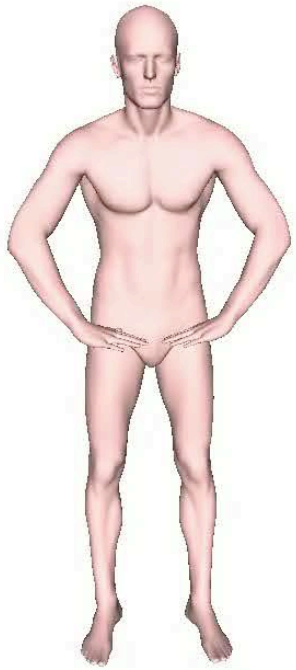
1996



2006



2016



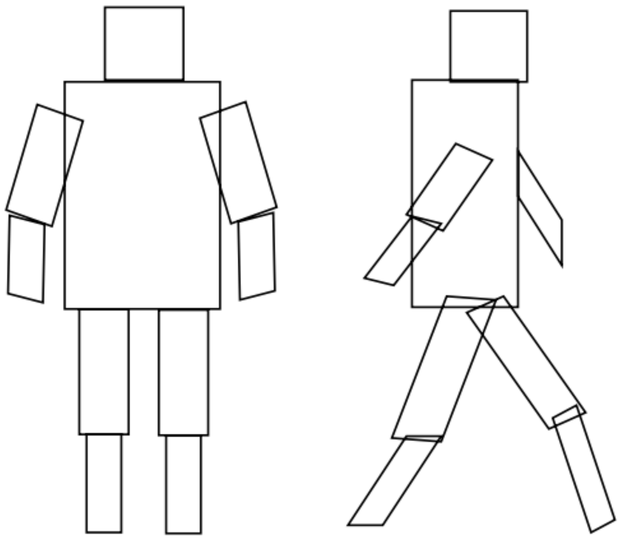
2018

Infants are harder to capture because you can’t direct them and scanning is complicated

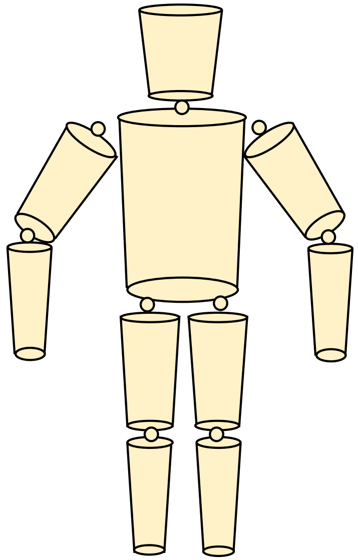
Hesse, et al., Learning an Infant Body Model from RGB-D Data for Accurate Full Body Motion Analysis, MICCAI 2018

The evolution of body models

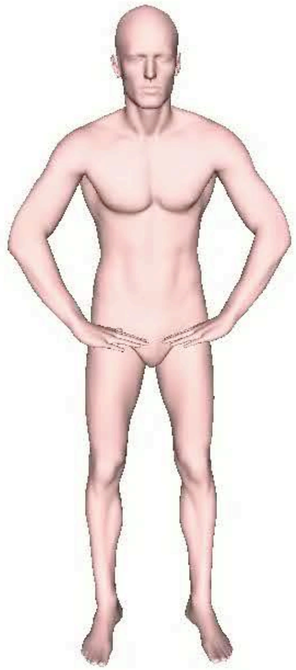
1996



2006



2016

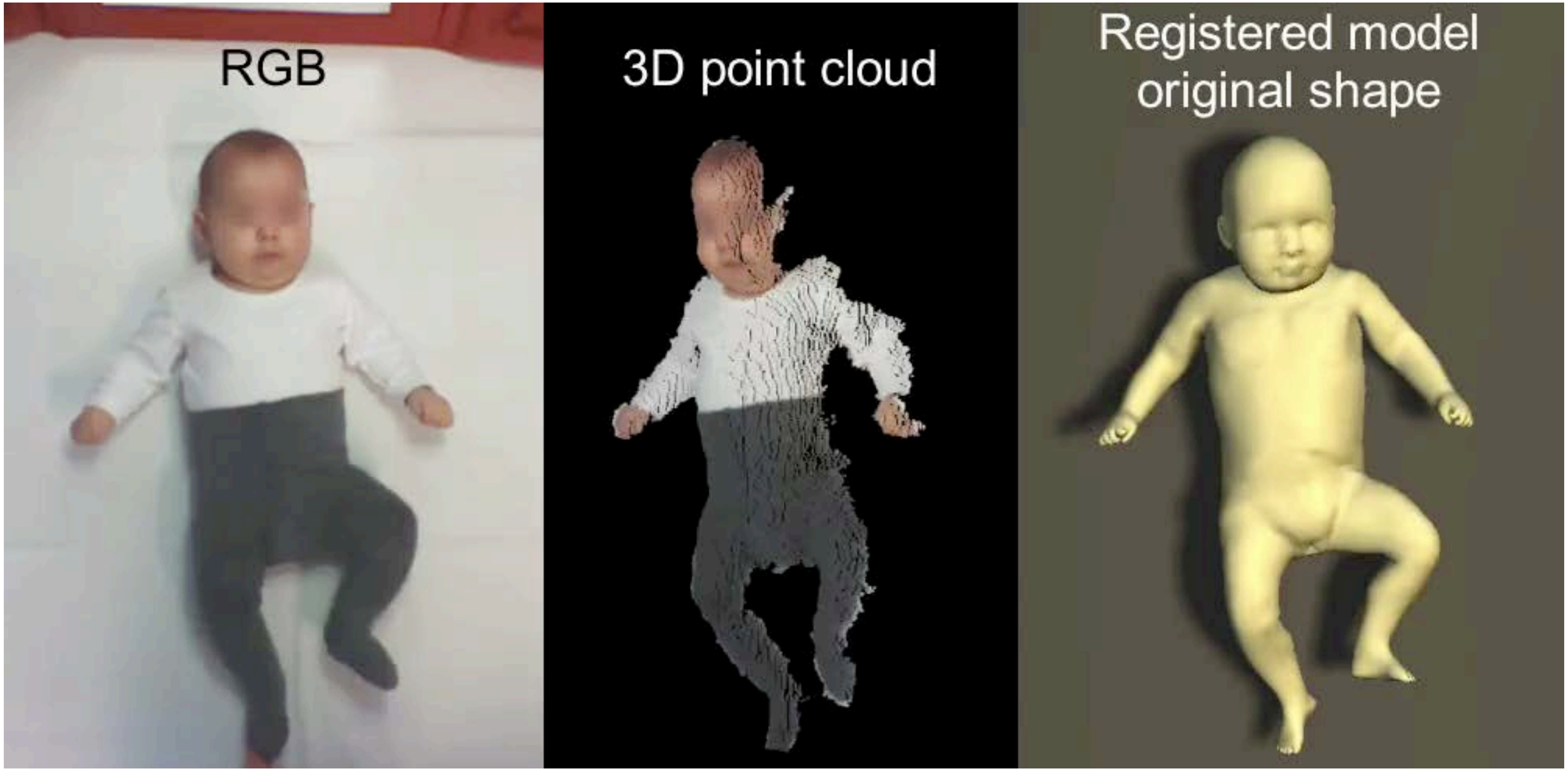


2018

Use RGB-D sequences to track and learn the model.

Hesse, et al., Learning an Infant Body Model from RGB-D Data for Accurate Full Body Motion Analysis, MICCAI 2018

