BROWN

Body Shape from Video

Alexandru Balan Michael Black
Brown University

What’s constant?

- Video is interesting because it gives us information about the “structure” of the world.
- What’s constant is important (explanation).
- What changes may be more important (violation of assumptions).

What's constant?

Constant:

- camera pose and focal length
- identity
- height, weight
- limb lengths
- global scene illumination
- albedo

Changing

- pose (joint angles and spine)
- visibility
- soft tissues
- drape of clothing
- cast shadows and shading

What’s constant?

Approach: Model what’s constant using a model of 3D body shape.

Problem: Pose changes shape. Need a pose-invariant shape model.

Why a graphics model?

Goals

- Provide strong constraints for interpreting video.
- Combine information
  - Across views
  - Across non-rigid pose changes
- Support inference of gender, height, age, etc.
- Explain changes due to illumination.

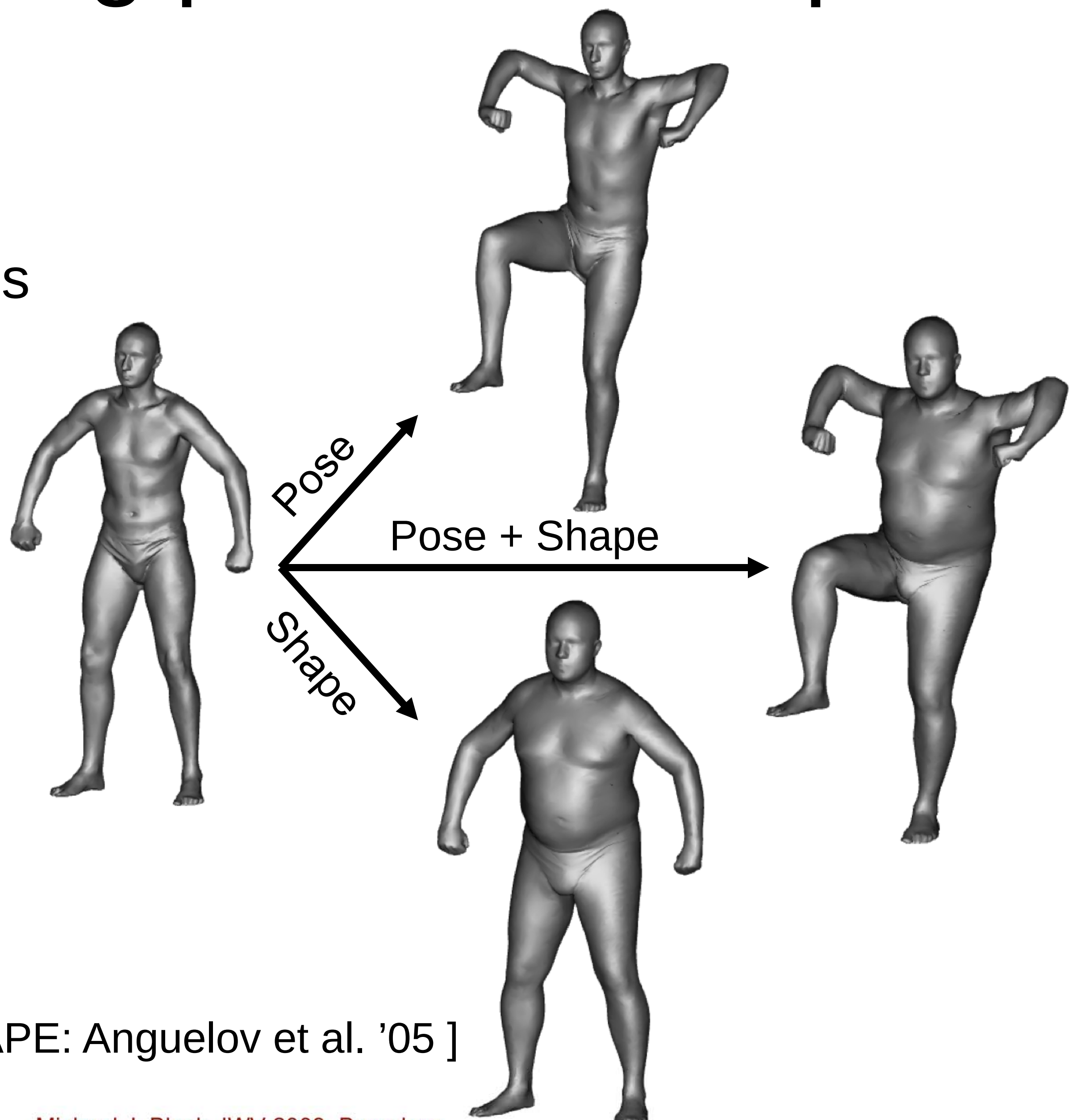
Problem: must factor changes due to shape and pose

- Learn statistics of shape variation across people and poses
- Model what you know, learn the statistics of the rest.

Factoring pose and shape

Low dimensional parameterization learned from examples

We use an "intrinsic" shape representation *invariant* to pose



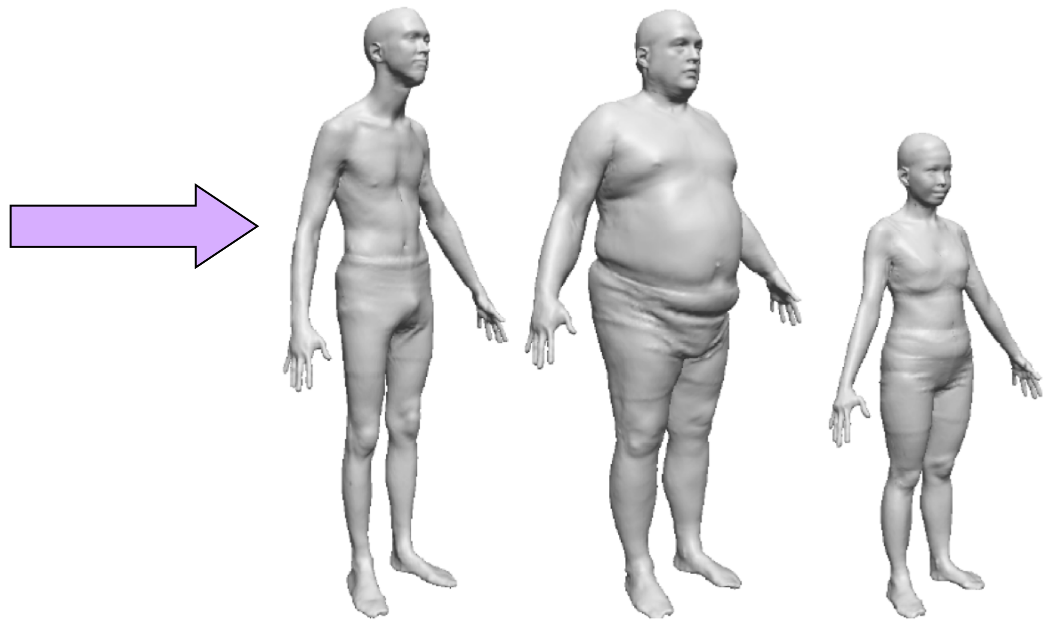
[SCAPE: Anguelov et al. '05]

SCAPE

Shape Completion and Animation of PEople

[Anguelov et al. Siggraph ’05]

[Cyberware]

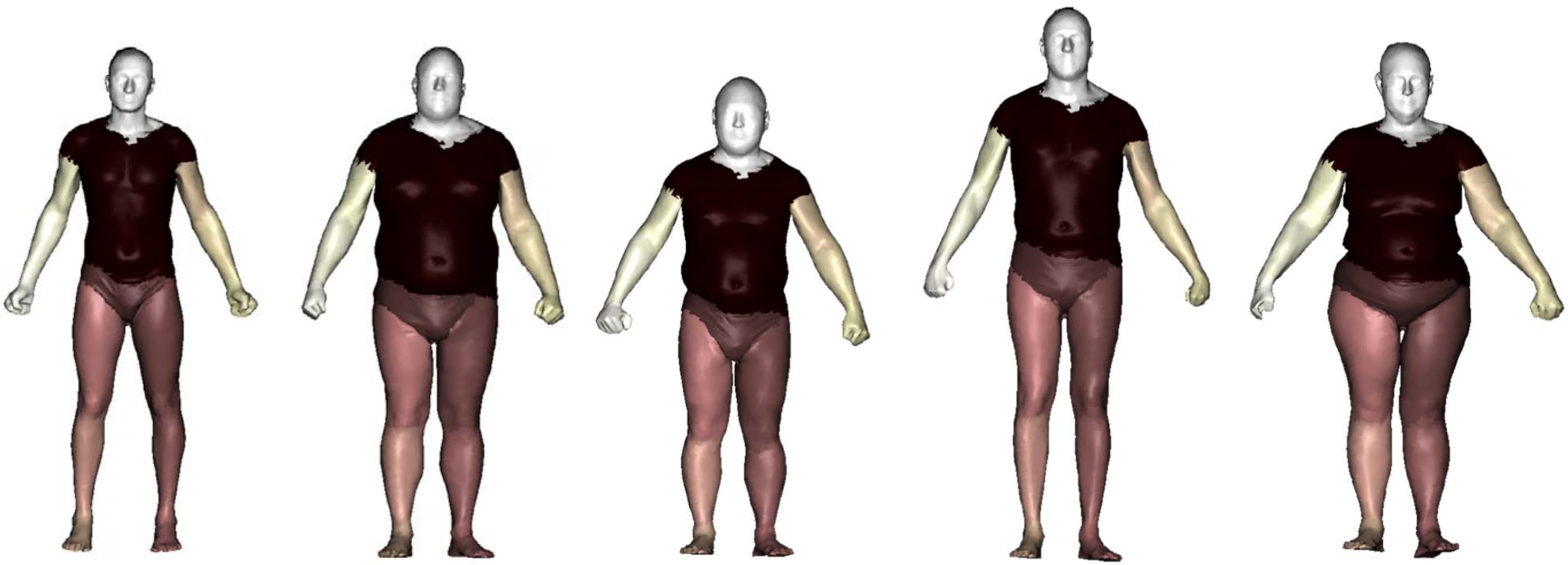


[Allen et al. ’03]

CAESAR dataset (SAE Int.): 3D mesh models of over 2000 North American adults

Shape space

- Align a “template mesh” to each scan using an iterative closest point method.
- 25,000 triangles and 12,500 vertices.



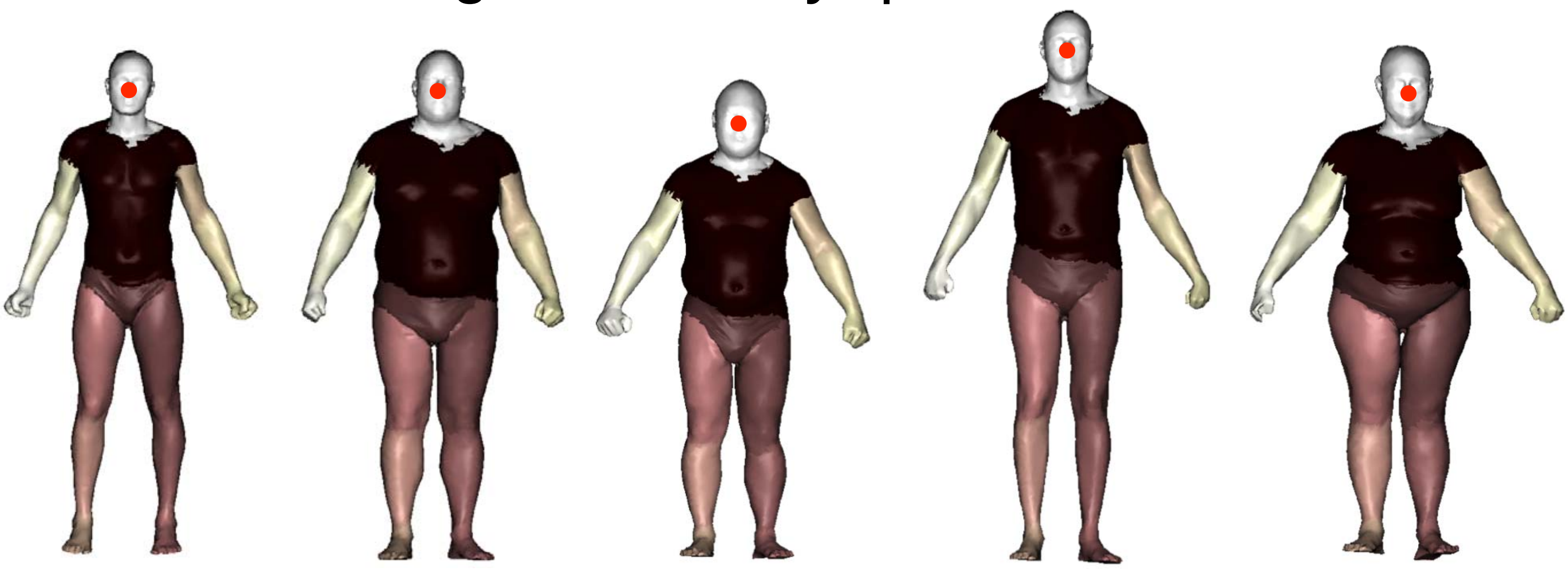
Example scans of different people.

[Anguelov et al. ’05]

BROWN

Shape space

- All vertices are in correspondence.
- All bodies in the canonical pose.
- Vertices assigned to body "parts".

