A history of human motion in Computer Vision: From puppets to large language models

Michael J. Black

Max Planck Institute for Intelligent Systems

June 2024

The beginning: 48 years ago

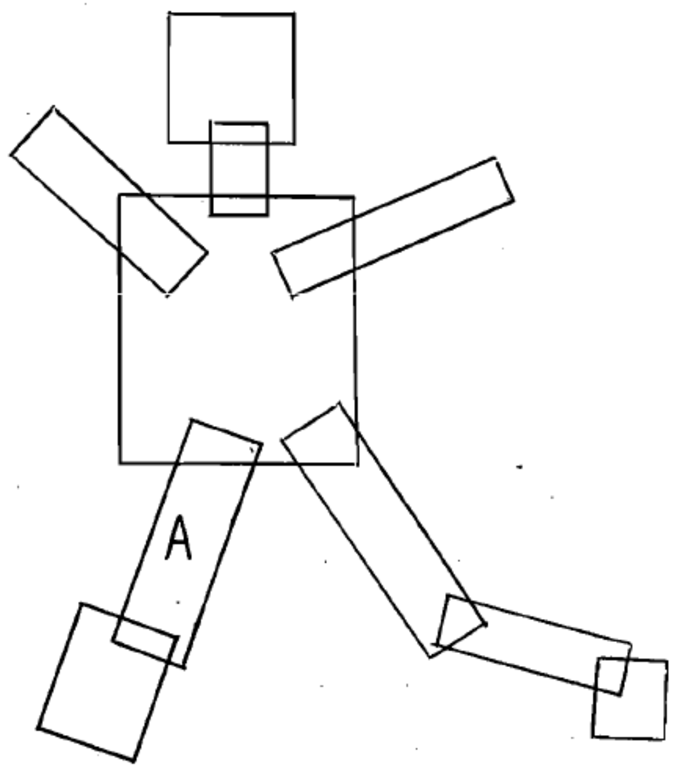


Figure 4. Relaxation picks out the interpretation of A as a thigh even though a calf is a locally better alternative.

The beginning: 48 years ago

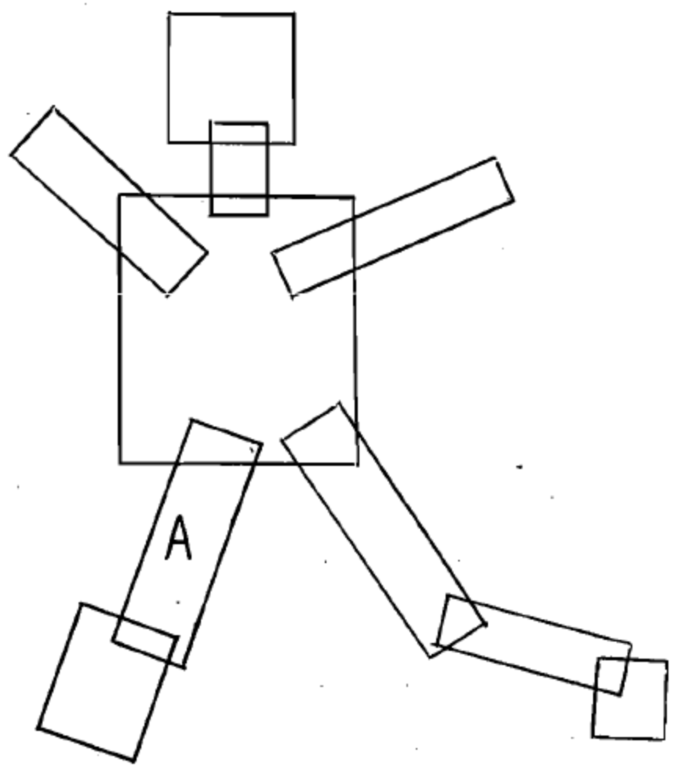


Figure 4. Relaxation picks out the interpretation of A as a thigh even though a calf is a locally better alternative.

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!

Dealing with ambiguity in image evidence:

- Body representation
  - 2D & 3D primitives
  - Learned, parametric
- Search methods
  - Relaxation
  - Optimization
  - Stochastic search
- Priors
  - Hand constructed
  - Learned from data
- Machine learning (training data)
  - Classification
  - Regression

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.

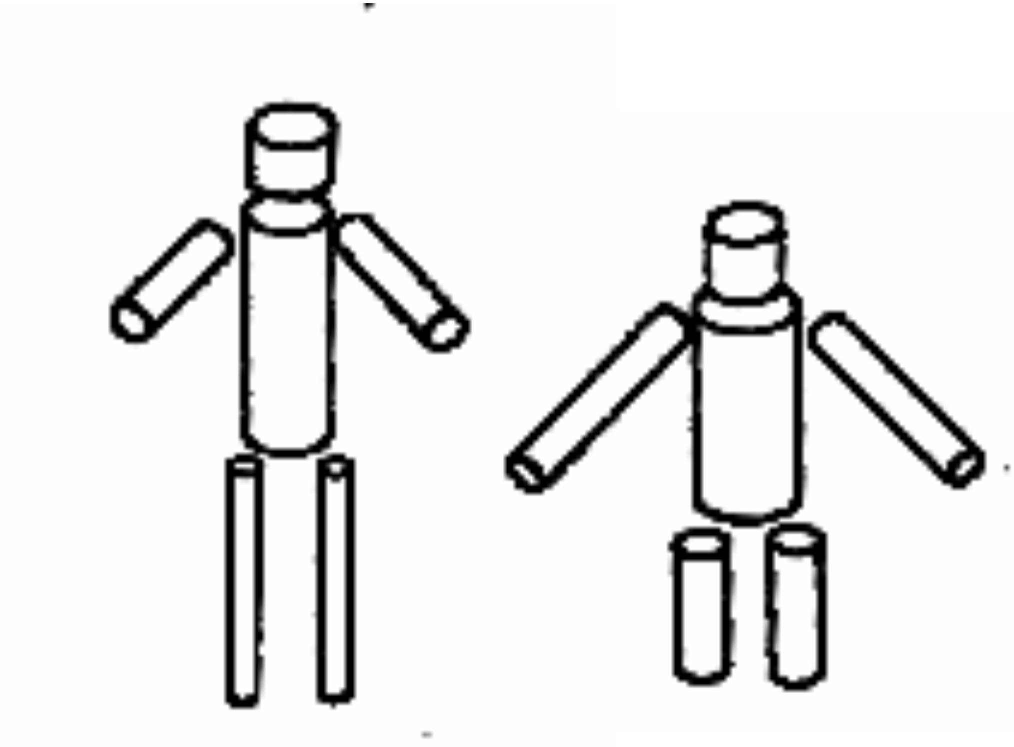
3. It produces the best global interpretation, not just a good one as in heuristic search.

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

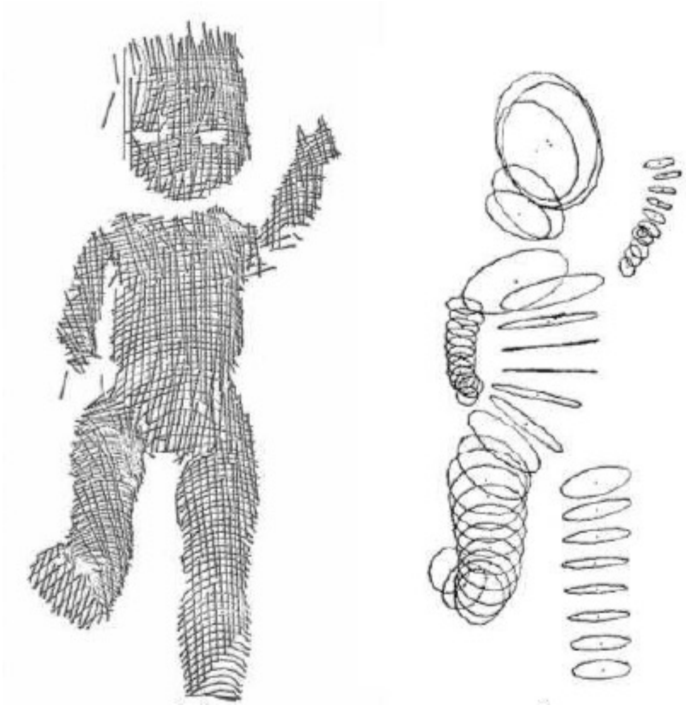
The puppet, which is always depicted in side view, consists of fifteen rectangular parts having the following properties and

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 in 1973!

David Hogg, 1983

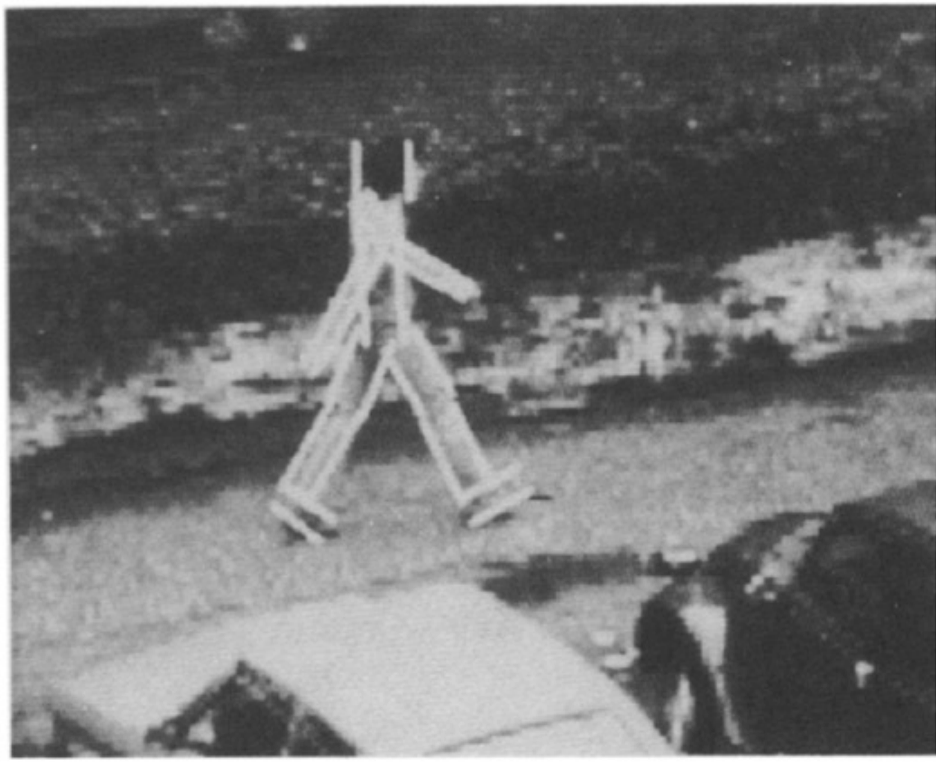


Figure 12. Set of lines which correspond to the imag projections of occluding surfaces. They represent the image in Figure 4

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

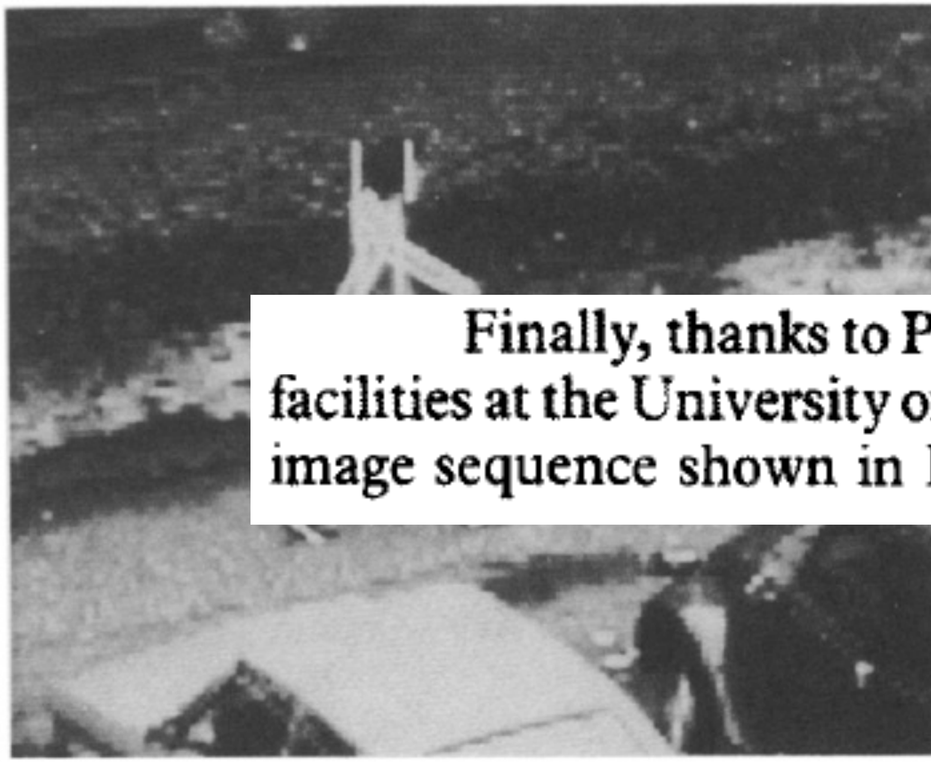


Figure 12. Set of lines which correspond to the imag[e] projections of occluding surfaces. They represent the image in Figure 4

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

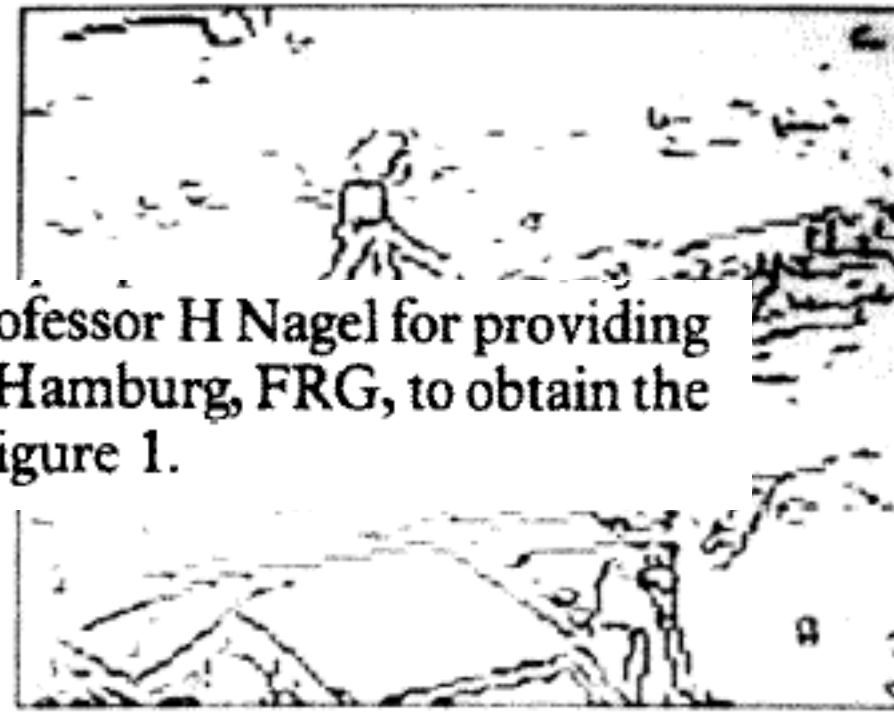


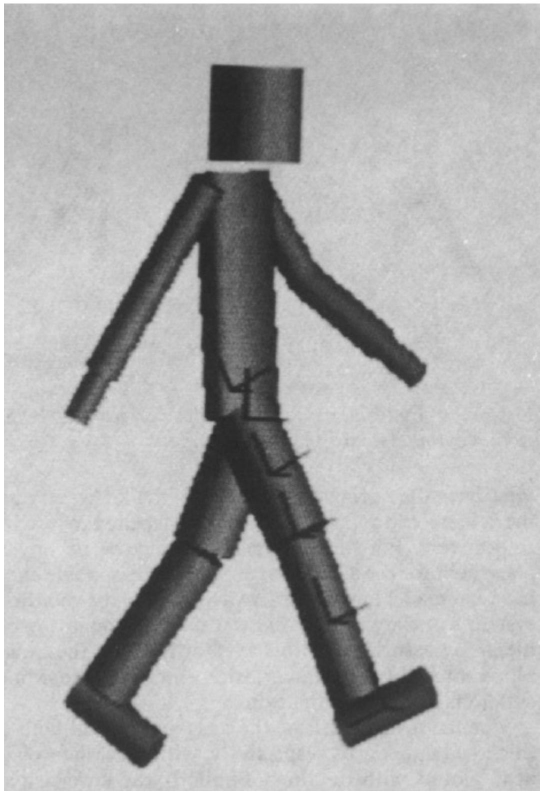
Figure 5. Edge-finding operation applied to the image in Figure 4

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
            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

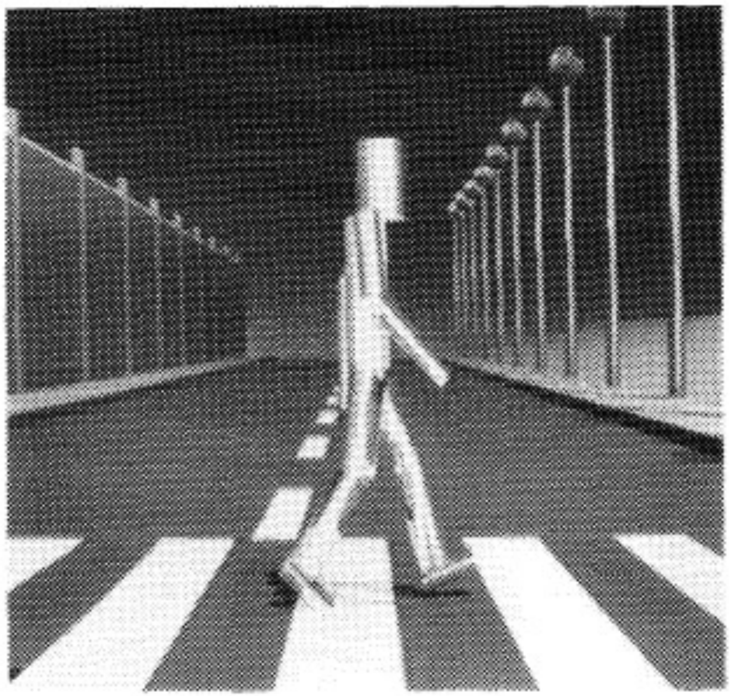


FIG. 4. Model of the human body.