SMIL: Skinned Multi-Infant Linear model

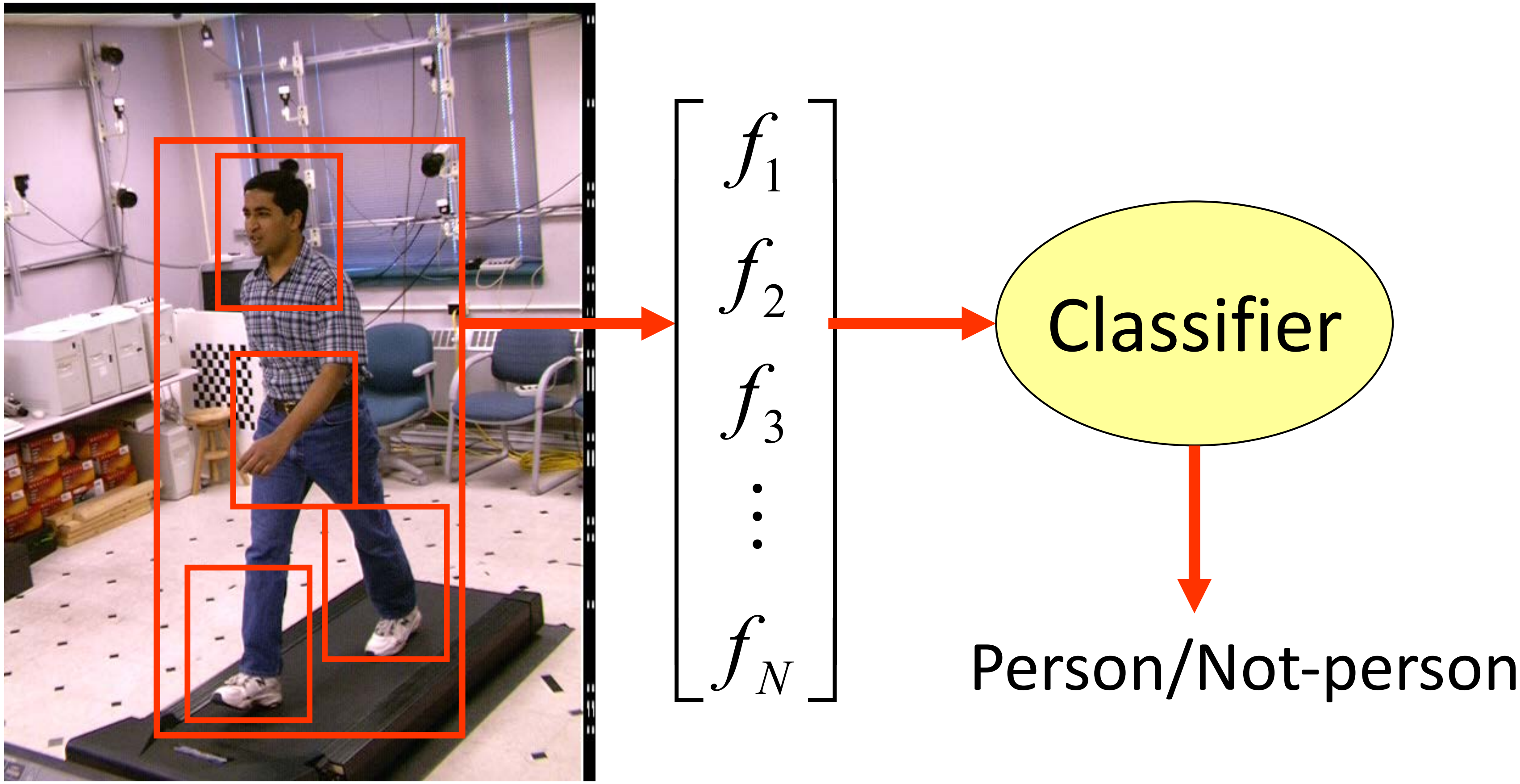


Goal: early detection of cerebral palsy from movement.

Hesse, et al., Learning an Infant Body Model from RGB-D Data for Accurate Full Body Motion Analysis, MICCAI 2018

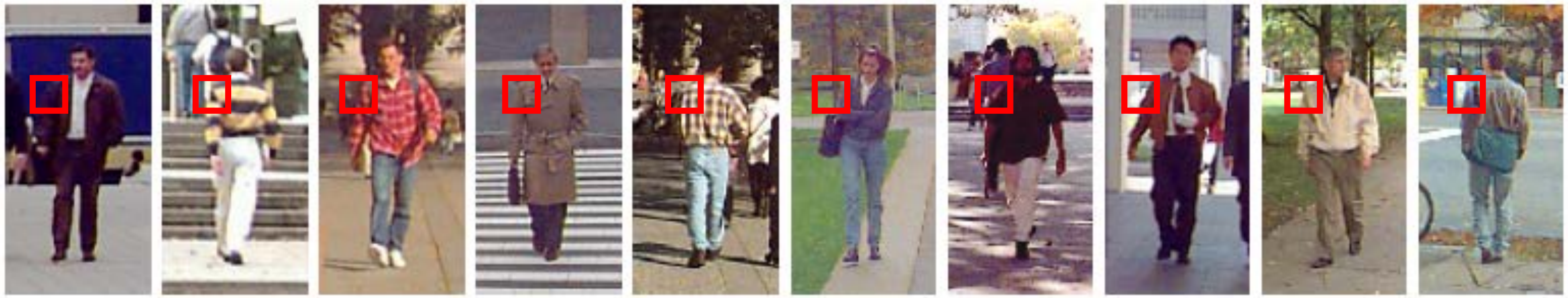
An alternative thread emerges
1997 - today

Detection: The Pure ML Approach



Single image

Support Vector Machines



Multiply the pixel values in the region by this "mask" or "filter":

1	-1

Average the resulting absolute responses.

"Pedestrian detection using wavelet templates," Oren *et al* CVPR'97.

Support Vector Machines

"Pedestrian detection using wavelet templates," Oren *et al* CVPR'97.

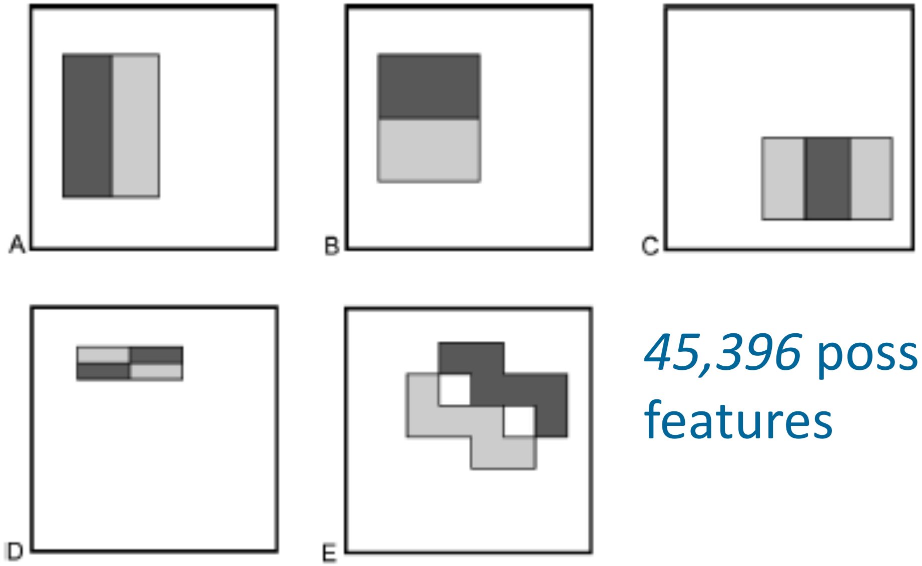
Support Vector Machines

Product of wavelet templates and filtered image regions gives a vector of responses for each region.

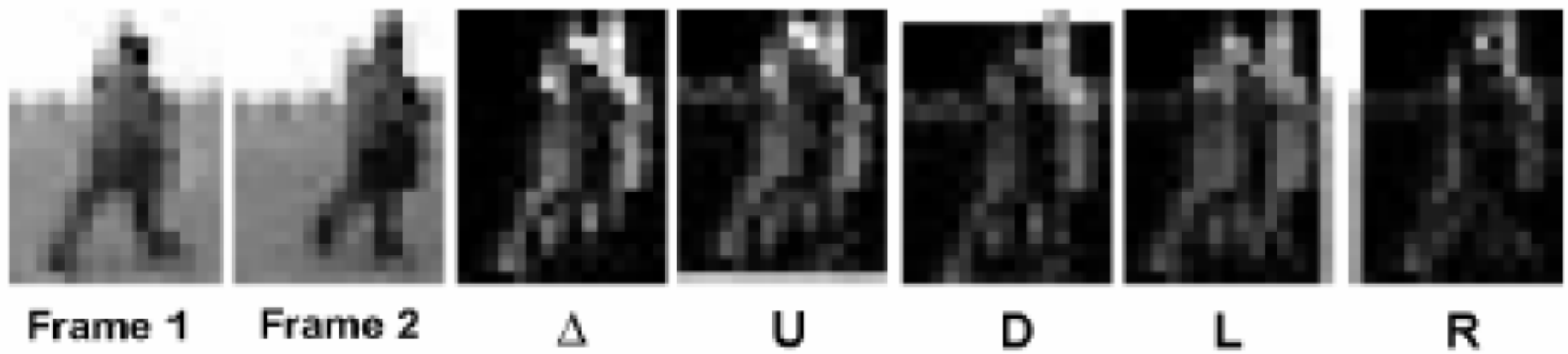
Bootstrapped SVM learns the classify pedestrian/background.

"Pedestrian detection using wavelet templates," Oren *et al* CVPR'97.

AdaBoost



45,396 possible features



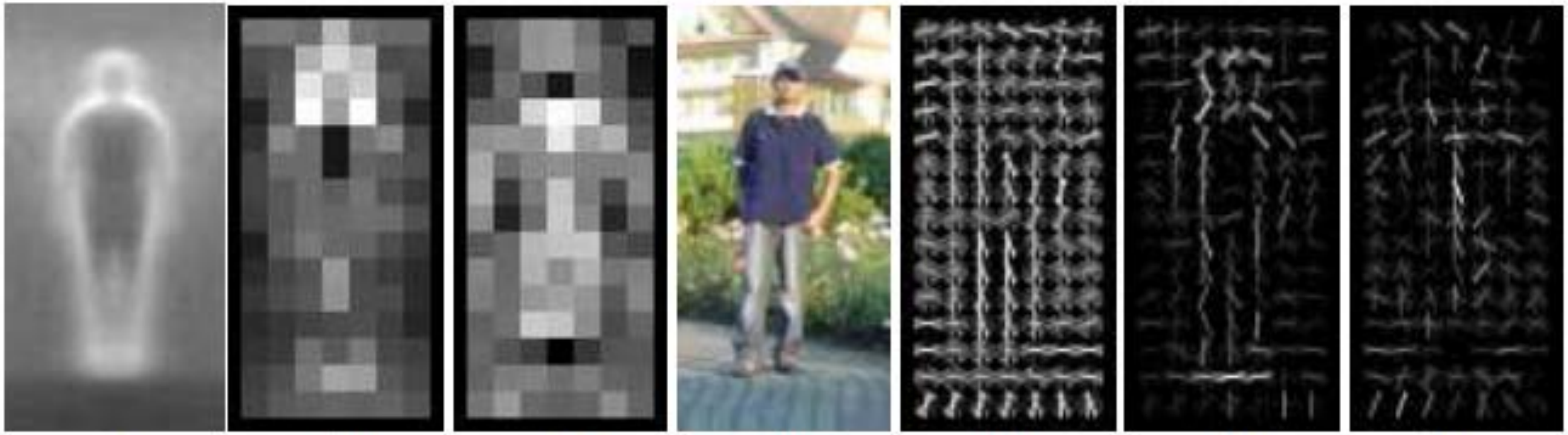
Viola, Jones and Snow, ICCV’03

Pedestrian Detection

Viola, Jones and Snow, ICCV'03

Hogg features

Histograms of Oriented Gradients for Human Detection Navneet Dalal and Bill Triggs, CVPR 2005