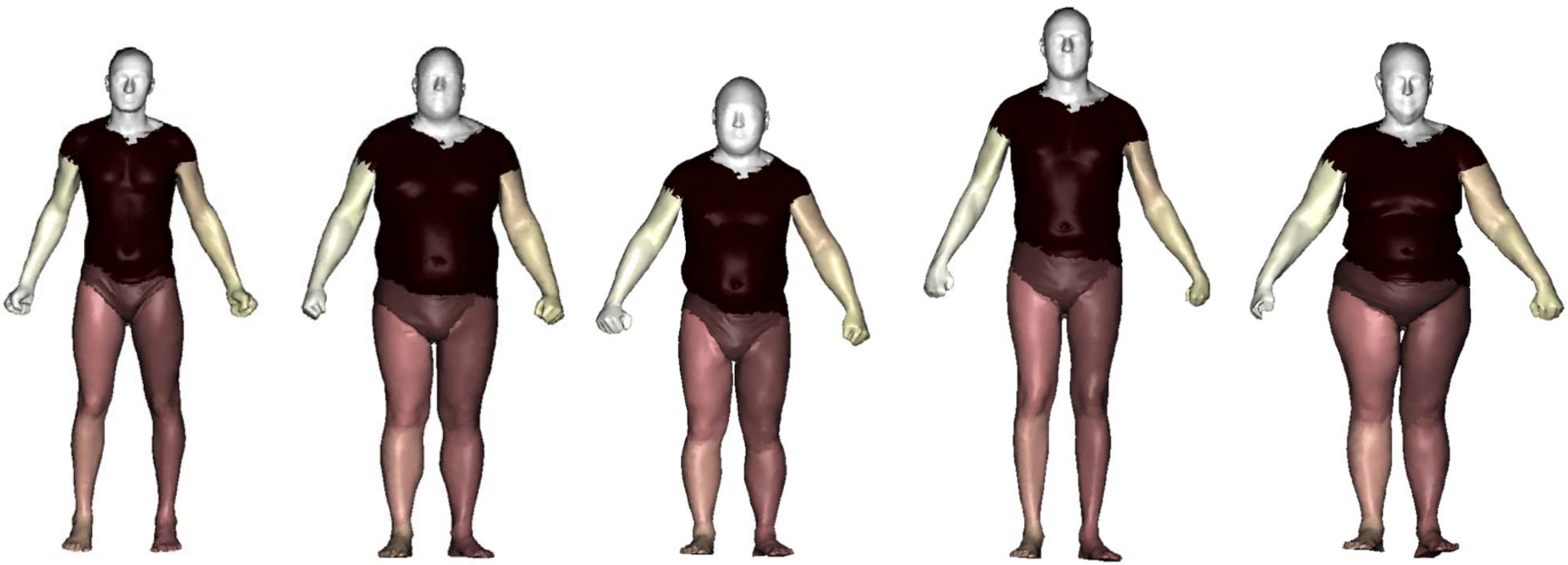


Example scans of different people.

[Anguelov et al. '05]

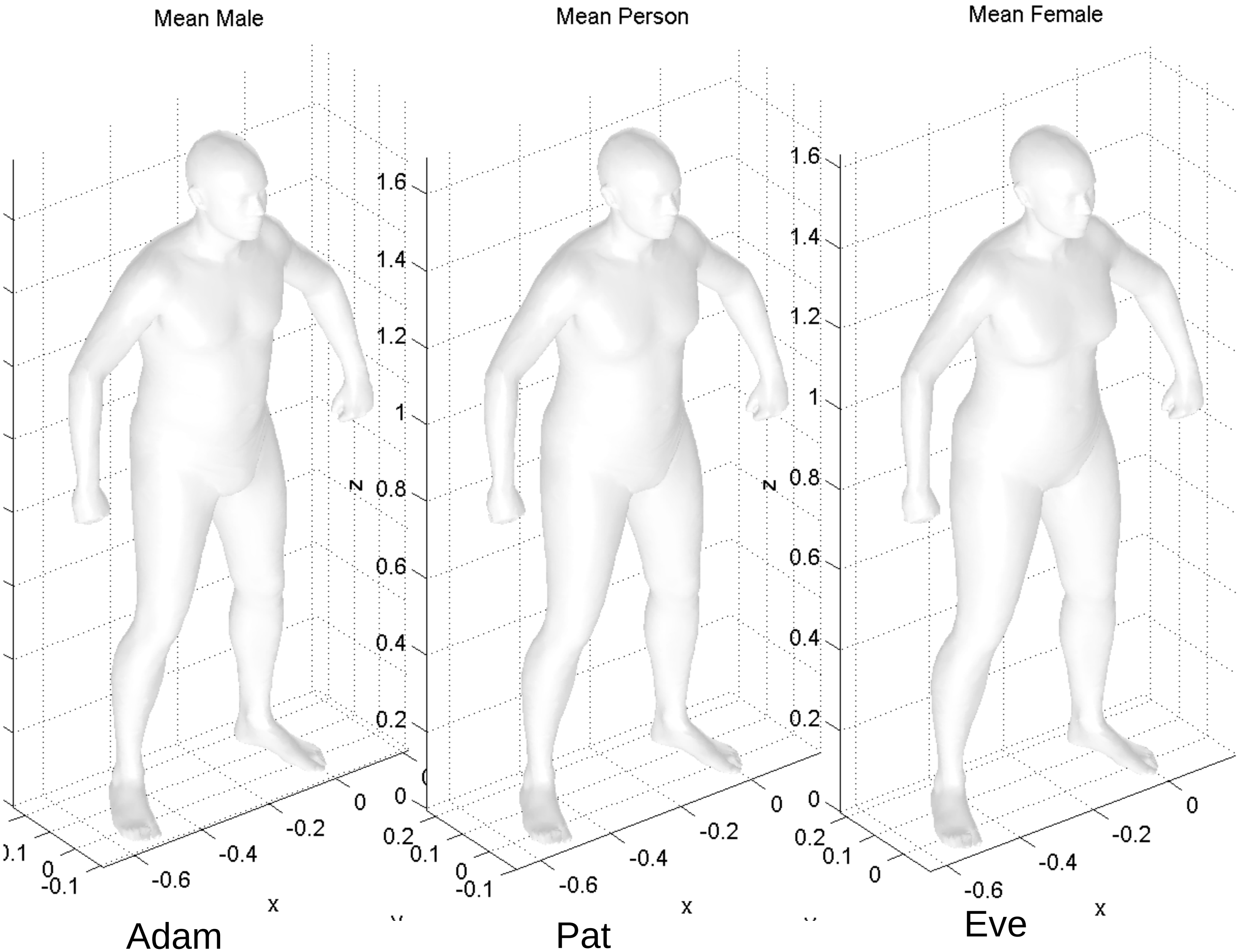
Shape space

- Learn low dimensional shape deformation model using incremental PCA (Brand, ECCV'02).
  - applied to deformations (i.e. rotations) of the triangles.

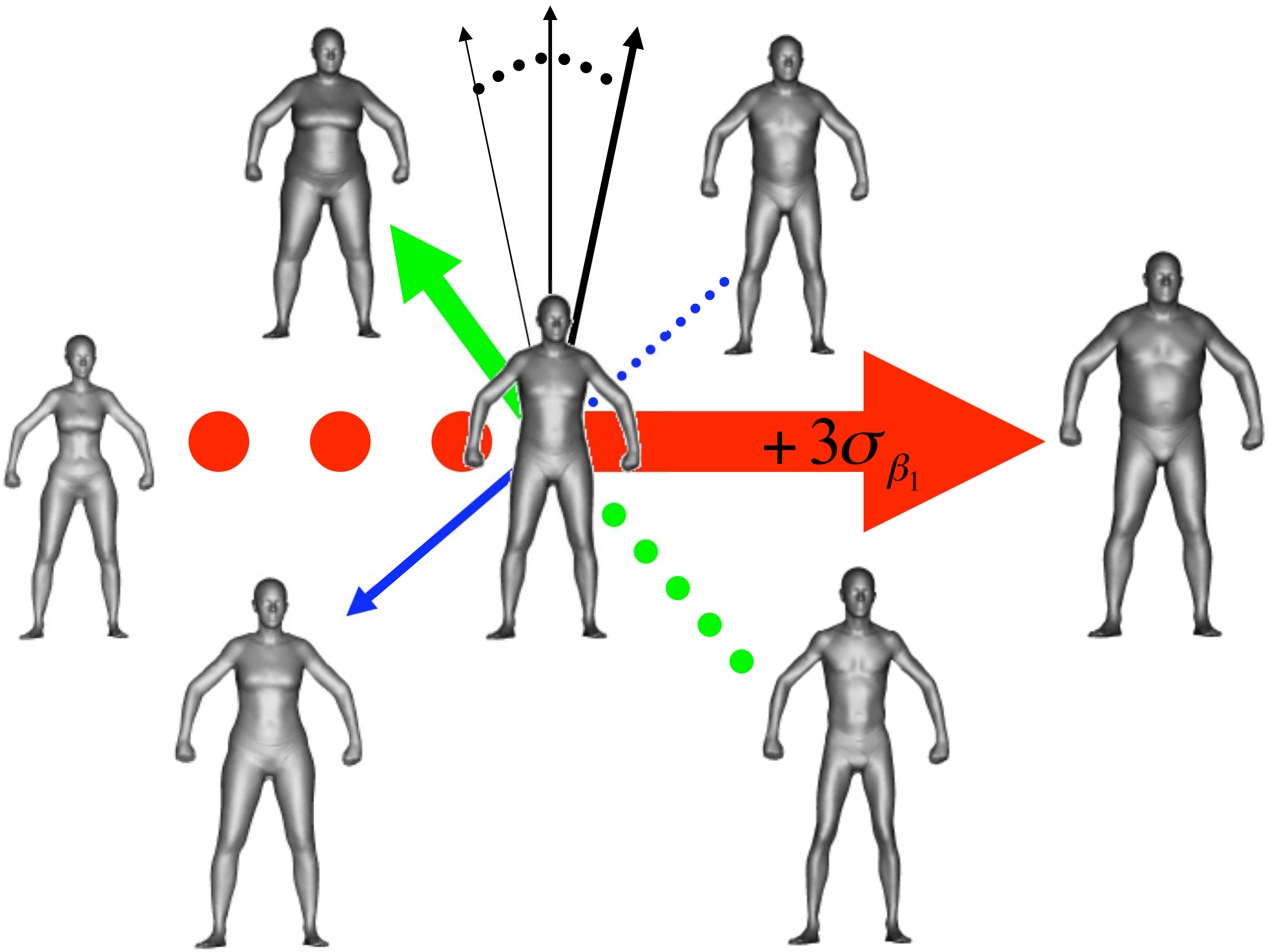


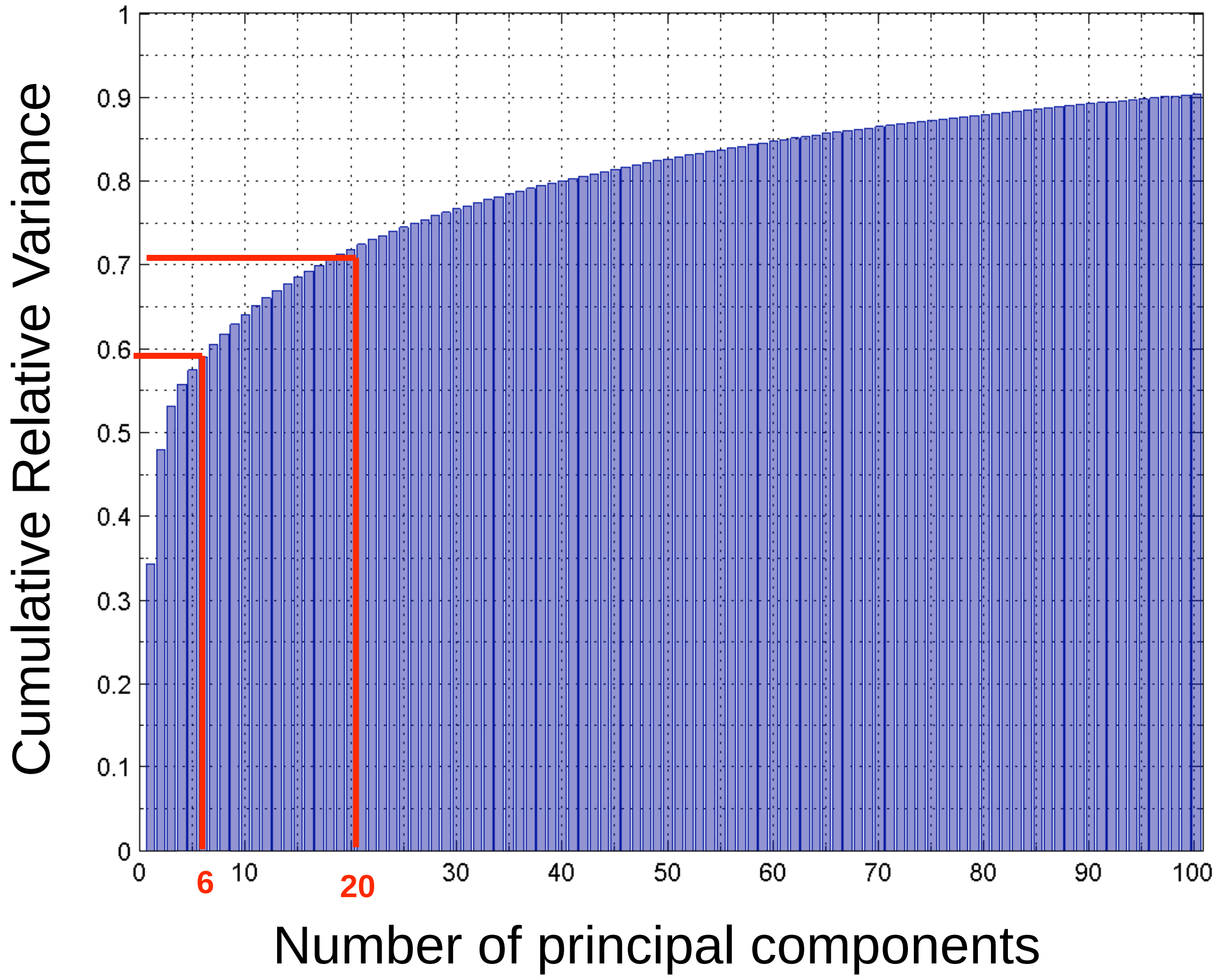
Example scans of different people.