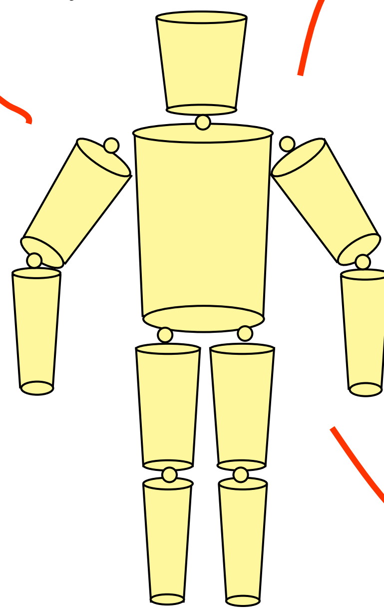
FIG. 20. Determined motion state.

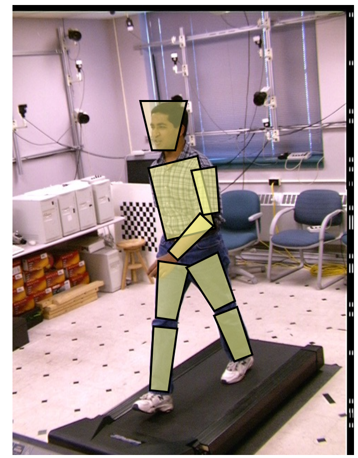
Rohr, Towards Model-Based Recognition of Human Movements in Image Sequences, CVGIP, 1994

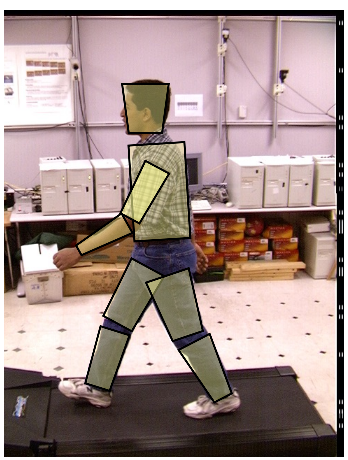
Analysis by synthesis

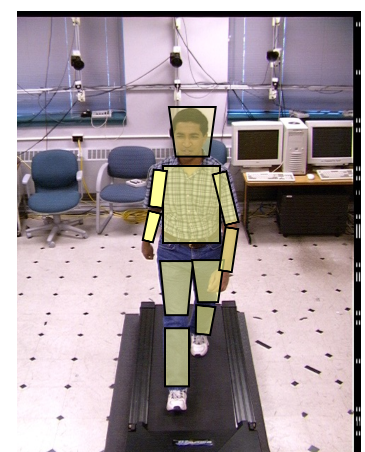
Optimize the pose

$$\theta_t$$



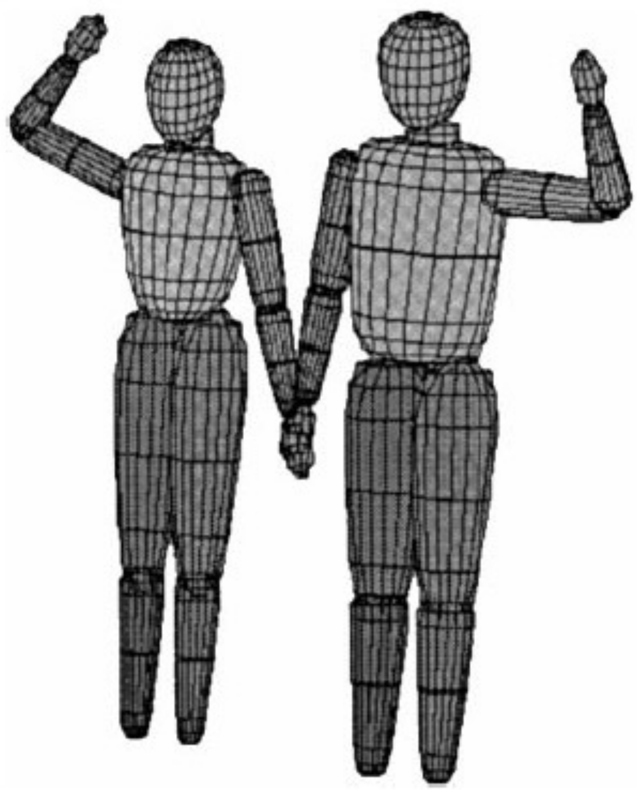


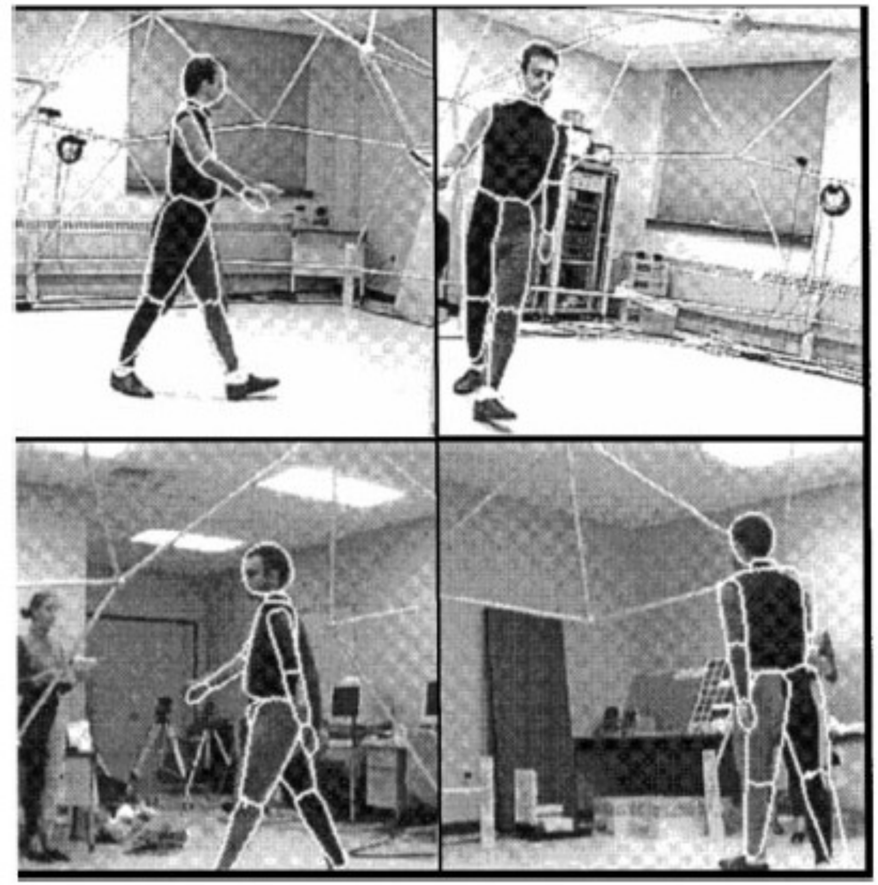




such that the projection "matches" the image data (edges, regions, color, texture…).

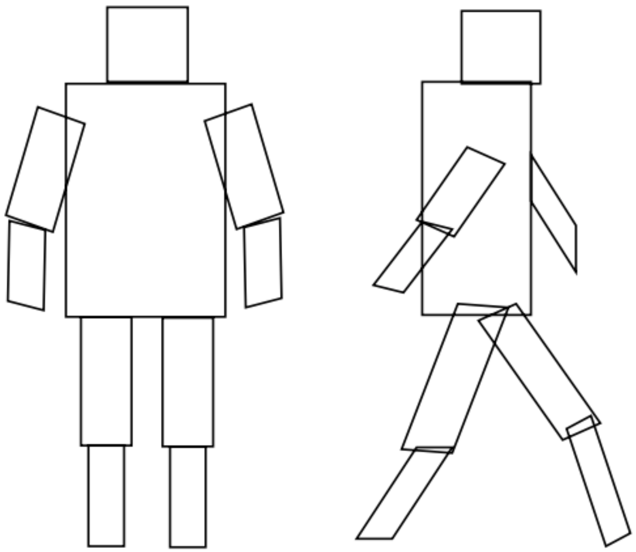
Multi-camera, markerless, mocap





Simple shapes, multi-camera, special clothing.

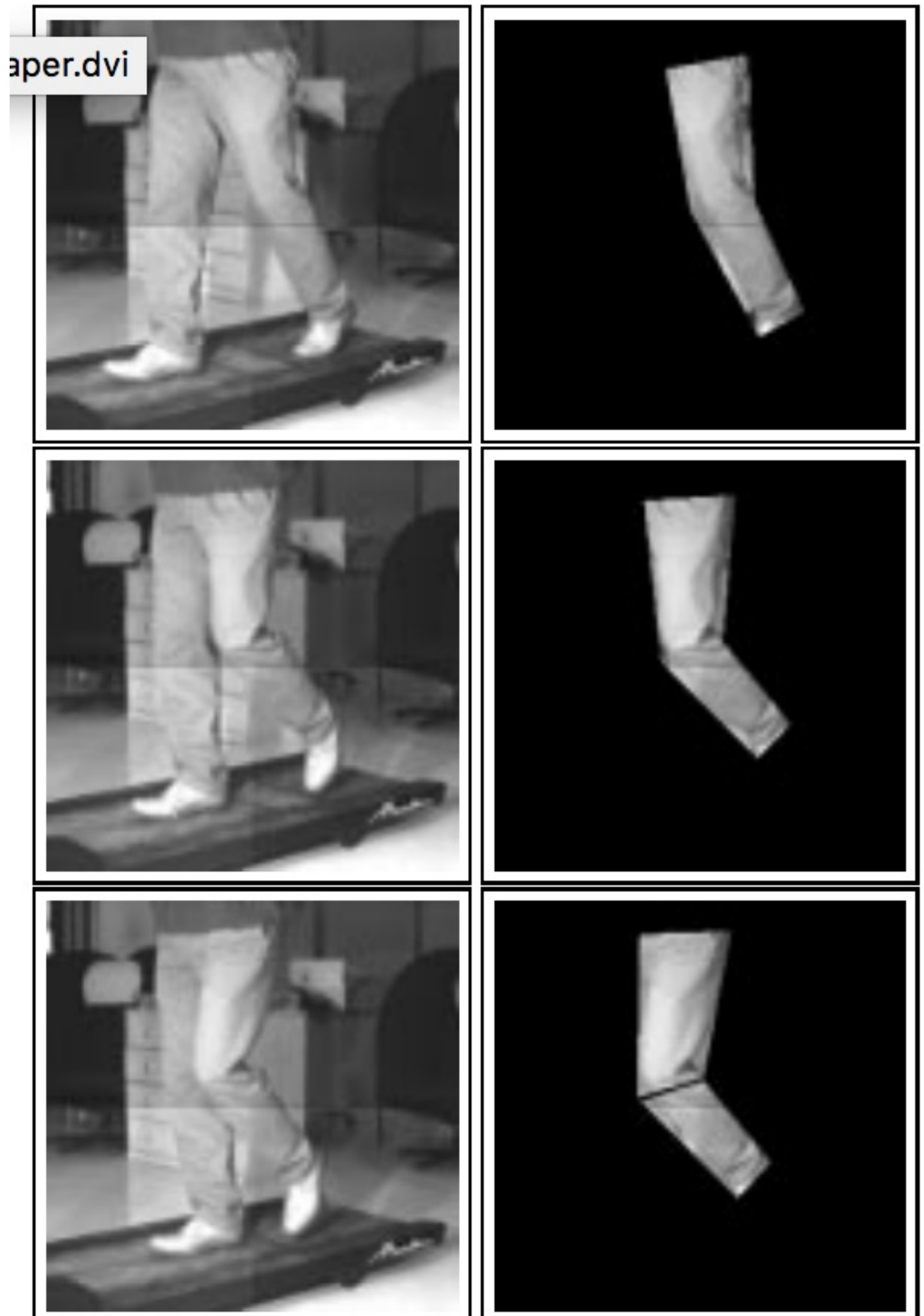
D. Gavrila, Vision-based 3-D Tracking of Humans in Action, Ph.D. thesis, 1996.



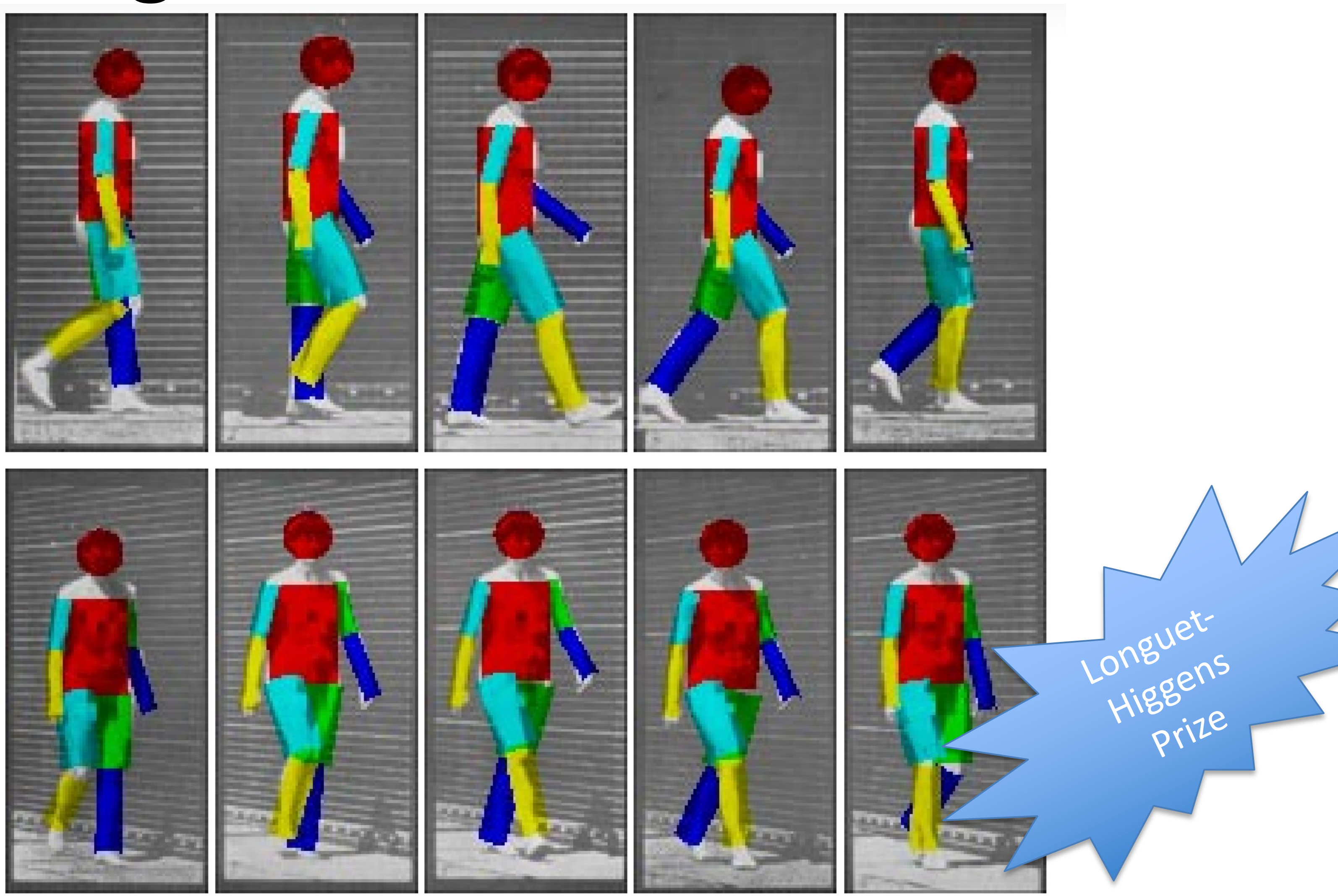
Treat the problem as an optical flow problem

