ENGN 8501: Week #7

Human Pose and Shape

2021 S2

What are we now

- This week is Week-7;

- Topics for the rest of this semester:
  - **Human pose and body shape** (w7)
  - **Graphical Models** (w8 + w9; Guest lecture given by Dr Zhiwei Xu)
  - Indoor Scene Understanding: room layout (w10; Guest lecture given by Dr Miaomiao Liu)
  - Indoor Scene Understanding: 3D object model fitting (w11)
  - **Outdoor Scene Understanding** for autonomous Driving (w12)

Many thanks to Class Reps for collecting feedback, comments

- Afsar Ahamed Asaraf Ali AfsarAhamed.AsarafAli@anu.edu.au
- Shreya Chawla <Shreya.Chawla@anu.edu.au>
- Ning Lai Ning.Lai@anu.edu.au

- Feedback and Issue identified:
  - Time-table clash
  - Workload: 7~9 hours per week spent on reading 2~4 papers
  - Too many diverse topics/subjects.
  - More feedback to Reports.
  - GPU access ?   (DUG no longer an option, but MLCV_GPU1 is open for our students).

This week:

- Monday Lecture: Human pose and body shape
- Thursday Tutorial: Python/Pytorch, and how to use MLCV1 GPU Cluster.
- Remaining assignments:
  - Reading Report-4 and Reading Report-5.

  - Final Project Report (PDF) 20%.
  - Seminar Presentation (PPT + MP4 video) 20%
  - Source Code (Zip file) 5%.

Outline for today’s lecture

- 1. Background introduction: 20 minutes
- 2. supervised paper reading session: about 30 minutes
- 3. breakout room: group discussion and report back to all: about 40 minutes.

Background Introduction:

Human pose, motion capture and body shape

Contents

- 2D Human pose estimation (2D match stick figure extraction)
- 3D Human Pose estimation (3D mocap)
- 3D Human body shape fitting (3d body shape)

Traditional Human Motion Caption Technologies

- Optical – uses video capture
  - passive – markers just reflect light
  - active – markers emit light

- Magnetic – active sensors sense their position and orientation in magnetic field

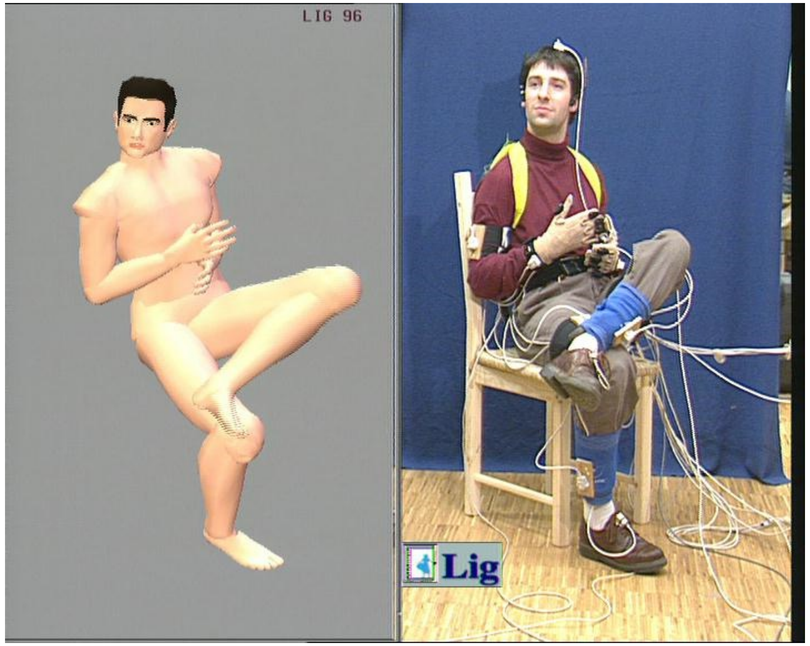
- Mechanical – rotors connected to limb-aligned rods record their status – for hands, optical sensors used sometimes

Mechanical

Ex: Metamotion
http://www.metamotion.com

Magnetic

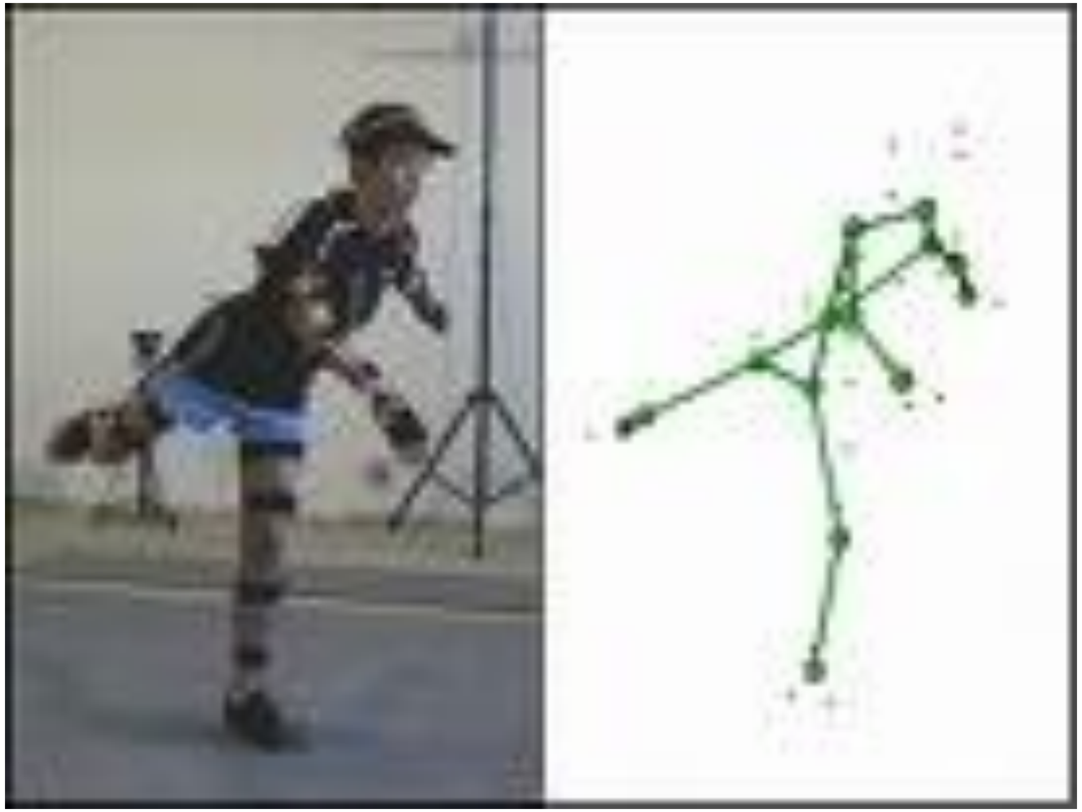
Ex: Ascension technology
http://www.ascension-tech.com/



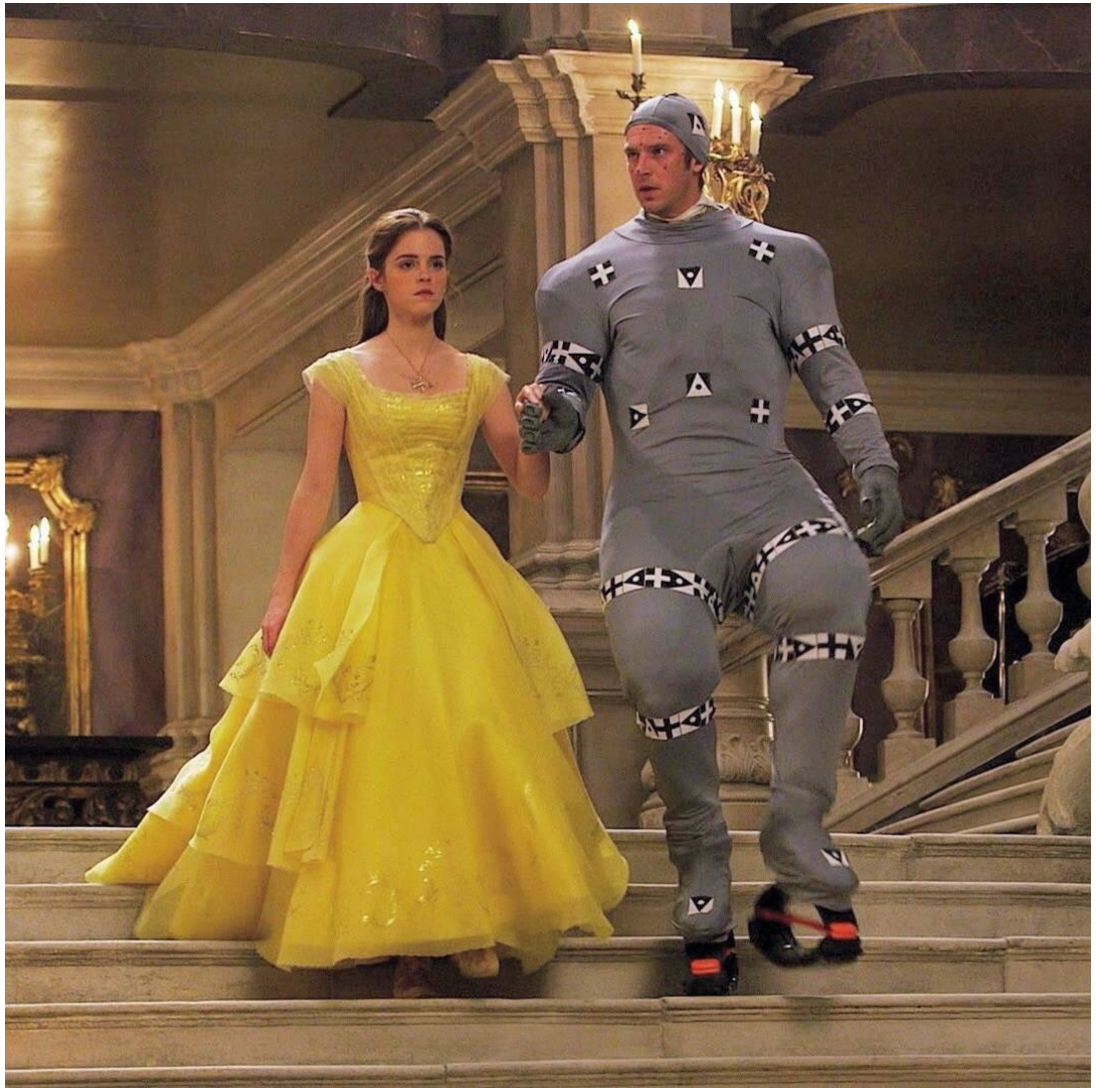
Optical - Active

Optical- Active

Optical - Passive



Optical- Passive



Body and Facial Motion Capture

Motion capture lab

Modern StoA MoCap techniques are mostly based on computer vision and machine learning (AI)

One of my 3D Mo-cap Systems (CMU & ANU)

Structure from Recurrent Motion: From Rigidity to Recurrency

Xiu Li[1,2] Hongdong Li[2,3] Hanbyul Joo[2] Yebin Liu[1] Yaser Sheikh[2]

Tsinghua University[1] Carnegie Mellon University[2] Australian National University[3]

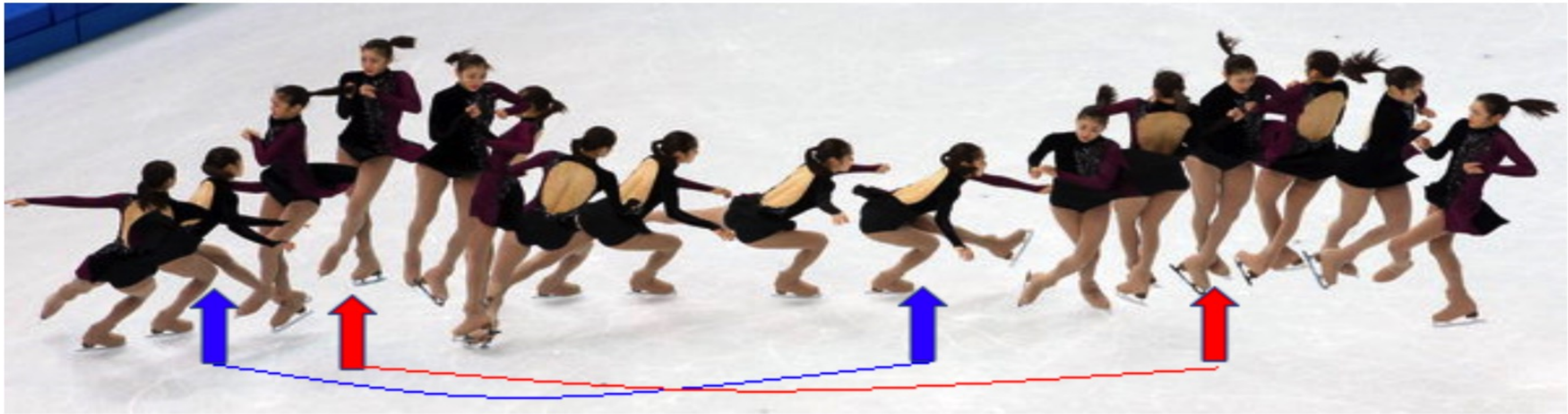
Abstract

This paper proposes a new method for Non-Rigid Structure-from-Motion (NRSfM) from a long monocular video sequence observing a non-rigid object performing recurrent and possibly repetitive dynamic action. Departing from the traditional idea of using linear low-order or low-rank shape model for the task of NRSfM, our method exploits the property of shape recurrency (i.e., many deforming shapes tend to repeat themselves in time). We show that

in rapid deformation, which are however common in reality.

This paper presents a new method for non-rigid structure from motion. Contrary to the traditional wisdom for NRSfM, we do not make a linear model assumption. Instead, we describe how to exploit shape *recurrency* for the task of non-rigid reconstruction. Specifically, we observe that in our physical world many deforming objects (and their shapes) tend to repeat themselves from time to time, or even only occasionally. In the context of SfM if a shape reoccurs in the video we say it is *recurrent*.

The main idea: finding recurrent human poses



Observation: Objects usually deform with some recurrent patterns.

$$\boldsymbol{S}(\boldsymbol{t}+\boldsymbol{\delta}) \simeq \boldsymbol{S}(\boldsymbol{t})$$

Task: Automatically identify those recurrent pairs and perform rigid SFM on them.

Example result: 3D human pose tracking

Example video clip(6s out of 3min)

2D skeleton detection

(Camera calibration data,both instinct and extrinsic are unavailable)

Combine human pose and SLAM

CMU walking sequence with Static Background

Camera pose is calculated through rigid-SfM using background points.
Dynamic human body is estimated using SFRM.