[Anguelov et al. '05]


Mean Male
Mean Person
Mean Female
Adam
Pat
Eve

Eigen-People

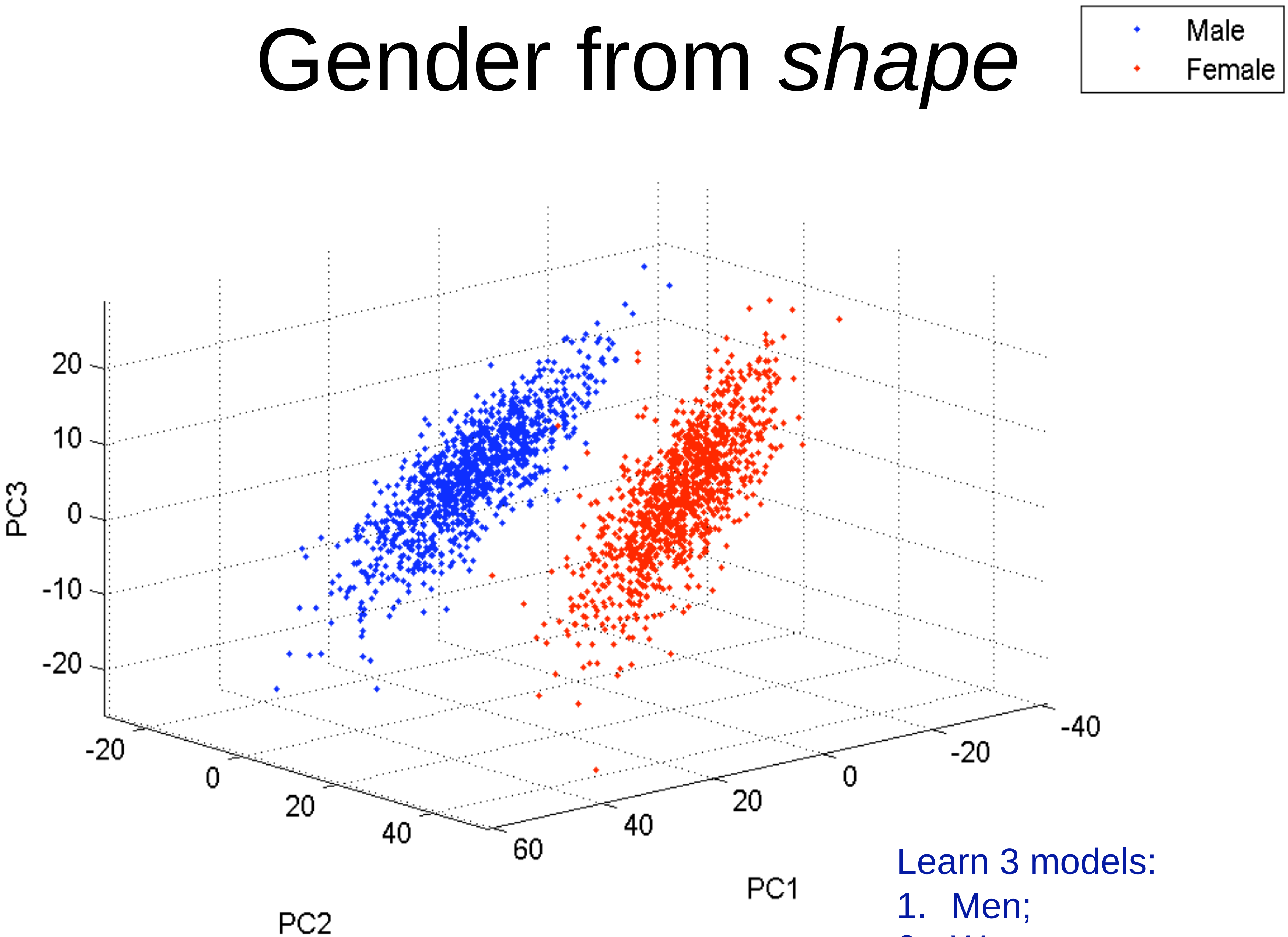




2124 scans of various people

[CAESAR dataset]

Gender from *shape*



Learn 3 models:

1. Men;
2. Women;
3. Both together.

What about pose changes?

For each part, apply a rigid rotation to each triangle.

R_{part}

(3x3)

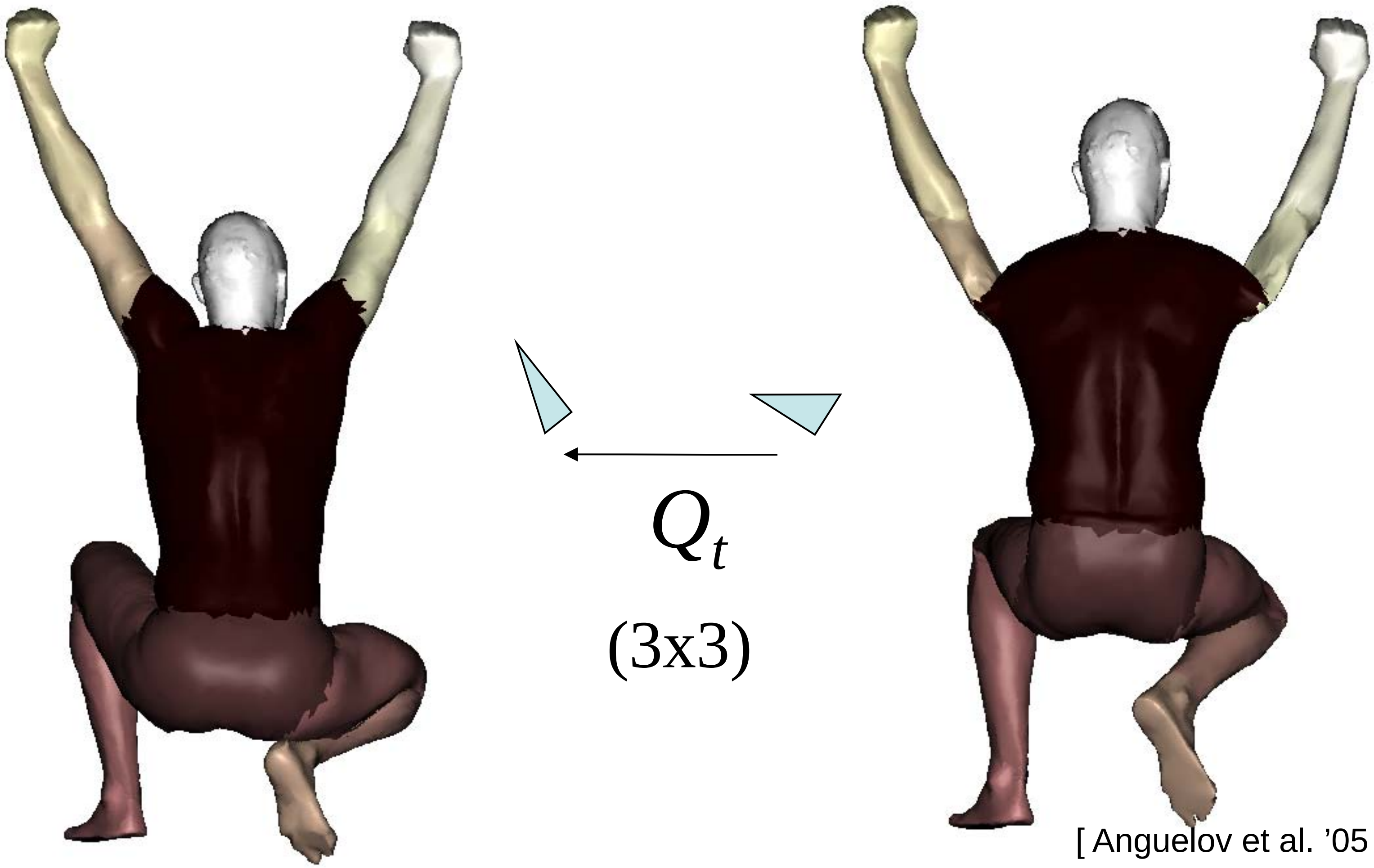
x 15 parts

[Anguelov et al. ’05]

Preserve desired orientation and scale of edges in a least-squares sense

What about pose changes?

Triangles are all in correspondence.



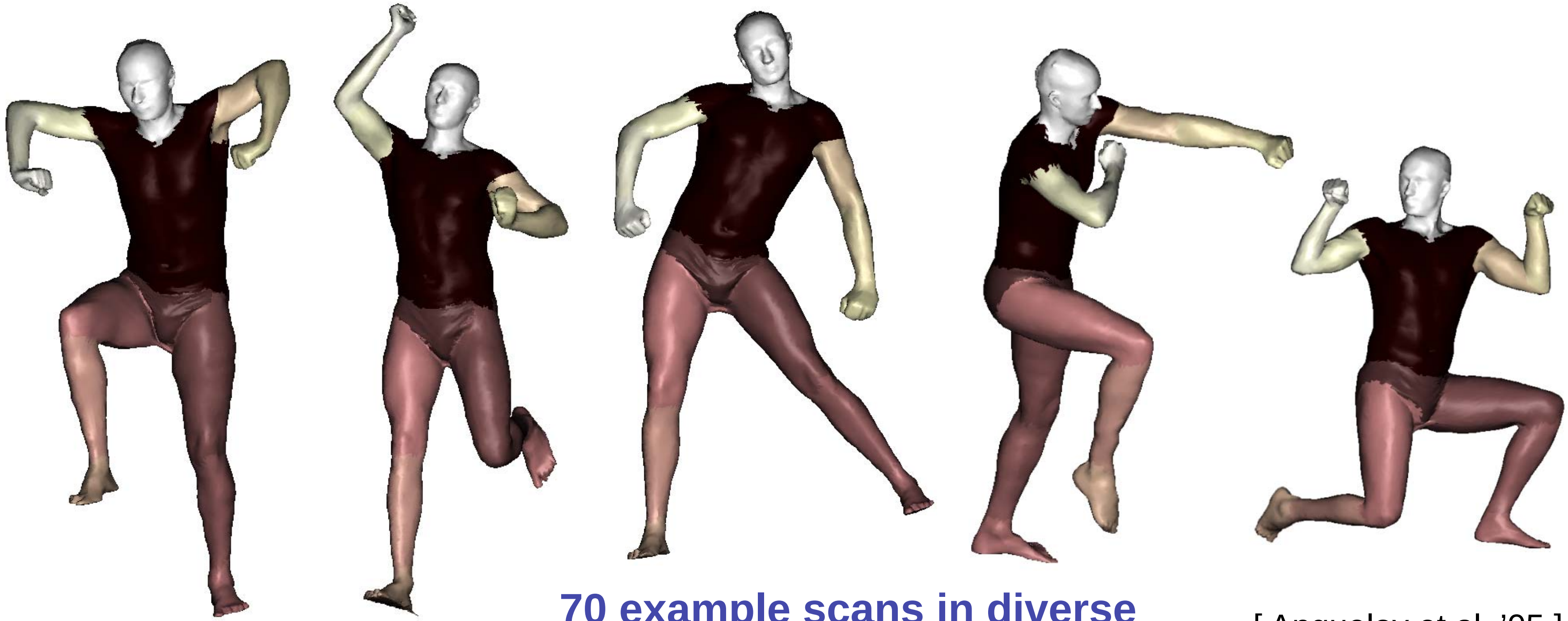
[Anguelov et al. ’05]

Deformation of each triangle from predicted model to training example.

Pose deformation space

Model articulated and non-rigid deformations

– Non-rigid deformations learned as a function of relative part orientations.