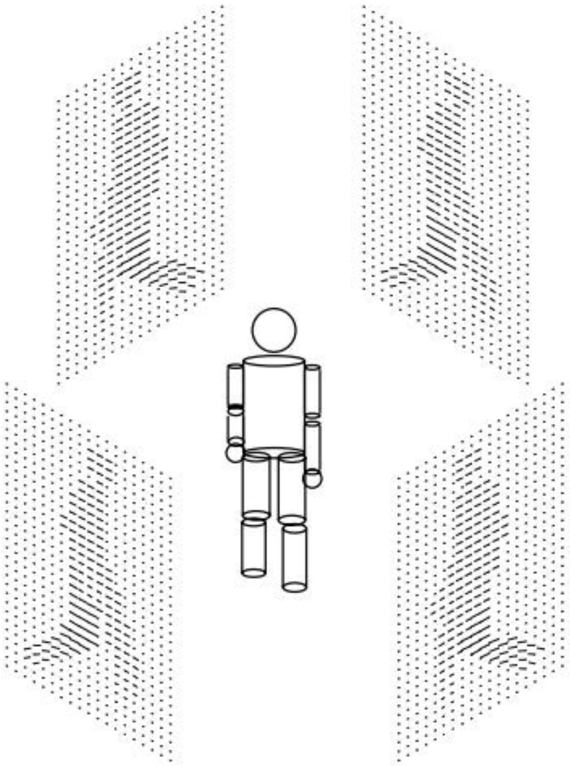


(a) (b) (c) (d) (e) (f) (g)

Figure 6. Our HOG detectors cue mainly on silhouette contours (especially the head, shoulders and feet). The most active blocks are centred on the image background just *outside* the contour. (a) The average gradient image over the training examples. (b) Each "pixel" shows the maximum positive SVM weight in the block centred on the pixel. (c) Likewise for the negative SVM weights. (d) A test image. (e) It's computed R-HOG descriptor. (f,g) The R-HOG descriptor weighted by respectively the positive and the negative SVM weights.

Synthetic data for training

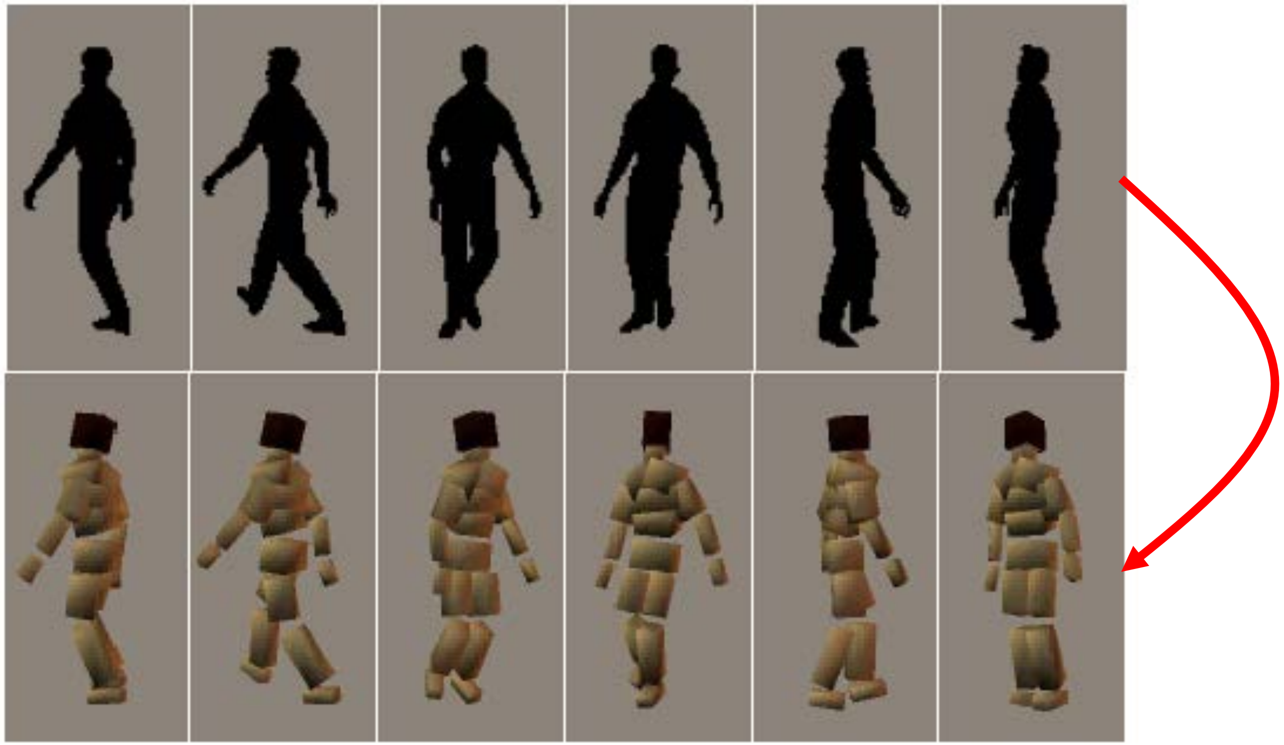
Use graphics to generate data



Learn a view-based model of optical flow and detect human motion, which is different from background motion.

Automatic detection and tracking of human motion with a view-based representation, Fablet, R., Black, M. J.
In European Conf. on Computer Vision, ECCV 2002

Single View to 3D Pose



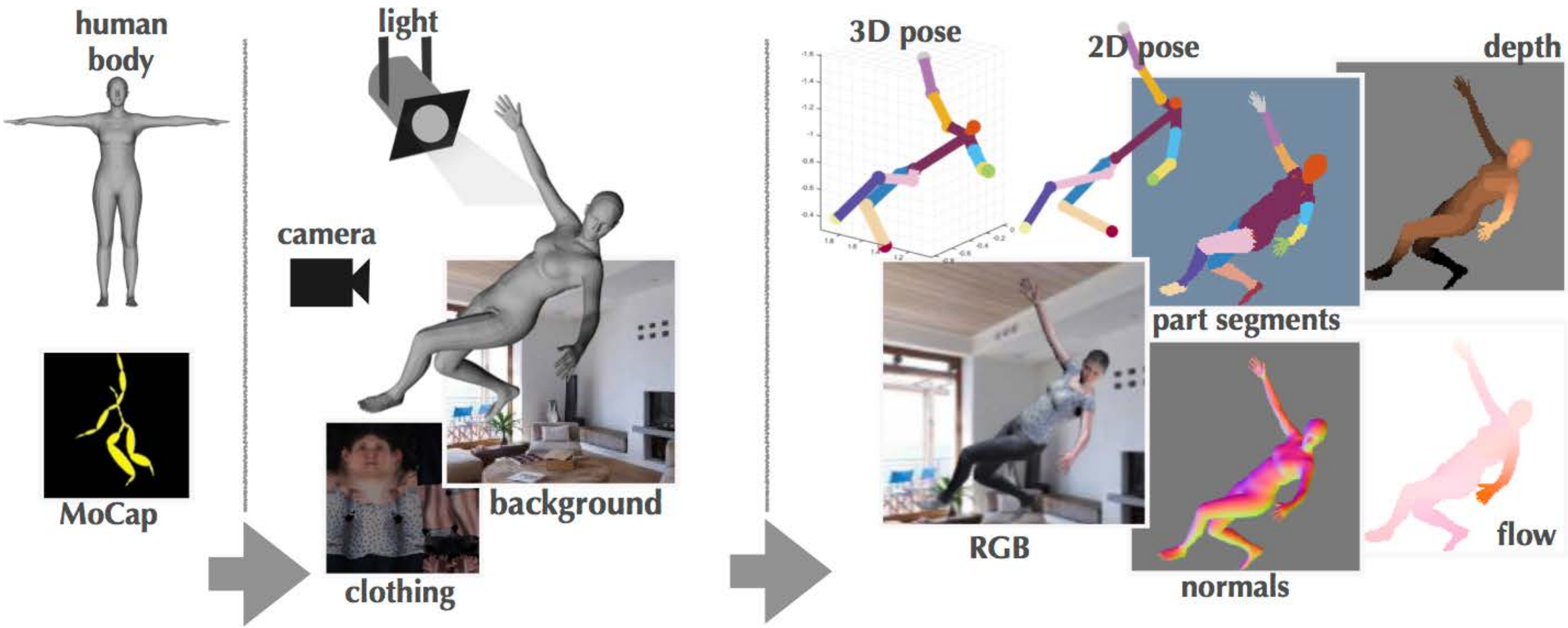
Given synthetic training data, learn the mapping from silhouette contours to 3D pose.

"Gaussian kernel RVM", Agarwal and Triggs CVPR04

"Fast Pose Estimation with Parameter Sensitive Hashing", Shakhnarovich, G., Viola, P., & Darrell, T. CVPR'03.