Human 3D body shape capture

Video Demo-1: CMU OpenPose

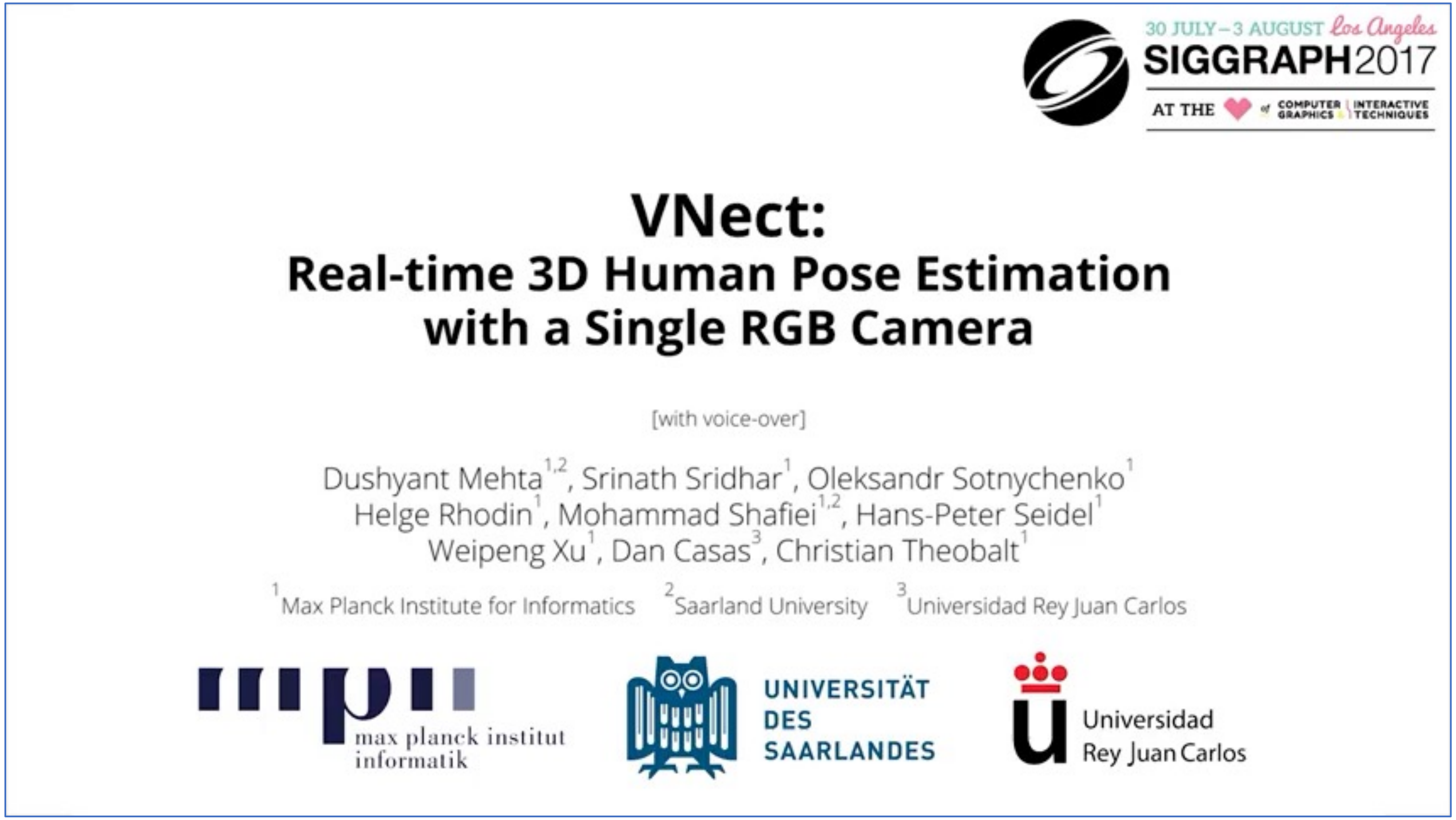
Real-time Multi-Person 2D Pose Estimation Using Part Affinity Fields

Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh

Carnegie Mellon University

Another demo of wrnchAI vs openPose

Video Demo-2:


30 JULY – 3 AUGUST Los Angeles
SIGGRAPH2017
AT THE ♥ of COMPUTER GRAPHICS & INTERACTIVE TECHNIQUES
VNect:
Real-time 3D Human Pose Estimation
with a Single RGB Camera
[with voice-over]
Dushyant Mehta1,2, Srinath Sridhar1, Oleksandr Sotnychenko1
Helge Rhodin1, Mohammad Shafiei1,2, Hans-Peter Seidel1
Weipeng Xu1, Dan Casas3, Christian Theobalt1
1Max Planck Institute for Informatics 2Saarland University 3Universidad Rey Juan Carlos
max planck institut informatik
UNIVERSITÄT DES SAARLANDES
Universidad Rey Juan Carlos

Video Demo-3: add 3D **body shape** and cloth

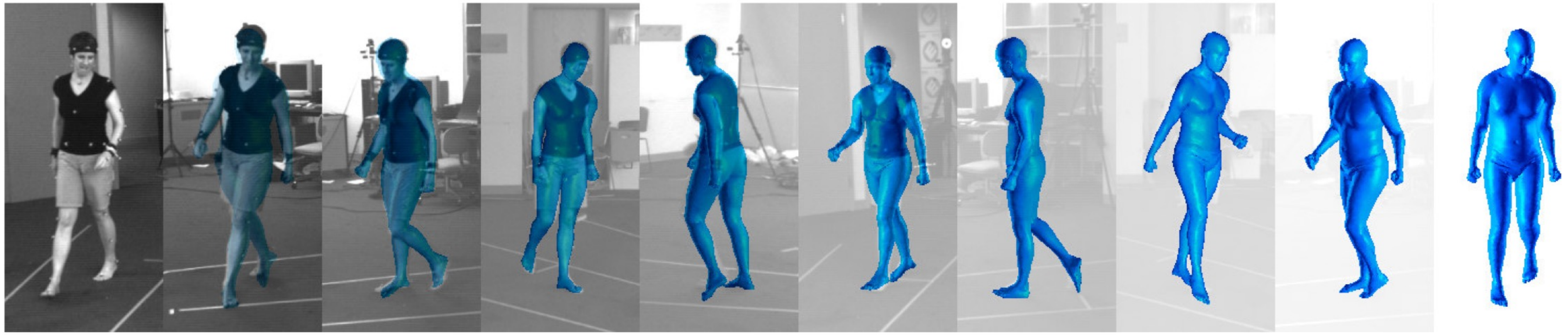
Papers to read today

- Optimization based approach:
  - Paper-1 (Brown University)
  - Paper-2 (Brown University)
- Deep learning based approach:
  - Paper-3 (Keep it SMPL: SMPLify)
  - Paper-4 (end-to-end human body mesh **regression**)

Paper-1 to read today

CVPR' 2008

Detailed Human Shape and Pose from Images



[1]Alexandru O. Bălan [1]Leonid Sigal [1]Michael J. Black [2]James E. Davis [3]Horst W. Haussecker
[1]Department of Computer Science, Brown University, Providence, RI 02912, USA
[2]Computer Science Department, UC Santa Cruz, Santa Cruz, CA 95064, USA
[3]Intel Research, Santa Clara, CA 95054, USA

Algorithm Overview

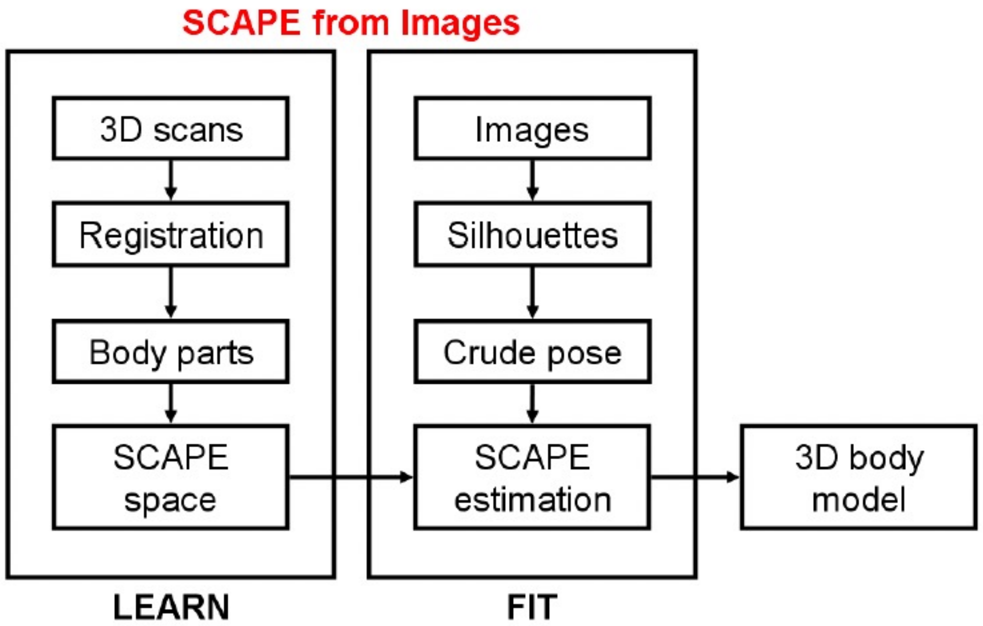


Figure 2. **Algorithm Overview.** A learning phase is used to build the 3D body model from range scans and follows the approach proposed in [1]. Our contribution provides a method for fitting the pose and shape parameters of the model to image data.

SCAPE: a parametric body model

SCAPE Overview. The *template mesh* acts as a reference mesh that is morphed into other poses and body shapes to establish correspondence between all meshes. Let (x_1, x_2, x_3) be a triangle belonging to the template mesh and (y_1, y_2, y_3) be a triangle from an instance mesh. We define the two edges of a triangle starting at x_1 as $\Delta x_j = x_j - x_1,\ j = 2, 3$.

The deformation of one mesh to another is modeled as a sequence of linear transformations applied to the triangle edges of the template mesh:

$$\Delta y = RDQ\Delta x \tag{1}$$

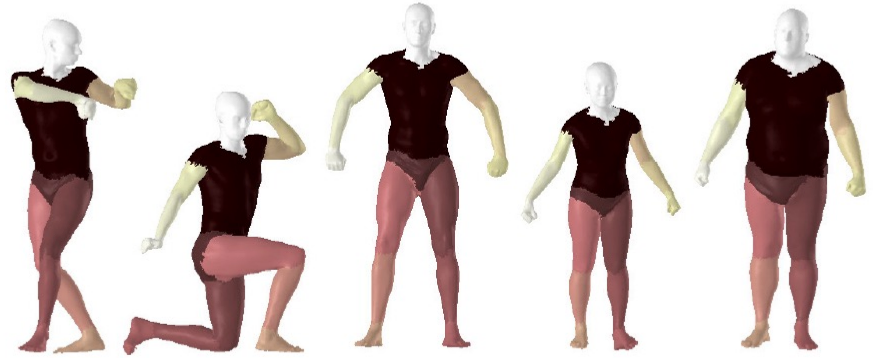
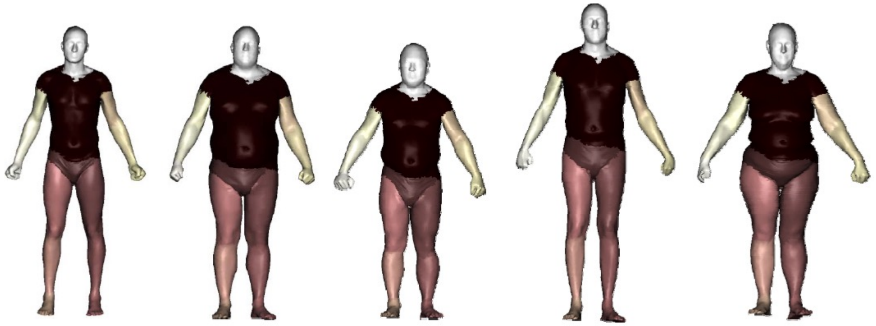


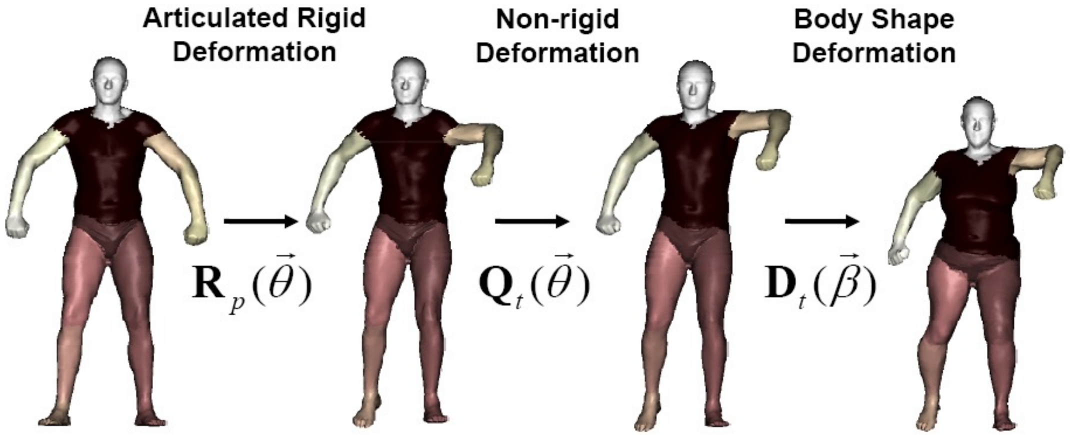
Figure 3. **3D Body Meshes.** Two example meshes from the pose set, the template mesh, and two example meshes from the body shape set (left to right).

SCAPE Body Model

Pose Training Set (70 poses)



Shape Training Set (2100 bodies)



$\vec{\theta}$ – pose parameters (40D)

$\vec{\beta}$ – shape parameters (20D)

Parameters (state): $\Theta_B(\vec{\theta}, \vec{\beta})$

D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape completion and animation of people. *SIGGRAPH, 24(3):408–416, 2005.*

Stochastic Optimisation

To avoid becoming trapped in local optima, predicted particles are re-weighted using an annealed version of the likelihood function: $f^{(k)}(s) = \left(p(I|s)\right)^{t^{(k)}} p(s)$, where $t^{(k)}$ is an annealing temperature parameter optimized so that approximately half the samples get re-sampled.

4.1. Initialization

There exist a number of techniques that can be used to initialize the stochastic search; for example, pose prediction from silhouettes [19], voxel carving skeletonization [5], or loose-limbed body models [18]. Here we employ an existing human tracking algorithm [2] based on a cylindrical body model. The method is initialized in the first frame from marker data, and the position and joint angles of the body are automatically tracked through subsequent frames. The method uses an annealed particle filtering technique for inference, uses fairly weak priors on joint angles, enforces non-interpenetration of limbs and takes both edges and silhouettes into account. The recovered position and joint angles together with the mean body shape parameters are used to initialize the stochastic search of the SCAPE parameters.

5. Image Cost Function

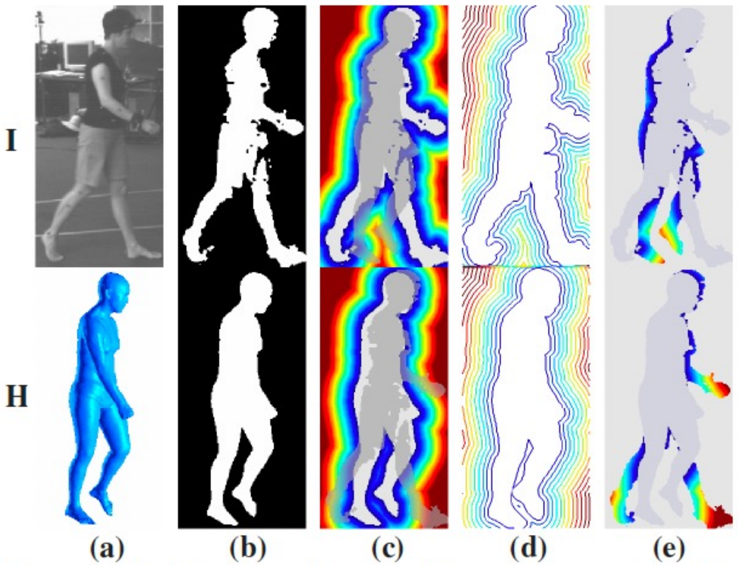


Figure 4. **Cost function.** **(a)** original image I (top) and hypothesized mesh H (bottom); **(b)** image foreground silhouette F^I and mesh silhouette F^H, with 1 for foreground and 0 for background; **(c)** Chamfer distance maps C^I and C^H, which are 0 inside the silhouette; the opposing silhouette is overlaid transparently; **(d)** contour maps for visualizing the distance maps; **(e)** per pixel silhouette distance from F^H to F^I given by $\sum_p F^H_p \cdot C^I_p$ (top), and from F^I to F^H given by $\sum_p F^I_p \cdot C^H_p$ (bottom).