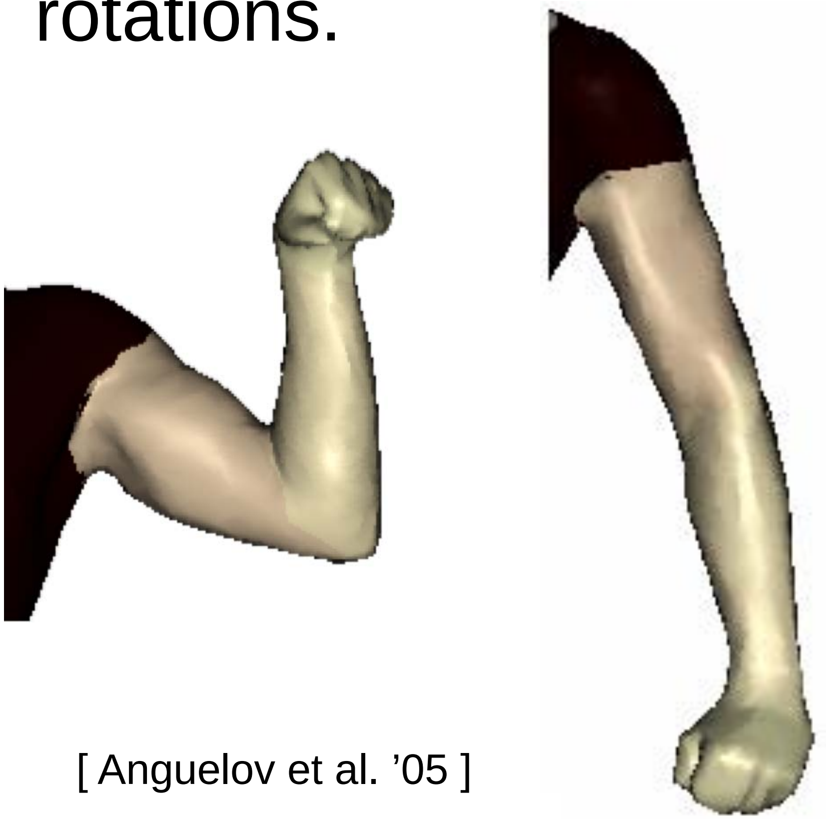
70 example scans in diverse poses

[Anguelov et al. ’05]

Pose deformation space

Model non-rigid deformations.

Linear prediction from relative part rotations.



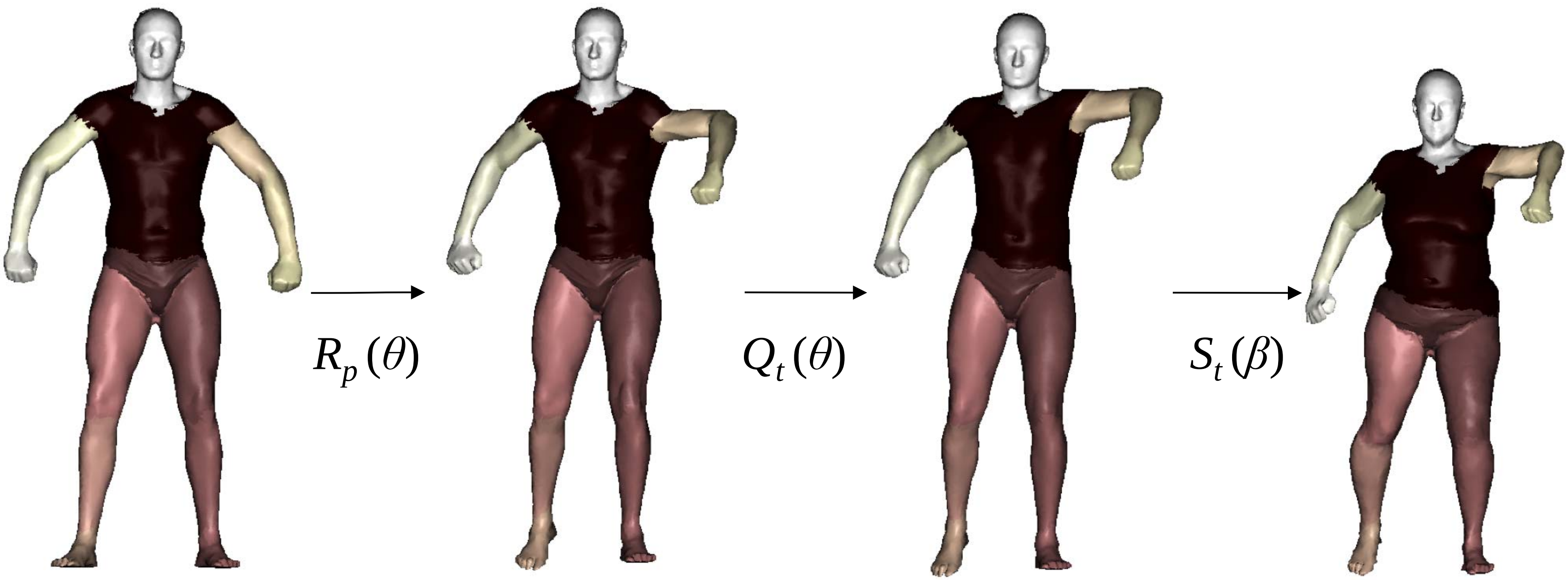
[Anguelov et al. ’05]

$$Q_t = A_t \cdot \begin{bmatrix} \theta_{p_t^1} \\ \theta_{p_t^2} \\ 1 \end{bmatrix}$$

Learn a matrix A_t for each triangle t in the mesh via linear regression.

BROWN

SCAPE deformations



Articulated Rigid Deformation

Non-rigid Deformation

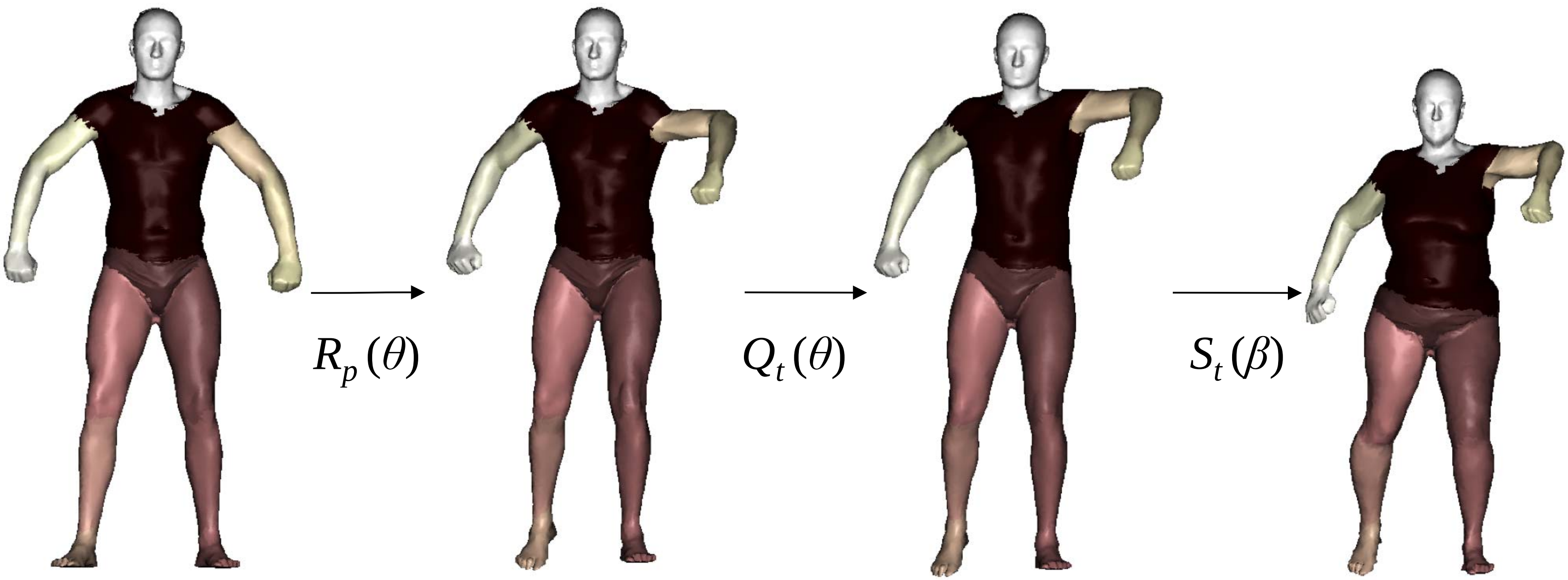
Body Shape Deformation

θ – part rotations

β – shape parameters

[Anguelov et al. ’05]

SCAPE deformations

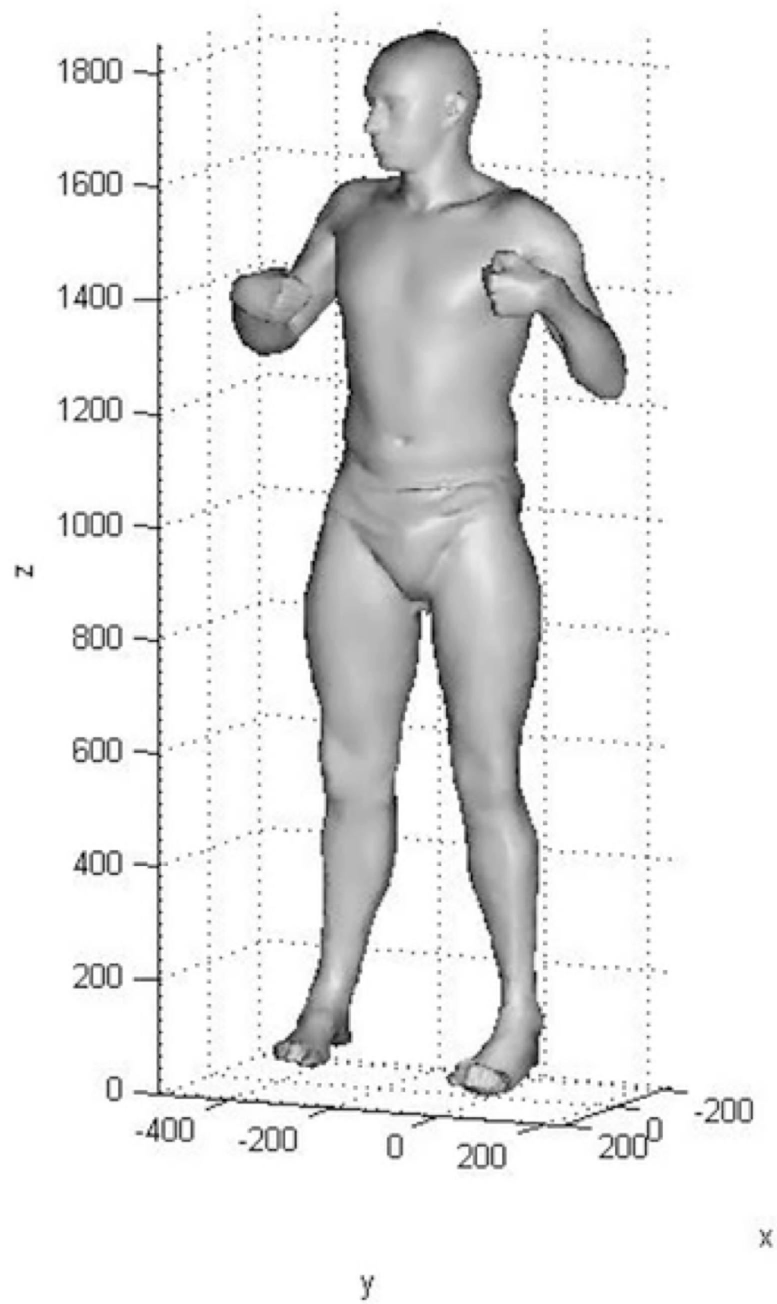


θ – part rotations (37D) τ – global position (3D) β – shape parameters (6-20D)

Parameters (state): $s = (\theta, \tau, \beta)$

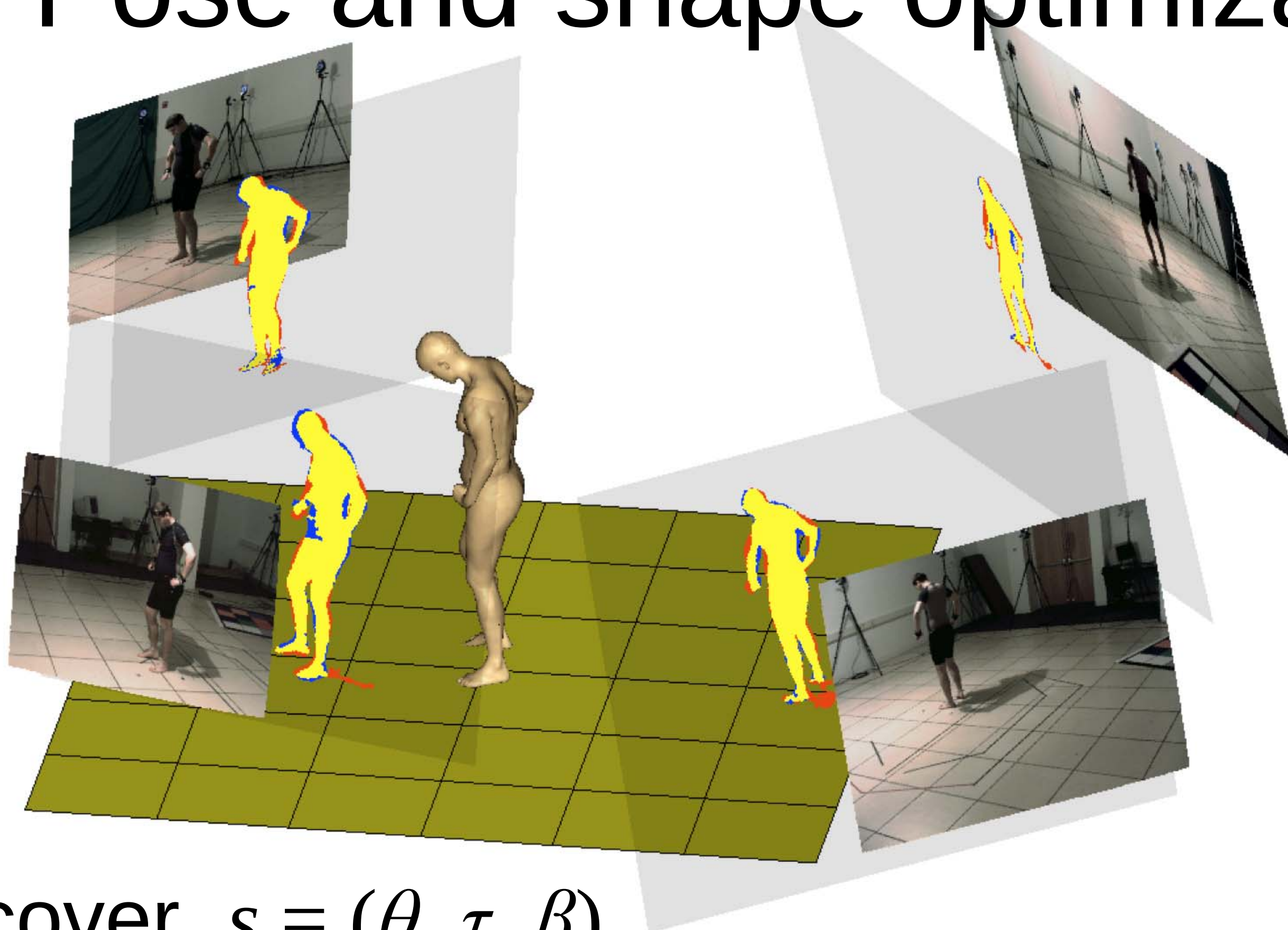
Shape parameters can be gender specific.