SURREAL Dataset

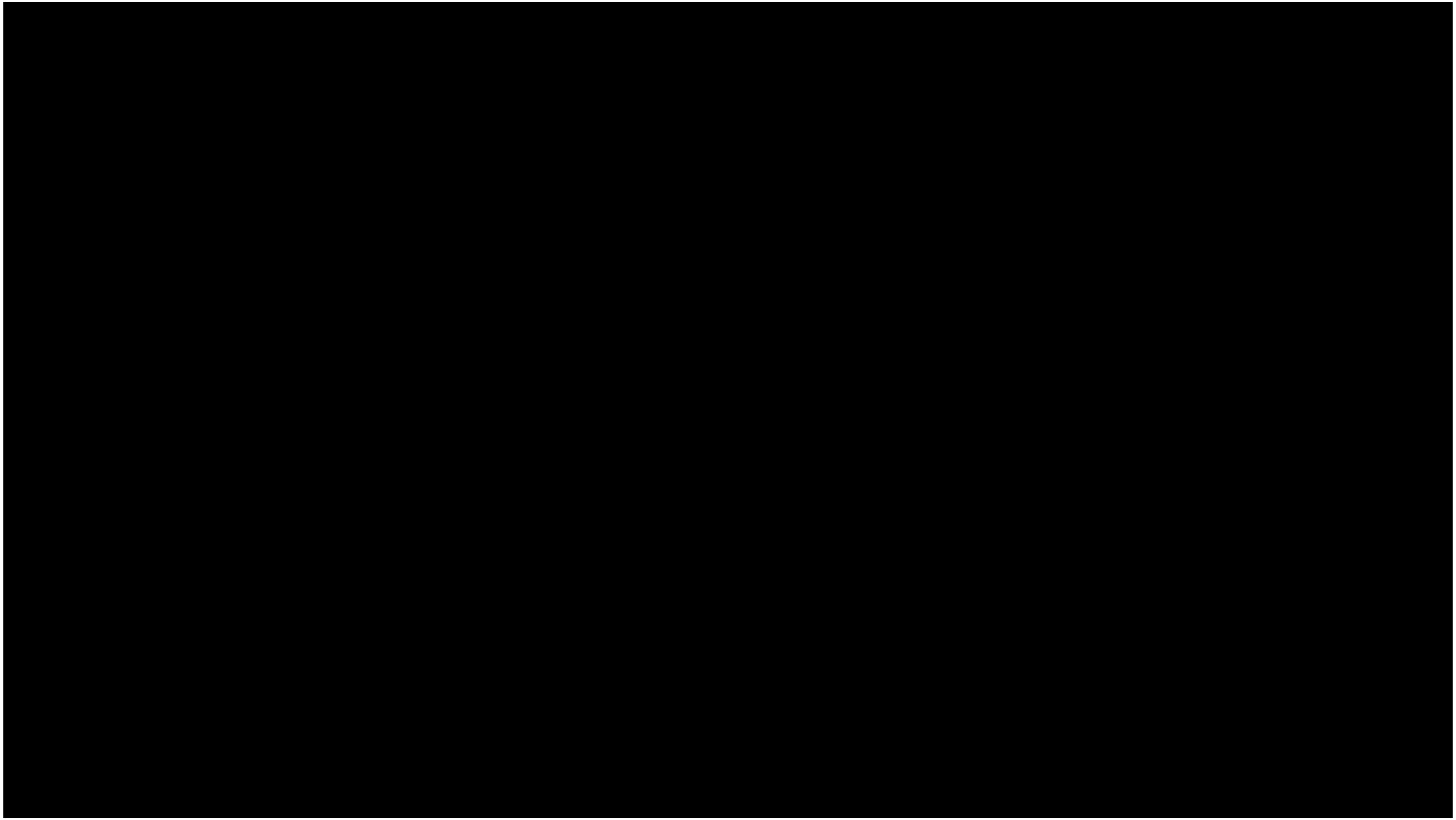
Synthetic hUmans foR REAL tasks



Varol, Romero, Martin, Mahmood, Black, Laptev, Schmid, “Learning from synthetic humans,” CVPR 2017



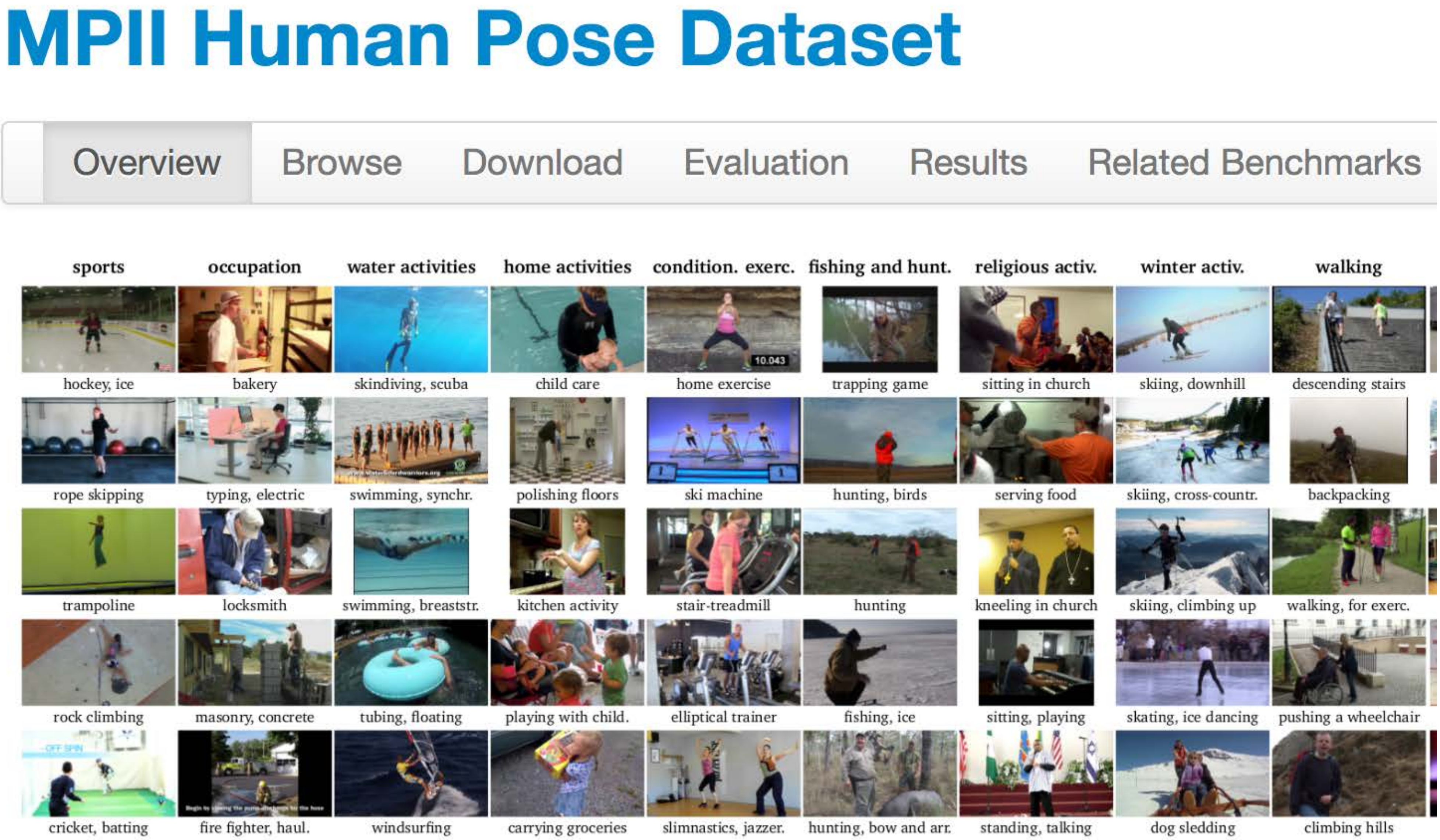
SURREAL Dataset



Varol et al, CVPR 2017

http://www.di.ens.fr/willow/research/surreal

Key innovation:
Mechanical Turk
Have people click on joints



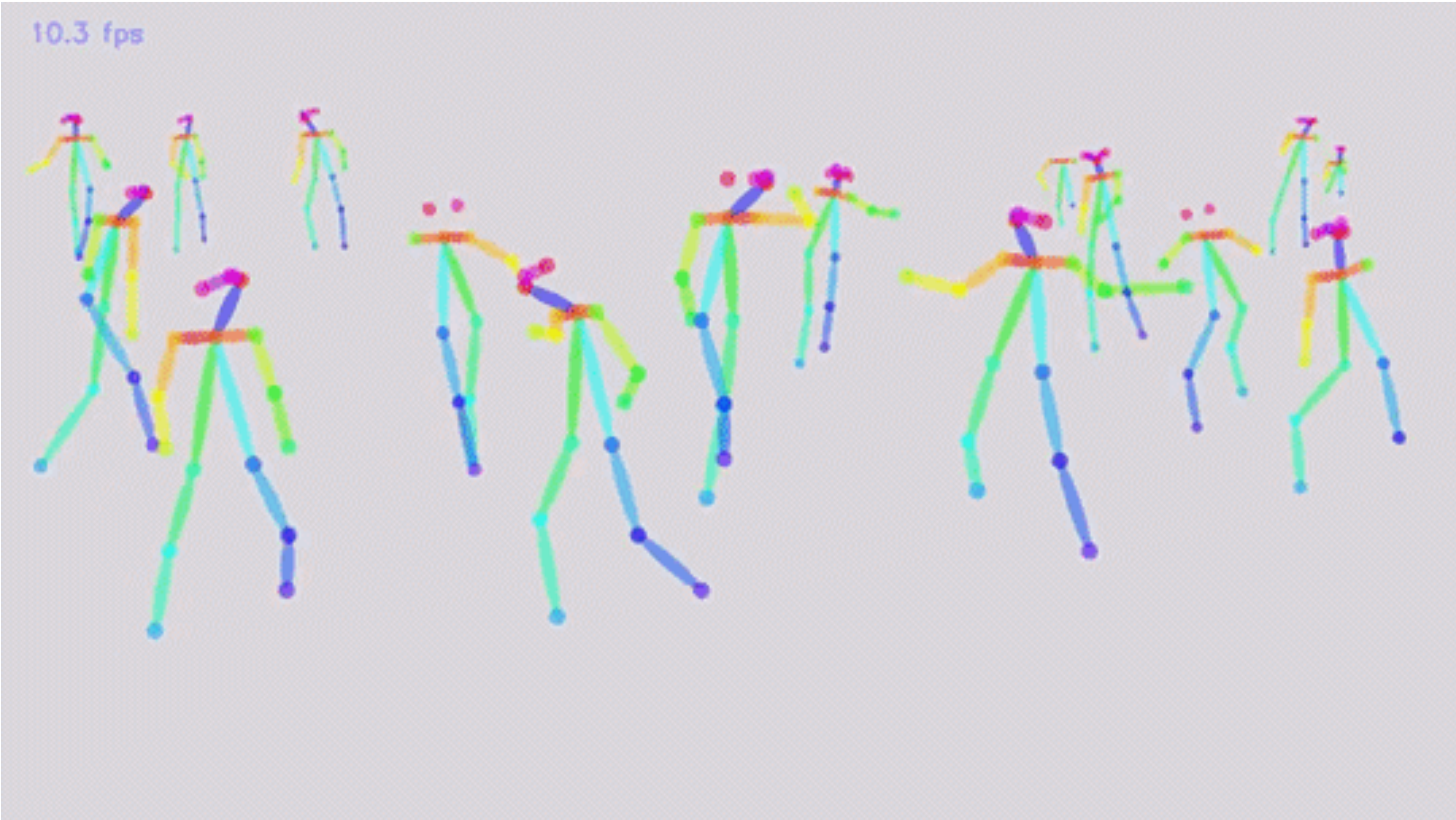
2D Human Pose Estimation: New Benchmark
and State of the Art Analysis, CVPR 2014