Cardboard people: A parameterized model of articulated motion

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



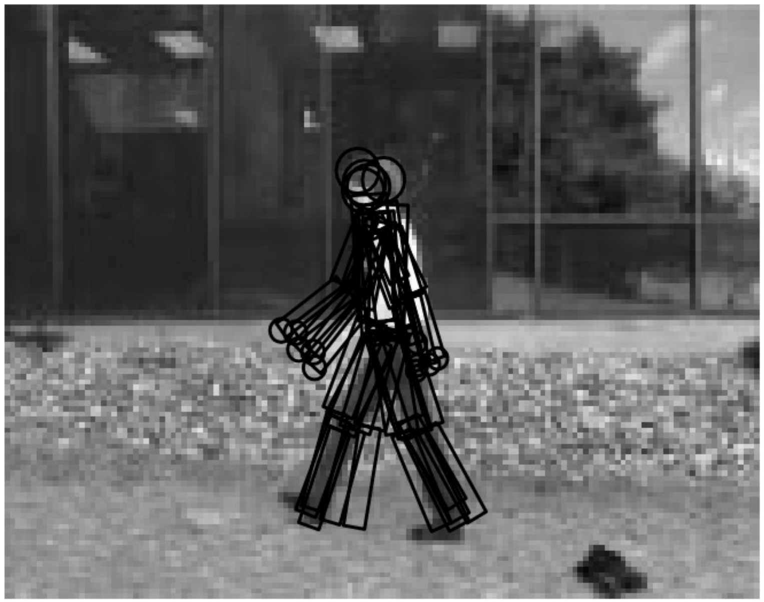
Bregler & Malik CVPR 1998



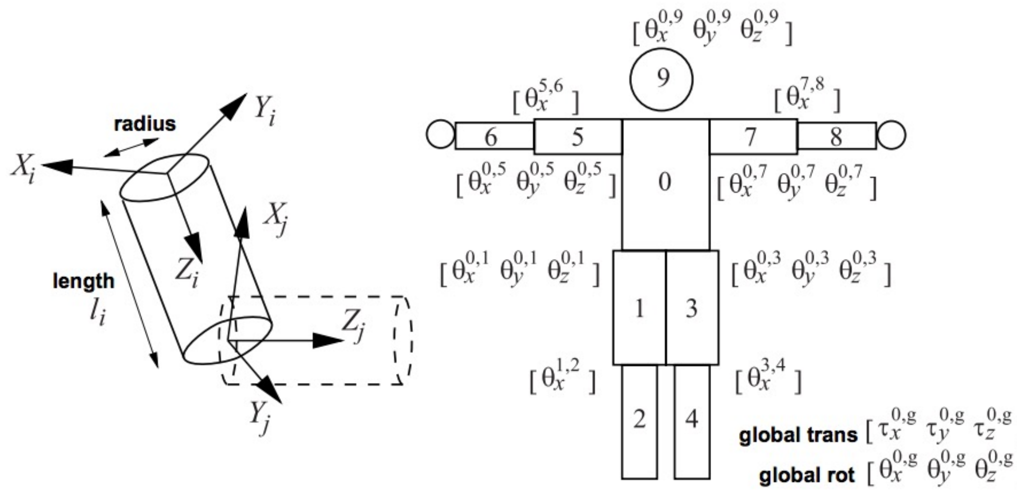
Tracking People with Twists and Exponential Maps

Stochastic search to deal with ambiguity

Represent a distribution over poses



a



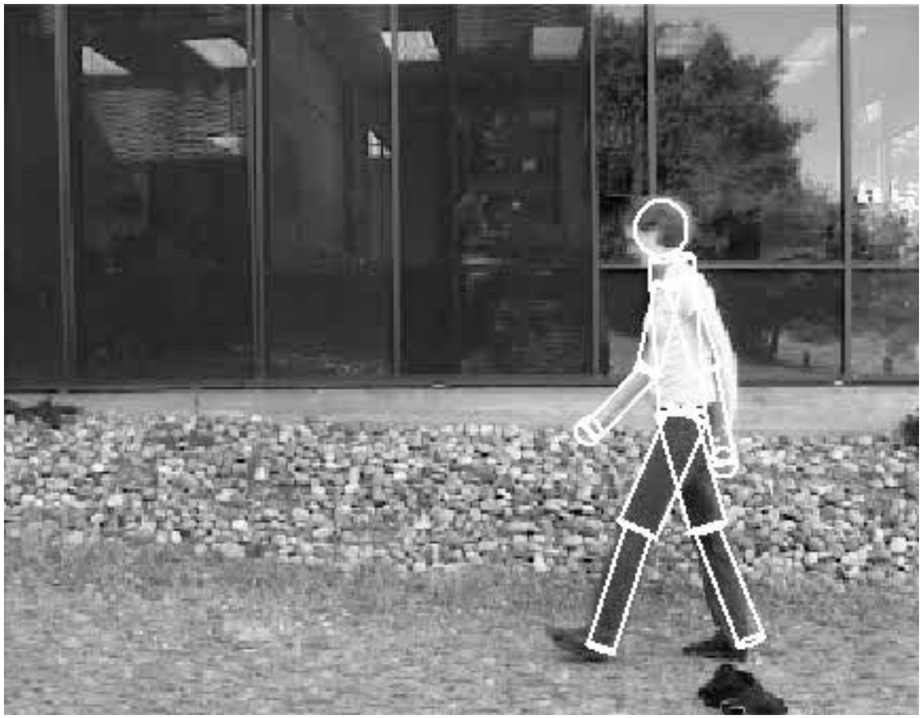
b

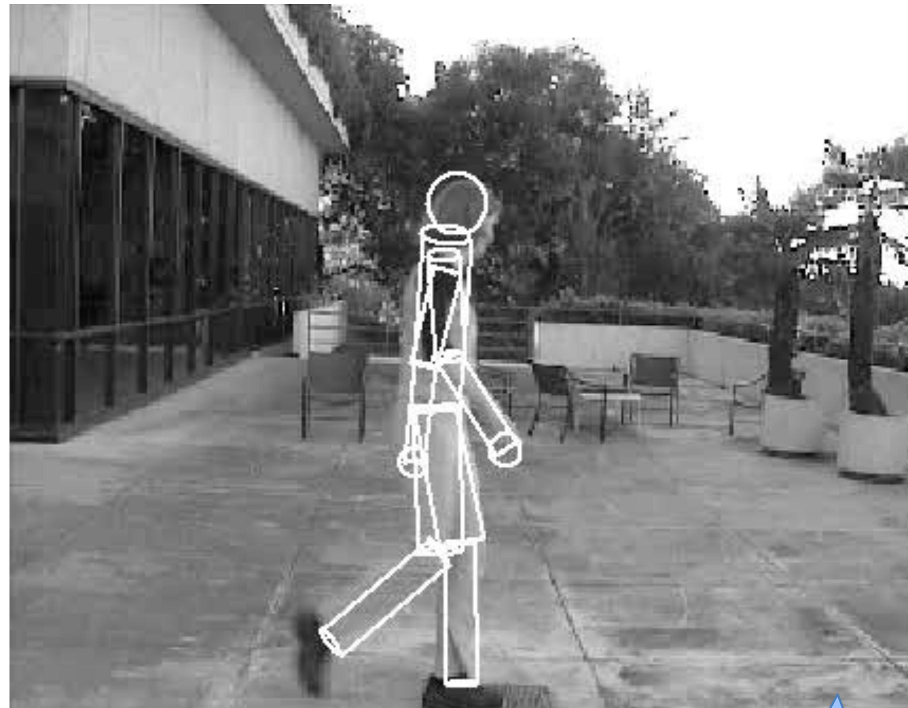
- Particle filter to propagate over time

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

Koenderink Prize, 2010

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

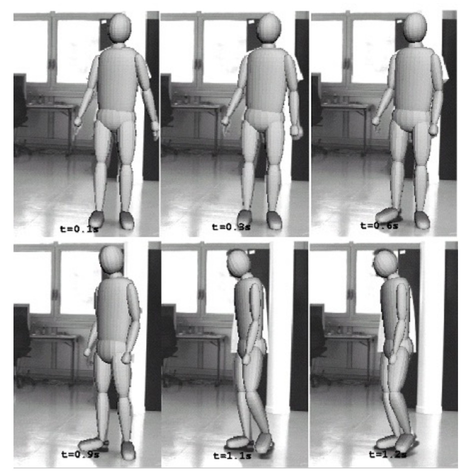
Koenderink Prize, 2010

Stochastic search and tracking


Deutscher, North, Bascle, & Blake '99


Sidenbladh, Black and Fleet, '00


Cham and Rehg '99

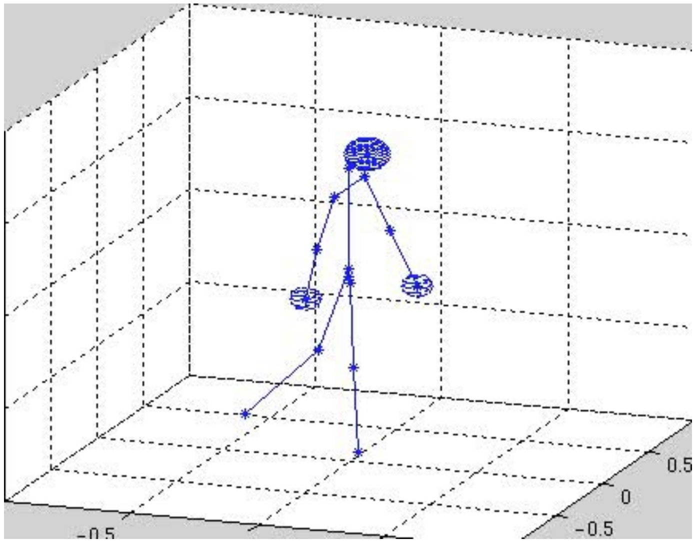

Sminchisescu &Triggs '01

Nothing really works.
Add a prior.

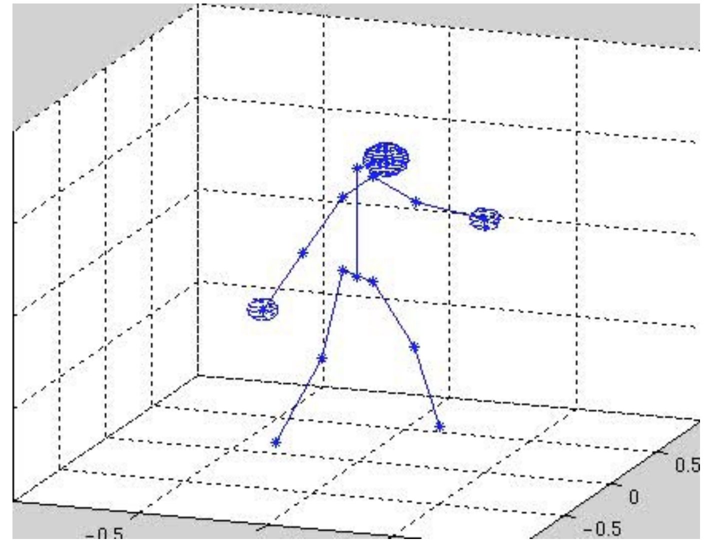
Early generative models of human motion.

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

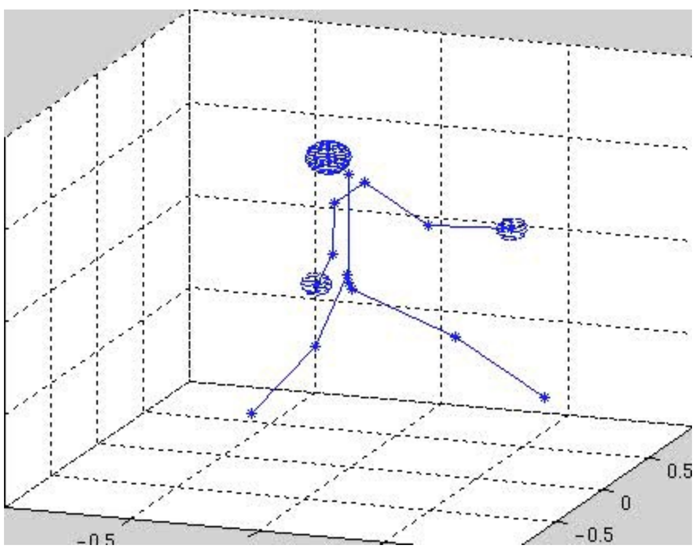
PC1



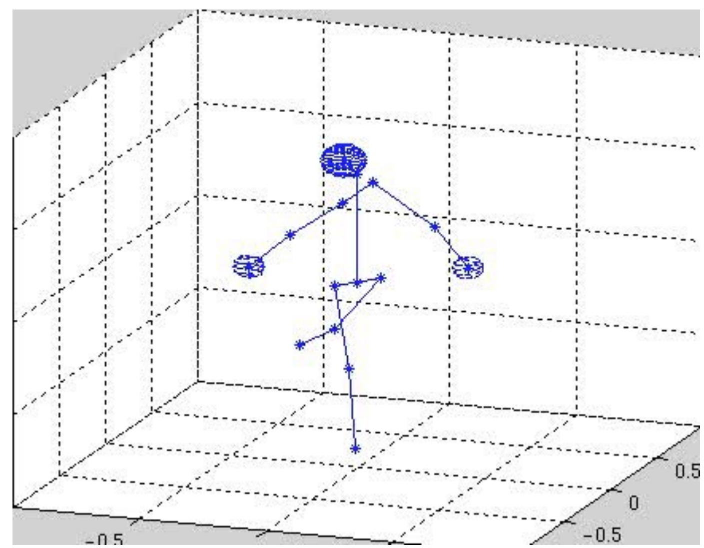
PC2



PC3



PC4



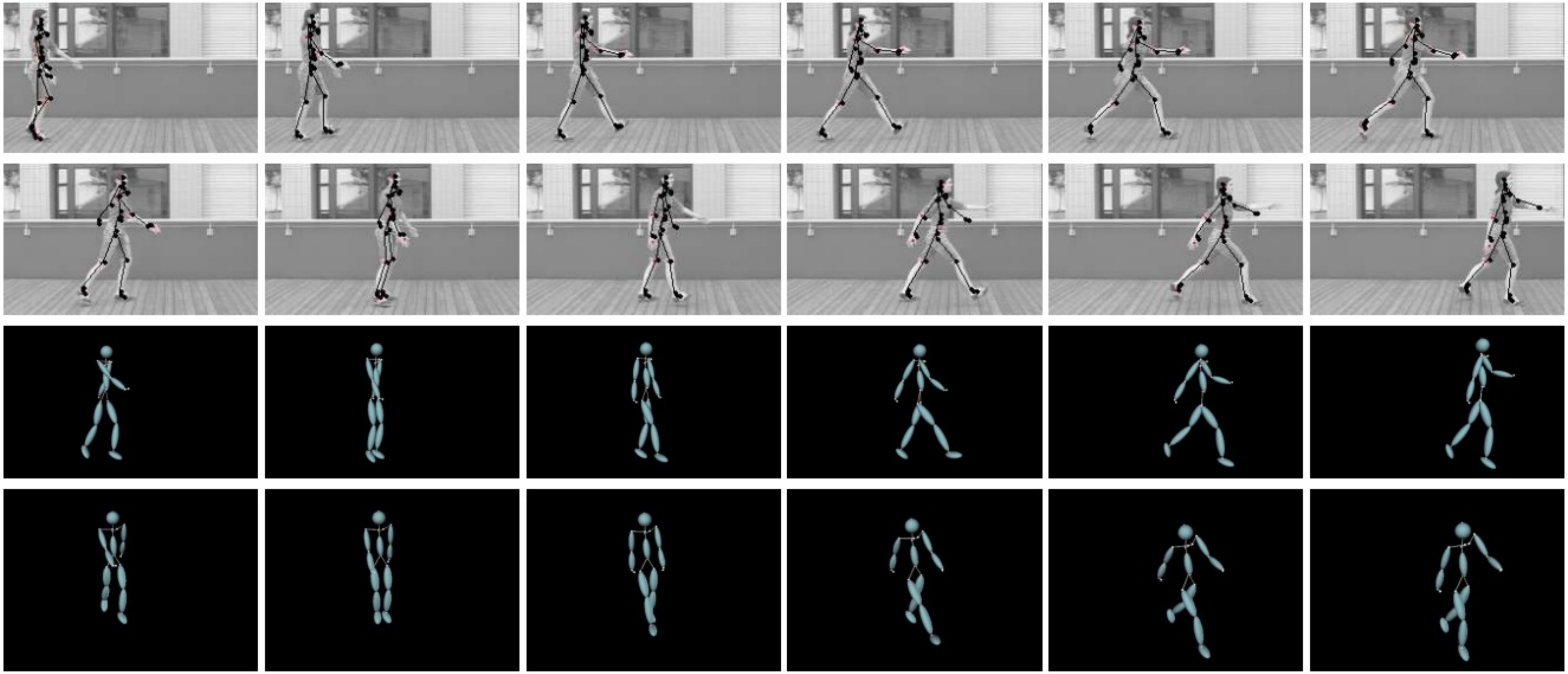
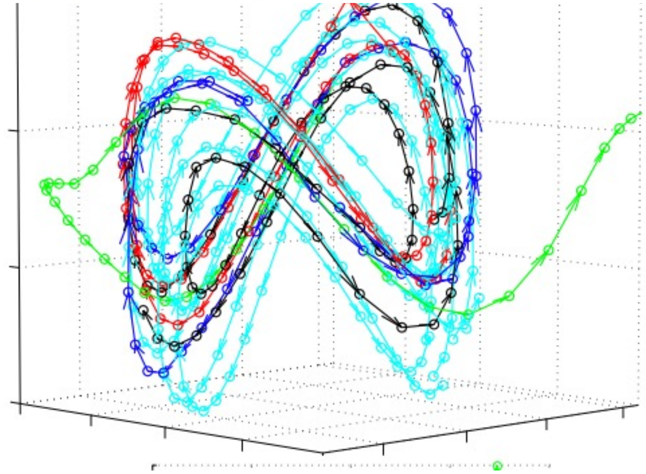


Figure 9. Tracking 37 frames of an exaggerated gait. Note that the results are very accurate even though the style is very different from any of the training motions. The last two rows depict two different views of the 3D inferred poses of the second row.