Results

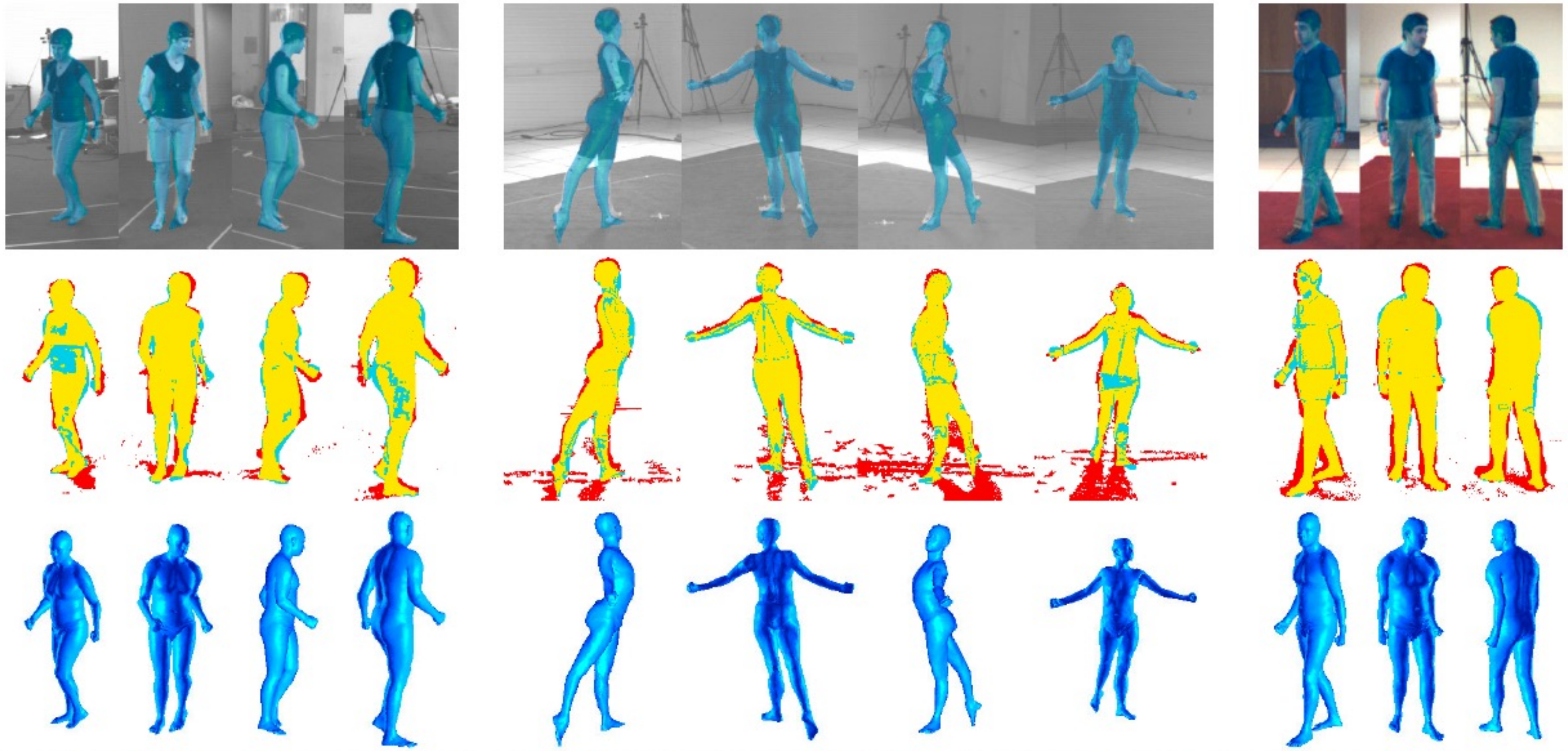


Figure 5. **SCAPE-from-image results.** Reconstruction results based on the views shown for one male and two female subjects, in walking and ballet poses, wearing tight fitting as well as baggy clothes. *(top)* Input images overlaid with estimated body model. *(middle)* Overlap (yellow) between silhouette (red) and estimated model (blue). *(bottom)* Recovered model from each camera view.

Results

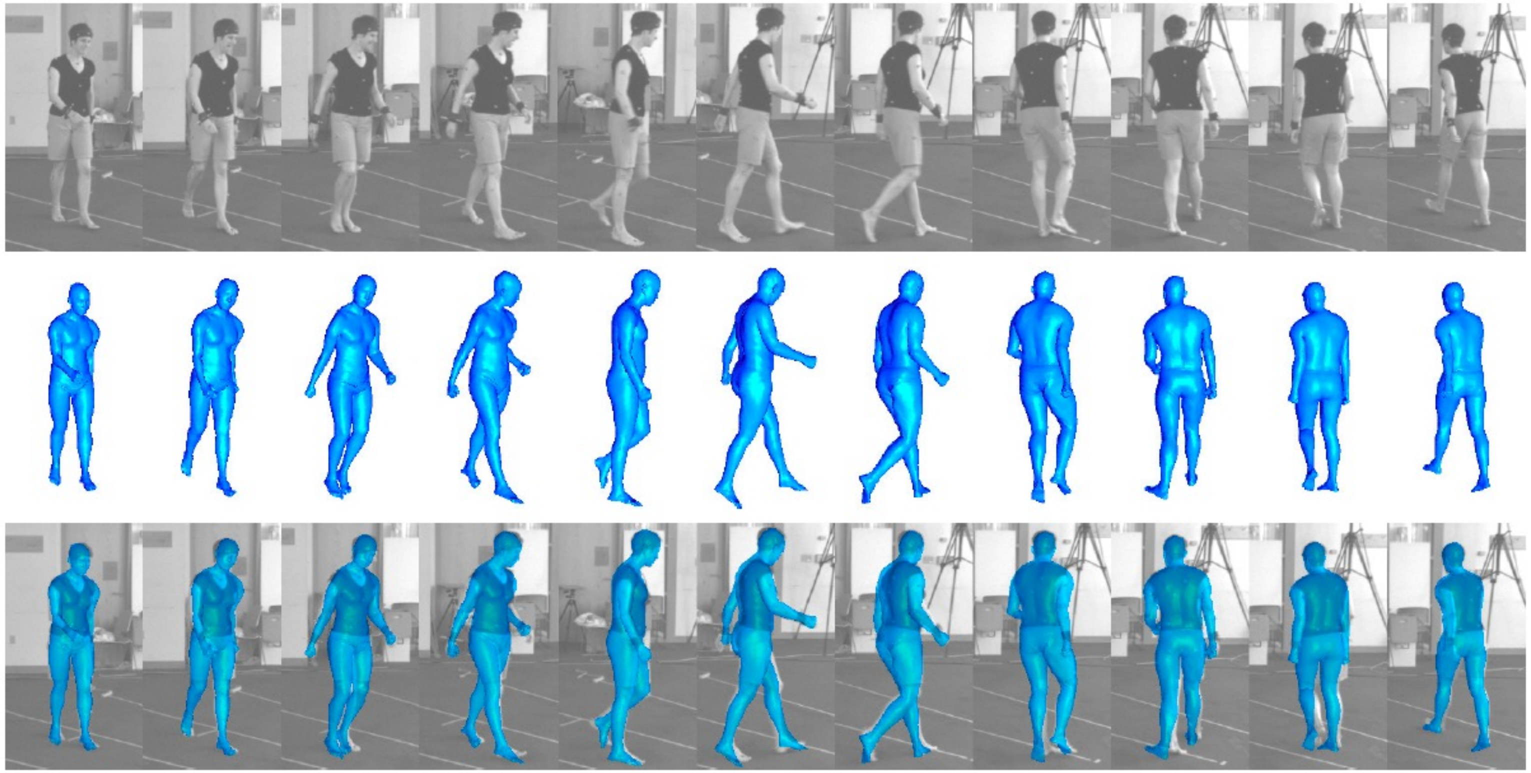


Figure 6. **First row:** Input images. **Second row:** Estimated mesh models. **Third row:** Meshes overlaid over input images. By applying the shape parameters recovered from 33 frames to the template mesh placed in a canonical pose, we obtained a shape deviation per vertex of $8.8 \pm 5.3mm$, computed as the mean deviation from the average location of each surface vertex.

Paper-2 to read today

ICCV' 2009

Estimating Human Shape and Pose from a Single Image

Peng Guan Alexander Weiss Alexandru O. Bălan Michael J. Black
Department of Computer Science, Brown University, Providence, RI 02912, USA
{pguan, aweiss, alb, black}@cs.brown.edu

Abstract

We describe a solution to the challenging problem of estimating human body shape *from a single photograph or painting. Our approach computes shape and pose parameters of a 3D human body model directly from monocular image cues and advances the state of the art in several directions. First, given a user-supplied estimate of the subject's height and a few clicked points on the body we estimate an initial 3D articulated body pose and shape. Second, using this initial guess we generate a tri-map of regions inside, outside and on the boundary of the human, which is used to segment the image using graph cuts. Third, we learn a low-dimensional linear model of human shape in which variations due to height are concentrated along a single dimension, enabling height-constrained estimation of body shape. Fourth, we formulate the problem of* parametric human shape from shading. *We estimate the body pose, shape and reflectance as well as the scene lighting that produces a synthesized body that robustly matches the image evidence. Quantitative experiments demonstrate how smooth shading provides powerful constraints on human shape. We further demonstrate a novel application in which we extract 3D human models from archival photographs and paintings.*

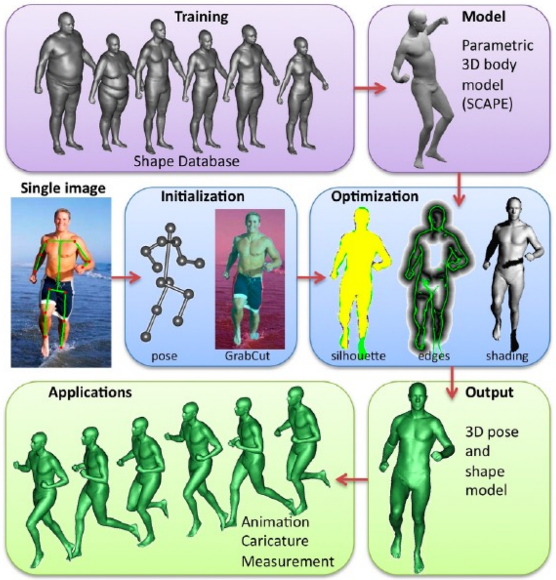


Figure 1. **Overview.** Given a single image and minimal user input, we compute an initial pose, light direction, shape and segmentation. Our method optimizes 3D body shape using a variety of image cues including silhouette overlap, edge distance, and smooth shading. The recovered body model can be used in many ways; animation using motion capture data is illustrated.

Body shape and pose from 1 image?

Introduction

What others do

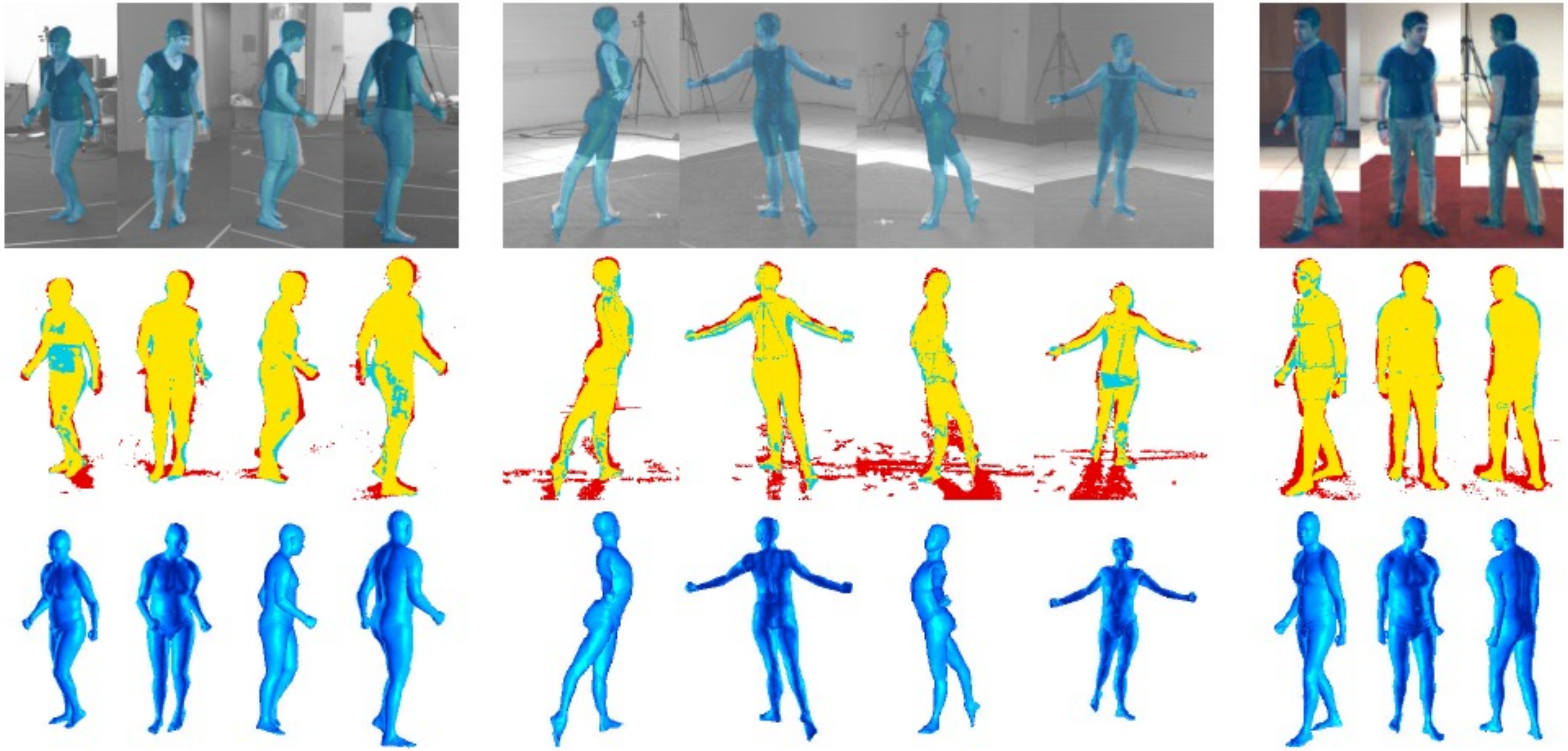
- Estimating 3D human pose in uncalibrated monocular imagery
- Use silhouette in multi-camera setting to recover 3D body shape
- Most work assumes the existence of a known background to extract foreground silhouette
- In previous body models, height is correlated with other shape variations

What we do

- Estimating both 3D shape and pose in uncalibrated monocular imagery
- Use additional monocular cues including smooth shading
- Use GrabCut (graph-cut) to produce foreground region
- Make height variation concentrated along one shape basis vector, which allows "height constrained fitting"

Previous Work

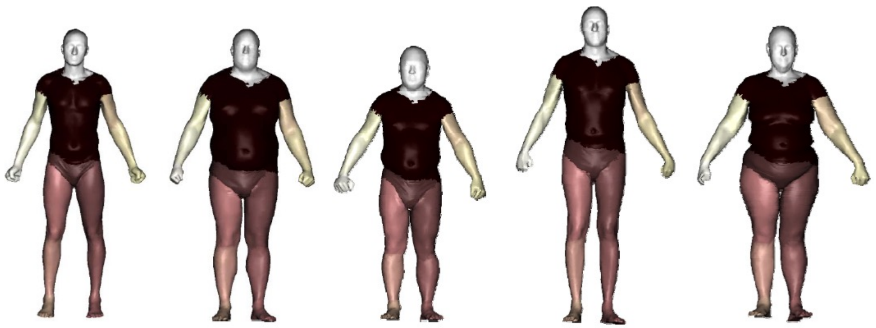
3D pose and shape estimation from multiple, calibrated, cameras



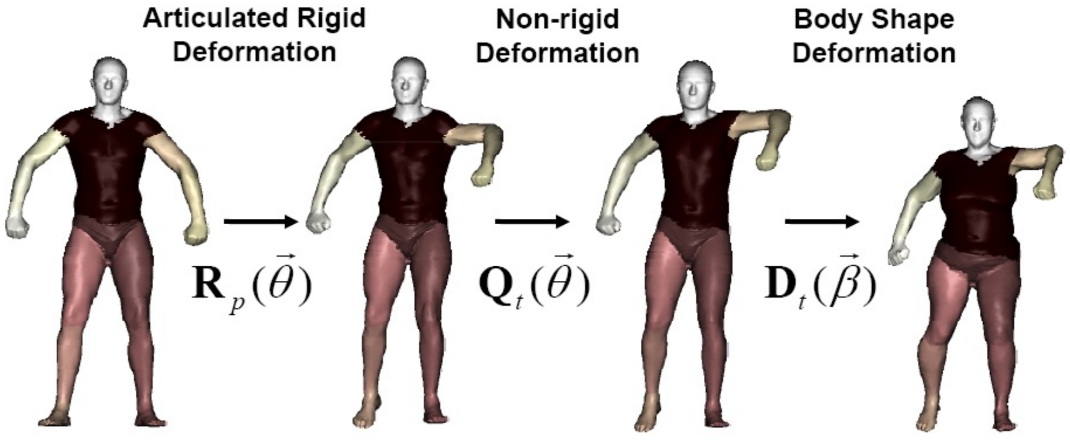
Balan, A., Sigal, L., Black, M. J., Davis, J., Haussecker, H, "Detailed human shape and pose from images", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, *CVPR*, Minneapolis, June 2007

SCAPE Body Model

Pose Training Set (70 poses)



Shape Training Set (2100 bodies)



$\vec{\theta}$ – pose parameters (40D)

$\vec{\beta}$ – shape parameters (20D)

Parameters (state): $\Theta_B(\vec{\theta}, \vec{\beta})$

D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape completion and animation of people. *SIGGRAPH, 24(3):408–416, 2005.*