Deep learning: 2014-now

- MoDeep: A Deep Learning Framework Using Motion Features for Human Pose Estimation, Jain, Tompson, LeCun, Bregler

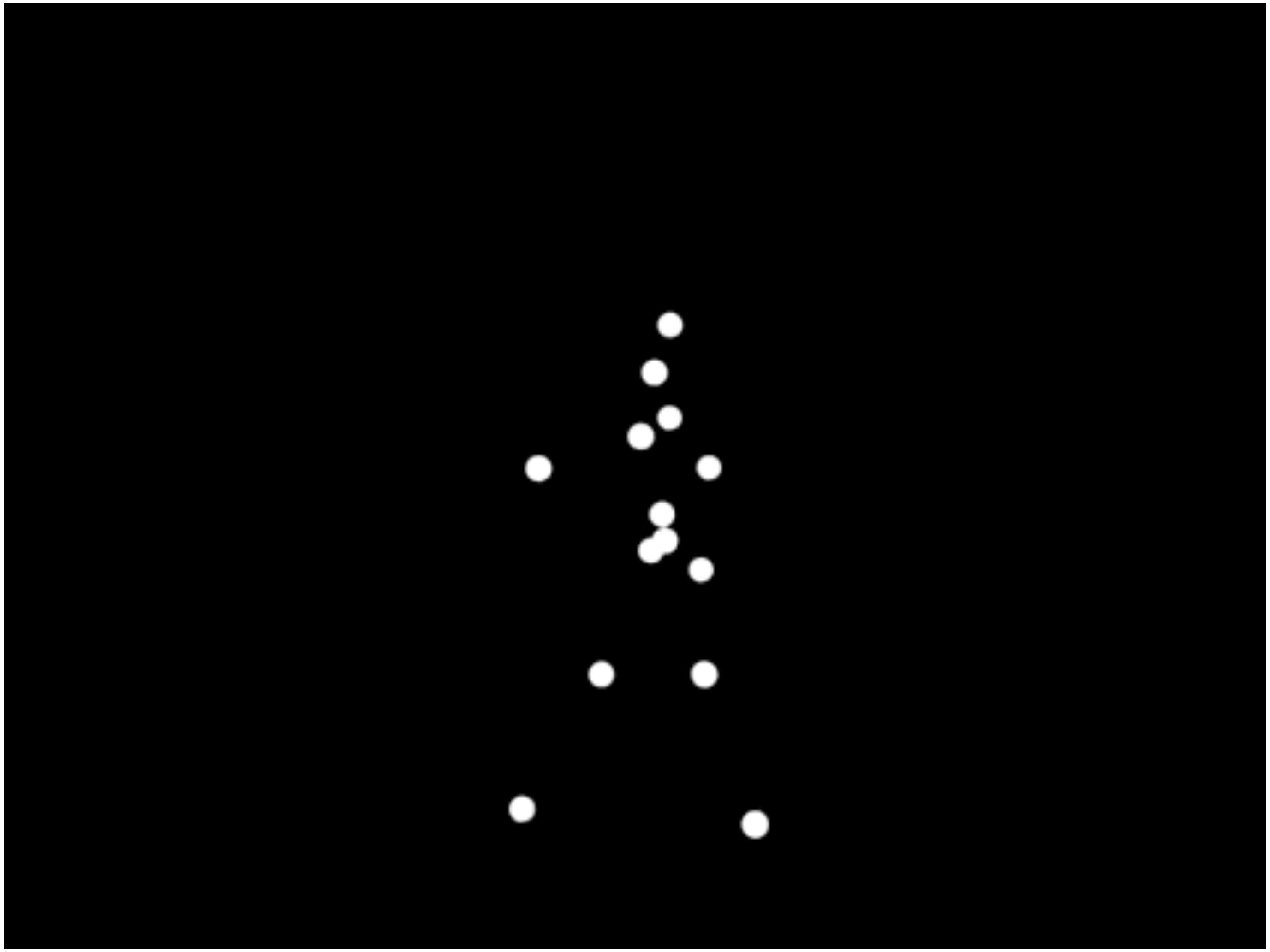
- DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation, Pischulin et al. CVPR 2016

Progress: Bodies as 2D joints



OpenPose, CMU 2017.

Are we our 2D joints?



"…. 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."

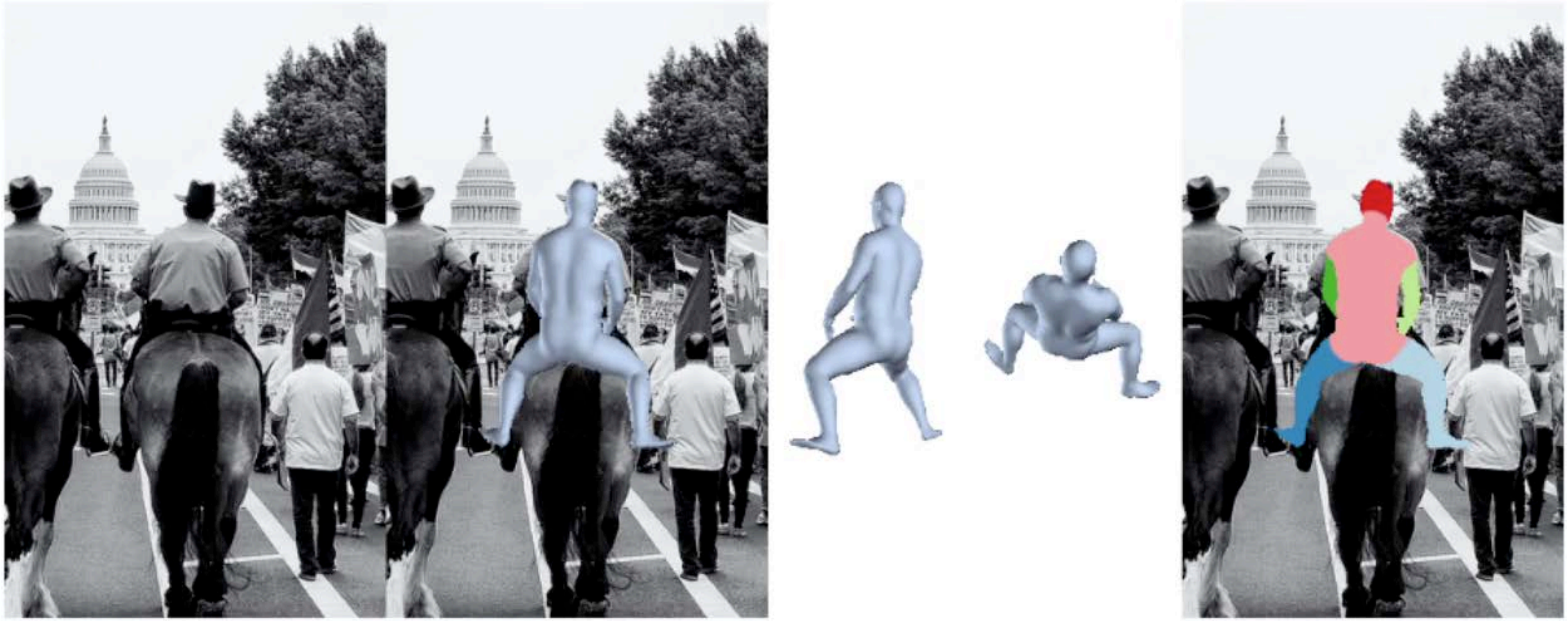
Gunnar Johansson, Visual perception of biological motion and a model for its analysis, Perception & Psychophysics, 1973.