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



Early deep network prior

Restricted Boltzmann machine

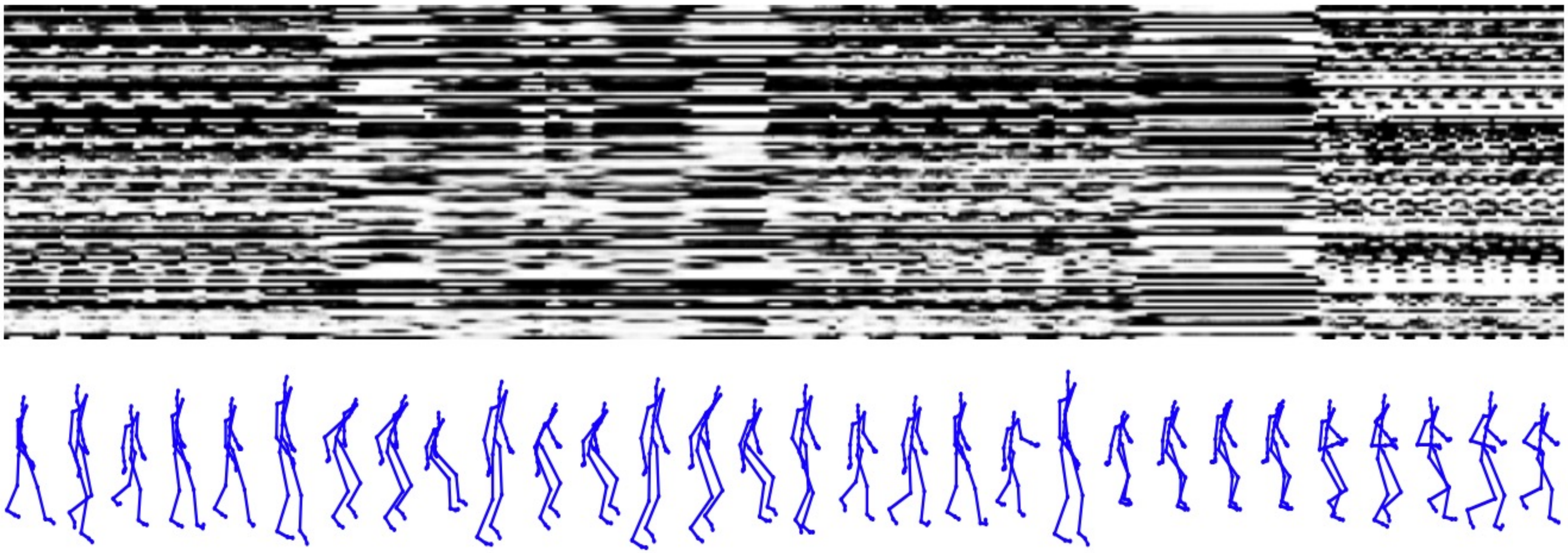
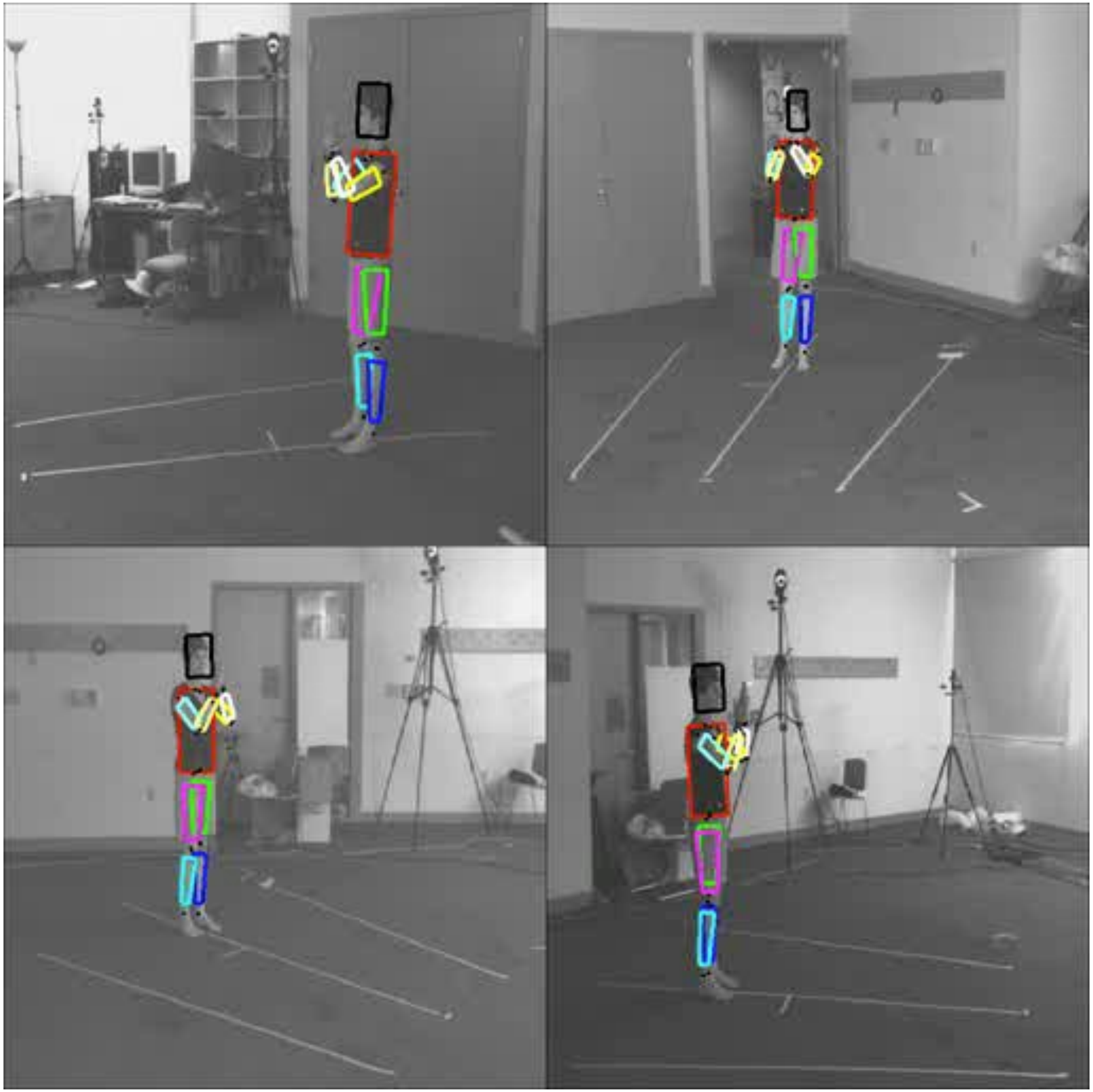


Figure 1: In a trained model, probabilities of each feature being “on” conditional on the data at the visible units. Shown is a 100-hidden unit model, and a sequence which contains (in order) walking, sitting/standing (three times), walking, crouching, and running. Rows represent features, columns represent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

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

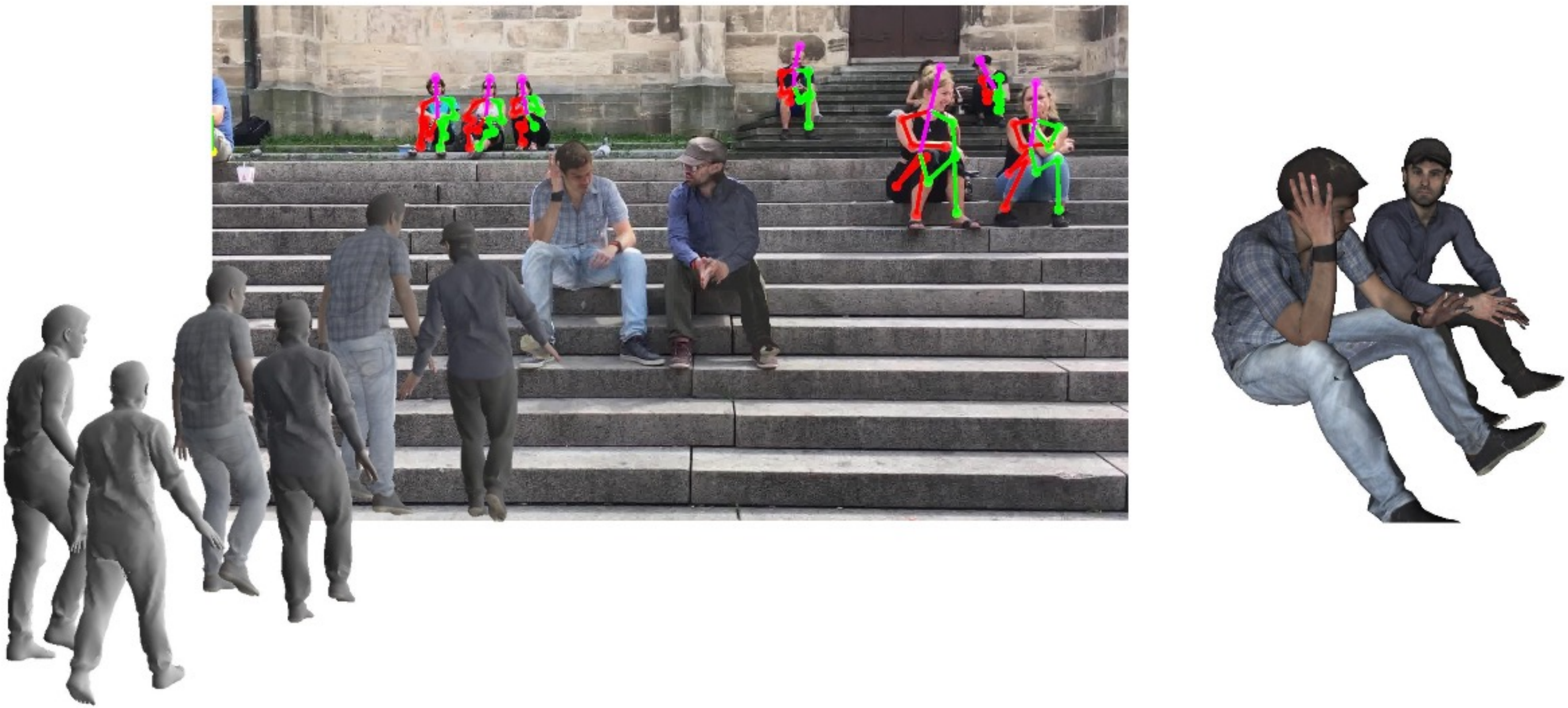
Ground truth.
There was none.
Were we making progress?

First tests in around 2001 before the HumanEva dataset



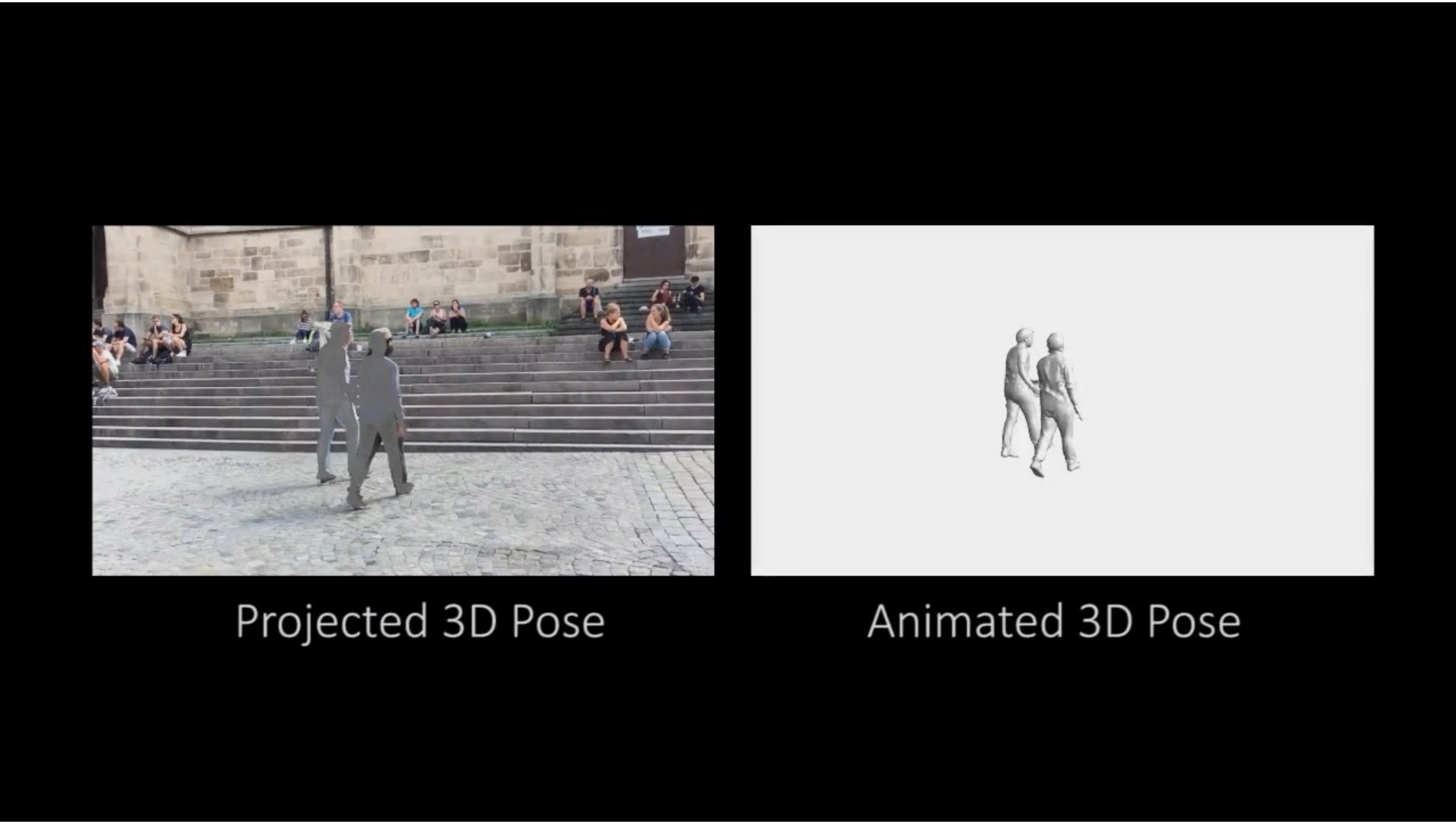
Sigal, et al., HumanEva, 2006 and IJCV 2009. And 3.6M, Ionescu, et al., PAMI 2014

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

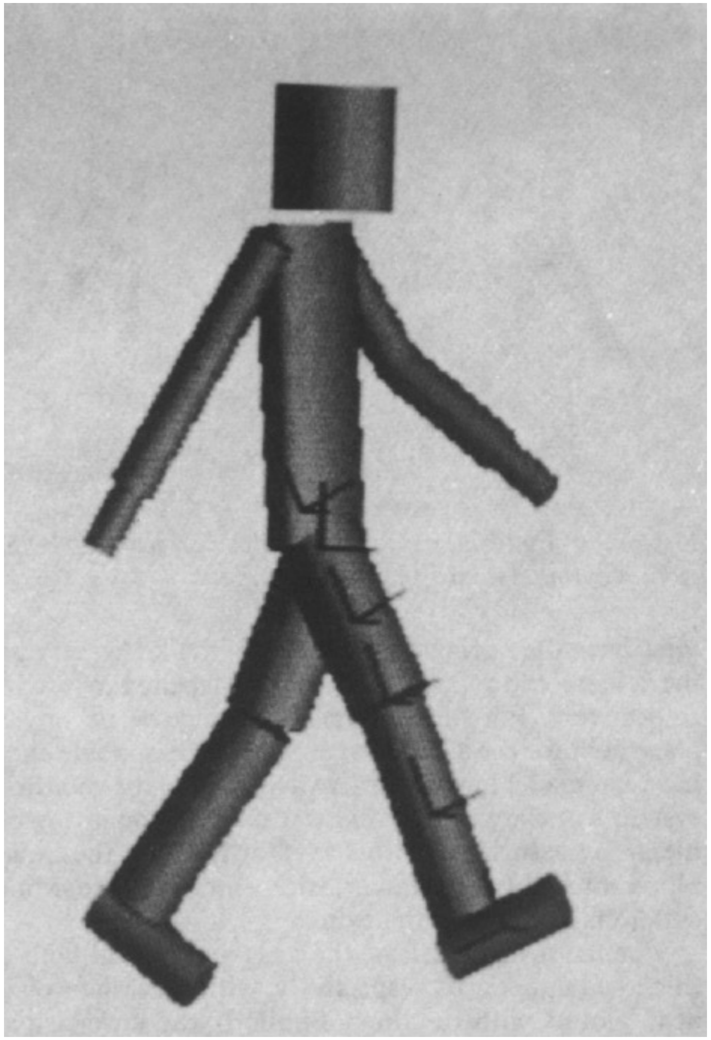
Real video with "ground truth"



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

Body representation.