[Anguelov et al. '05]



1800
1600
1400
1200
1000
800
600
400
200
0
z
-400
-200
0
200
y
-200
0
200
x

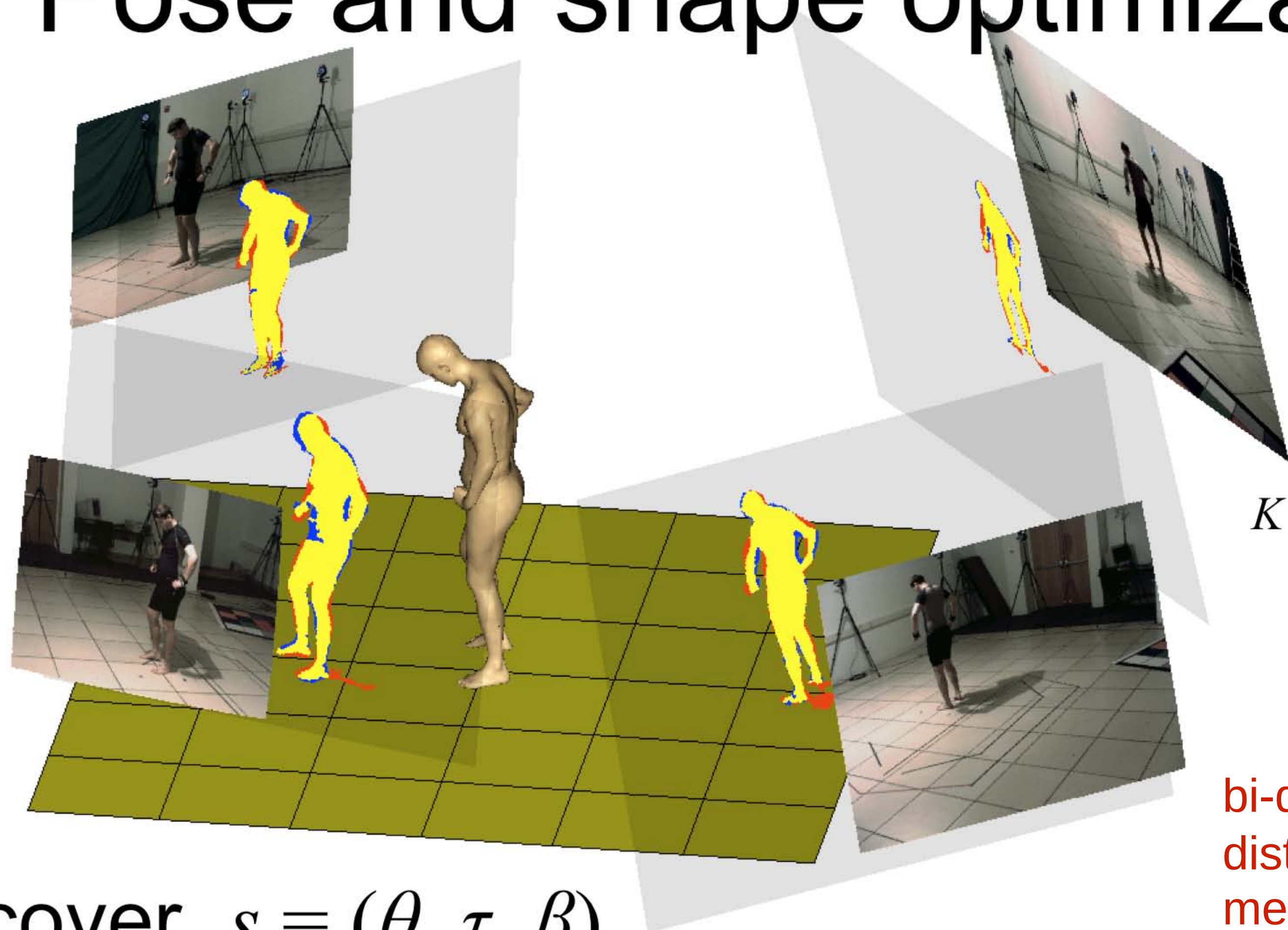
Pose and shape optimization



Recover $s = (\theta, \tau, \beta)$

(Initialization: Sigal *et al.* NIPS '07)

Pose and shape optimization



K cameras

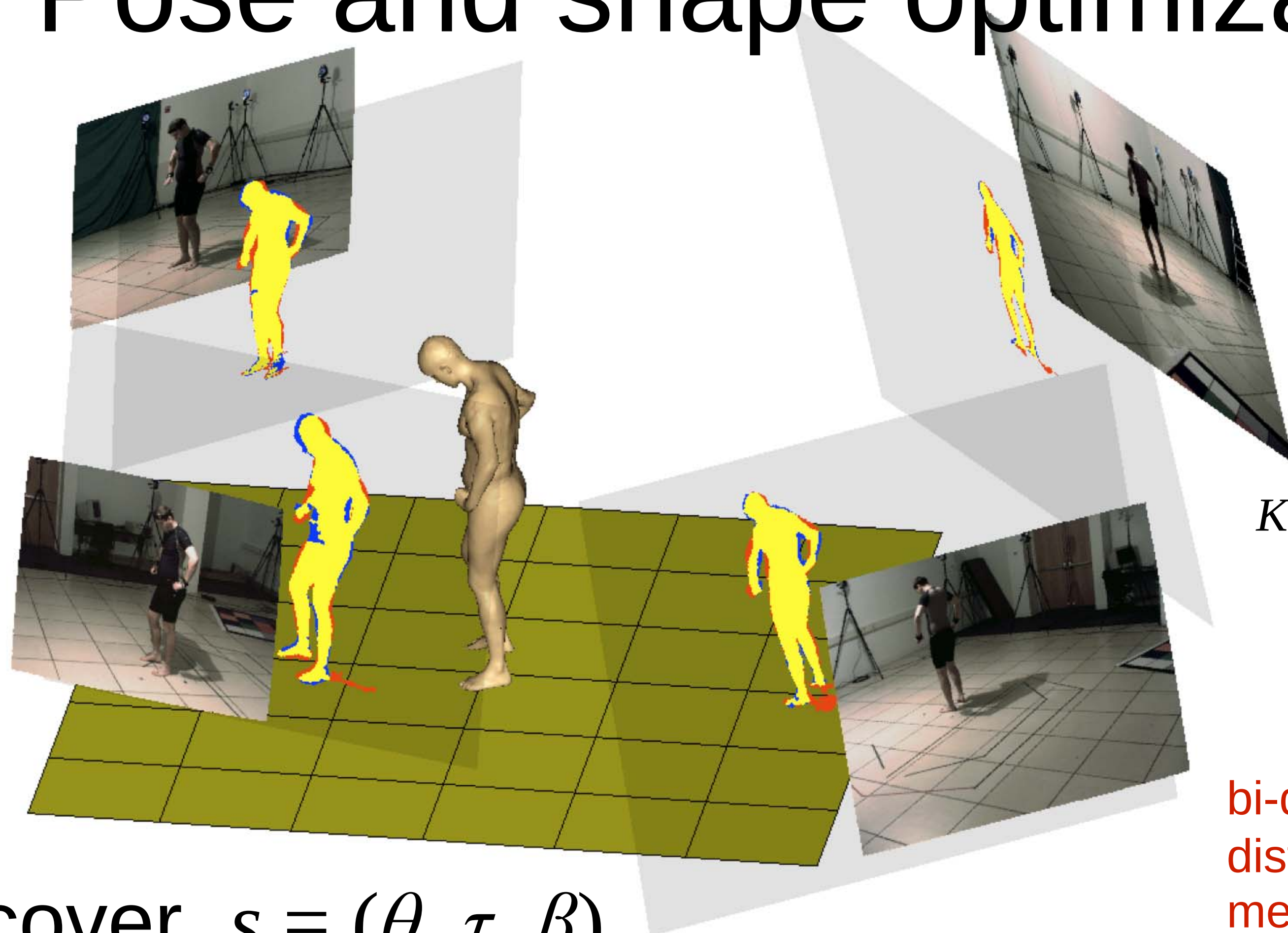
Recover $s = (\theta, \tau, \beta)$

Minimize $$E(s) = \sum_{i=1}^{K} D\left(\mathrm{F}_i^e(s), \mathrm{F}_i^o\right)$$

bi-directional distance measure between silhouettes

Penalize interpenetration.

Pose and shape optimization



K cameras

bi-directional distance measure between silhouettes

Recover $s = (\theta, \tau, \beta)$

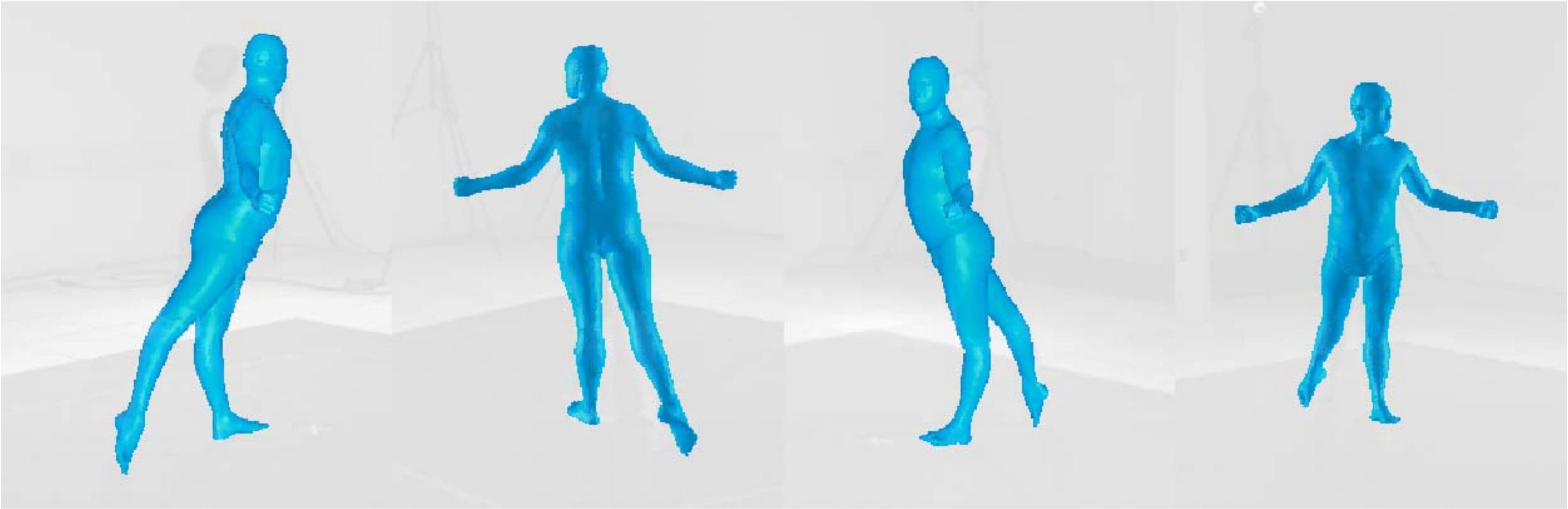
Minimize $E(s) = \sum_{i=1}^{K} \alpha\, d(\mathrm{F}_i^e(s), \mathrm{F}_i^o) + (1-\alpha)\, d(\mathrm{F}_i^o, \mathrm{F}_i^e(s))$

Penalize interpenetration.