Body shape/pose from 1 image: **Problems**

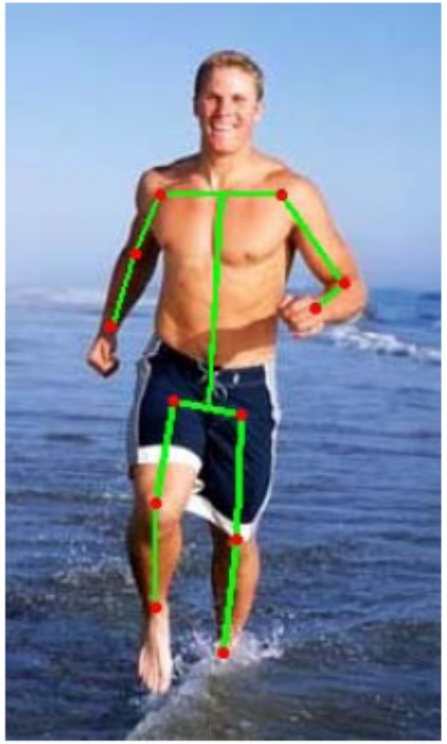
1. High dimensional body model (shape and pose) – initialization problem.
2. Background unknown
3. Single, monocular image
    1. poorly constrained
    2. Shape/Pose ambiguities
4. Silhouette insufficient

Previous Work (CJ Taylor 2000, Lee & Chen 1985)

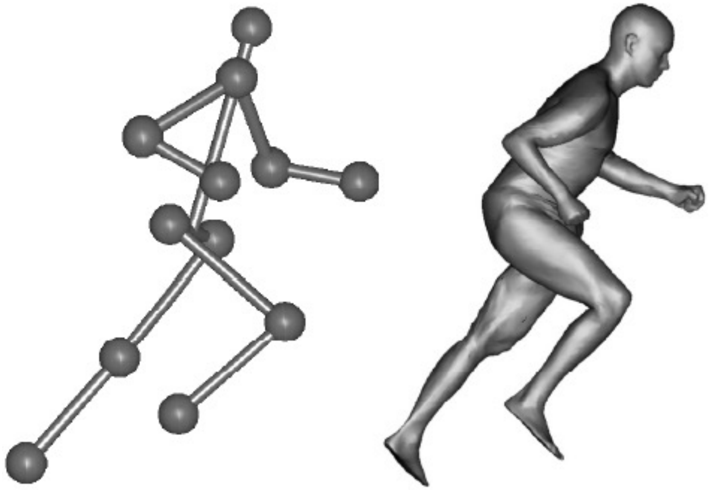
- 3D pose estimation using orthographic camera assumption

C. J. Taylor, “Reconstruction of Articulated Objects from Point Correspondences in a Single Uncalibrated Image”, Computer Vision and Image Understanding, Vol: 80, No: 10, Pgs: 349-363, October 2000

Solution 1: Pose Initialization

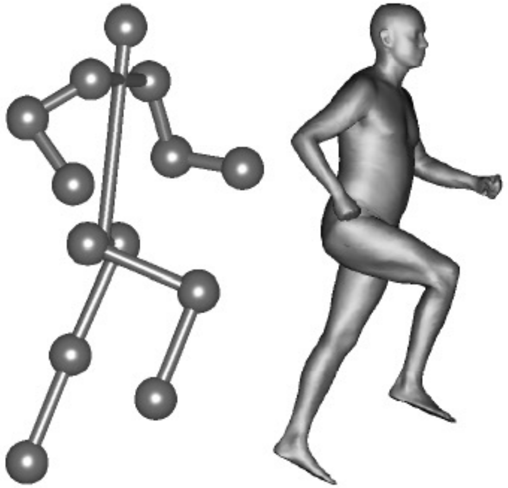


Clicked Points



Orthographic Projection
(Taylor 2000)

Better



Perspective Projection
(Lee & Chen 1985)

Shape: initialized to mean body shape.

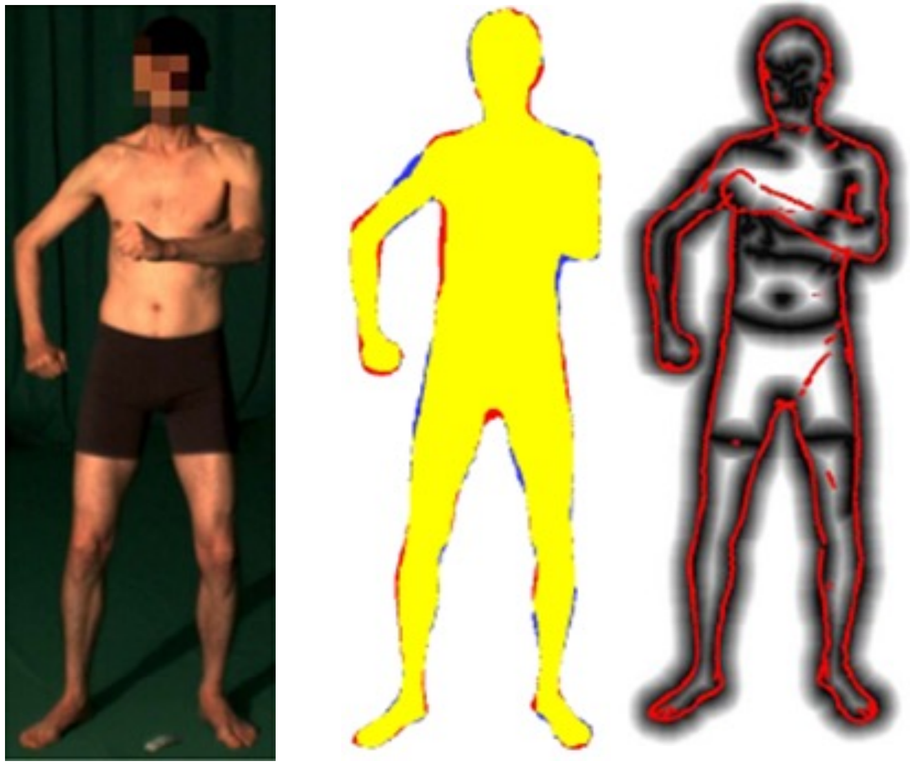
Solution 2: Segmentation

Initial Silhouette **Tri-map** **GrabCut** **Segmentation Result**

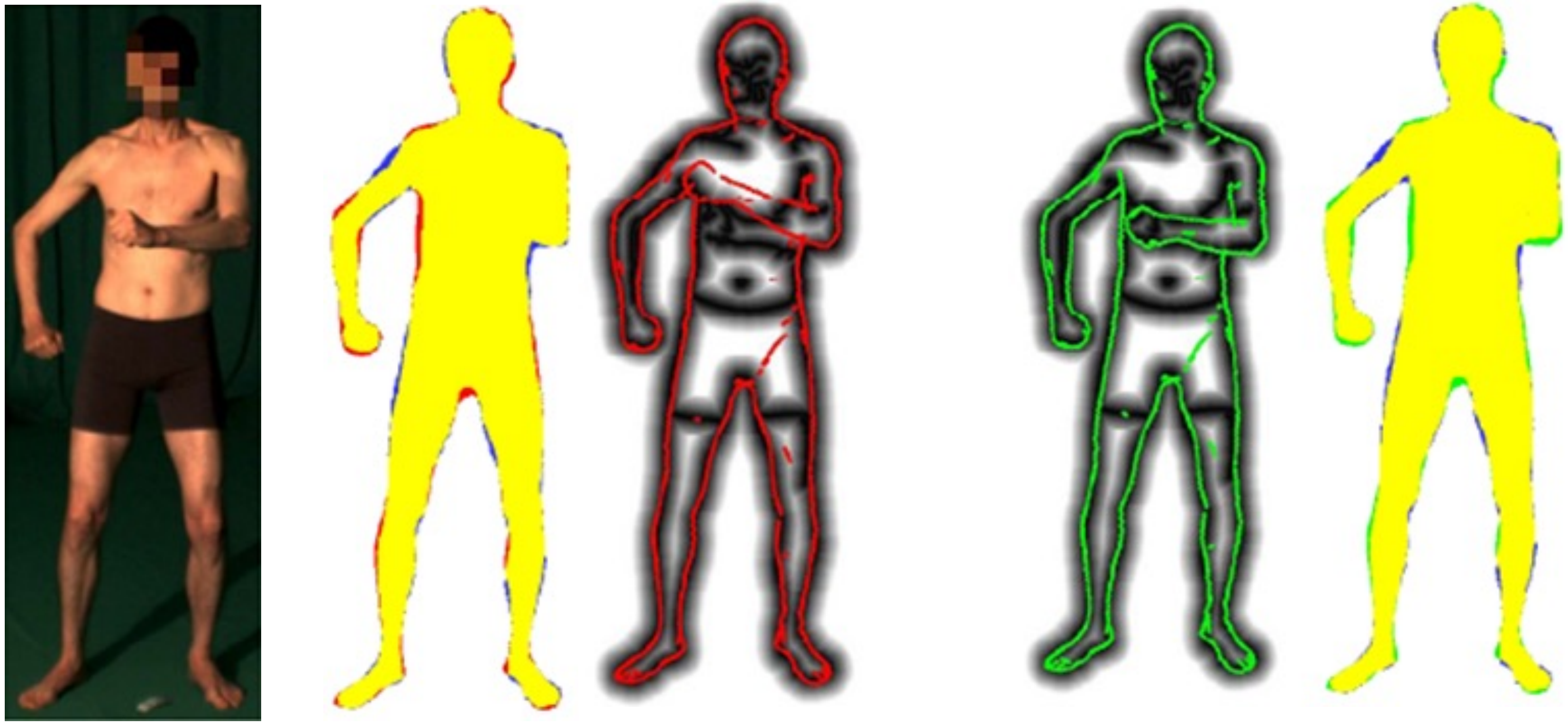
$$E_{Silhouette}(\vec{\theta}, \vec{\beta}) = \sum_{p \in allPixels} (I_p - \hat{I}_p(\vec{\theta}, \vec{\beta}))$$

C. Rother, V. Kolmogorov, and A. Blake. "GrabCut": Interactive foreground extraction using iterated graph cuts. *SIGGRAPH,* 23(3):309–314, 2004.

Problem: Silhouette not sufficient

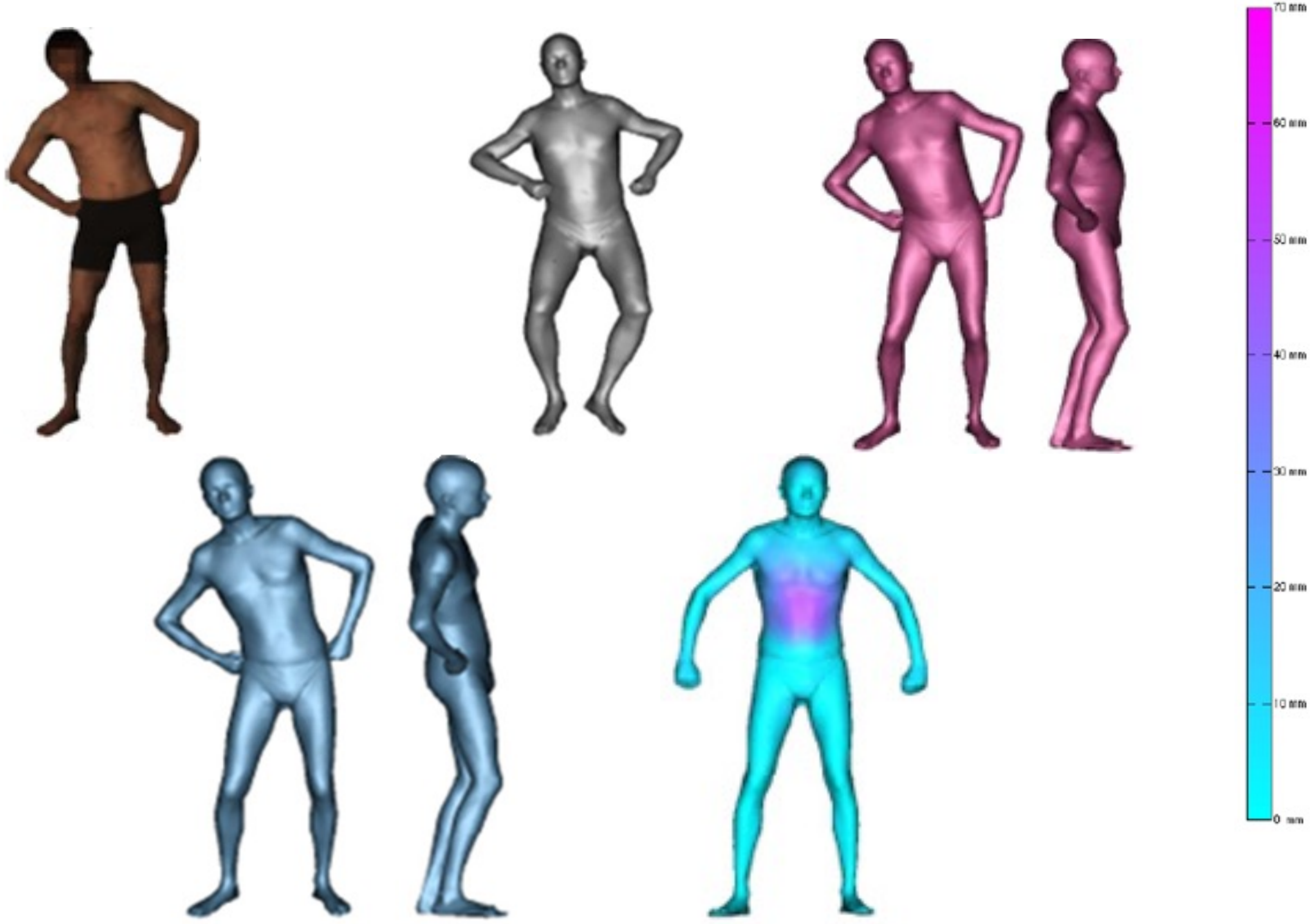


Solution 3: Edge Cues



$$E_{Edge}(\vec{\theta},\vec{\beta}) = \sum_{p \in m\_c} (C_p(\vec{\theta},\vec{\beta}))$$

Problem: Shape not well constrained



Solution 4: Shape from Shading

Light parameters: azimuth, elevation, distance $\Theta_L = [\gamma, \phi, z]$

Body parameters: shape, pose $\Theta_B = [\vec{\beta}, \vec{\theta}]$

Body reflectance properties $\Theta_R = [\vec{a}, \vec{s}, b, \alpha]$

Set of visible vertices by camera $\{i\}$

Light intensity l

Light direction $\vec{\ell}_i(\Theta_B, \Theta_L)$

Normal $\vec{n}_i(\Theta_B)$

Vertex $\vec{x}_i$

Albedo a_i

Specularity s_i

Half vector $\vec{h}_i(\Theta_B, \Theta_L)$

Specular exponent α

Background illumination b

Model appearance $\hat{r}_i(\Theta_B, \Theta_L, \Theta_R)$

Image appearance r_i

$$\hat{r}_i(\Theta_B, \Theta_L, \Theta_R) = \begin{cases} b + a_i(\vec{\ell}_i \cdot \vec{n}_i) l + s_i(\vec{h}_i \cdot \vec{n}_i)^{\alpha} l & \text{Direct illumination} \\ b & \text{In shadow} \end{cases}$$

M. de la Gorce, N. Paragios and David Fleet. Model-Based Hand Tracking with Texture, Shading and Self-occlusions. IEEE Conference in Computer Vision and Pattern Recognition (CVPR), Anchorage 2008.