The problem



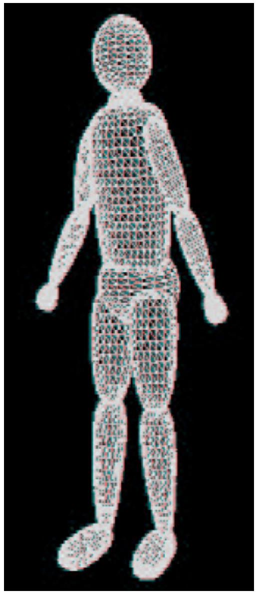
We don’t look like this.

Models don’t match the data.

Systems using such models tend to be brittle.

We argued that we need a better model of human shape and motion.

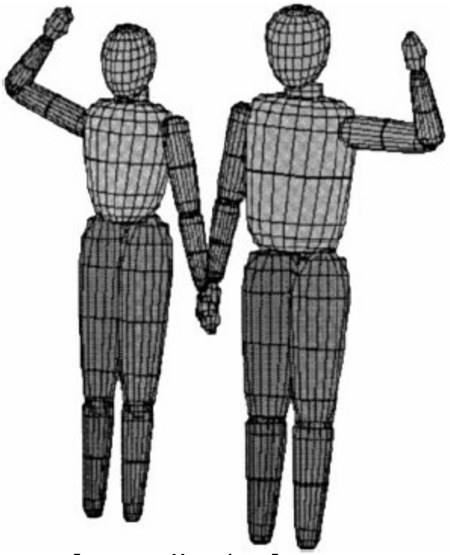
Early body models



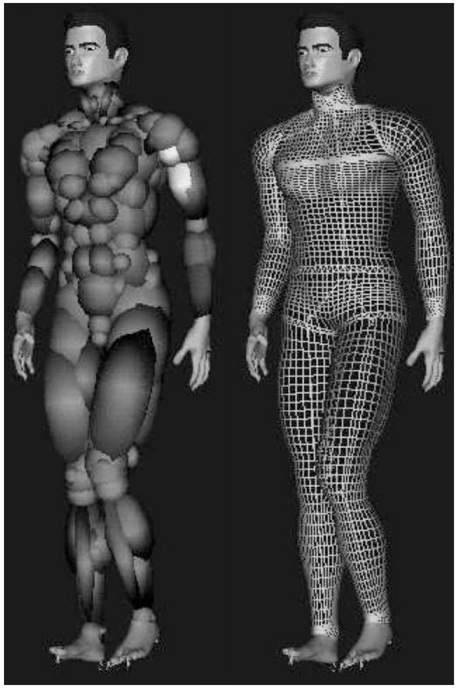
[Sminchisescu
and Triggs ’03]

[Terzopoulos
and Metaxas ’93]

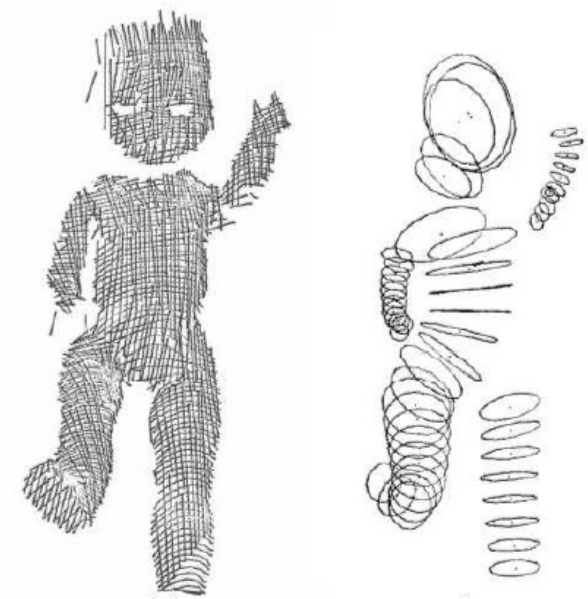
[Kakadiaris and Metaxas ’00]



[Gavrilla, ‘96]

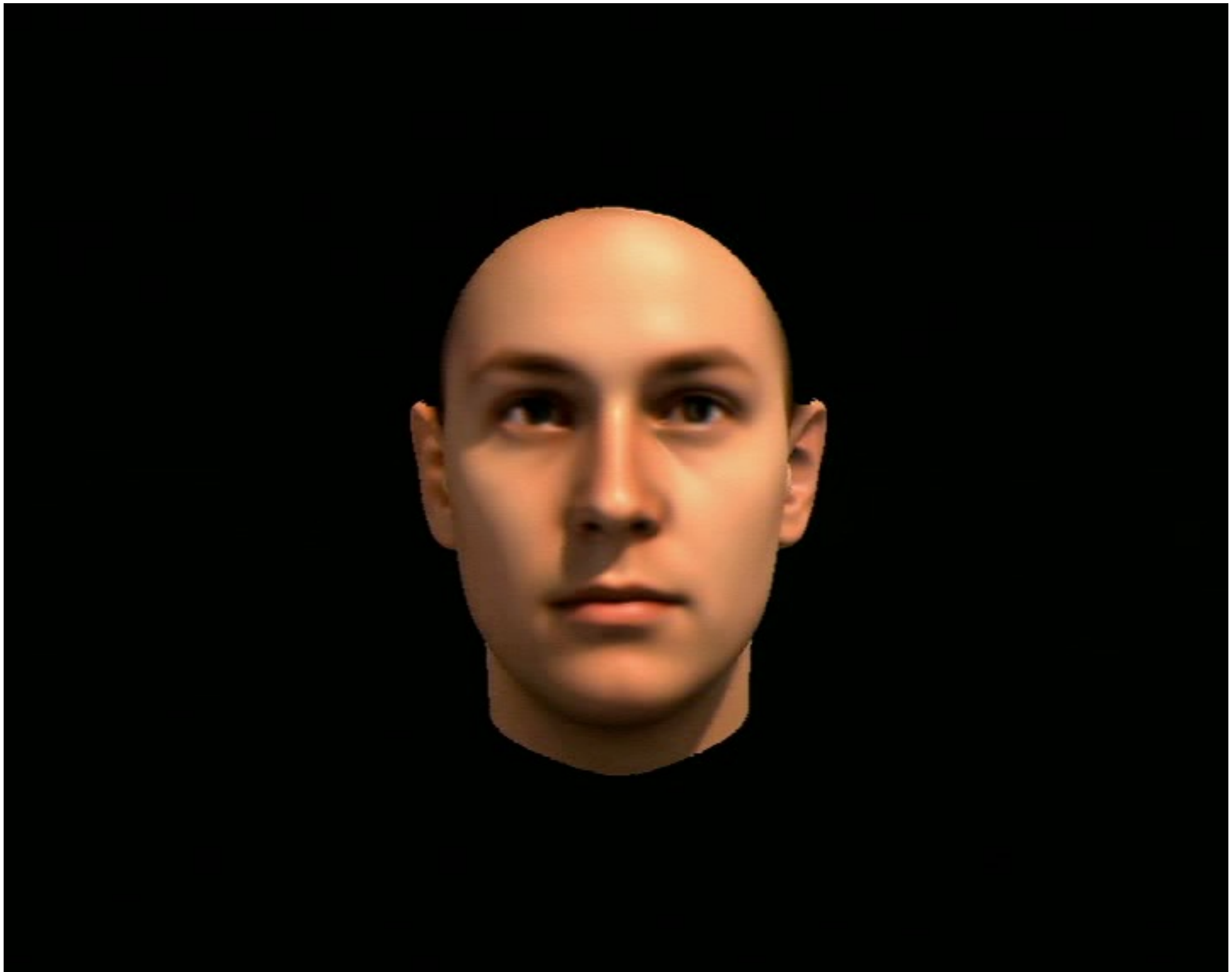


[Plänkers and
Fua ’01]



Nevatia & Binford ’73

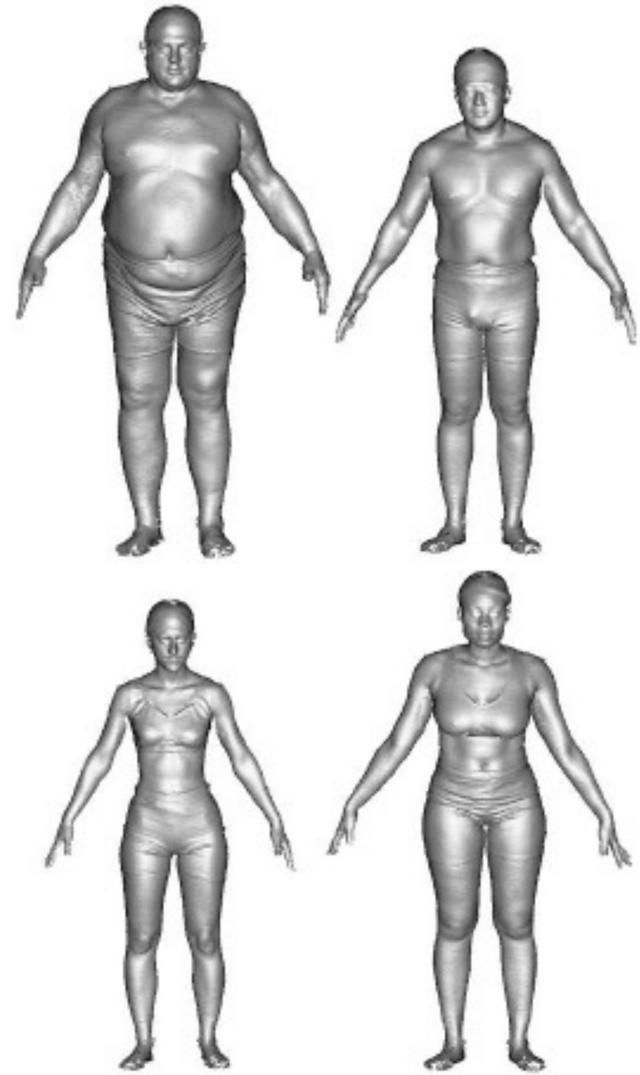
Learning face shapes



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

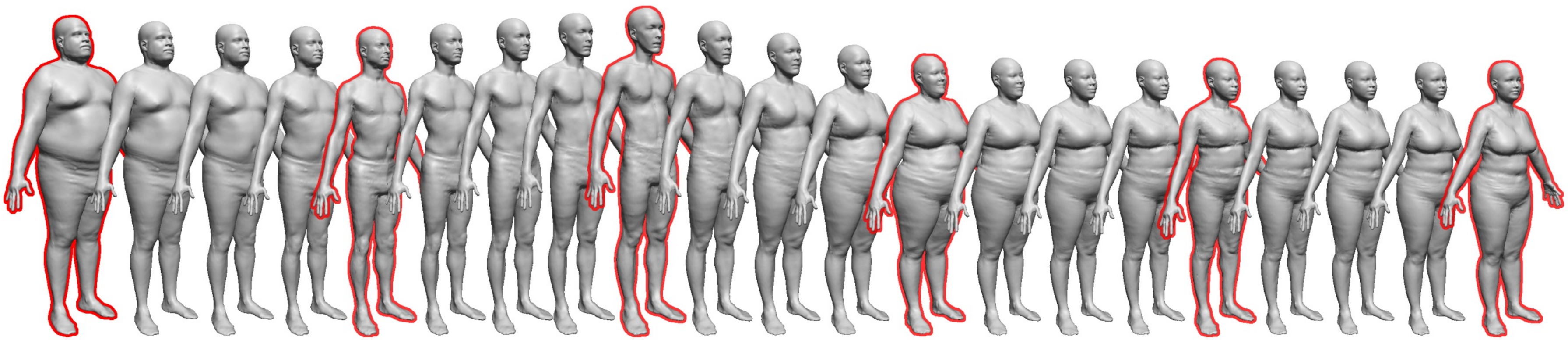
Learning a body model

[Cyberware]



CAESAR dataset – 2001.