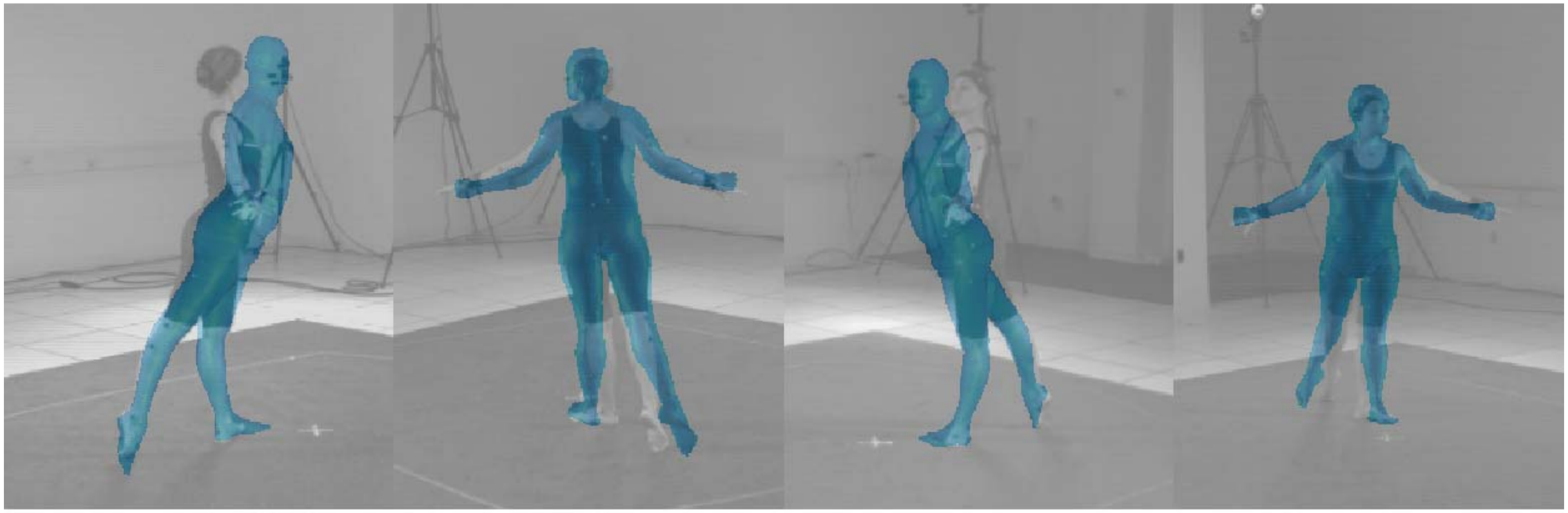
Optimization – Step by Step

Input images and foreground silhouettes

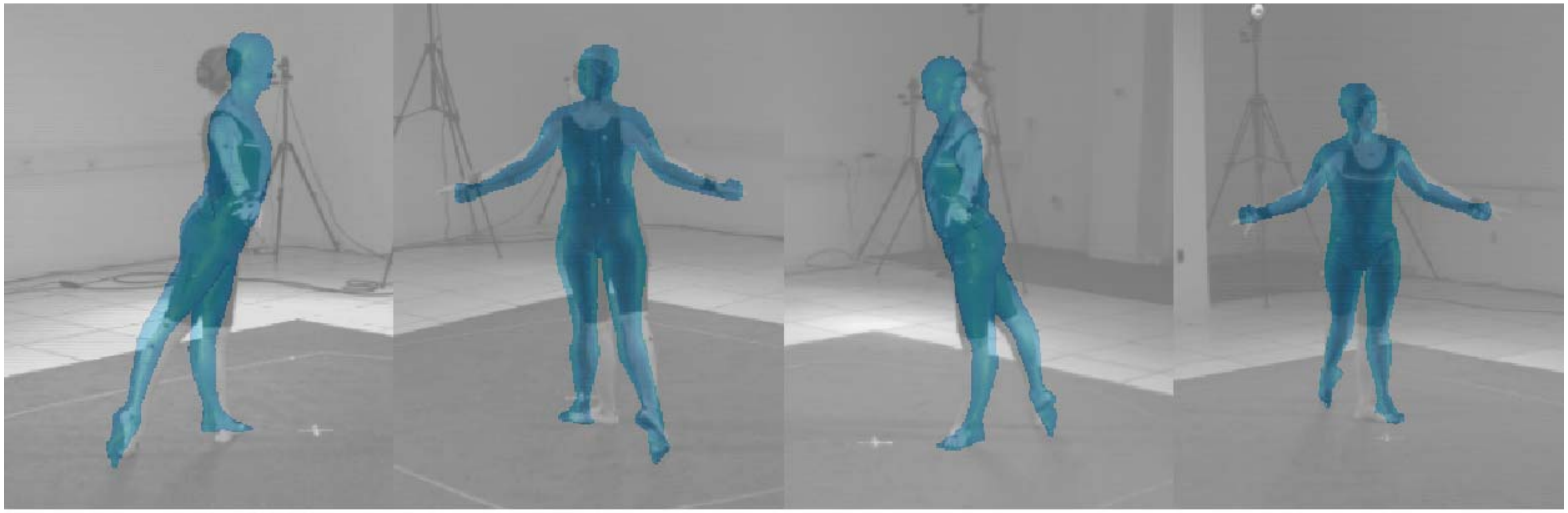
Optimization – Step by Step



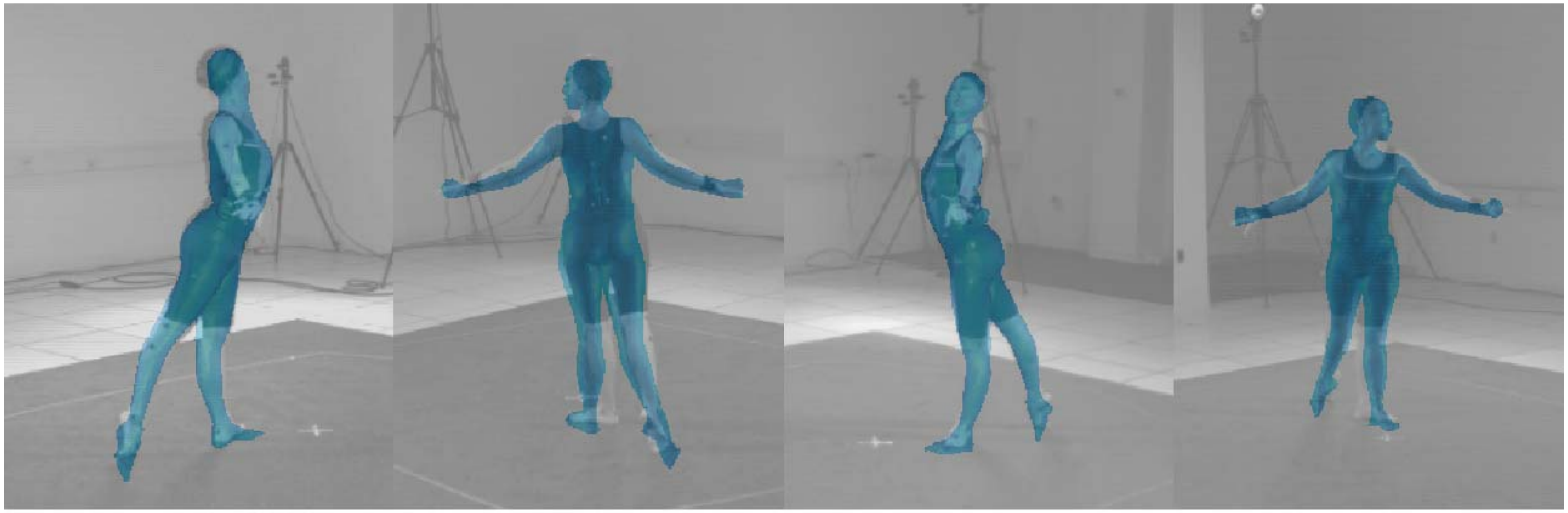
Optimization – Step by Step



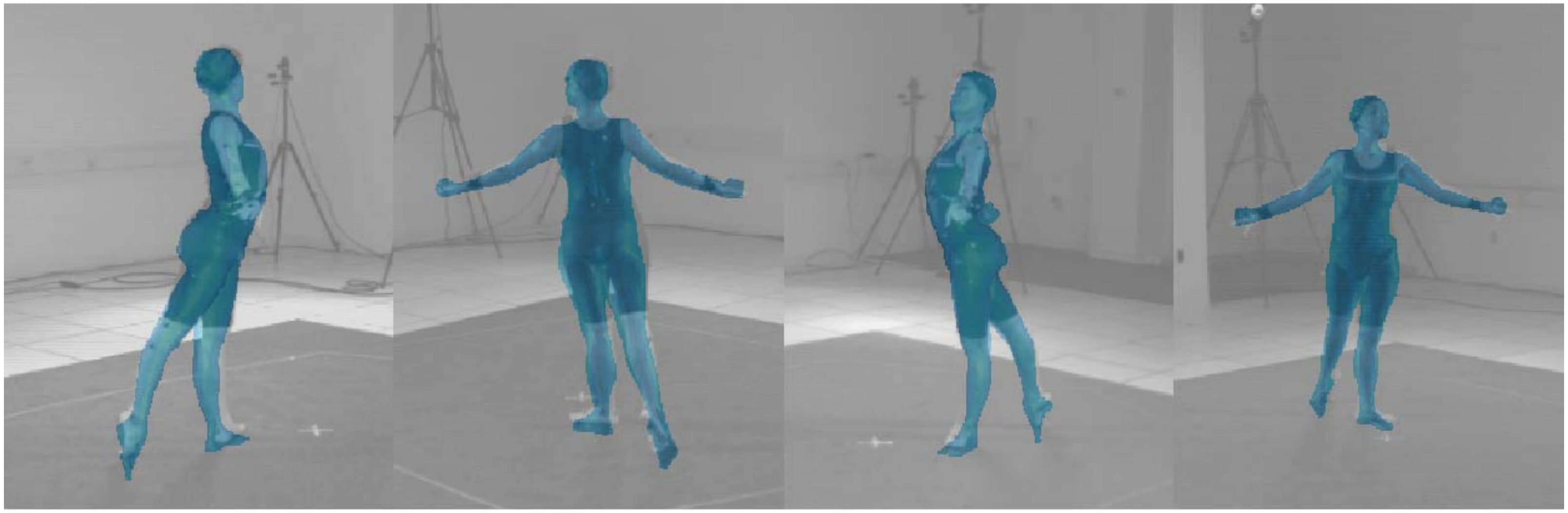
Optimization – Step by Step



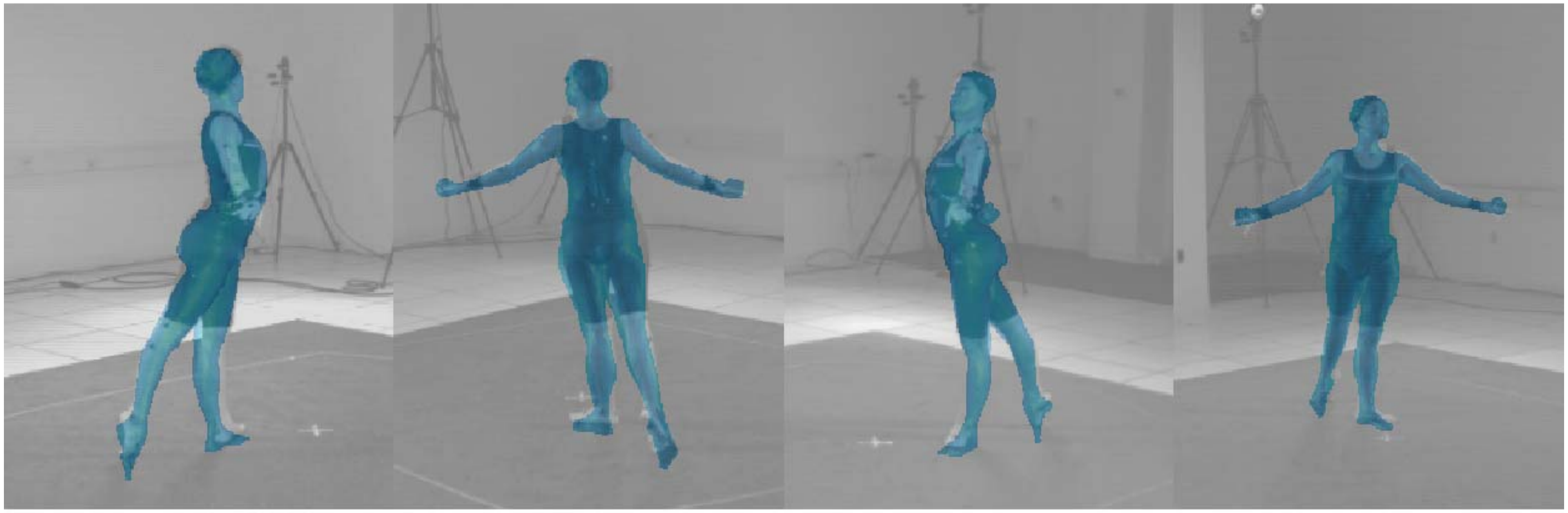
Optimization – Step by Step

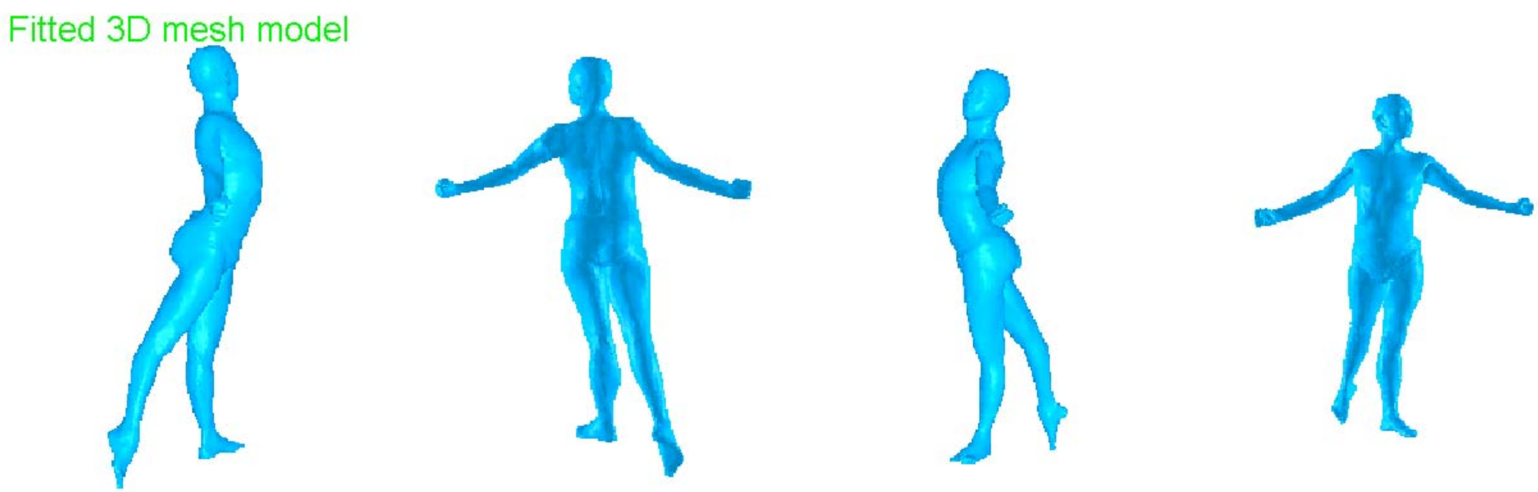


Optimization – Step by Step



Optimization – Step by Step





The "naked" case

20 bases.
Subject not in training set.