Shading/Overall Cost function

Shading cost function:

Minimize:

$$E_{Shading}(\Theta_B, \Theta_R, \Theta_L) =$$

Data Term

$$\sum_{i \in visible} \left\{ \rho_{\eta 1}(\hat{r}_i(\Theta_B, \Theta_R, \Theta_L) - r_i) + \delta_1 \sum_{j \in N(i)} \frac{\rho_{\eta 2}(a_j - a_i)}{\left\| \vec{x}_j - \vec{x}_i \right\|} + \delta_2 \sum_{j \in N(i)} \frac{\rho_{\eta 3}(s_j - s_i)}{\left\| \vec{x}_j - \vec{x}_i \right\|} \right\}$$

Spatial Terms: Piece-wise Constancy

Robust Error Function: $\rho_\eta(x) = \frac{x^2}{\eta^2 + x^2}$

Overall cost function:

Minimize:

$$E_{total} = \lambda_1 E_{Silhouette}(\Theta_B) + \lambda_2 E_{Edge}(\Theta_B) + \lambda_3 E_{Shading}(\Theta_B, \Theta_R, \Theta_L)$$

Experiment: Lab Images

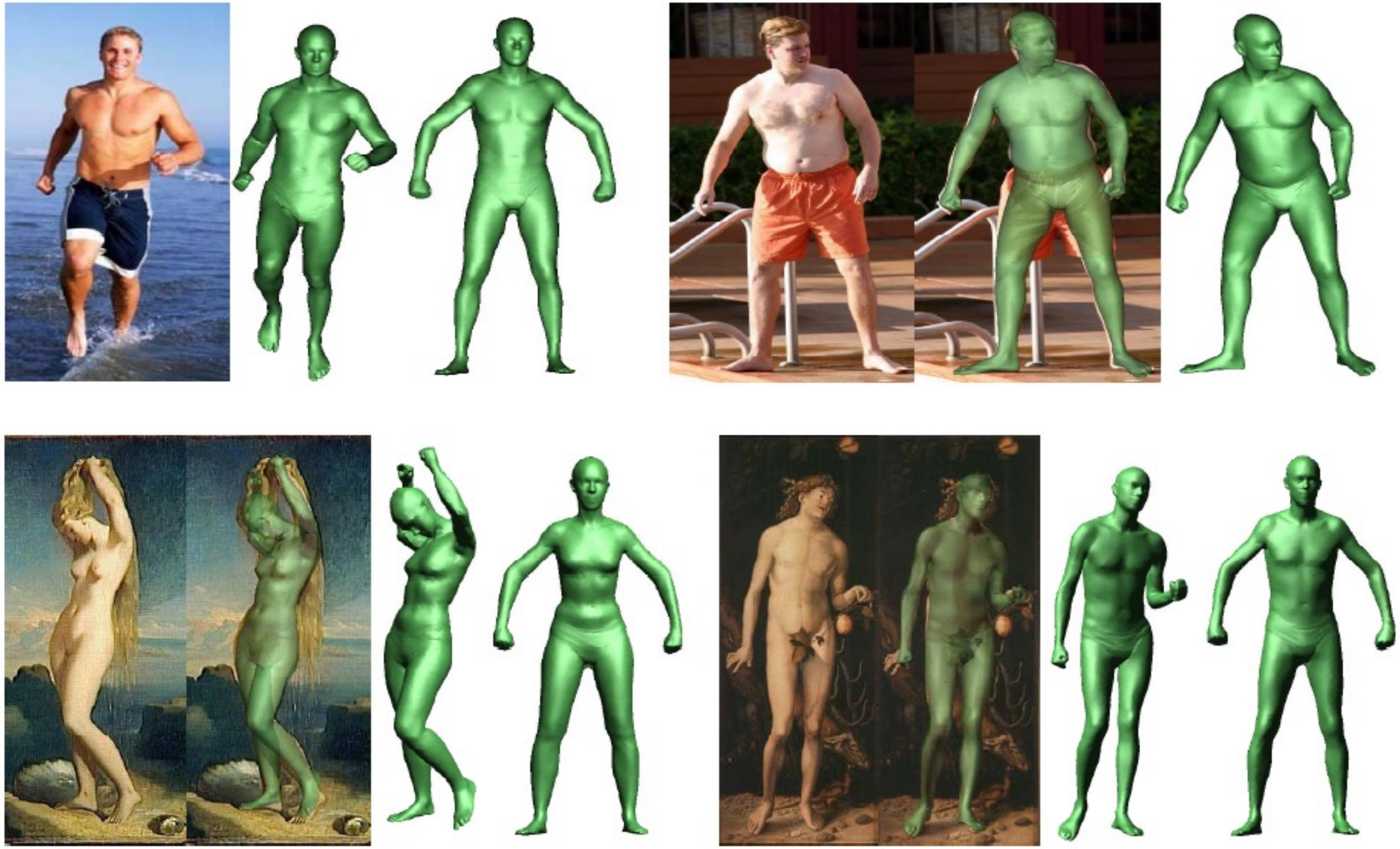
Quantitative Comparison

	Chest Size (cm)		
	SE	***SES***	***GT***
Pose 1	95.7 (+3.1)	92.7 (+0.1)	92.6
Pose 2	84.3 (-7.3)	87.1 (-4.5)	91.6
Pose 3	95.4 (+4.0)	91.9 (+0.5)	91.4

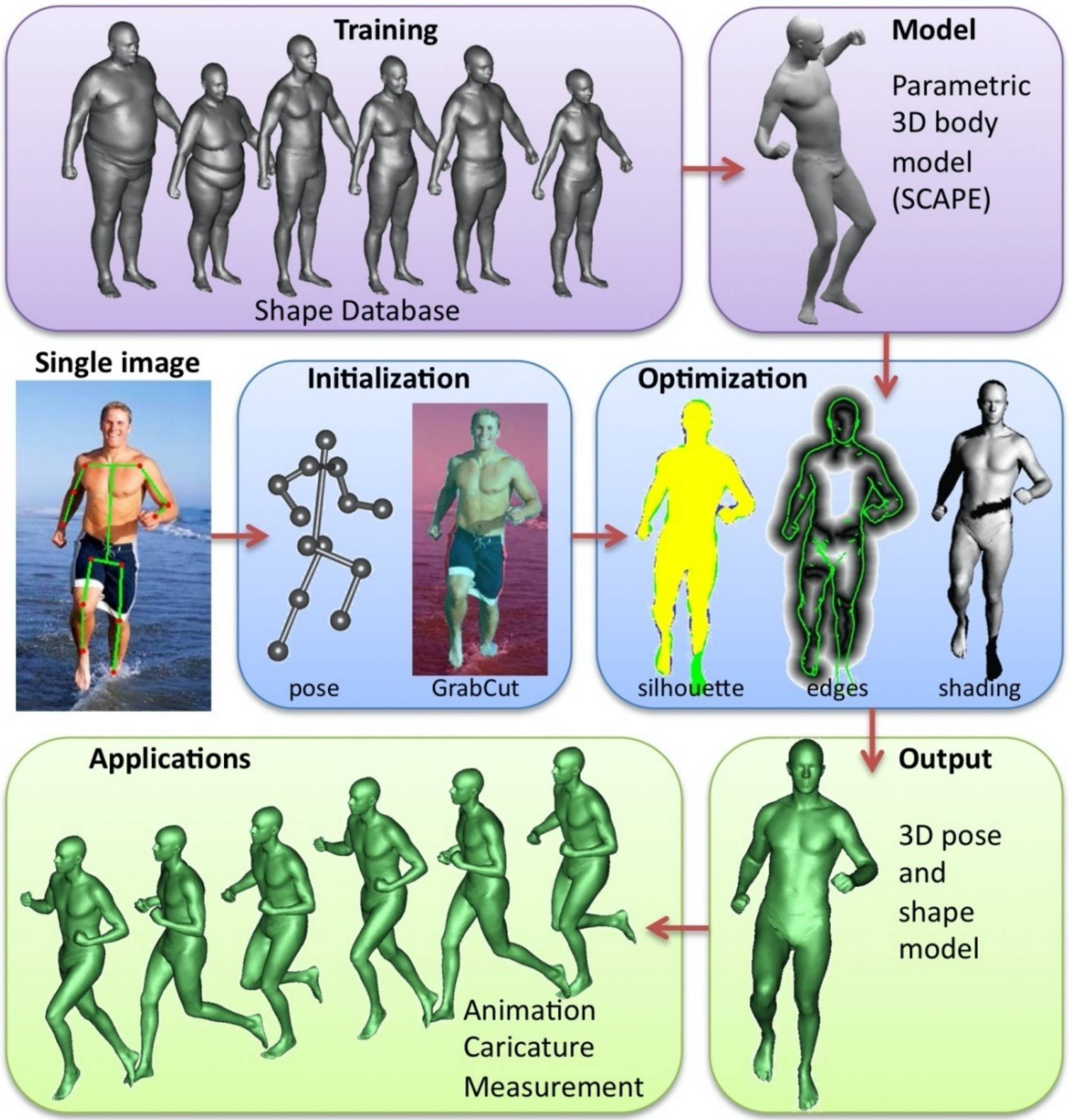
Waist Size (cm)		
SE	***SES***	***GT***
86.4 (+6.2)	79.6 (-0.6)	80.2
83.7 (+4.3)	78.5 (-0.9)	79.4
88.0 (+7.7)	76.9 (-3.4)	80.3

Body Weight (kg)		
SE	***SES***	***GT***
72.0 (+8.2)	65.4 (+1.6)	63.8
62.5 (-0.7)	62.4 (-0.8)	63.2
70.8 (+8.2)	63.5 (+0.9)	62.6

Experiments



Summary


Training
Shape Database
Model
Parametric 3D body model (SCAPE)
Single image
Initialization
pose
GrabCut
Optimization
silhouette
edges
shading
Applications
Animation
Caricature
Measurement
Output
3D pose and shape model

Paper-3 to read today

Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image

Federica Bogo[2,*], Angjoo Kanazawa[3,*], Christoph Lassner[1], Peter Gehler[1,4], Javier Romero[1], Michael J. Black[1]

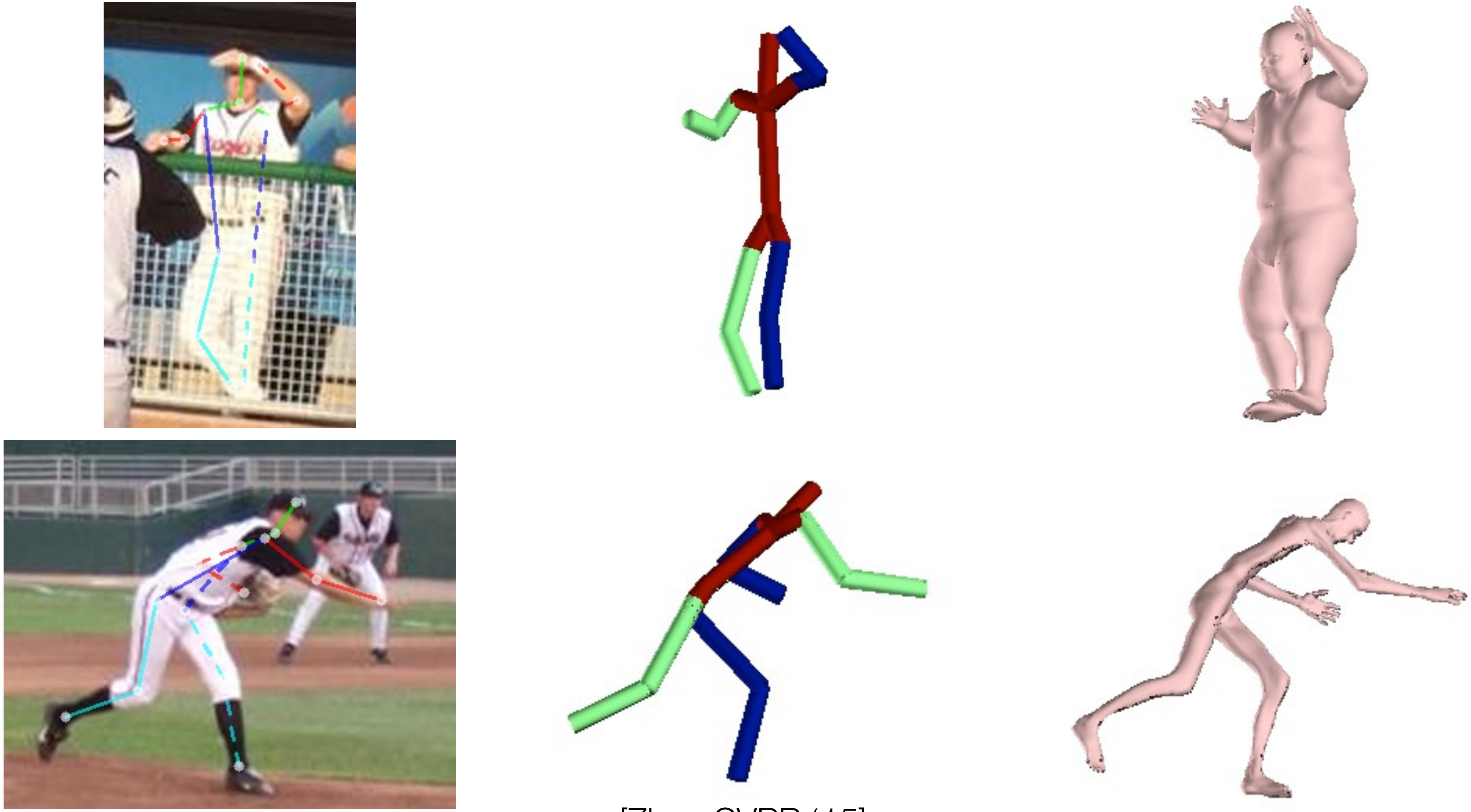
[1]Max Planck Institute for Intelligent Systems, Tübingen, Germany
[2]Microsoft Research, [3]University of Maryland, [4]University of Tübingen
febogo@microsoft.com, kanazawa@umiacs.umd.edu
{christoph.lassner, pgehler, jromero, black}@tue.mpg.de

Project Page

- https://ps.is.mpg.de/publications/bogo-eccv-2016

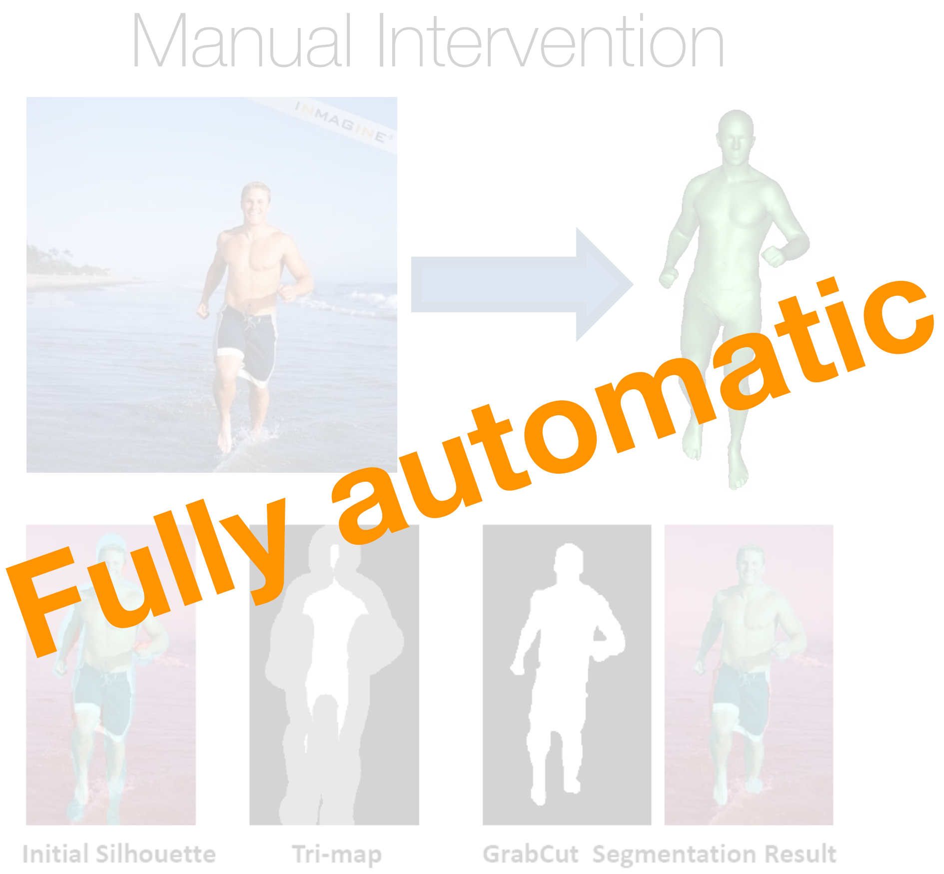
3D Shape and Pose from a Single Image

3D Pose Estimation



[Zhou CVPR '15]

Pose and Shape, but …



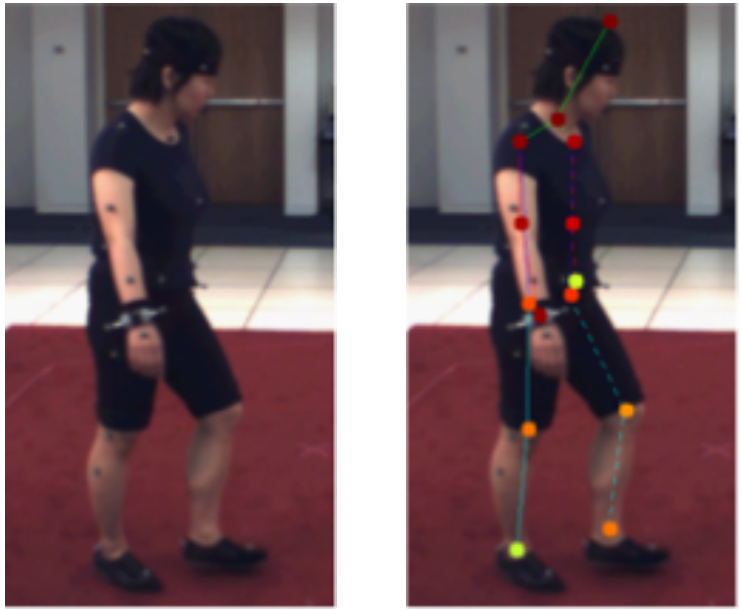
[Guan et al. ICCV '09, Hasler et al. CVPR '10, Zhou et al. SIGGRAPH '10]