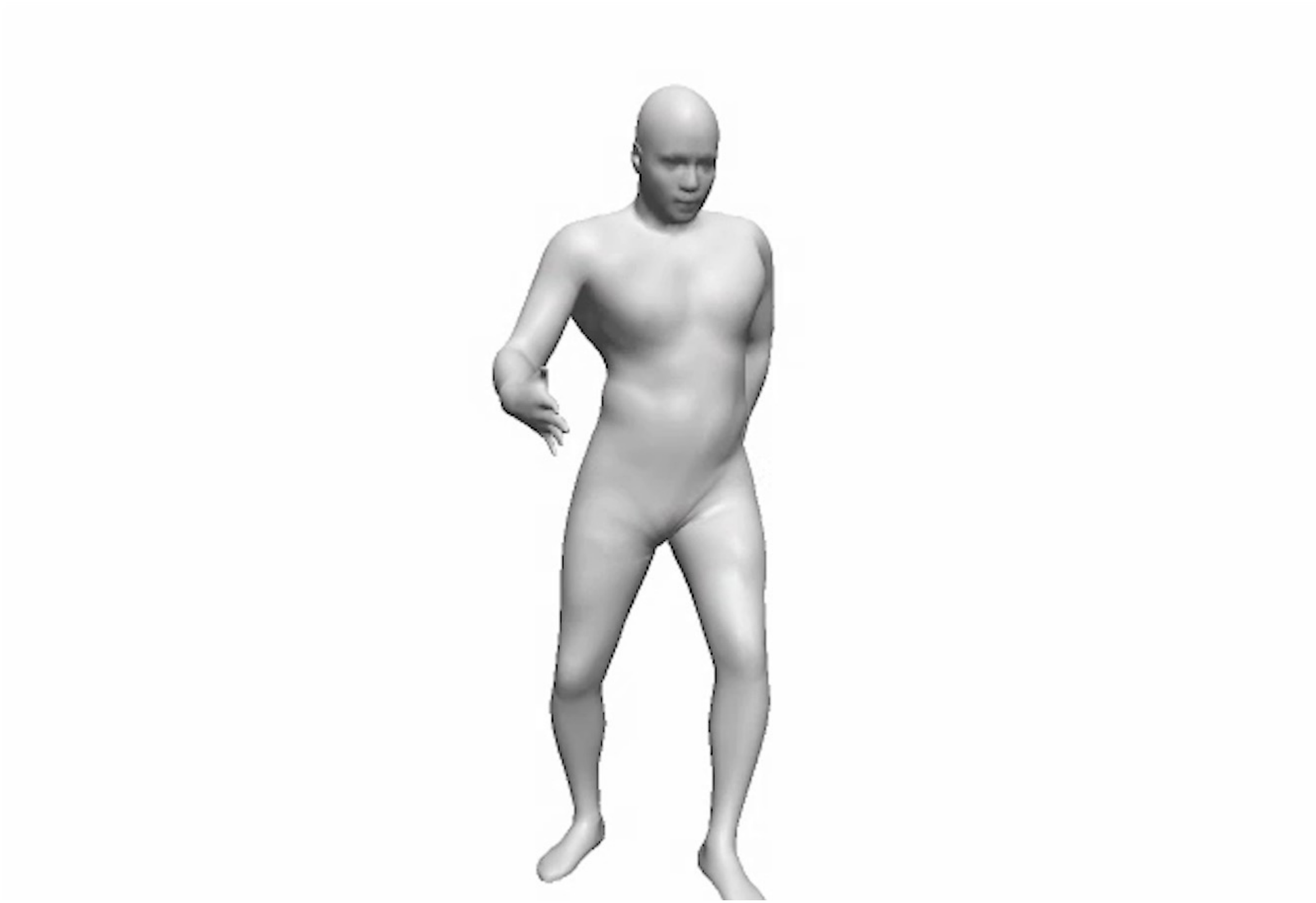
First generative model of body shape



- Register a template mesh to CAESAR scans
- PCA on bodies in a canonical pose

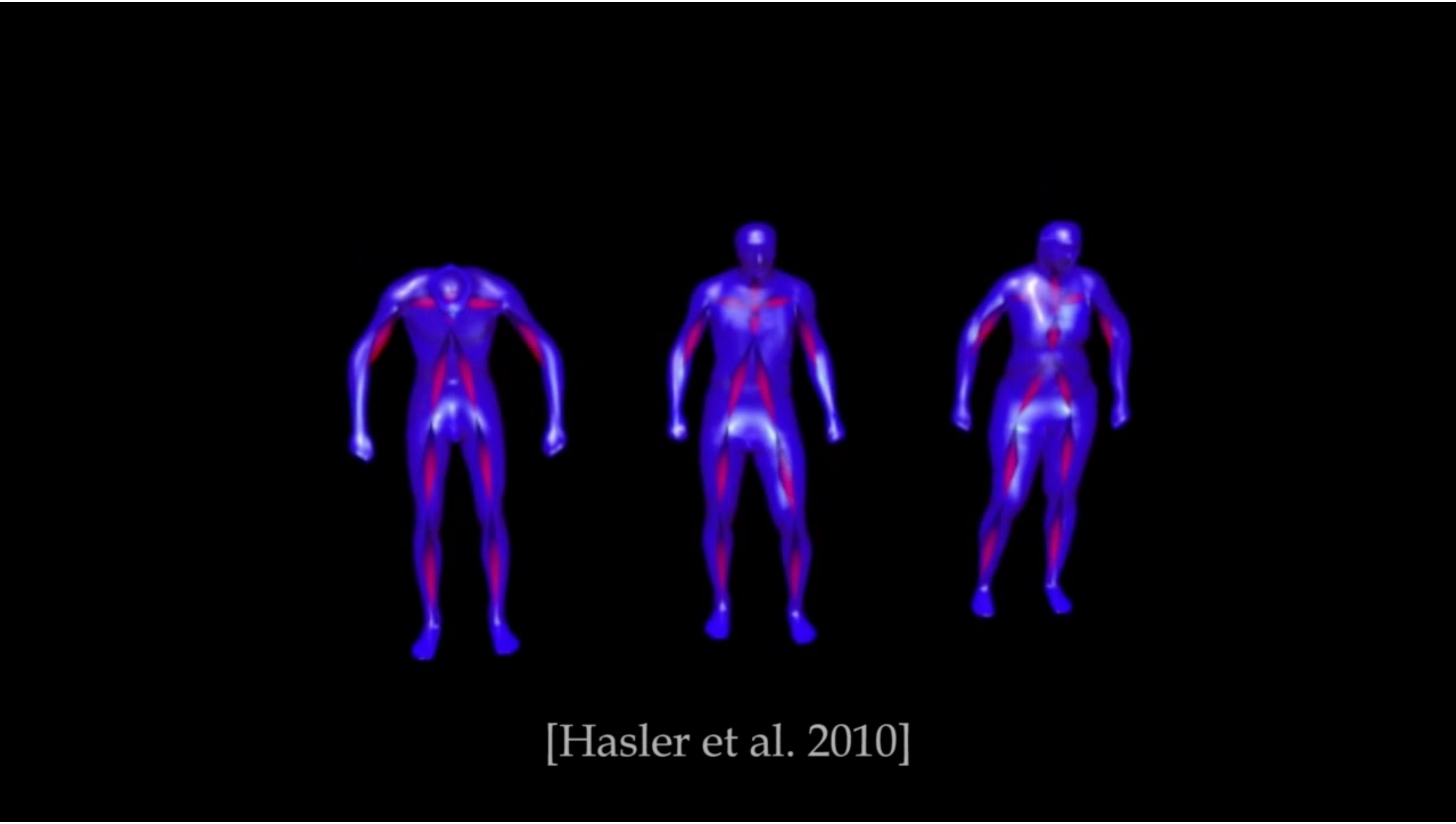
The space of human body shapes: reconstruction and parameterization from range scans, Allen, Brian, Popovic´, 2003

Learning body models (2003-2013)

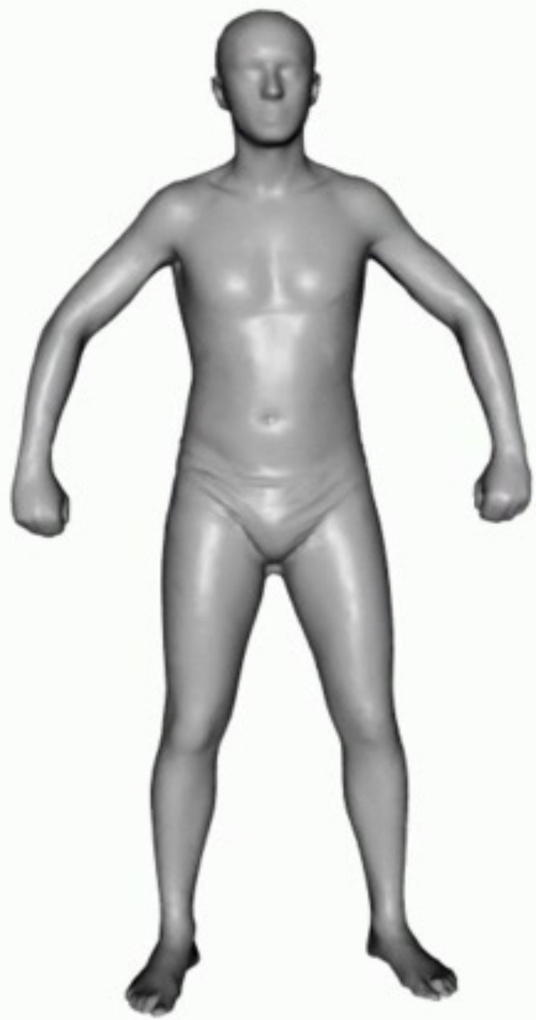


Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. B Allen, B Curless, Z Popović, A Hertzmann, 2006

Learning body models (2003-2013)



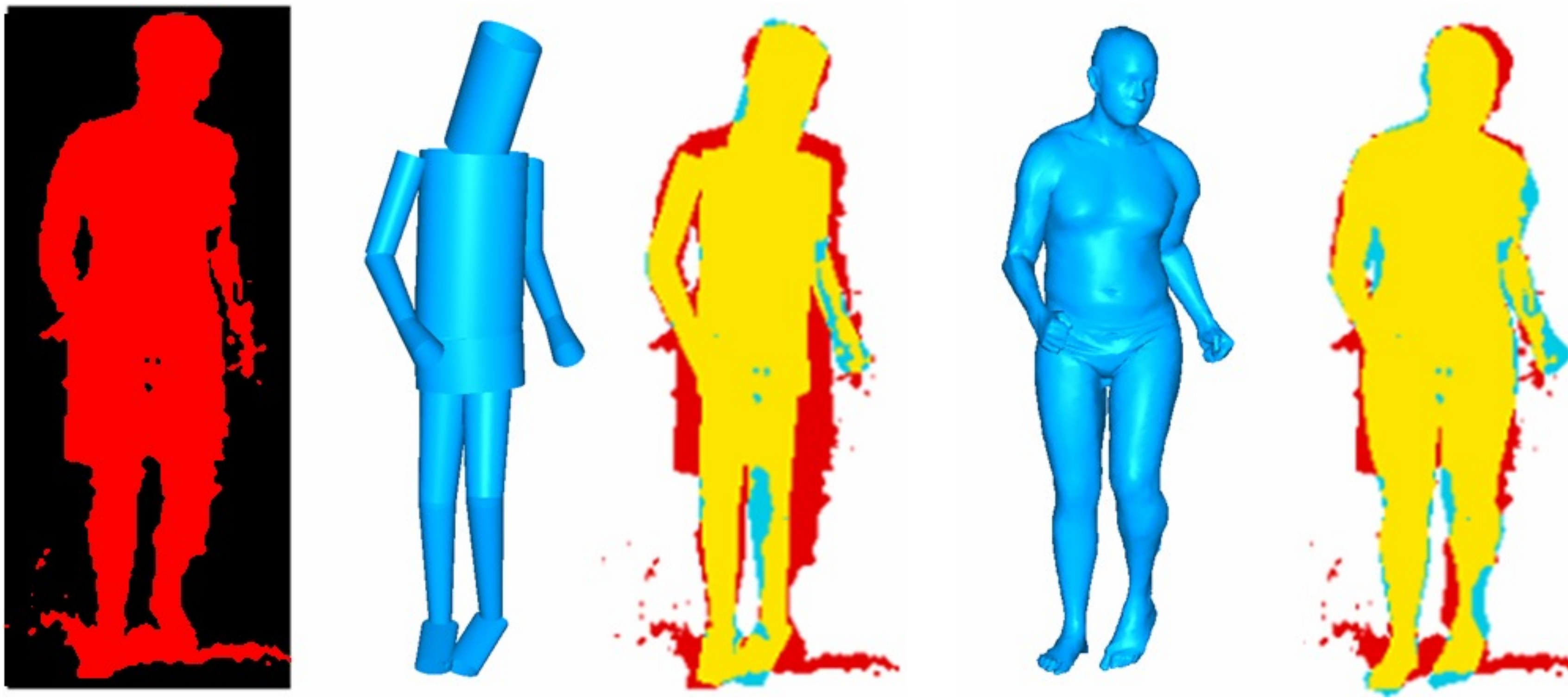
Learning body models (2003-2013)



First to factor pose and shape

Anguelov et al., SCAPE, 2005

First parametric human fit to images



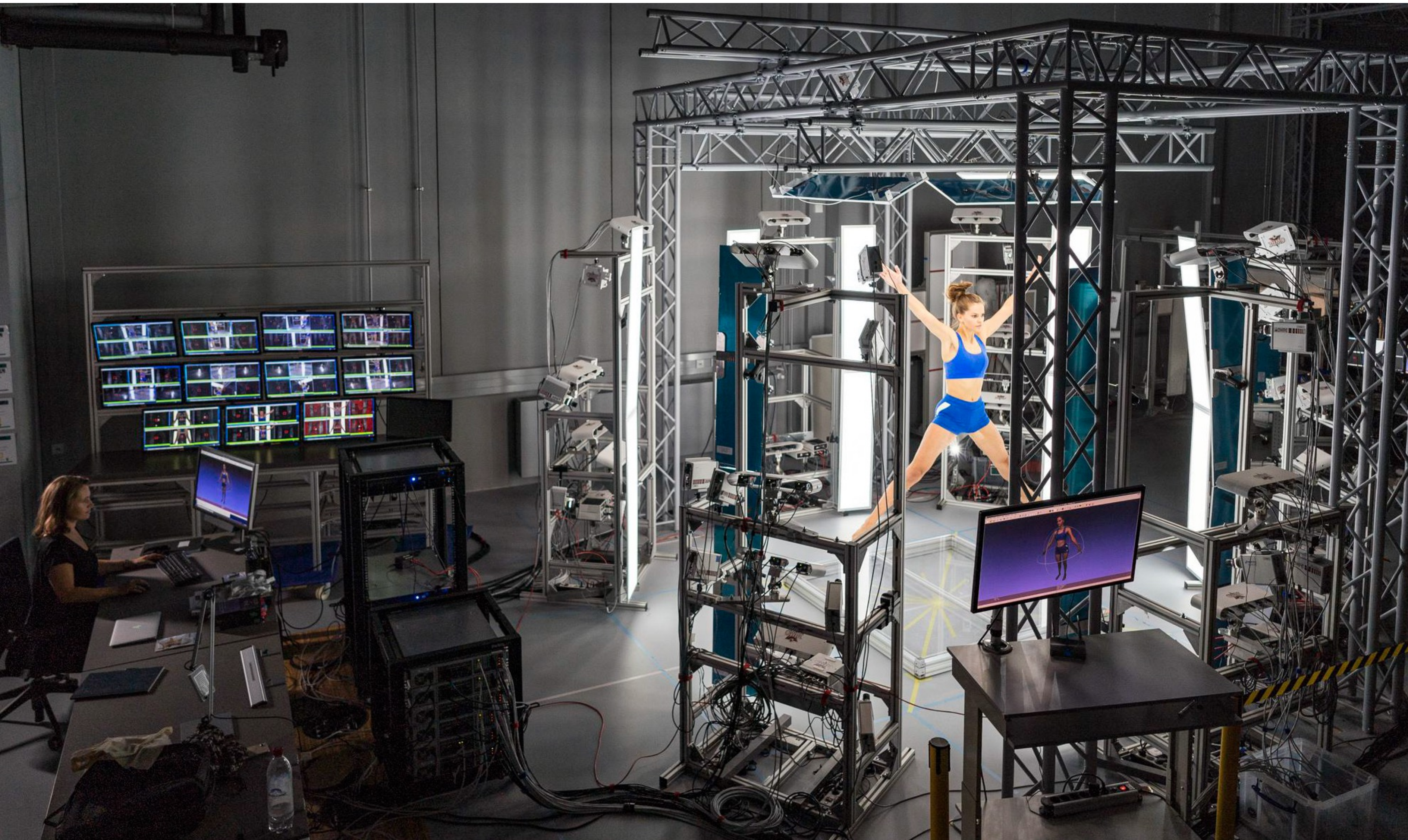
Traditional model　　　　　　Proposed model

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

Problems with SCAPE

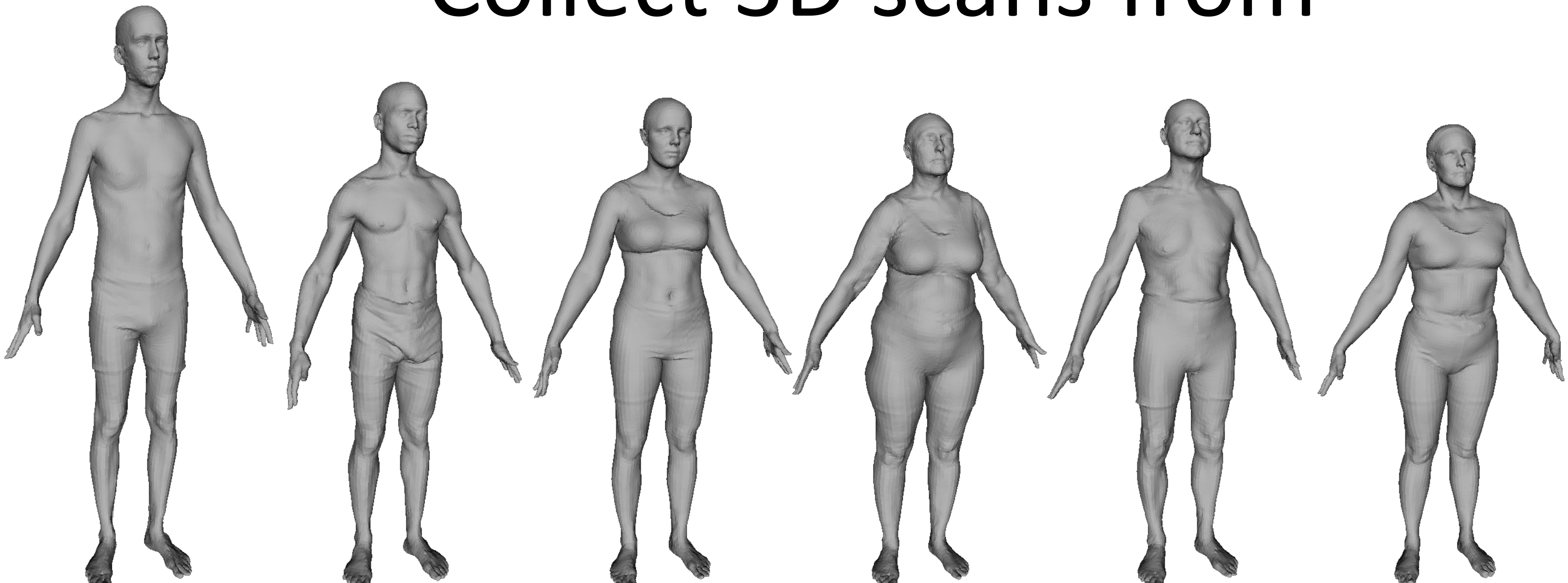
- Every triangle is deformed independently using a 3x3 transformation
- Over parameterized
- Have to solve a least squares optimization to get a mesh
- No skeleton and no joints
- Limb lengths change with pose
- Not compatible with anything