Today: 3D pose and shape

3D pose and shape from 1 image

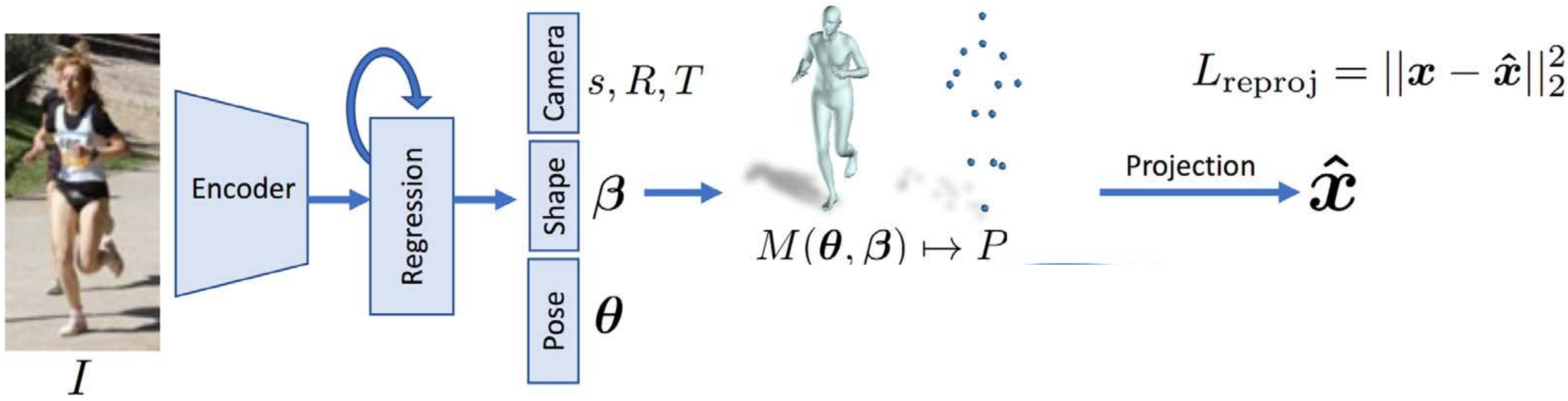
Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image, Bogo, F., et al., ECCV 2016

Problem: No 3D ground truth



Kanazawa, Black, Jacobs, Malik, "End-to-End Recovery of Human Shape and Pose," CVPR 2018

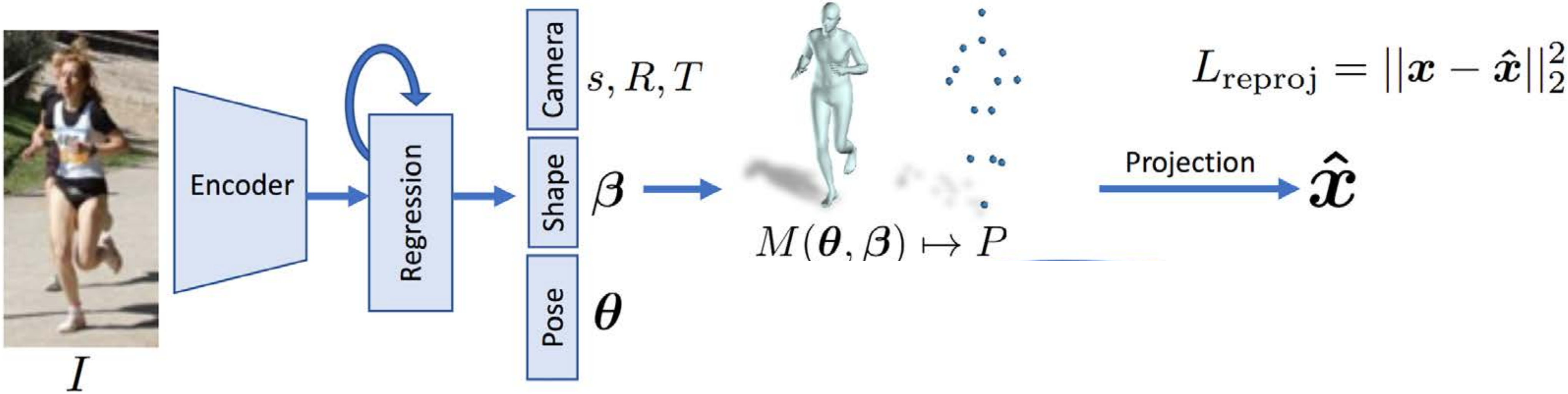
Learning 3D from 2D annotations



- 2D annotations of major joints are easy to get.
- Use them to learn 3D pose and shape from pixels?

Kanazawa, et al., End-to-End Recovery of Human Shape and Pose, CVPR'18

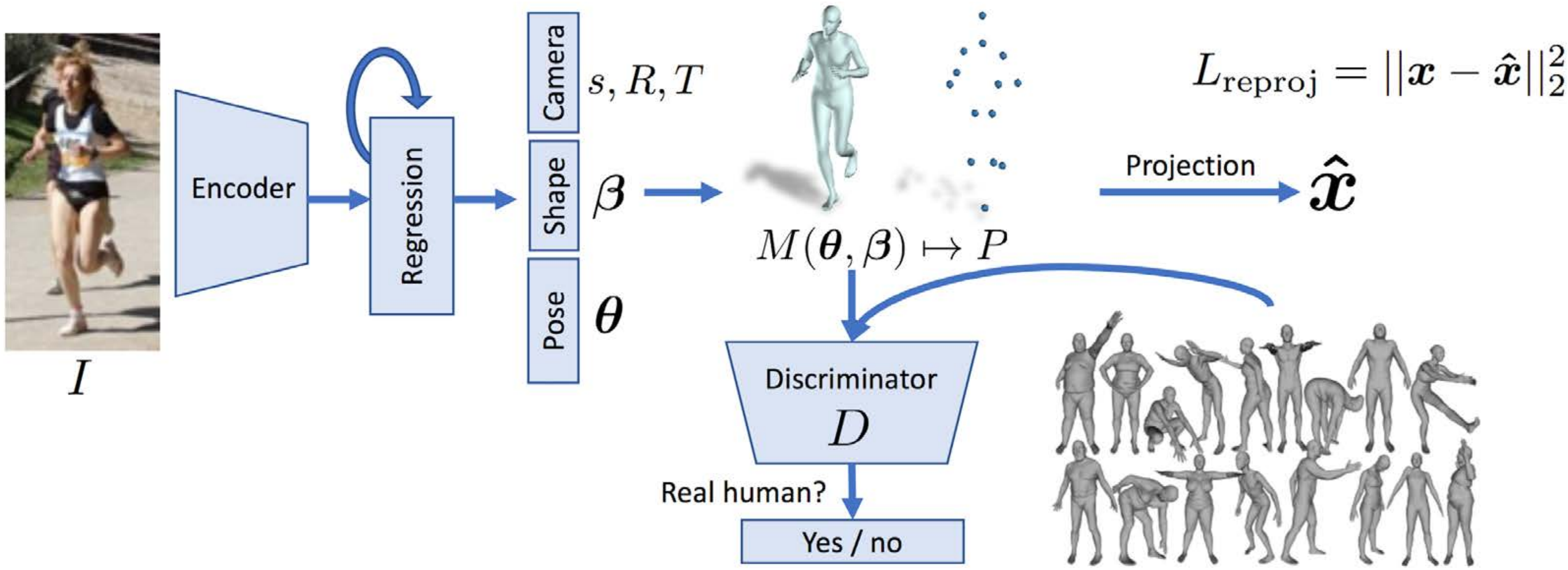
Learning 3D from 2D annotations



Produces monsters.

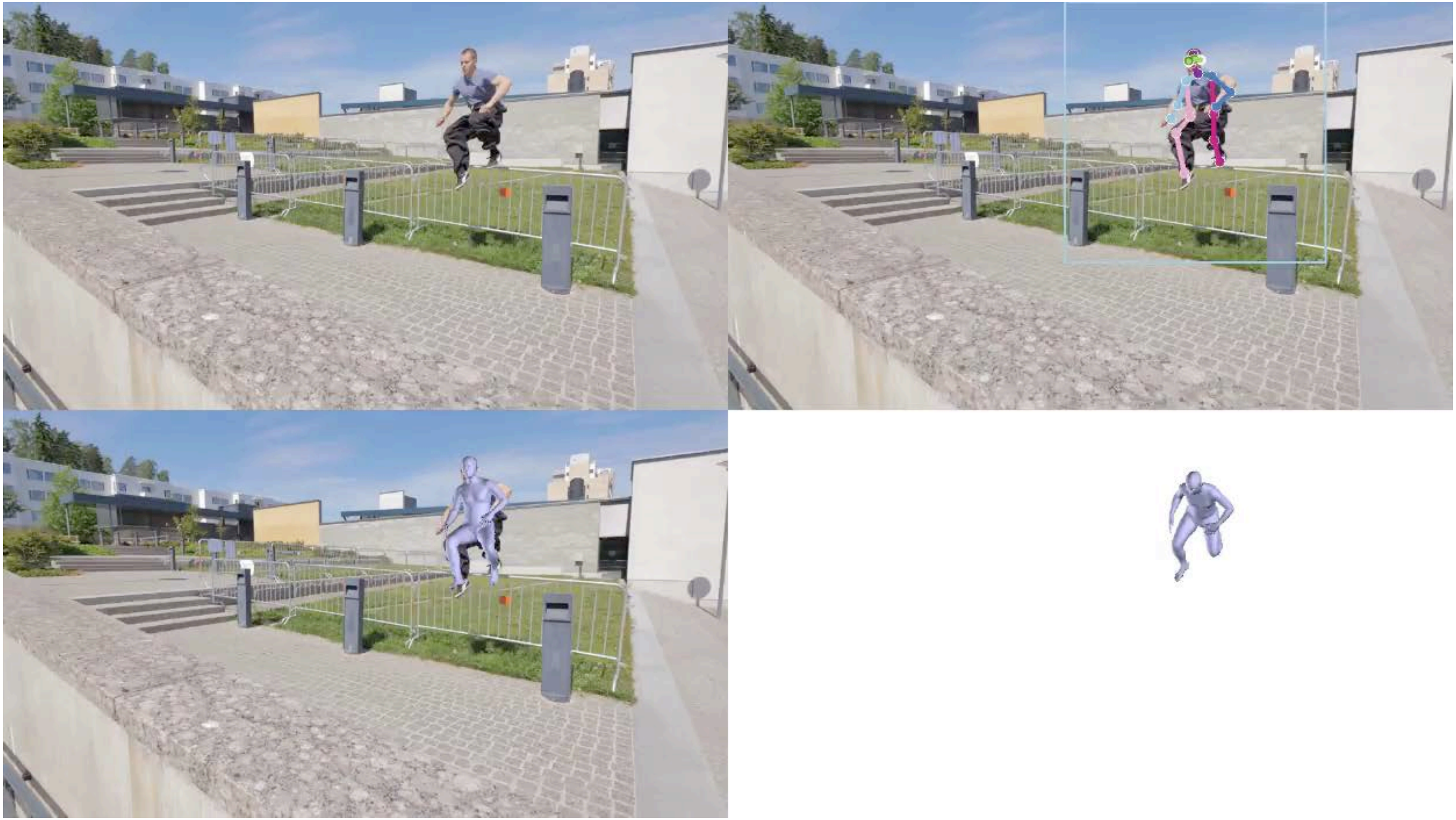
Kanazawa, et al., End-to-End Recovery of Human Shape and Pose, CVPR'18

Learning 3D from 2D annotations



Knowing what humans are (i.e. having a body model) lets you solve pixels to 3D pose without any 3D training data.