[Bălan et al. CVPR '07, Jain et al. SIGGRAPH '10,
Rhodin et al. ECCV '16]

Our Results

Overview

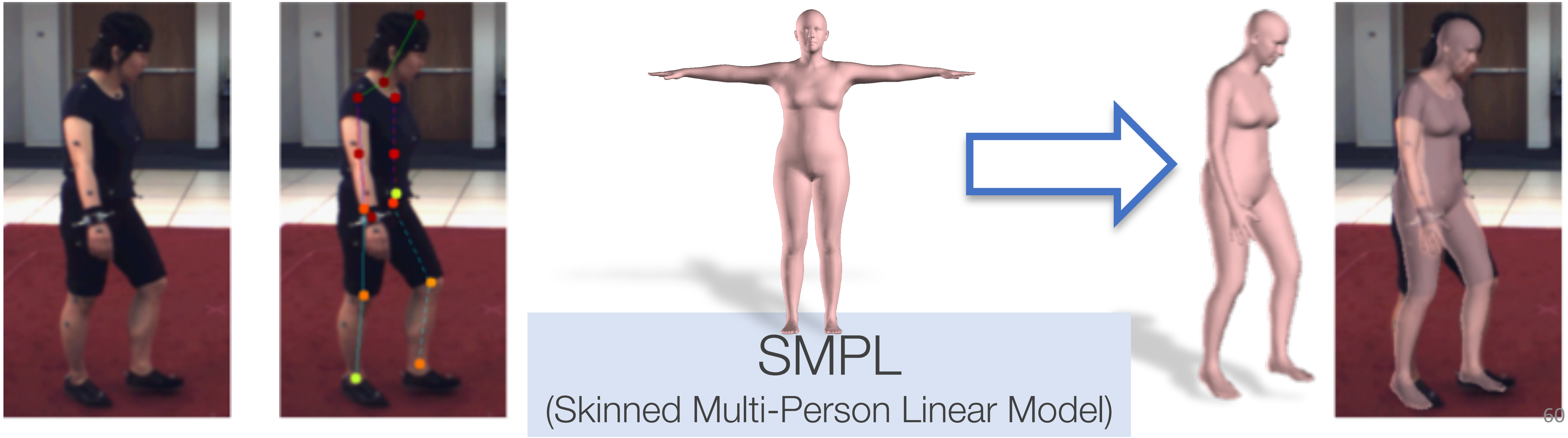
1. Automatic joint detection via CNNs



Overview

1. Automatic joint detection via CNNs

2. A good generative model of 3D pose and shape



Overview SMPLify

1. Automatic joint detection via CNNs — BOTTOM UP

2. A good generative model of 3D pose and shape — TOP DOWN

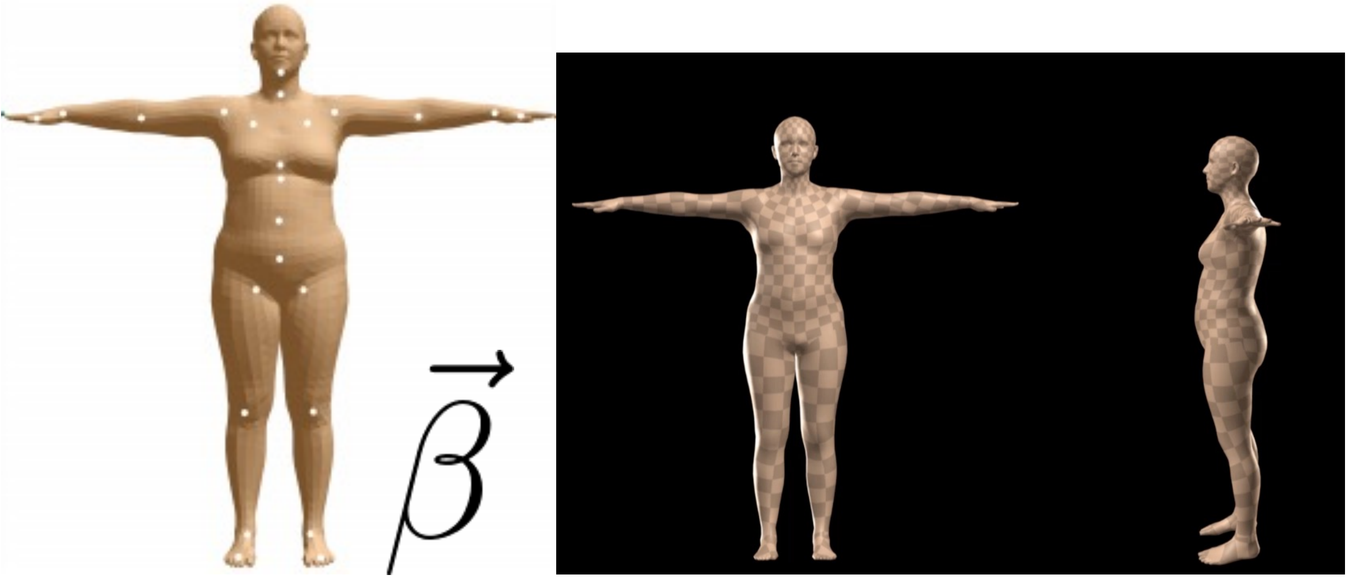
2D Joint Detection

Supervised end-to-end training to detect joints with confidence values.

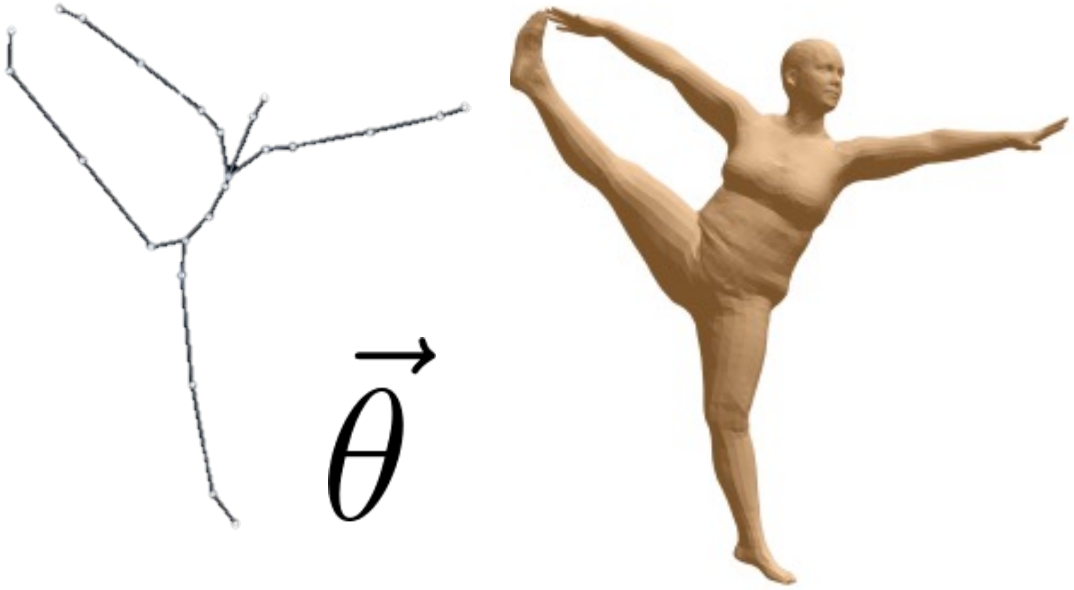
[Pishchulin et al. DeepCut CVPR 2016]

SMPL body model, improves over SCAPE

Shape: PCA coefficients

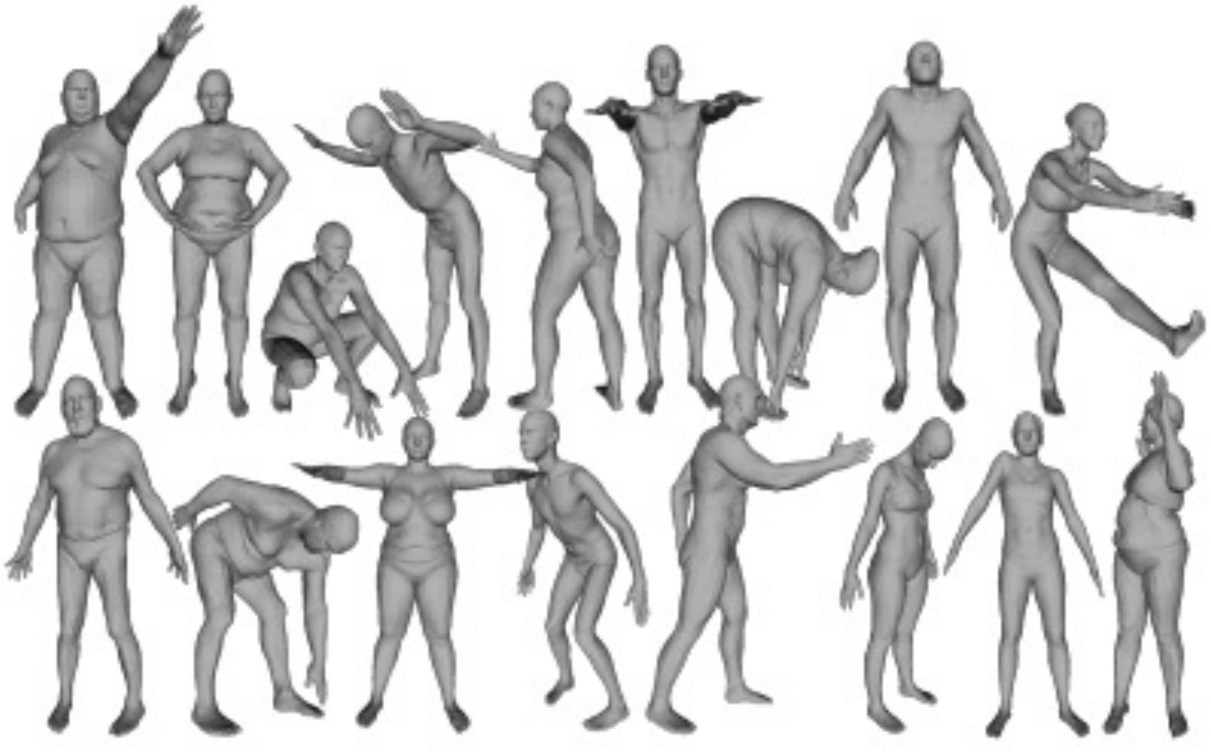


Pose: Rotation of 23 joints

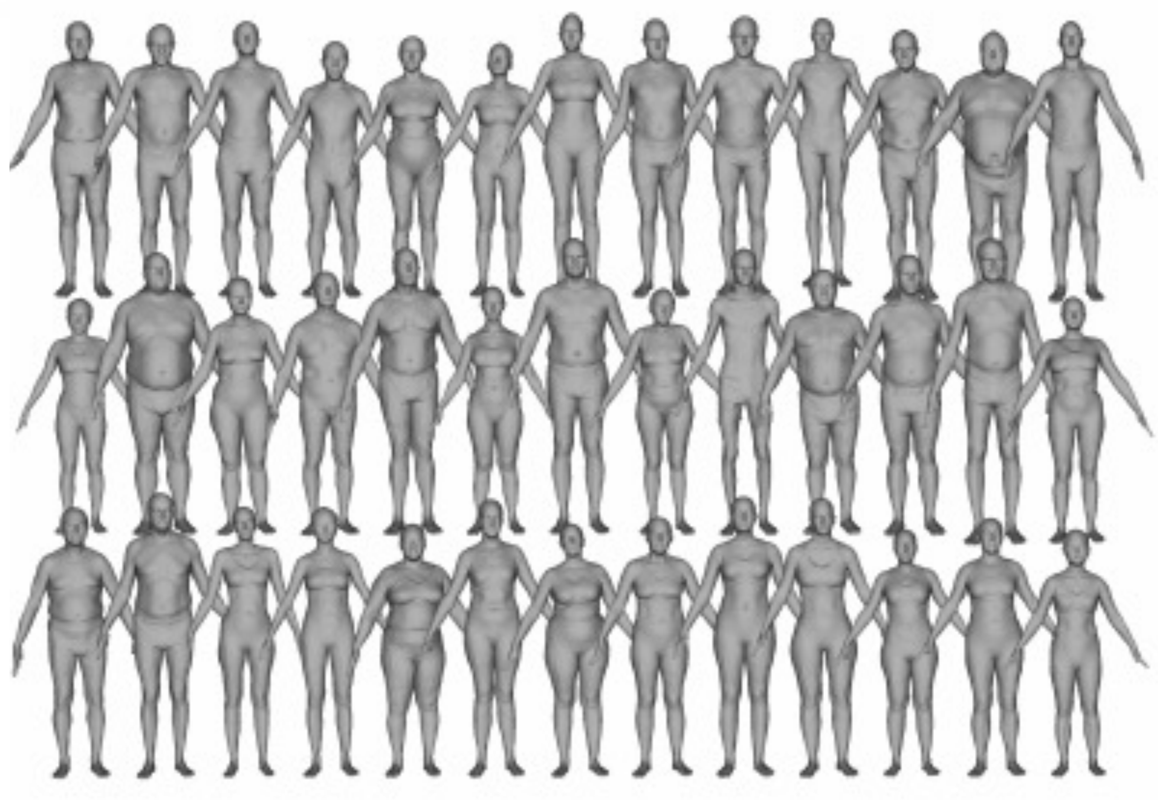


[Loper et al. SIGGRAPH Asia ‘15]

Pose and Shape Priors

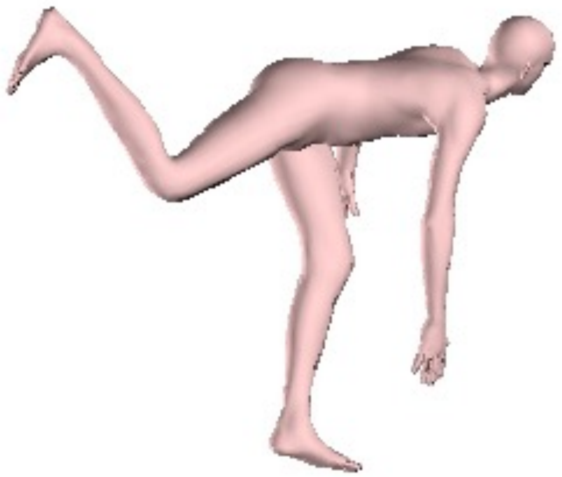


$$E_\theta(\vec{\theta})$$ Pose Prior



$$E_\beta(\vec{\beta})$$ Shape Prior

SMPLify Objective Function



camera joints

$$E(\vec{\beta}, \vec{\theta}, K; J_{est}) =$$

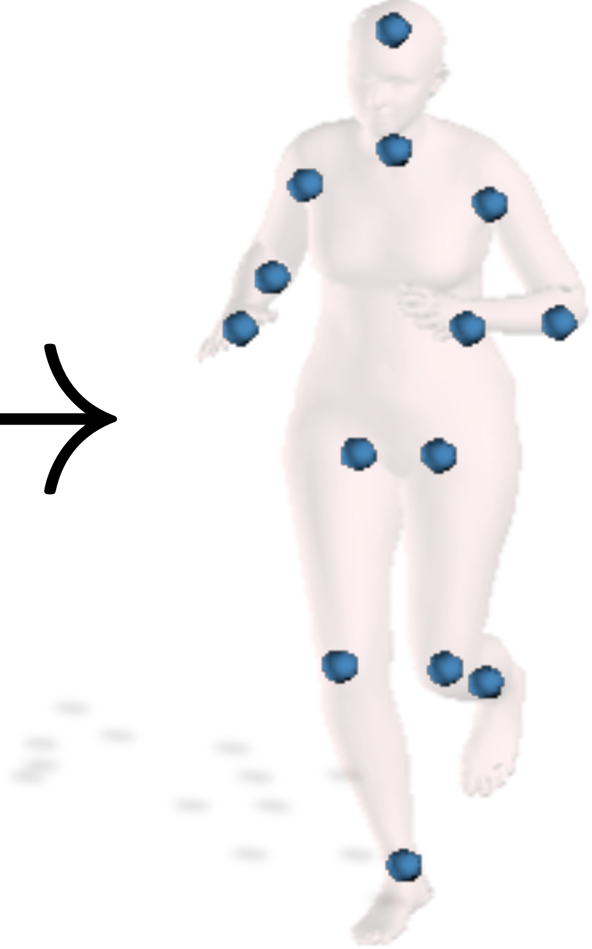
$$E_J(\vec{\beta}, \vec{\theta}, K; J_{est}) + E_a(\vec{\theta}) + E_\theta(\vec{\theta}) + E_{sp}(\vec{\theta}, \vec{\beta}) + E_\beta(\vec{\beta})$$