Scale up: First 4D body scanner

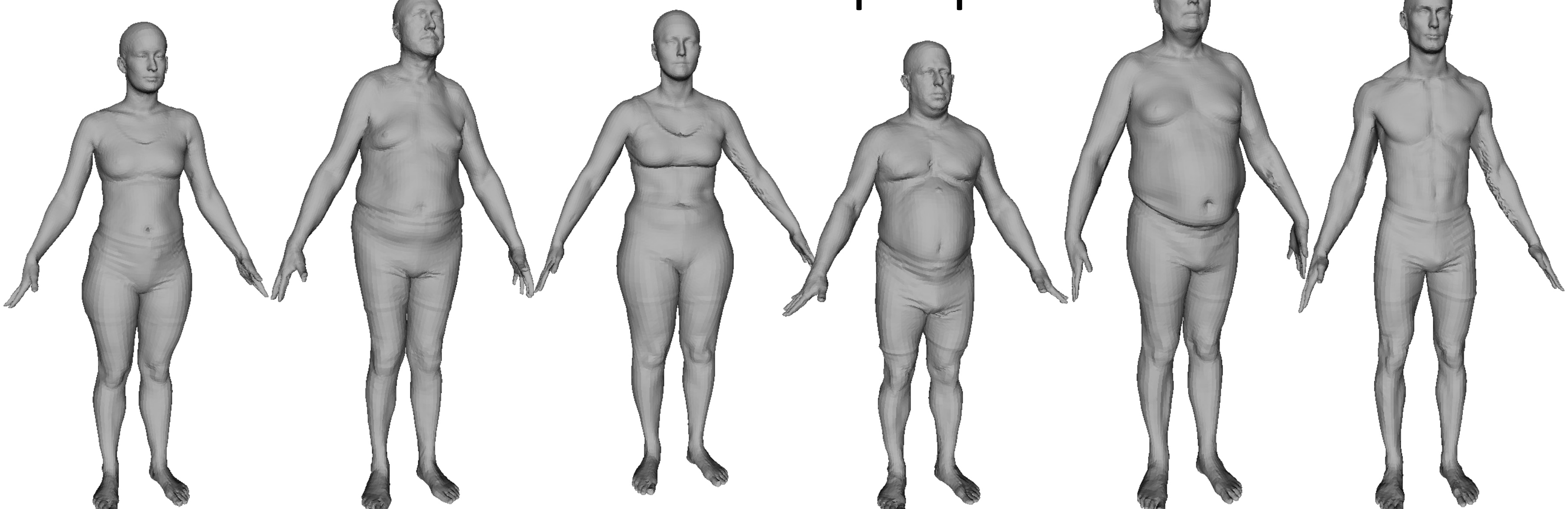


4D scanner:3D at 60 fps

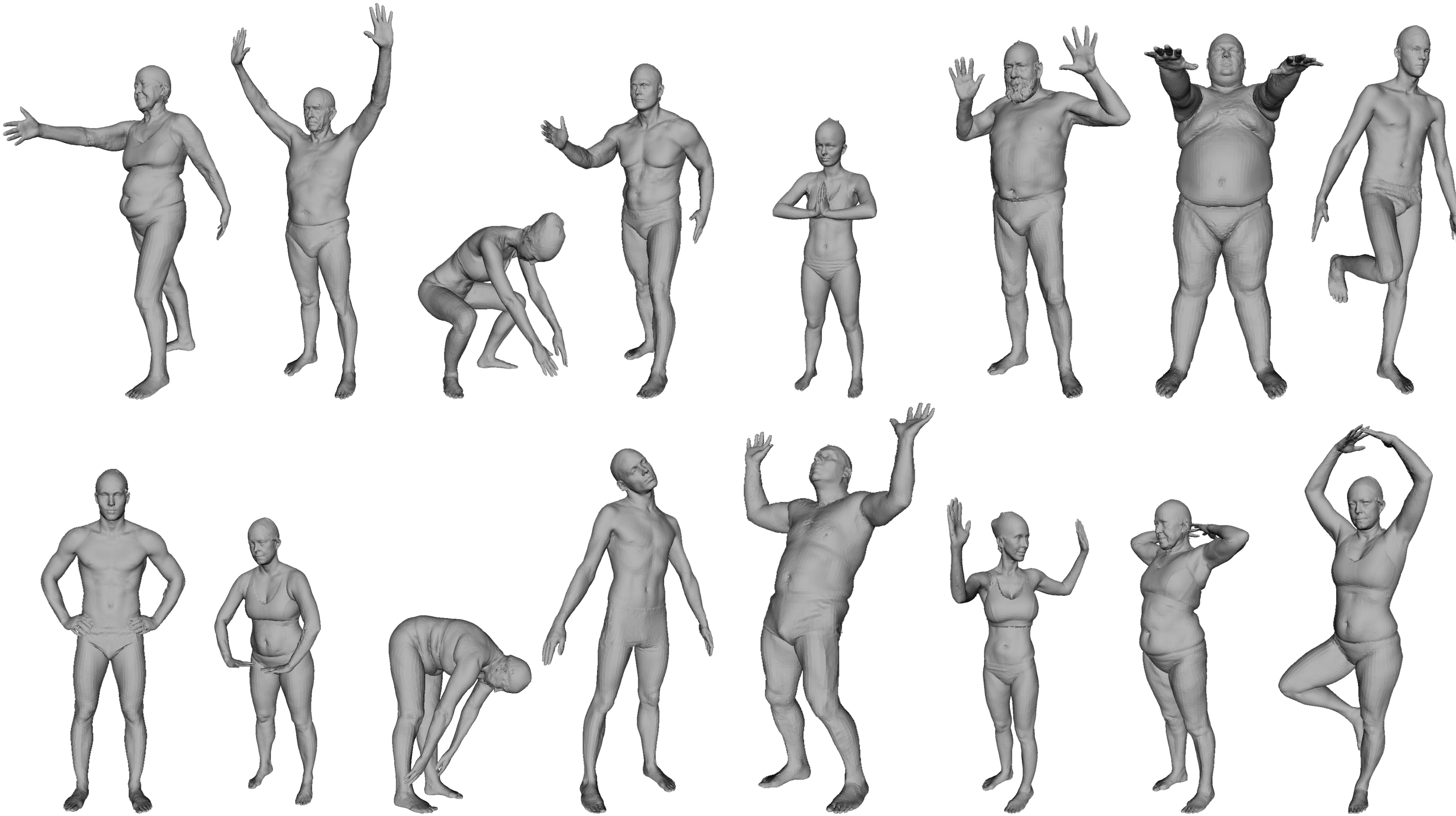
Collect 3D scans from



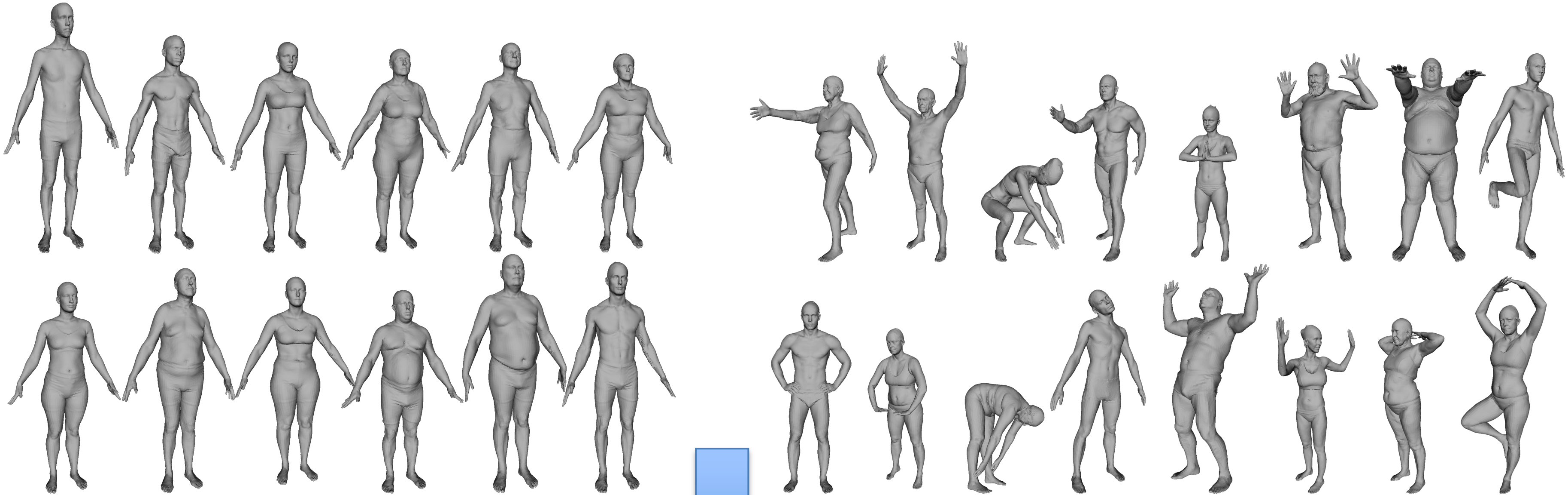
thousands of people…



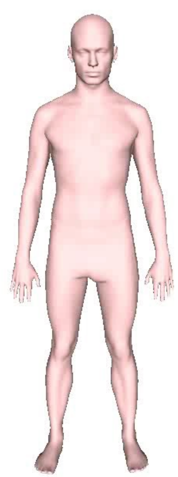
and thousands of poses



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

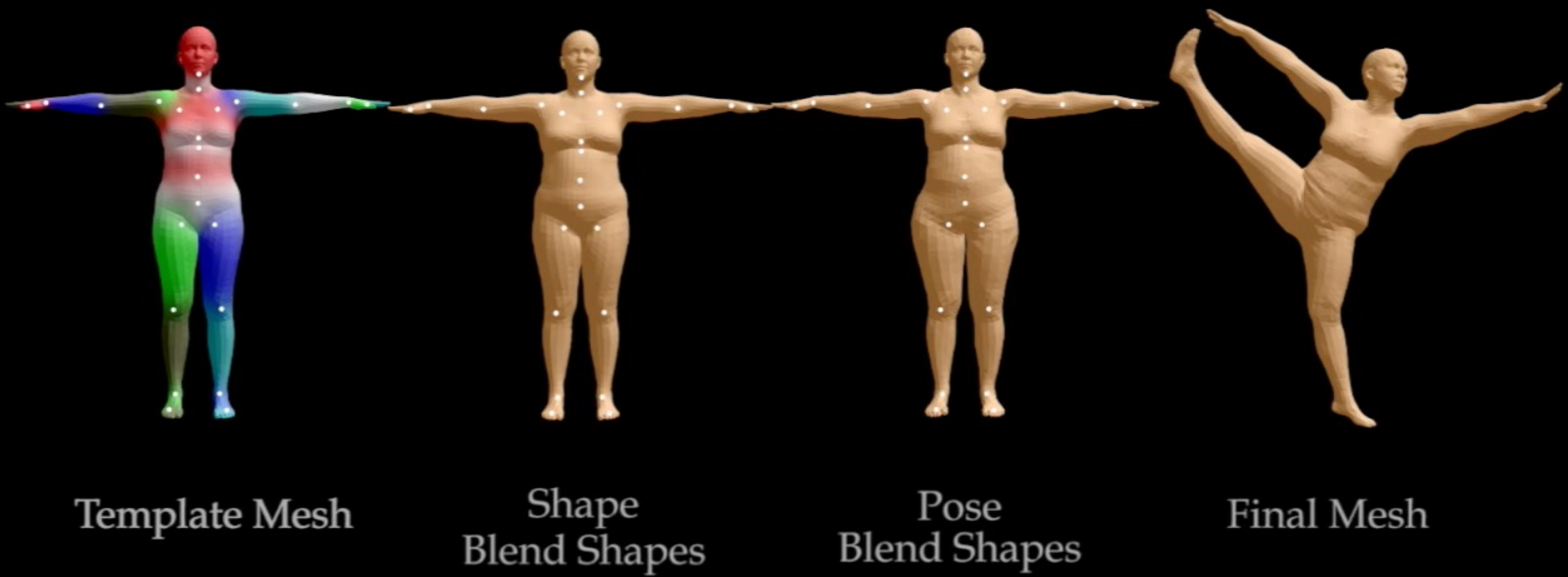


$$M(\theta, \beta, \delta, A)$$



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

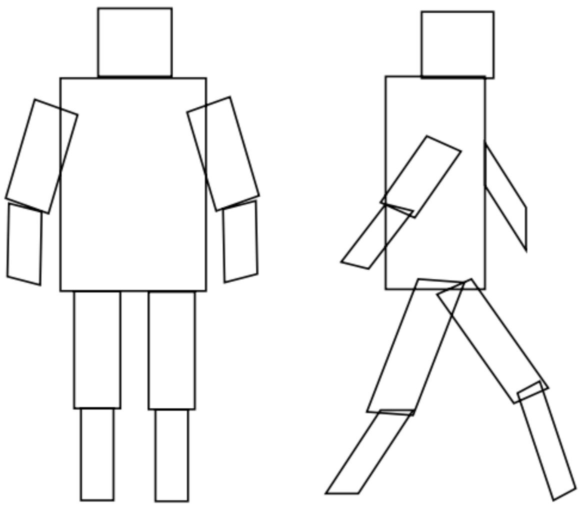
SMPL: SIGGRAPH Asia'15



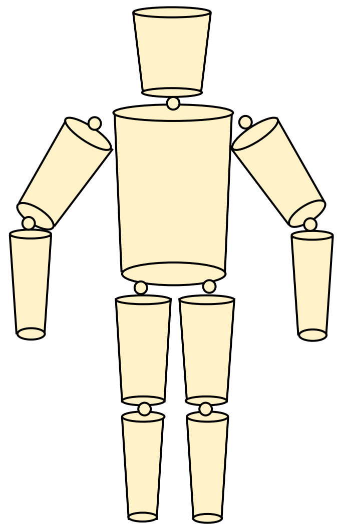
$$M(\vec{\theta}, \vec{\beta}) = W(\mathbf{T}_F(\vec{\beta}, \theta), \mathbf{J}(\vec{\beta}), \mathcal{W}, \vec{\theta}) \mapsto \text{vertices}$$

The evolution of body models

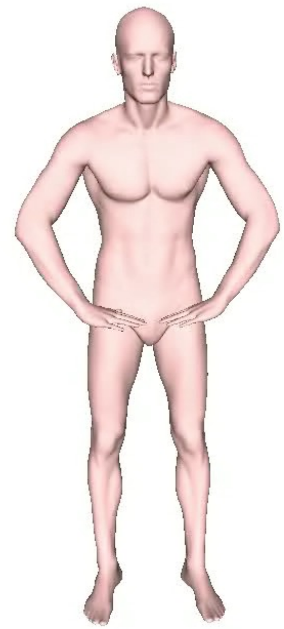
1976



1996



2016



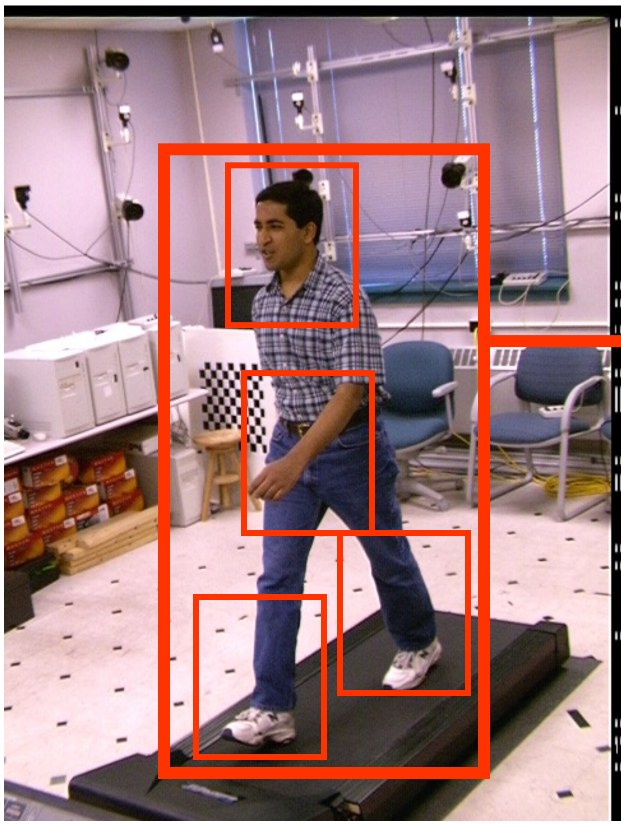
2024

HUGS: Human Gaussian Splats.
Kocabas et al, CVPR 2024

2036 (or much sooner)?
Fully realistic personalities.
Autonomous, intelligent,
interactive.

People from images.
The ML approach
1997 - today

Detection: The Pure ML Approach



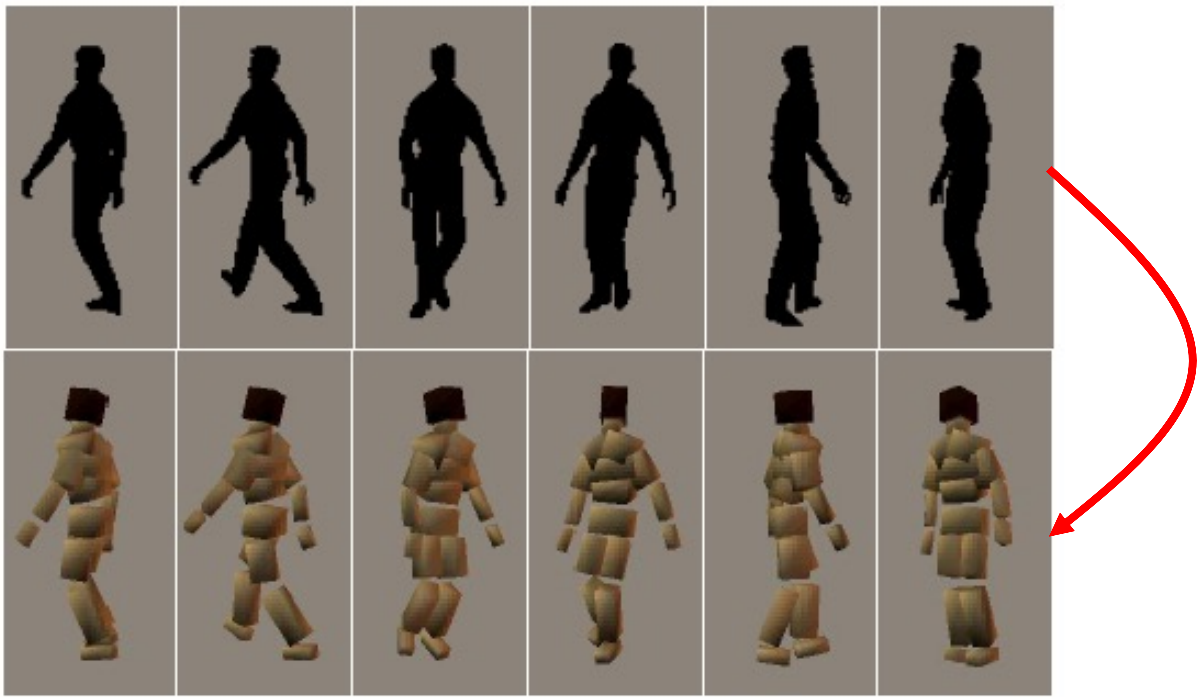
Single image

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_N \end{bmatrix} \rightarrow \text{Trained Classifier} \rightarrow \text{Person/Not-person}$$

Support Vector Machines

“Pedestrian detection using wavelet templates,” Oren *et al* CVPR’97.