Kanazawa, et al., End-to-End Recovery of Human Shape and Pose, CVPR'18



Kanazawa, Black, Jacobs, Malik, “End-to-End Recovery of Human Shape and Pose,” CVPR 2018

Vision is knowing what is where by looking.

Someone's summary of Aristotle

Vision is about perceiving what can’t be seen. It is for interpreting the meaning behind what is visible.

Paraphrased from something I heard Shimon Edelman say.

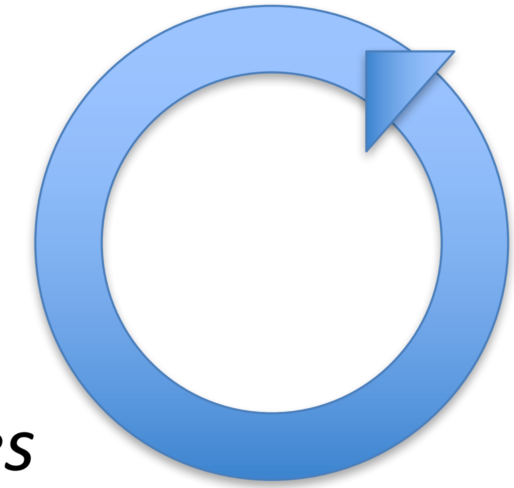







What's our real goal?

Bodies,
Scenes



Goals,
Affordances

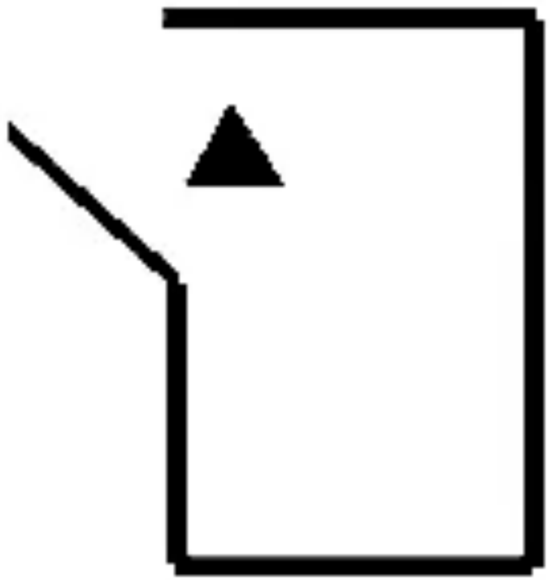
Actions,
Costs

We don't really care about pose per se. Our goal is to infer what can't be seen – the goals, emotions, and the "story".

Adapted from Shimon Edelman

Motion and emotion

Interaction between agents and of agents with the environment



Heider & Simmel, 1944

Speech

Human movement

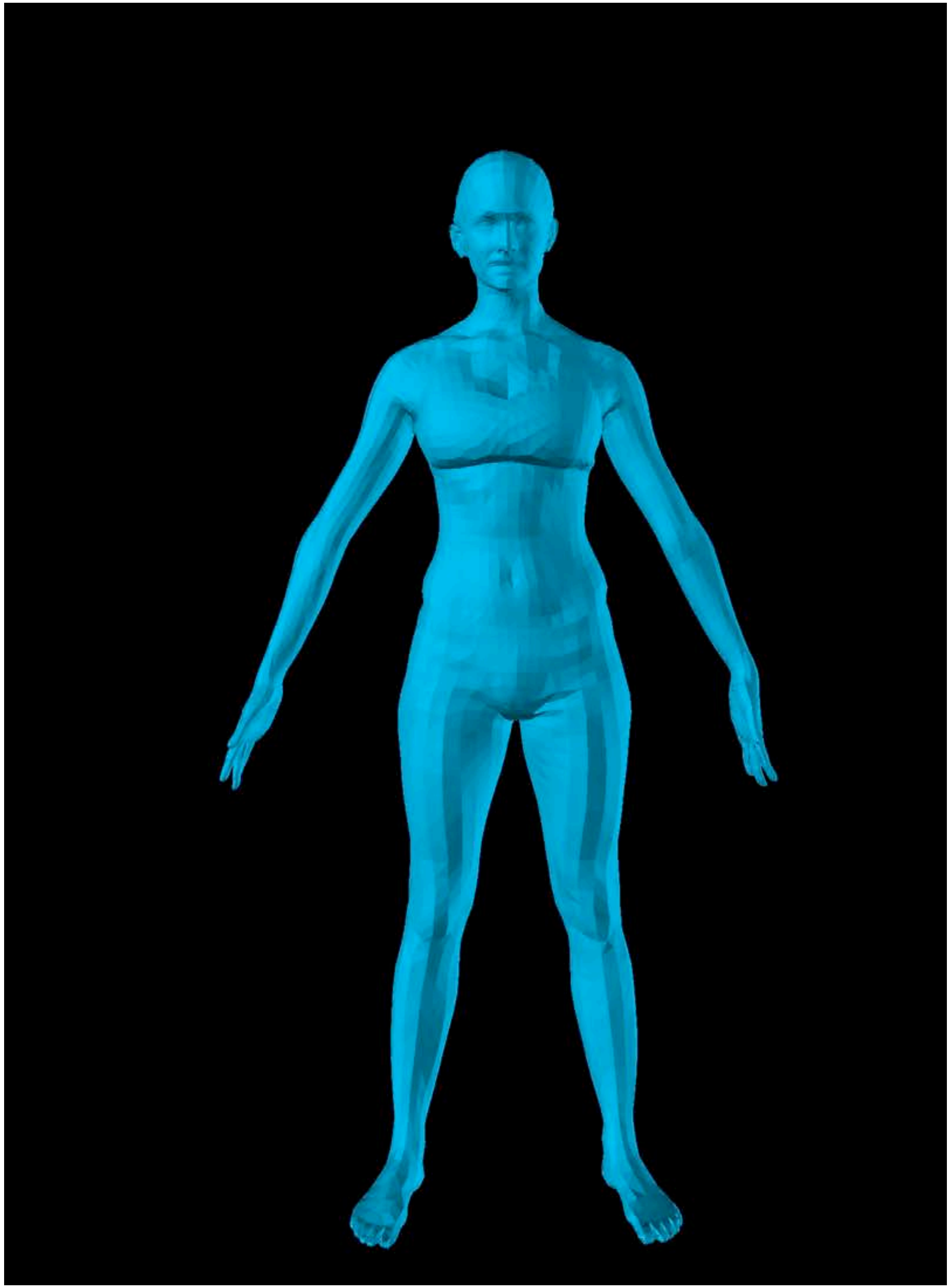
Scene context

B(goals, history, 3D scene, others) ⇨ {speech, movement}

(Warning: AI-complete)

2026

- Realistic bodies with expressive faces, eyes, hands, hair, and clothes.
- Photorealistic, detailed.
- Autonomous agents.
- Interaction with the 3D world and other agents.
- Goals, emotions, speech, communication.



Max Planck Institute for Intelligent Systems

Perceiving Systems Department

http://ps.is.tue.mpg.de

Tübingen, Germany

https://ps.is.tuebingen.mpg.de/code

Early work

- The Representation and Matching of Pictorial Structures, M.A. Fischler ; R.A. Elschlager, IEEE Transactions on Computers, Volume: C-22 , Issue: 1 , Jan. 1973
  - https://ieeexplore.ieee.org/document/1672195
- G. E. Hinton. Using relaxation to find a puppet. In Proc. of the A.I.S.B. Summer Conference, pages 148–157, July 1976.
  - http://files.is.tue.mpg.de/black/papers/HintonPuppet76.pdf
- Marr and Nisihara, Representation and recognition of the spatial organization of three-dimensional shapes, Proc. Royal Soc. B., 1978
  - http://www.cog.brown.edu/courses/cg195/pdf_files/CG195MaNi78.pdf
- Nevatia and Binford, Sturctured descriptions of complex objects, IJCAI 1973
  - https://www.semanticscholar.org/paper/Structured-Descriptions-of-Complex-Objects-Nevatia-Binford/638693c63b7788133b0d0541cd65550ce91c20dd

Early work

- Alex Pentland and Bradley Horowitz, Recovery of Nonrigid Motion and Structure, PAMI, VOL. 13, NO. 7, JULY 1991
  - https://www.computer.org/csdl/trans/tp/1991/07/i0730.pdf
- K. Rohr, Towards Model-Based Recognition of Human Movements in Image Sequences, CVGIP: Image Understanding, Volume 59, Issue 1, January 1994, Pages 94-115
  - https://www.sciencedirect.com/science/article/pii/S1049966084710060?via%3Dihub
- Wachter & Nagel, Tracking of Persons in Monocular Image Sequences, Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE
  - https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=609843
- Bregler & Malik, Tracking People with Twists and Exponential Maps Christoph Bregler and Jitendra Malik, CVPR 1998.
  - https://people.eecs.berkeley.edu/~malik/papers/bregler-malik98.pdf