Data term Priors

Data Term: Joint Projection Error

$$(\vec{\beta}, \vec{\theta})$$

$$\longmapsto$$



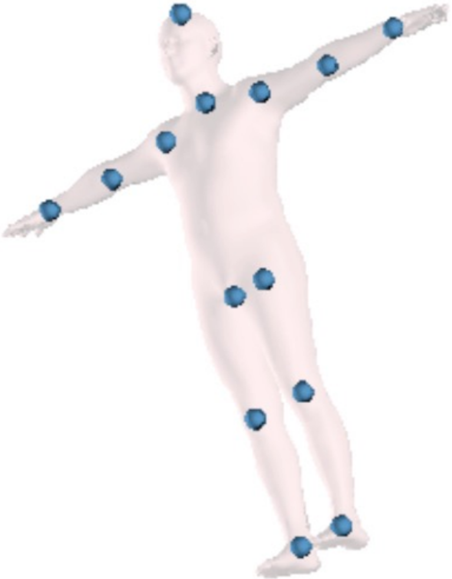
Data Term: Joint Projection Error

$$E_J(\vec{\beta}, \vec{\theta}, K; J_{est}) =$$

Camera Projection

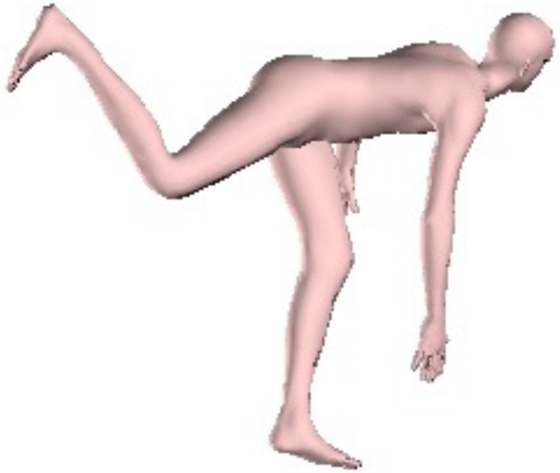
$$\Big\|$$

$$- \Pi_K \Big($$



$$\Big) \Big\|_2^2$$

SMPLify Objective Function



$$E(\vec{\beta}, \vec{\theta}, K; J_{est}) =$$

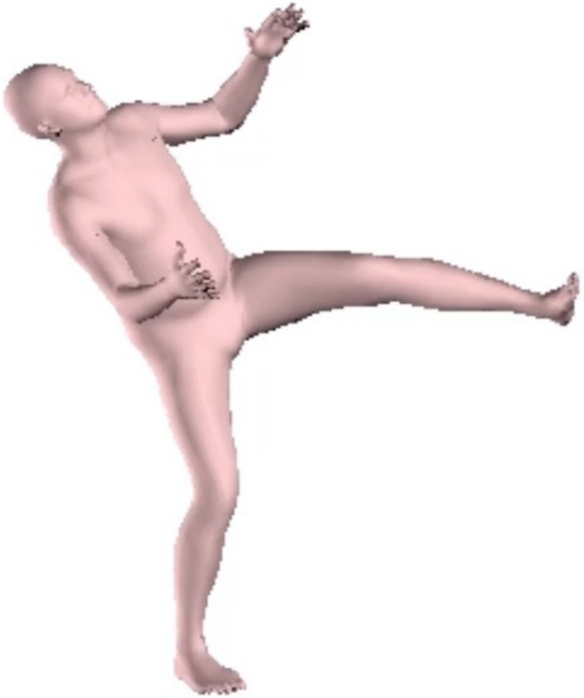
pose and shape priors

$$E_J(\vec{\beta}, \vec{\theta}, K; J_{est}) + E_a(\vec{\theta}) + E_\theta(\vec{\theta}) + E_{sp}(\vec{\theta}, \vec{\beta}) + E_\beta(\vec{\beta})$$

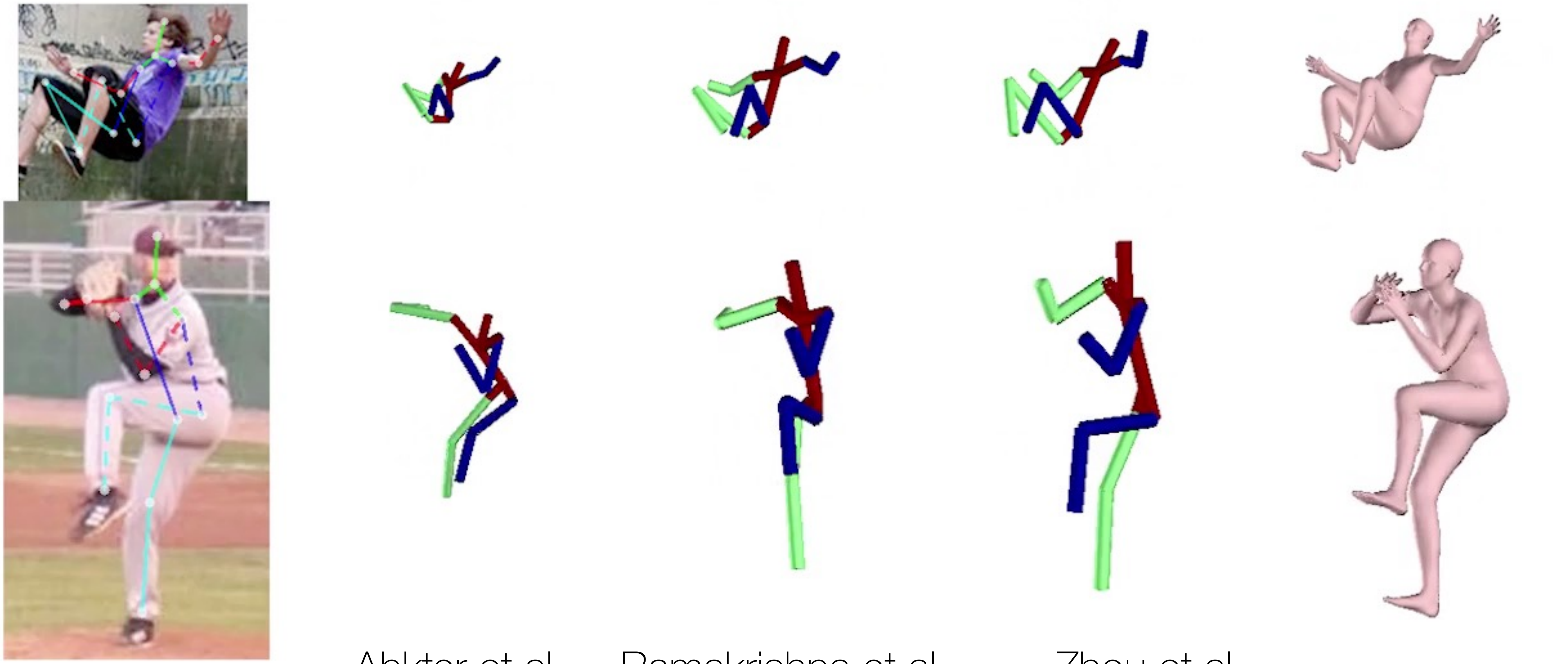
Joint projection error

interpenetration

Results on Leeds Sports Poses (LSP)



Results on Leeds Sports Poses (LSP)



Input

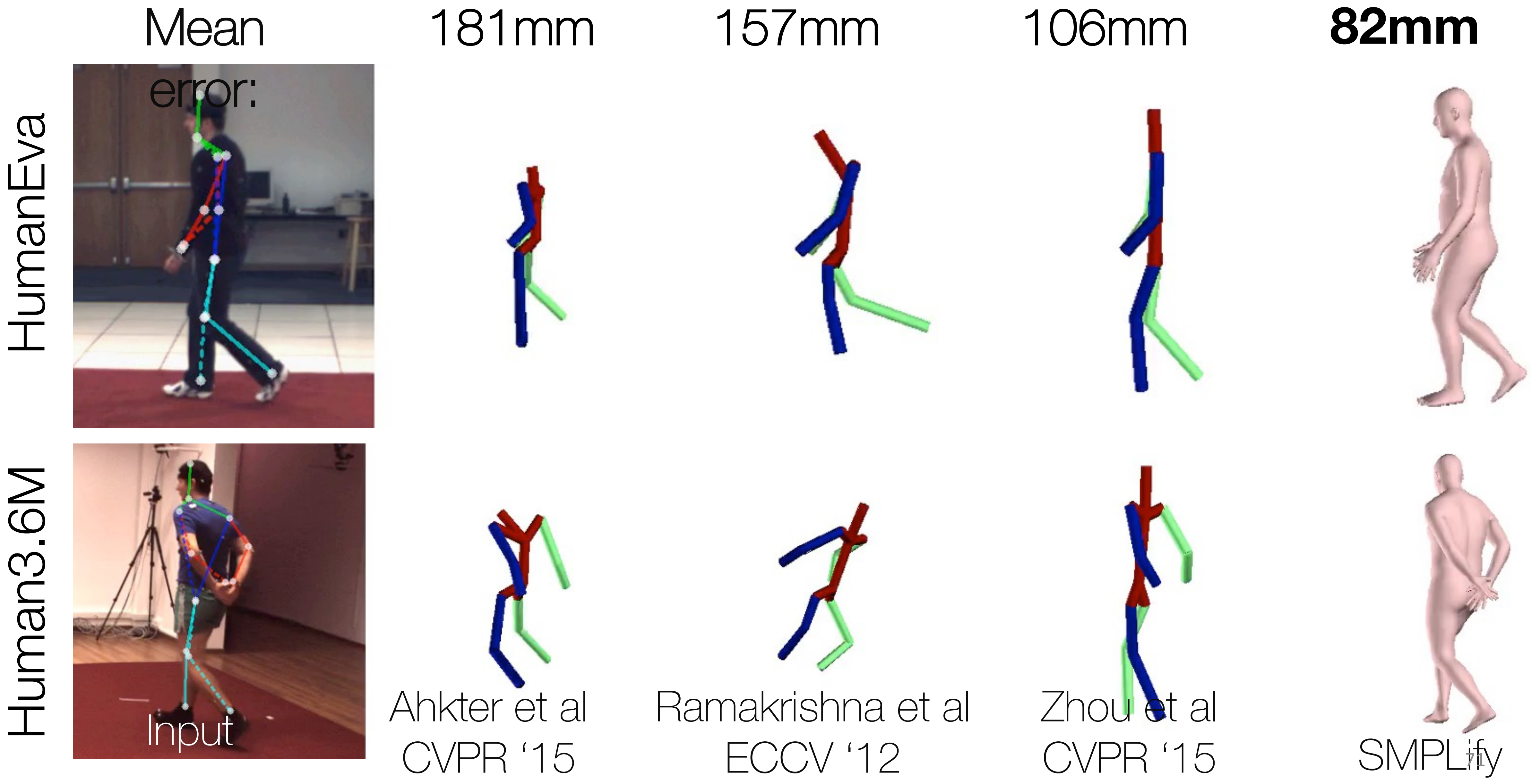
Ahkter et al CVPR '15

Ramakrishna et al ECCV '12

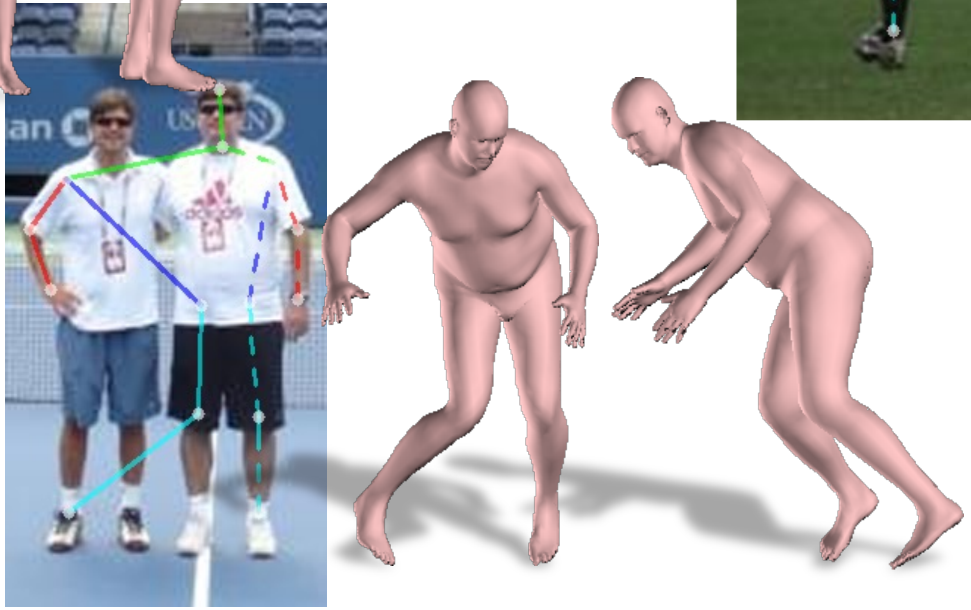
Zhou et al CVPR '15

SMPLify

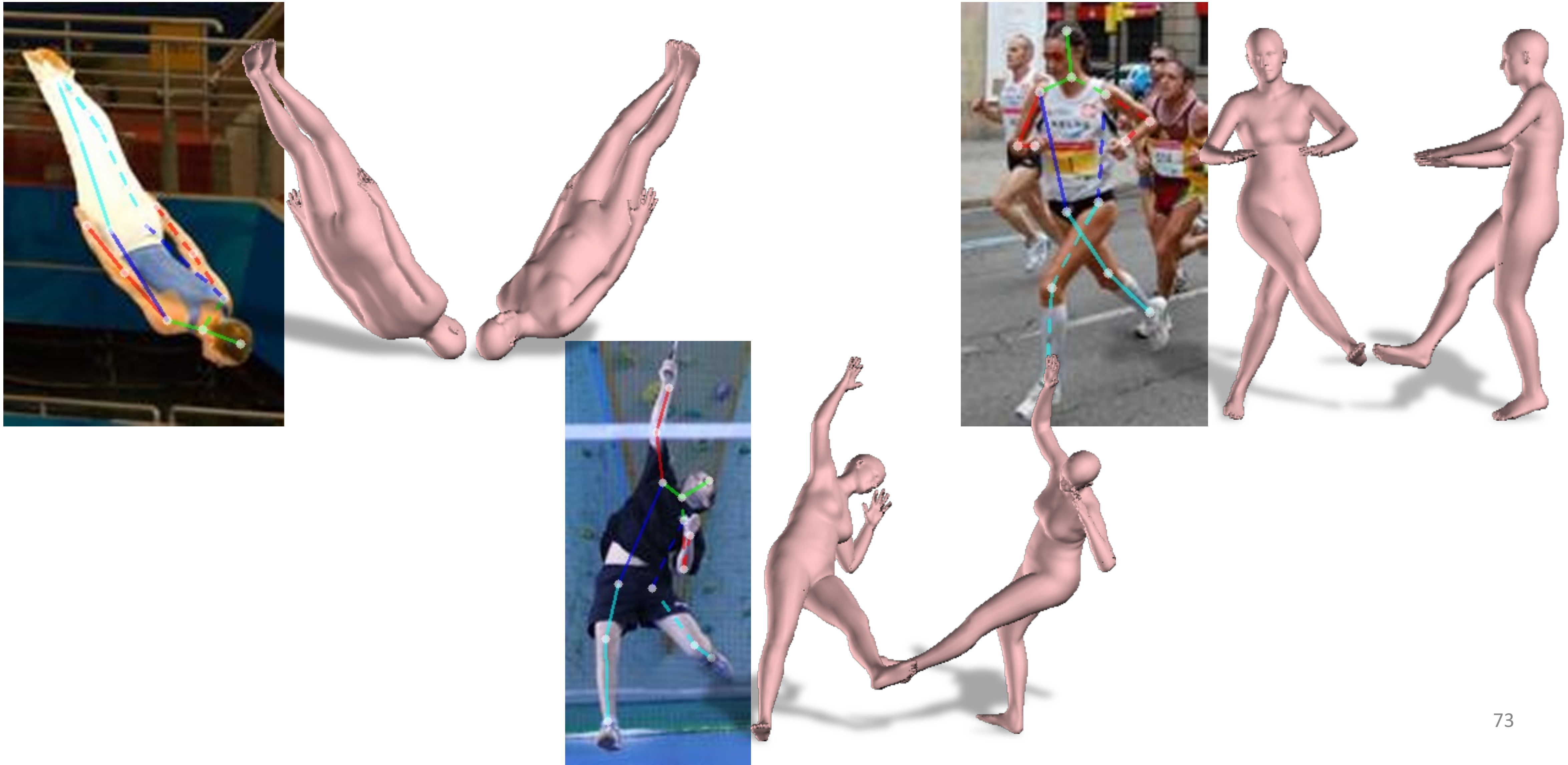
Quantitative Evaluation



Failure modes: caused by CNN failure



Failure modes: caused by Depth Ambiguity



Paper-4 to read today

End-to-end Recovery of Human Shape and Pose

Angjoo Kanazawa[1], Michael J. Black[2], David W. Jacobs[3], Jitendra Malik[1]
[1]University of California, Berkeley
[2]MPI for Intelligent Systems, Tübingen, Germany, [3]University of Maryland, College Park
{kanazawa,malik}@eecs.berkeley.edu, black@tuebingen.mpg.de, djacobs@umiacs.umd.edu