Single View to 3D Pose



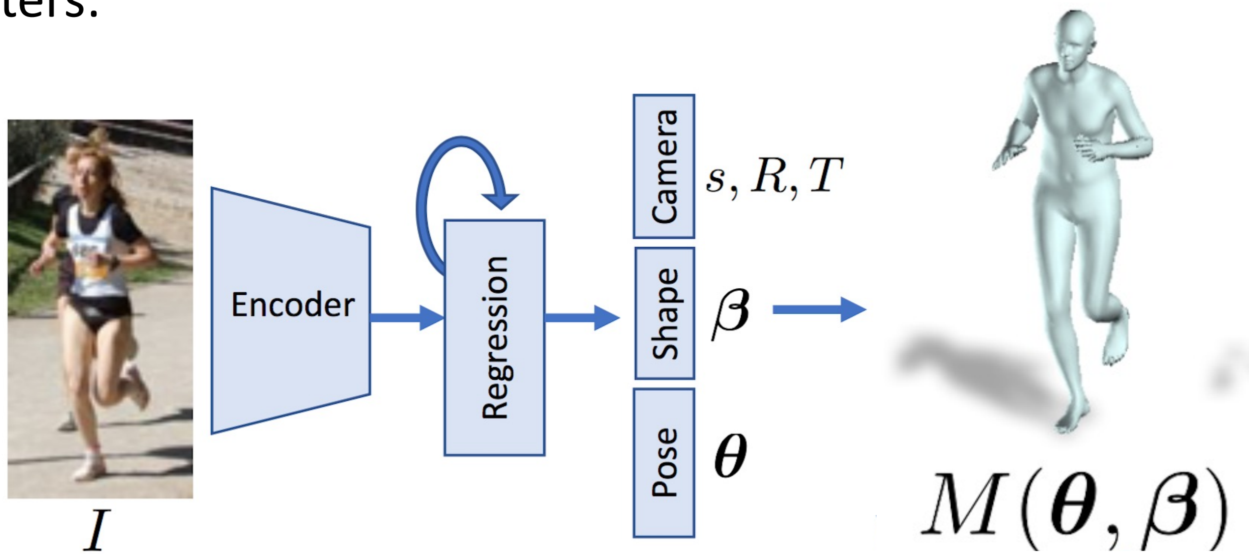
Given synthetic training data, learn the mapping from silhouette contours to 3D pose.

"Gaussian kernel RVM", Agarwal and Triggs CVPR04

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

Direct regression

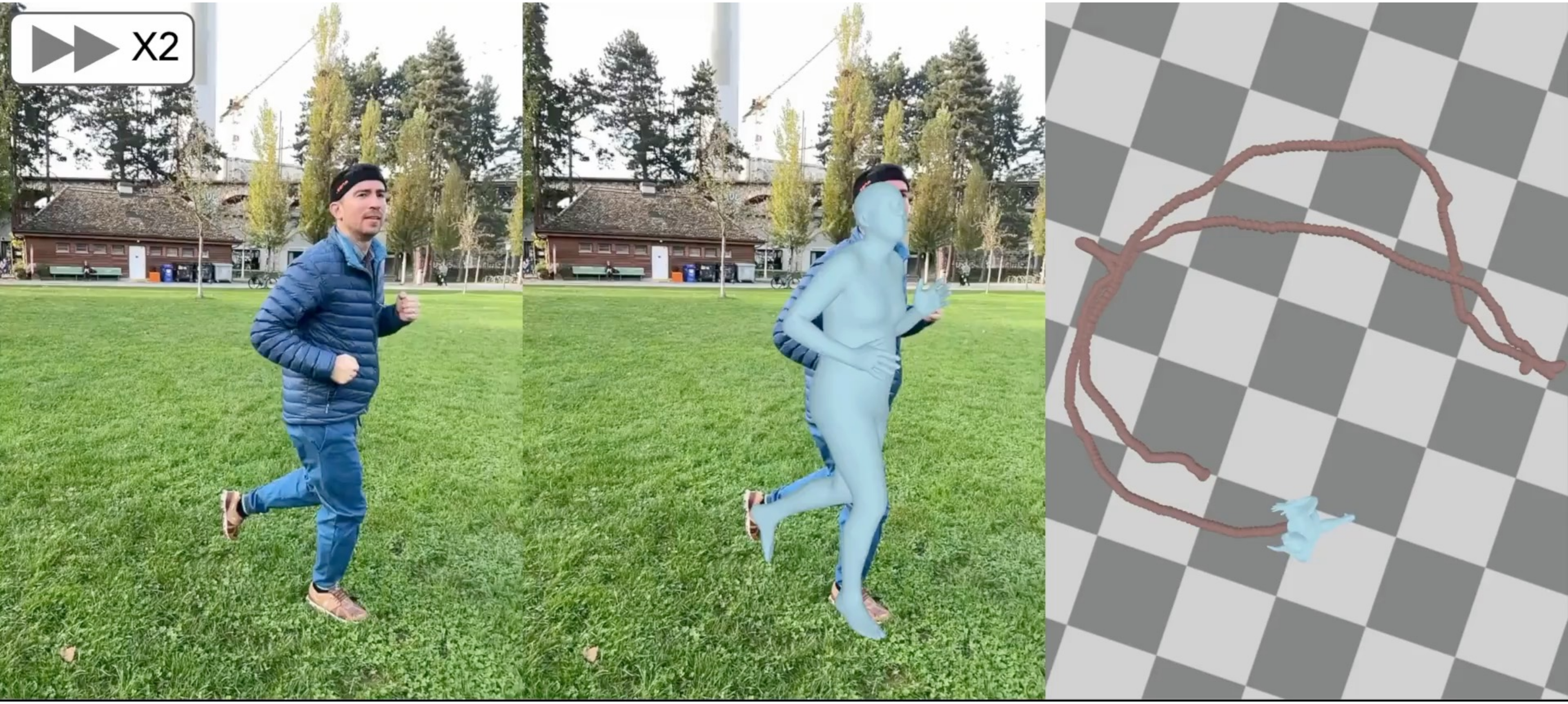
Given an image, directly regress the 3D human pose and shape (HPS) parameters:



HMR

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

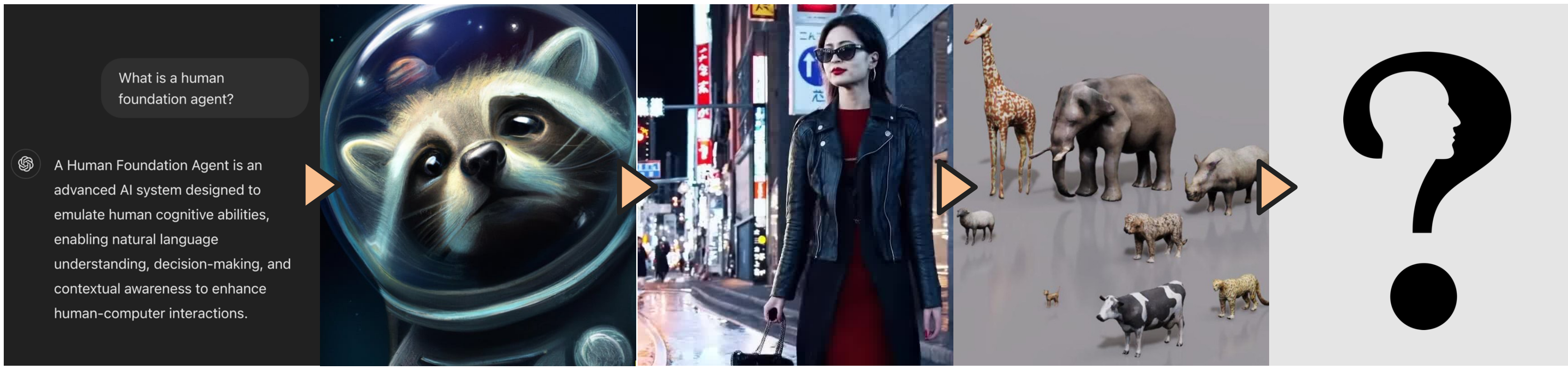
Current SOTA: WHAM.
3D humans in world coordinates efficiently



WHAM: Reconstructing World-grounded Humans with Accurate 3D Motion
Soyong Shin Juyong Kim Eni Halilaj Michael J. Black. CVPR 2024

What’s next?

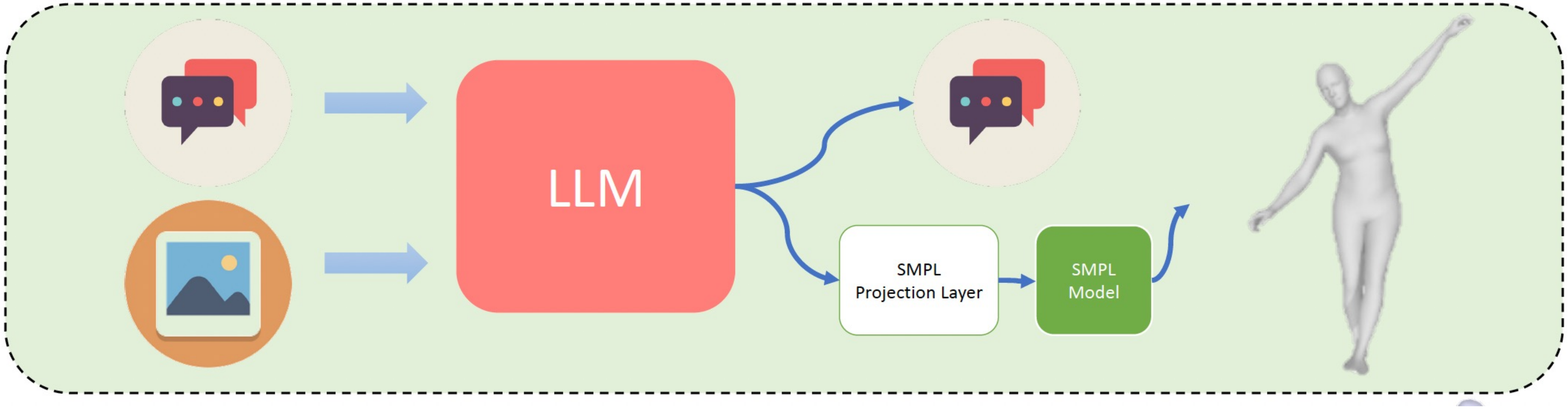
Combining LLMs with 3D human pose



Text | 2D | Video | 3D | 4D humans

- Large language models know a lot about the visual world.
- What do they know about people?
- Can we train them to reason about our 3D pose and motion?
- Can they infer 3D pose and motion from images and video?
- Can they combine this with their general world knowledge?

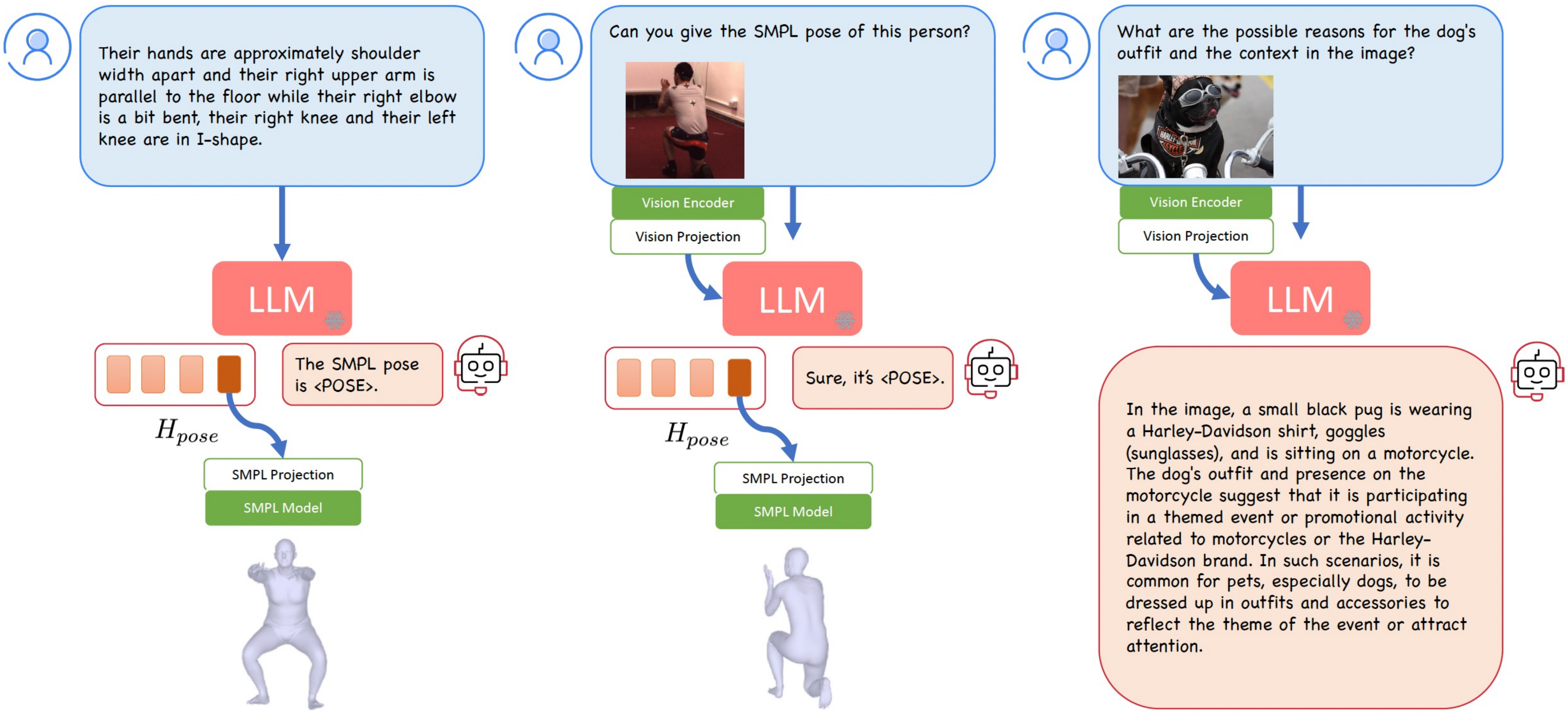
Chatting about humans in 3D



- ChatPose is a foundation model that understands 3D pose
- We fine-tune a multi-modal LLM, using LoRA to estimate and understand 3D pose

ChatPose: Chatting about 3D Human Pose, Yao Feng, Jing Lin, et al. CVPR 2024

Training



ChatPose: Chatting about 3D Human Pose, Yao Feng, Jing Lin, et al. CVPR 2024

Dealing with ambiguity in image evidence:

- Body representation
  - 2D & 3D primitives
  - Learned, parametric
- Search methods
  - Relaxation
  - Optimization
  - Stochastic search
- Priors
  - Hand constructed
  - Learned from data
- Machine learning (training data)
  - Classification
  - Regression

Dealing with ambiguity in image evidence:

- Body representation
  - Maybe implicit/latent
- Search methods
  - Reasoning and regression
- Priors
  - World model
- Machine learning (training data)
  - Large scale.
  - Humans and the world

Try it. Read about it.

More info

meshcapade

https://me.meshcapade.com/

PERCEIVING SYSTEMS
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

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

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
- Tracking people with twists and exponential maps. Bregler and Malik, CVPR 1998
  – people.eecs.berkeley.edu/~malik/papers/**bregler**-malik98.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

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/

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/

Early 2D ML methods

- Baumberg and Hogg, Learning flexible models from image sequences, ECCV 1994
  - https://www.researchgate.net/publication/221304119_Learning_Flexible_Models_from_Image_Sequences
- "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/
- BEDLAM, Black et al., CVPR 2023
  - https://bedlam.is.tue.mpg.de/

Latest work

- HUGS: Human Gaussian Splats, M. Kocabas, R. Chang, J. Gabriel, O. Tuzel, A. Ranjan, CVPR 2024
  - https://machinelearning.apple.com/research/hugs
- WHAM: Reconstructing World-grounded Humans with Accurate 3D Motion, S. Shin, J. Kim, . Halilaj, M. J. Black, CVPR 2024
  - https://wham.is.tue.mpg.de/
- ChatPose: Chatting about 3D Human Pose, Y. Feng, J. Lin, S. K. Dwivedi, Y. Sun, P. Patel, M.J. Black, CVPR 2024
  - https://yfeng95.github.io/ChatPose/