Early work

- Model-based vision: A program to see a walking person, D Hogg, Image and Vision computing 1 (1), 5-20
  - https://www.sciencedirect.com/science/article/pii/0262885683900033?via%3Dihub
- D. Gavrila, Vision-based 3-D Tracking of Humans in Action, Ph.D. thesis
  - http://www.gavrila.net/thesis.pdf
- Cardboard people: A parameterized model of articulated motion, Ju, S. X., Black, M. J., Yacoob, Y. Face and Gesture 1996.
  - http://files.is.tue.mpg.de/black/papers/fg96.pdf

Stochastic estimation

- Hedvig Sidenbladh, Michael J. Black, David J. Fleet, Stochastic Tracking of 3D Human Figures Using 2D Image Motion, ECCV 2000
  - http://files.is.tue.mpg.de/black/papers/eccv00.pdf
- A multiple hypothesis approach to figure trackingTJ Cham, JM Rehg, CVPR 1999.
  - http://www.hpl.hp.com/techreports/Compaq-DEC/CRL-98-8.pdf
- Tracking through singularities and discontinuities by random sampling J. Deutscher, B. North, B. Bascle and A. Blake, ICCV 1144-1149 (1999).
  - http://www.robots.ox.ac.uk/~vdg/abstracts/iccv99-deutscher.html
- Covariance Scaled Sampling for Monocular 3D Body Tracking Cristian Sminchisescu, Bill Triggs, CVPR 2001
  - https://hal.inria.fr/file/index/docid/548273/filename/Sminchisescu-cvpr01.pdf

Pose priors

- Ormoneit, Sidenbladh, Black, Hastie, Learning and Tracking Cyclic Human Motion, NIPS 2001
  - http://files.is.tue.mpg.de/black/papers/NIPS13.pdf
- 3D People Tracking with Gaussian Process Dynamical Models, Urtasun, Fleet, Fua, CVPR 2006
  - http://ttic.uchicago.edu/~rurtasun/publications/urtasun_et_al_cvpr06.pdf
- Modeling Human Motion Using Binary Latent Variables Graham W. Taylor, Geoffrey E. Hinton and Sam Roweis, NIPS 2007
  - http://www2.egr.uh.edu/~zhan2/ECE6111_Fall2015/modeling%20human%20motion%20using%20binary%20latent%20variables.pdf

Belief propagation

- Pedro F. Felzenszwalb, Daniel P. Huttenlocher, Pictorial Structures for Object Recognition, IJCV, January 2005, Volume 61, Issue 1, pp 55–79
  - https://link.springer.com/article/10.1023/B:VISI.0000042934.15159.49
- Loose-limbed People: Estimating 3D Human Pose and Motion Using Non-parametric Belief Propagation, Sigal, L., Isard, M., Haussecker, H., Black, M. J., IJCV, 98(1):15-48, May 2011
  - http://www.springerlink.com/content/h6524h1n0qw5tv07/fulltext.pdf
  -

Ground truth datasets

- HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion, Sigal, L., Balan, A., Black, M. J.
  - http://files.is.tue.mpg.de/black/papers/EHuM_Journal_webversion.pdf
  - http://humaneva.is.tue.mpg.de/
- Catalin Ionescu, Dragos Papava, Vlad Olaru and Cristian Sminchisescu, Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments, PAMI 2014
  - http://vision.imar.ro/human3.6m/description.php
- 3D Poses in the Wild Dataset. Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera, von Marcard and Henschel and Black and Rosenhahn and Pons-Moll, ECCV 2018
  - http://virtualhumans.mpi-inf.mpg.de/3DPW/
- MPII Human Pose Dataset
  - http://human-pose.mpi-inf.mpg.de/

Early body shape models

- Tracking and Modeling People in Video Sequences, Ralf Plänkers and Pascal Fua, Computer Vision and Image Understanding, Volume 81, Issue 3, March 2001, Pages 285-302
  - https://www.sciencedirect.com/science/article/pii/S1077314200908919
- Model-Based Estimation of 3D Human Motion Ioannis Kakadiaris, and Dimitris Metaxas, PAMI VOL. 22, NO. 12, Dec 2000
  - http://www.cbim.rutgers.edu/dmdocuments/21%20Kakadiaris%20IEEE.pdf
- Blanz and Vetter, A Morphable Model For The Synthesis Of 3D Faces, SIGGRAPH 1999
  - http://gravis.dmi.unibas.ch/publications/Sigg99/morphmod2.pdf

Learning body shape

- CAESAR dataset
  - http://store.sae.org/caesar/
- The space of human body shapes:
  reconstruction and parameterization from range scans, Brett Allen, Brian Curless, Zoran Popović, SIGGRAPH 2003
  - http://grail.cs.washington.edu/projects/digital-human/pub/allen03space.html
- Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis, Brett Allen, Brian Curless, Zoran Popovic , and Aaron Hertzmann, SCA 2006
  - http://grail.cs.washington.edu/projects/digital-human/pub/allen06learning.html
- Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. 2005. SCAPE: shape completion and animation of people. ACM Trans. Graph. 24, 3 (July 2005)
  - https://ai.stanford.edu/~drago/Projects/scape/scape.html

Learning body shape

- A Statistical Model of Human Pose and Body Shape N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel, EUROGRAPHICS 2009
  - https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2009.01373.x
- (Tenbo) Tensor-Based Human Body Modeling Yinpeng Chen Zicheng Liu Zhengyou Zhang, CVPR 2013
- http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.679.7064&rep=rep1&type=pdf
- (BlendSCAPE) Coregistration: Simultaneous Alignment and Modeling of Articulated 3D Shape David A. Hirshberg, Matthew Loper, Eric Rachlin, and Michael J. Black, ECCV 2012
  - http://files.is.tue.mpg.de/black/papers/HirshbergECCV2012.pdf
- SMPL: A Skinned Multi-Person Linear Model, Loper et al., SIGGRAPH Asia 2015
  - http://smpl.is.tue.mpg.de/

Evolution of body models

- (soft tissue) Dyna: A Model of Dynamic Human Shape in Motion, Pons-Moll et al, SIGGRAPH 2015
  - http://files.is.tue.mpg.de/black/papers/dyna.pdf
- (clothing) ClothCap: Seamless 4D Clothing Capture and Retargeting, Pons-Moll, G., Pujades, S., Hu, S., Black, M.J., SIGGRAPH, 2017.
  - http://clothcap.is.tue.mpg.de/
- (infants) Hesse, et al., Learning an Infant Body Model from RGB-D Data for Accurate Full Body Motion Analysis, MICCAI 2018
  - http://files.is.tue.mpg.de/black/papers/miccai18.pdf

RGB-D

- Real-Time Human Pose Recognition in Parts from Single Depth Images, Shotton et al., CVPR 2011
  - https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/BodyPartRecognition.pdf
- Home 3D body scans from noisy image and range data, Weiss, A., Hirshberg, D., Black, M. ICCV 2011
  - http://files.is.tue.mpg.de/black/papers/KinectICCV2011.pdf
- Detailed Full-Body Reconstructions of Moving People from Monocular RGB-D Sequences, Bogo, F., Black, M. J., Loper, M., Romero, J., ICCV 2015
  - https://ps.is.tuebingen.mpg.de/uploads_file/attachment/attachment/235/2262.pdf

Shape and pose from images

- Detailed Human Shape and Pose from Images, Balan, A., Sigal, L., Black, M. J., Davis, J., Haussecker, H., CVPR 2007
  - http://files.is.tue.mpg.de/black/papers/balan07imscape.pdf
- Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image, Bogo et al., ECCV 2016,
  - http://smplify.is.tuebingen.mpg.de/
- End-to-end Recovery of Human Shape and Pose, Kanazawa, et al., CVPR 2018
  - https://akanazawa.github.io/hmr/

2D deep pose from images

- MoDeep: A Deep Learning Framework Using Motion Features for Human Pose Estimation, Arjun Jain, Jonathan Tompson, Yann LeCun and Christoph Bregler,
  - https://arxiv.org/pdf/1409.7963.pdf
- DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. Pishchulin et al. CVPR 2016
  - https://arxiv.org/abs/1511.06645
- OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, CVPR 2017
  - https://arxiv.org/abs/1611.08050

Early 2D ML methods

- “Pedestrian detection using wavelet templates,” Oren *et al* CVPR’97.
  - https://dl.acm.org/citation.cfm?id=794507
- Detecting pedestrians using patterns of motion and appearance, Viola, Jones and Snow, ICCV’03
  - https://ieeexplore.ieee.org/document/1238422
- Histograms of Oriented Gradients for Human Detection Navneet Dalal and Bill Triggs, CVPR 2005
  - https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf

Synthetic training

- Automatic Detection and Tracking of Human Motion with a View-Based Representation Ronan Fablet and Michael J. Black, ECCV 2002
  - http://files.is.tue.mpg.de/black/papers/23500476.pdf
- 3D Human Pose from Silhouettes by Relevance Vector Regression Ankur Agarwal, Bill Triggs, CVPR04
  - https://hal.inria.fr/inria-00548551/document
- “Fast Pose Estimation with Parameter Sensitive Hashing”, Shakhnarovich, G., Viola, P., & Darrell, T. ICCV’03.
  - http://ttic.uchicago.edu/~gregory/papers/iccv2003.pdf
- Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera, von Marcard et al., ECCV 2018
  - http://virtualhumans.mpi-inf.mpg.de/3DPW/
- Learning from Synthetic Humans, Varol, G. et al., CVPR 2017
  - http://www.di.ens.fr/willow/research/surreal/

Understanding behavior

- An Experimental Study of Apparent Behavior, Fritz Heider and Marianne Simmel, The American Journal of Psychology, Vol. 57, No. 2 (Apr., 1944), pp. 243-259
  - https://www.jstor.org/stable/1416950?seq=1#metadata_info_tab_contents
- G. Johansson, Visual perception of biological motion and a model for its analysis, Perception & Psychophysics, June 1973, Volume 14, Issue 2, pp 201–211
  - https://link.springer.com/article/10.3758/BF03212378