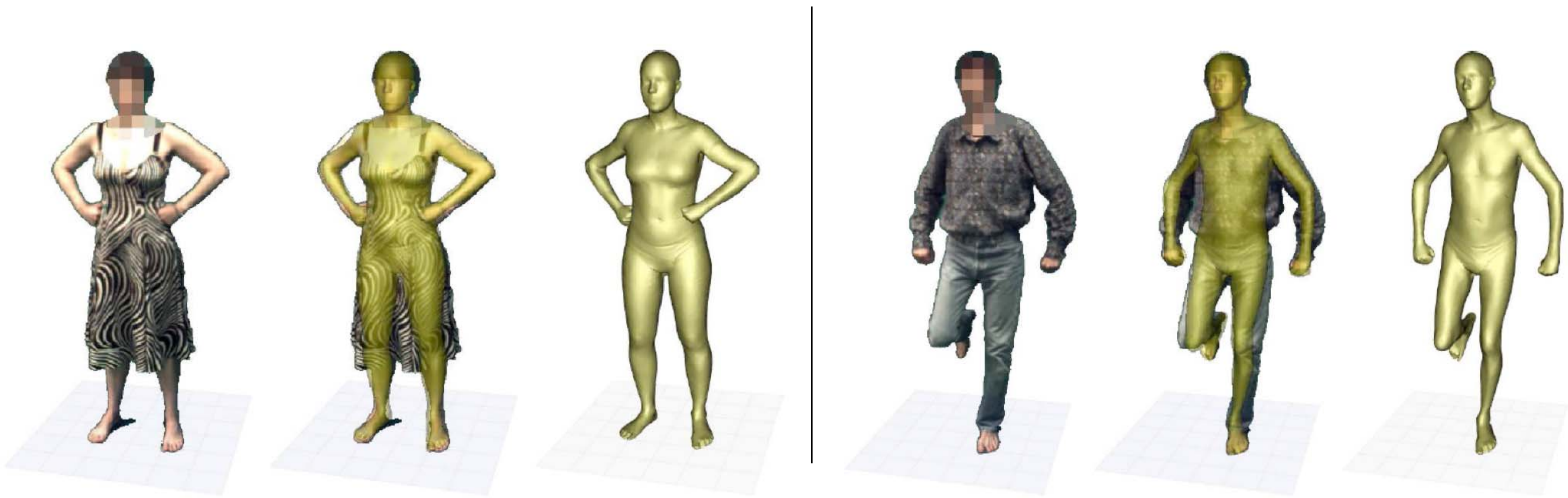
Problem: Clothing

Problem: Clothing

Shape under clothing

Can you guess what someone looks like under their clothes?

– Exploit the factorization of the model to combine constraints across poses.



Principle: Shape under clothing

- Silhouettes are larger when there is clothing

Principle: Shape under clothing

- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes

$$E_{\text{inside}}(s) = d\left(\mathrm{F}^{e}_{k,s}, \mathrm{F}^{o}_{k}\right)$$

Principle: Shape under clothing

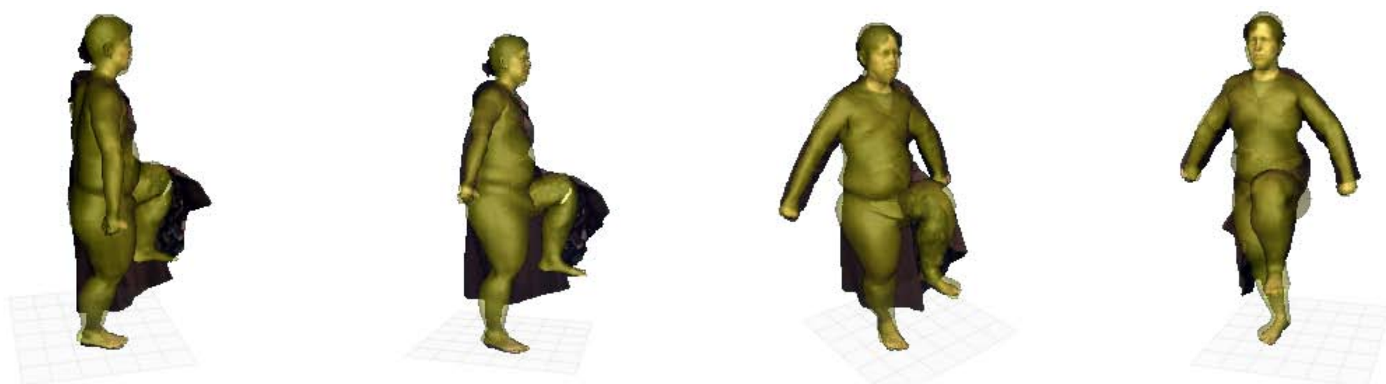
- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes

$$E_{\text{inside}}(s) = d\left(\mathrm{F}^{e}_{k,s}, \mathrm{F}^{o}_{k}\right)$$

  - Should not try to explain the entire image silhouette

$$d\left(\mathrm{F}^{o}_{k}, \mathrm{F}^{e}_{k,s}\right)$$

- NO

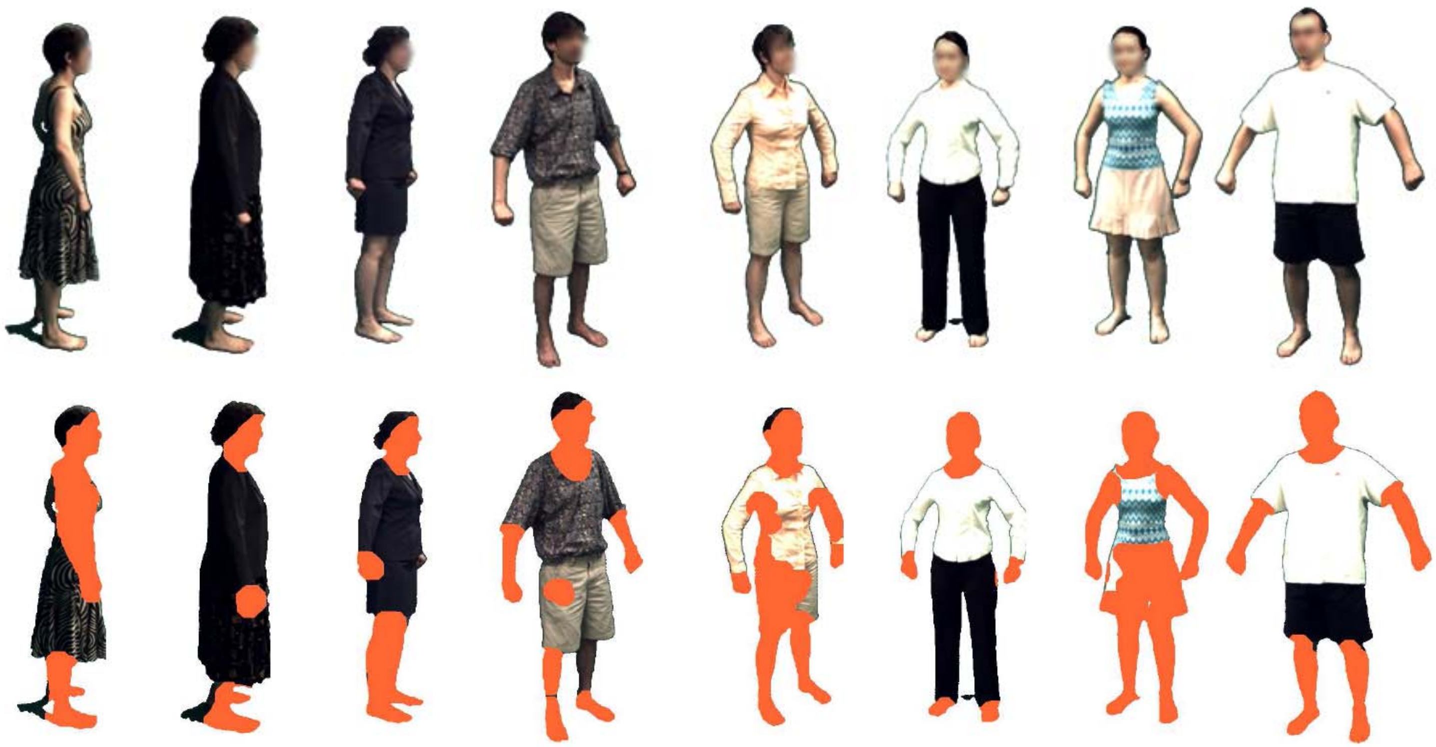


Principle: Shape under clothing

- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes
  - Constraints are tight in regions without clothes

Principle: Shape under clothing

- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes
  - Constraints are tight for skin regions
    Skin Detection



Principle: Shape under clothing

- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes
  - Constraints are tight for skin regions

$$E_{\text{expand}}(s) = d\left(\underbrace{S_k^o}_{\text{skin}}, \underbrace{F_{k,s}^e}_{\text{model}}\right) + \underbrace{\lambda}_{\lambda < 1} d\left(\underbrace{F_k^o \setminus S_k^o}_{\text{non-skin}}, \underbrace{F_{k,s}^e}_{\text{model}}\right)$$

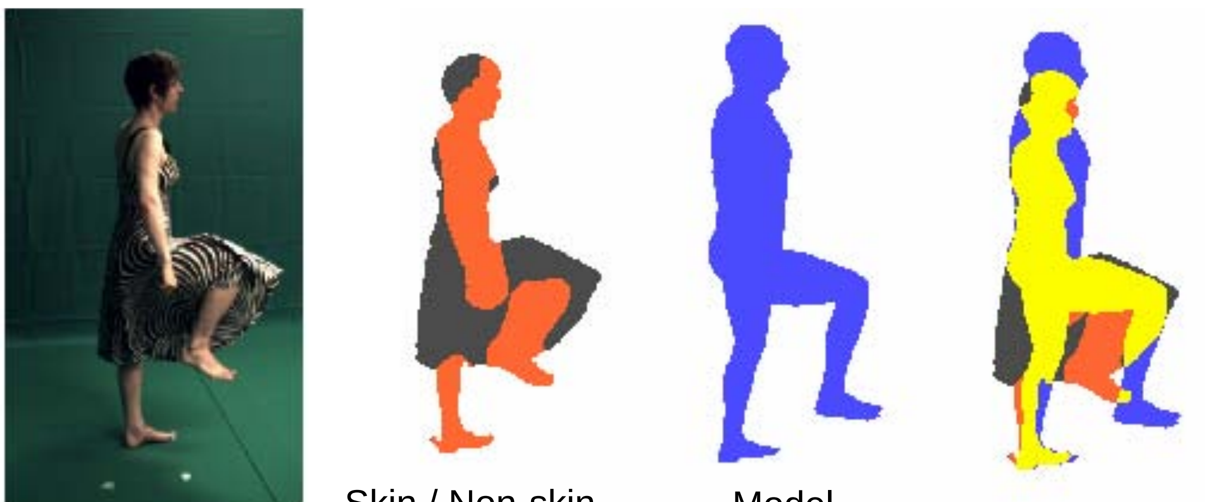


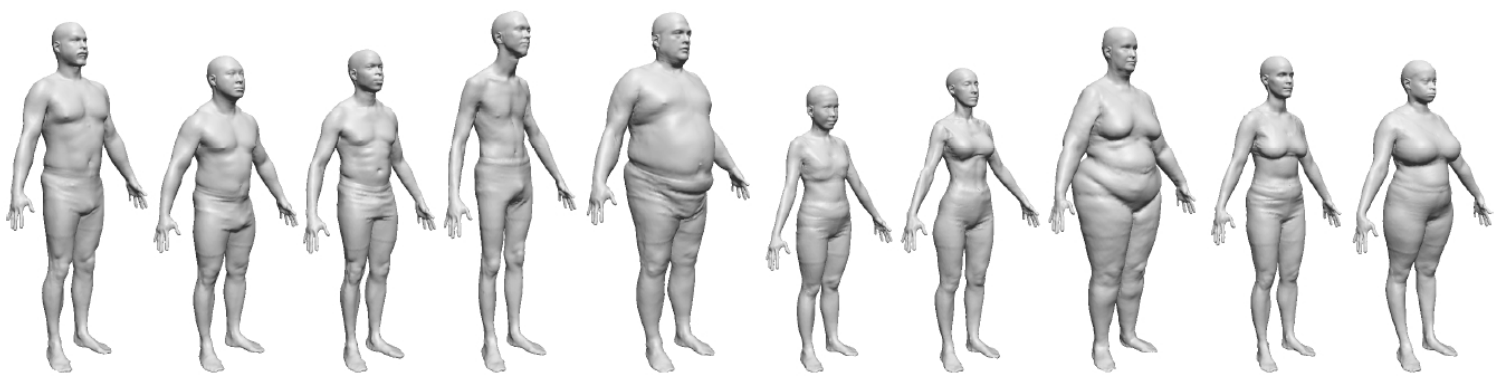
Image | Skin / Non-skin Silhouettes | Model Silhouette | Overlap