Project/code page

- https://akanazawa.github.io/hmr/
- https://github.com/akanazawa/hmr

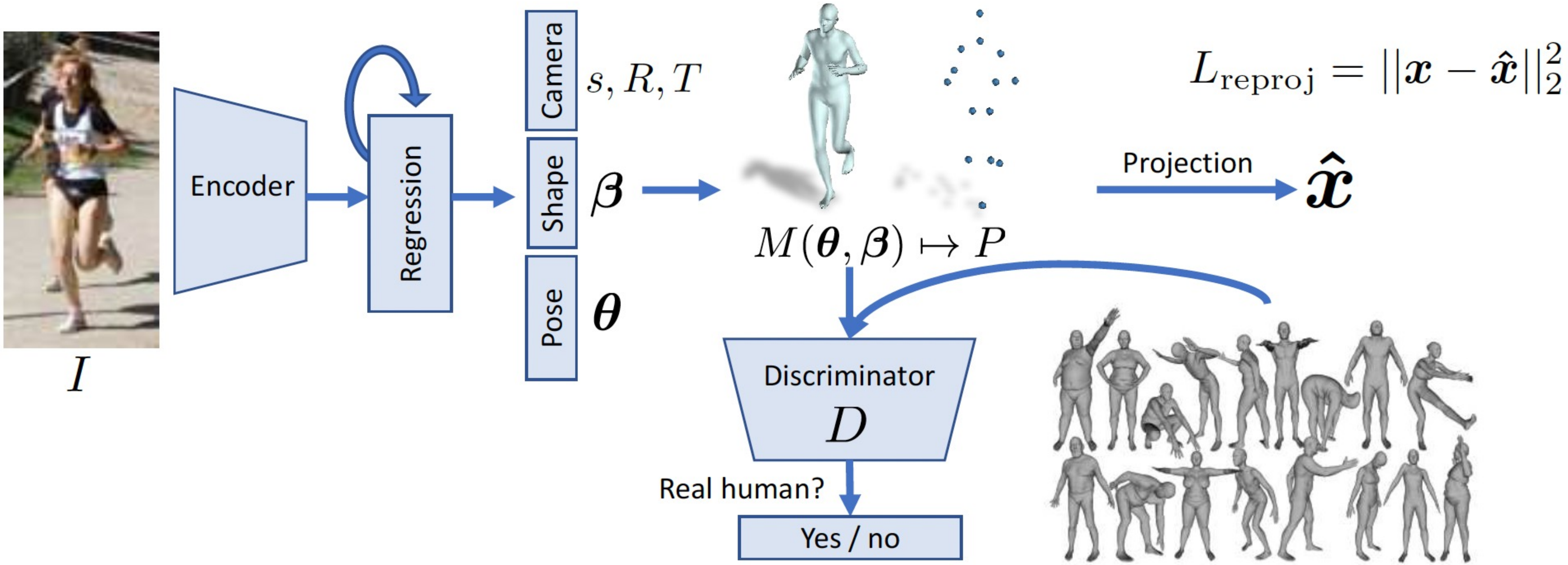


Figure 2: **Overview of the proposed framework.** An image I is passed through a convolutional encoder. This is sent to an iterative 3D regression module that infers the latent 3D representation of the human that minimizes the joint reprojection error. The 3D parameters are also sent to the discriminator D, whose goal is to tell if these parameters come from a real human shape and pose.

Loss function

$$L = \lambda(L_{\text{reproj}} + \mathbb{1}L_{\text{3D}}) + L_{\text{adv}} \qquad (1)$$

where λ controls the relative importance of each objective.

The goal of the 3D regression module is to output Θ given an image encoding ϕ such that the joint reprojection error

$$L_{\text{reproj}} = \Sigma_i ||v_i(\mathbf{x}_i - \hat{\mathbf{x}}_i)||_1, \qquad (3)$$

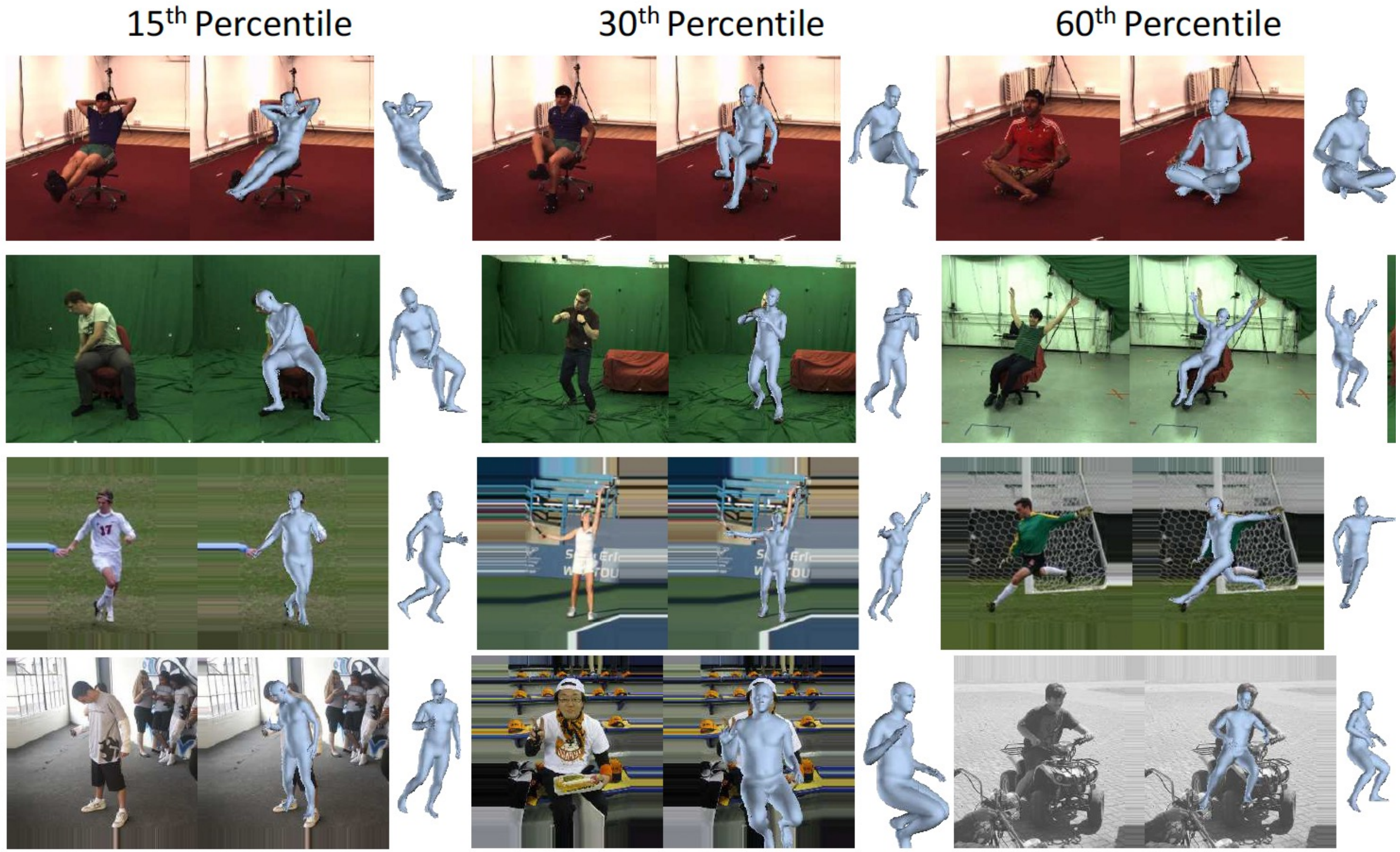
is minimized. Here $\mathbf{x}_i \in \mathbb{R}^{2\times K}$ is the ith ground truth 2D joints and $v_i \in \{0,1\}^K$ is the visibility (1 if visible, 0 otherwise) for each of the K joints.

$$L_{\text{3D}} = L_{\text{3D joints}} + L_{\text{3D smpl}} \qquad (4)$$

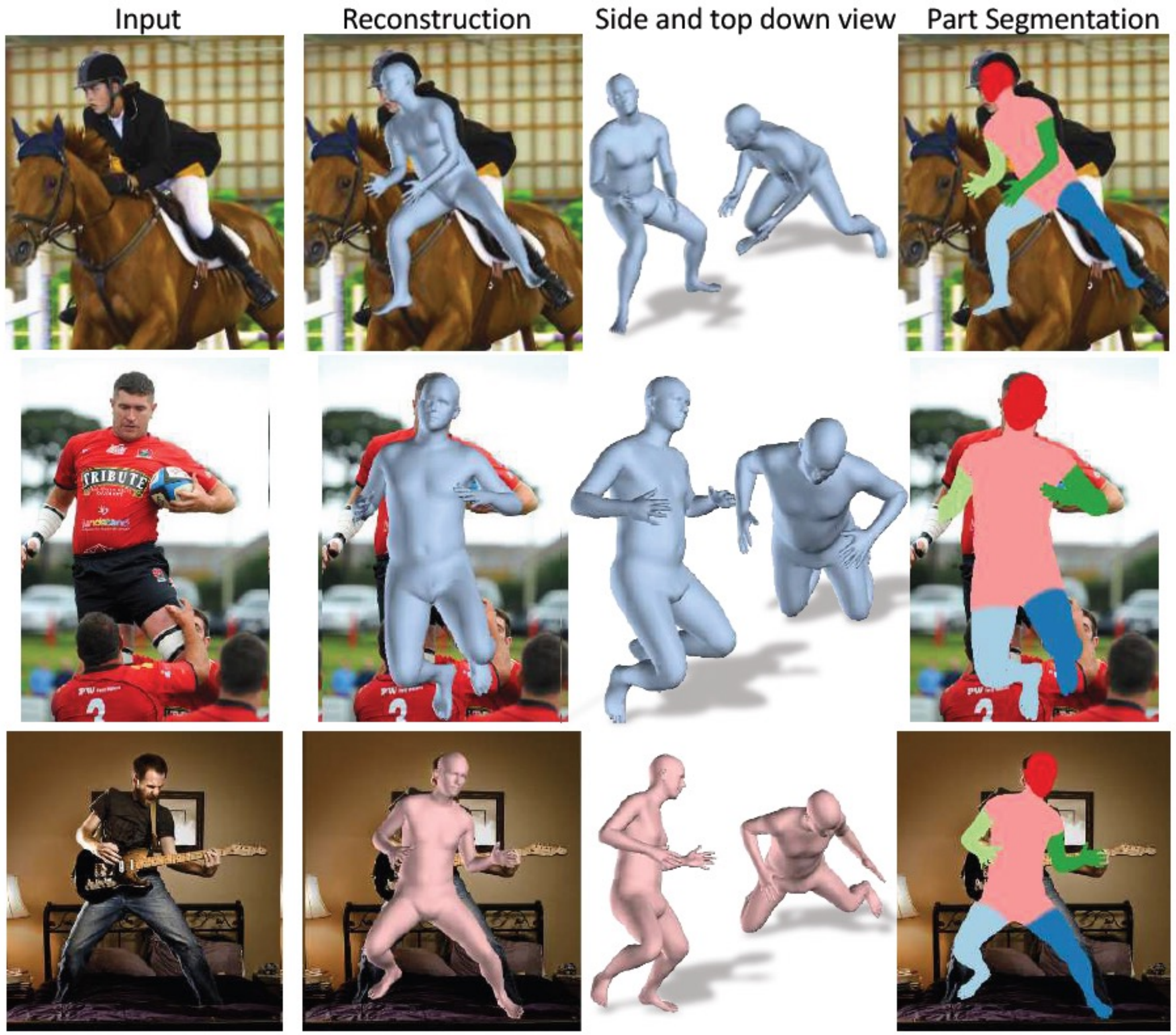
$$L_{\text{joints}} = ||(\mathbf{X_i} - \hat{\mathbf{X}}_\mathbf{i})||_2^2 \qquad (5)$$

$$L_{\text{smpl}} = ||[\boldsymbol{\beta}_i, \boldsymbol{\theta}_i] - [\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\theta}}_i]||_2^2. \qquad (6)$$

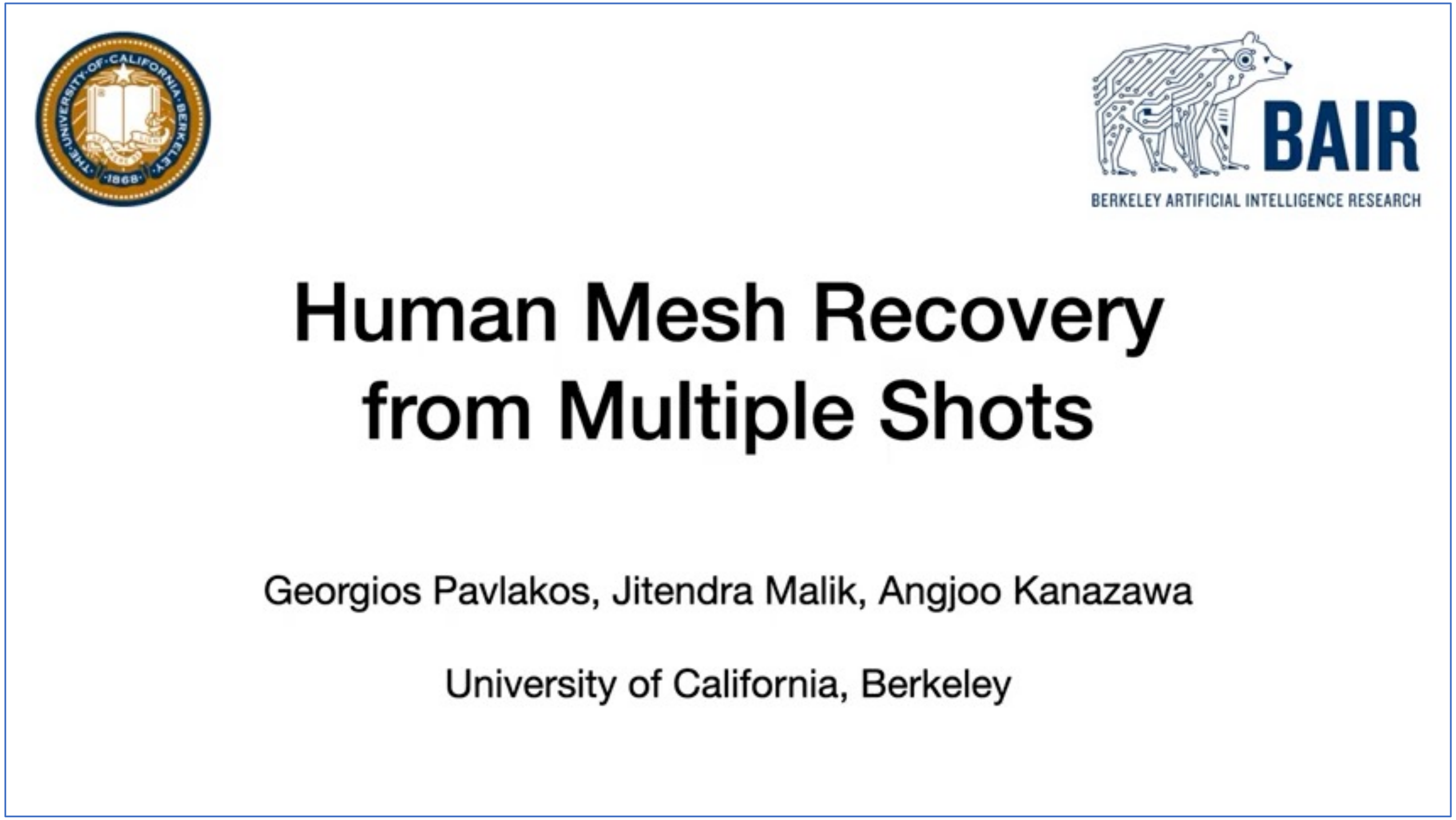
Sample Results



Sample Results



Video demo-4: human body mesh



Take a 5-minute break, then discussion.

Breakout room discussion

- Read and compare the following two papers; in particular, discuss the pros and cons between the “Optimisation” approach, and “deep Regression” approach.

- 1.

  Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image

- 2.

  End-to-end Recovery of Human Shape and Pose