Principle: Shape under clothing

- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes
  - Constraints are tight for skin regions
- True shape not observable
  - Family of human body shapes (known statistics)

$$E_{shape}(\beta) = \sum_j \max\left(0, \frac{|\beta_j|}{\sigma_{\beta,j}} - \sigma_{thresh}\right)^2$$

$$E_{pose}(\theta)$$



[Allen et al. '03]

Principle: Shape under clothing

- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes
  - Constraints are tight for skin regions
- True shape not observable
  - Family of human body shapes (known statistics)

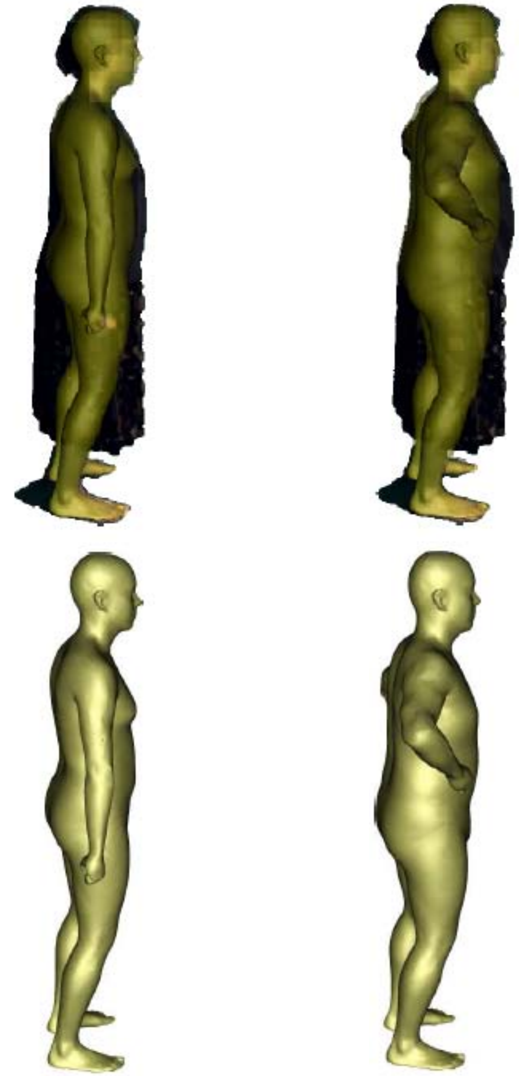
Objective Function

$$E_{\text{clothes}}(s) = \sum_{k=1}^{K} E_{\text{inside}}(s) + E_{\text{expand}}(s) + E_{\text{shape}}(\beta) + E_{\text{pose}}(\theta)$$

Comparison

Fitting as if Naked

Single-pose Fitting with Clothes



Principle: Shape under clothing

- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes
  - Constraints are tight for skin regions
- True shape not observable
  - Family of human body shapes (known statistics)
- **Body shape constant although pose may vary:**

Principle: Shape under clothing

- Silhouettes are larger when there is clothing
  - Body must fit inside silhouettes
  - Constraints are tight for skin regions
- True shape not observable
  - Family of human body shapes (known statistics)
- Combine constraints across pose

"Batch" objective function

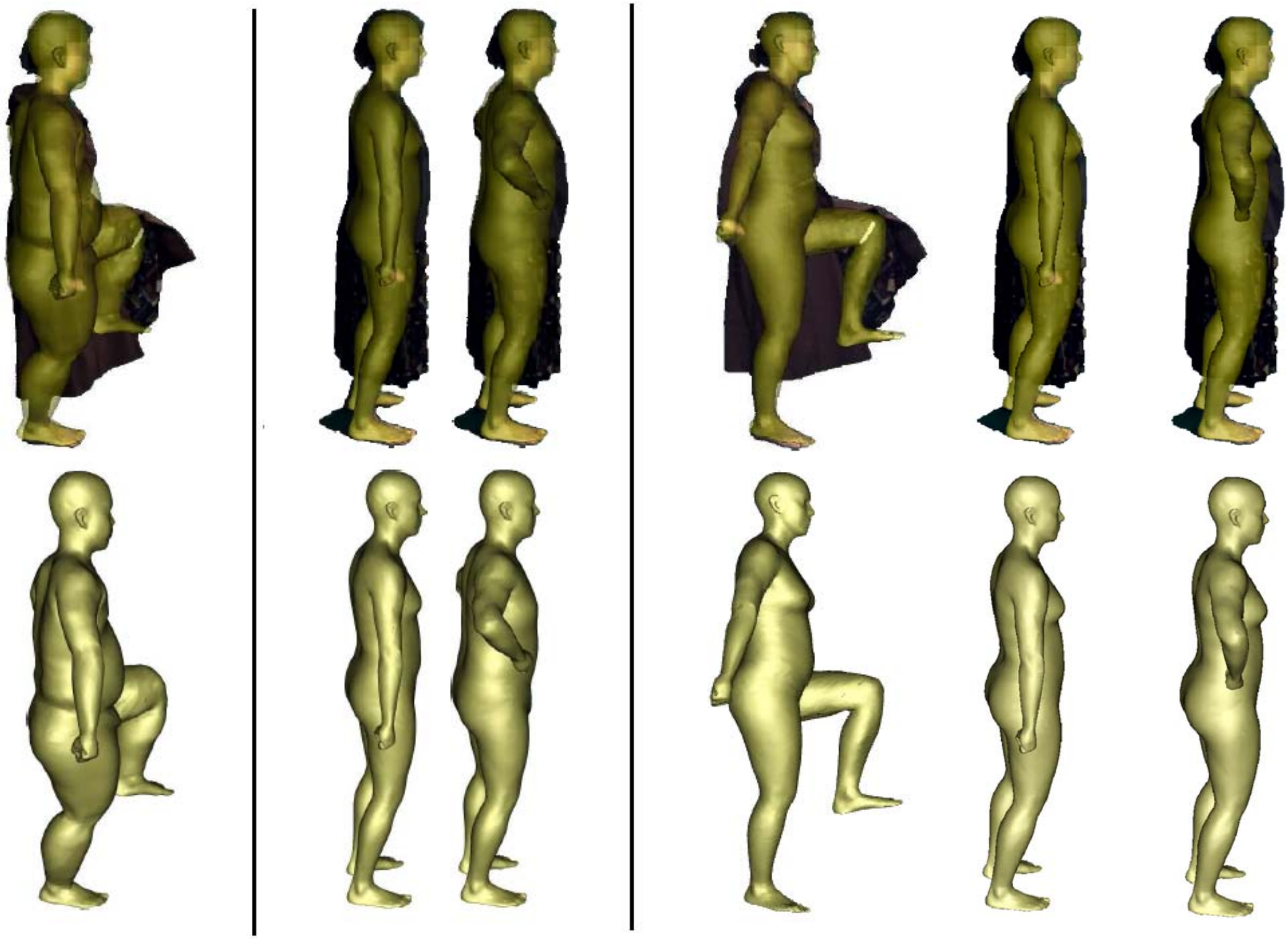
$$E_{\text{clothes}}(\beta, \Theta) = \sum_{p=1}^{P} \sum_{k=1}^{K} E_{\text{inside}}(\beta, \theta_p) + E_{\text{expand}}(\beta, \theta_p) + E_{\text{shape}}(\beta) + E_{\text{pose}}(\theta_p)$$

$$\Theta = \theta_1, \ldots, \theta_P$$

BROWN

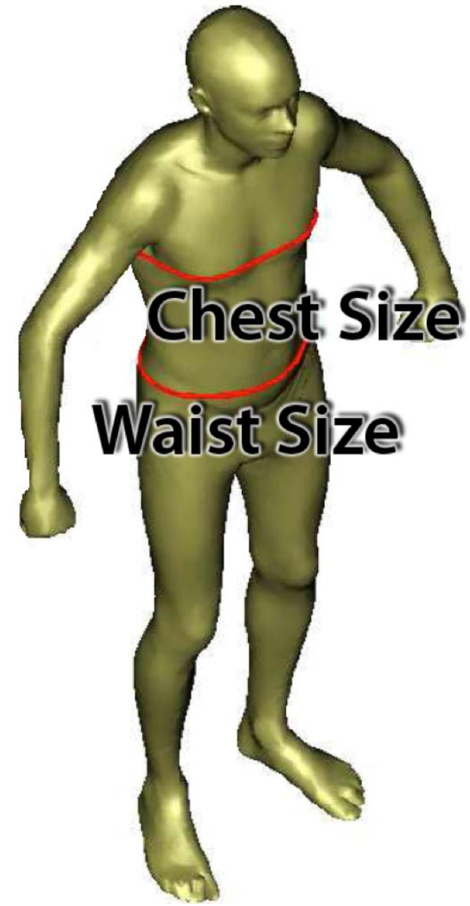
Comparison

Fitting as if Naked | Single-pose Fitting | Batch Fitting

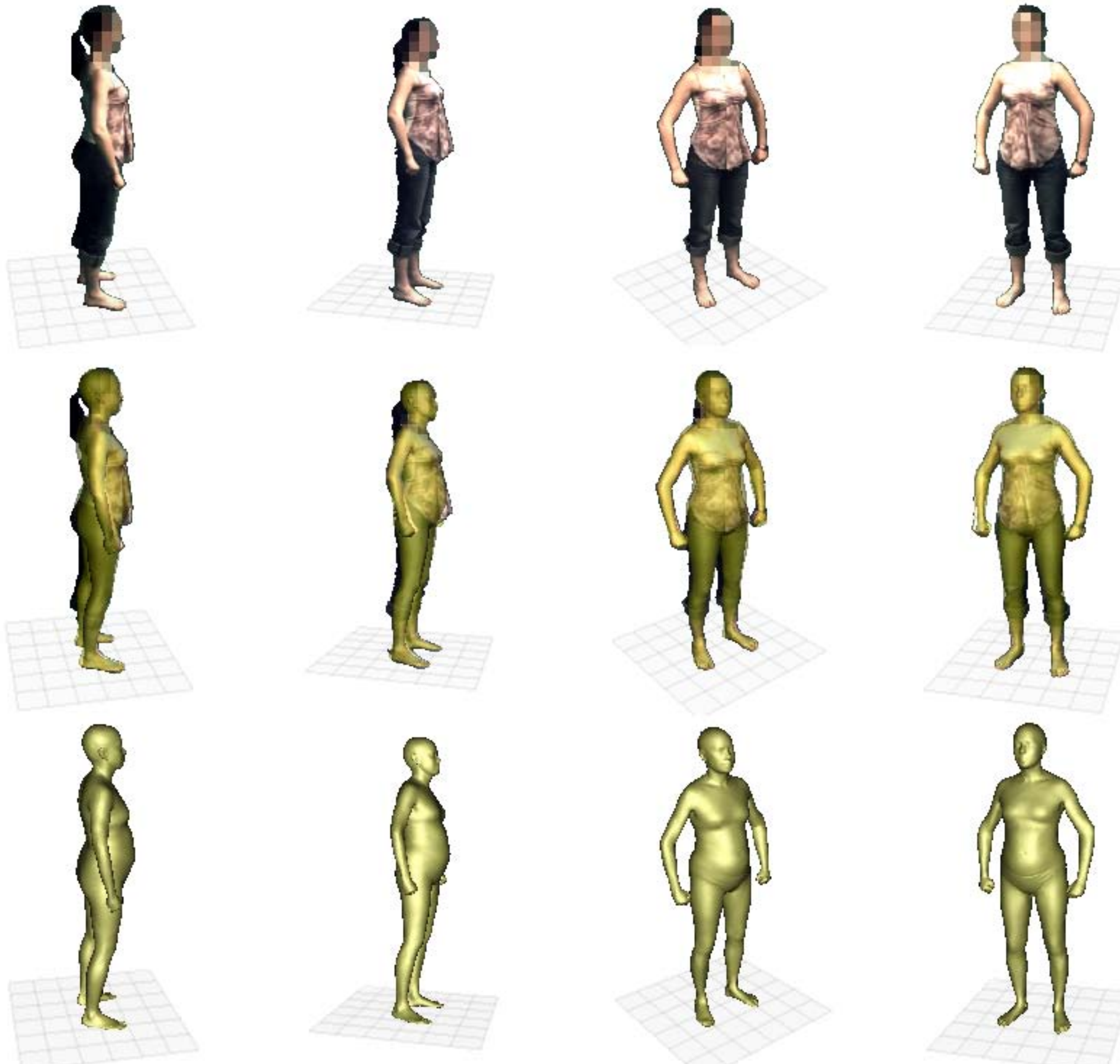


Shape under clothing

Quantitative Evaluation



Failure case



What about “regular” video?

- Everything here is multi-camera
  - See ICCV’07 for some monocular results (using cast shadows)
- Combine multiple poses to estimate shape
  - These poses can come from different cameras or different frames of the same camera.
    - Though pose estimation in a single frame is harder.
- There are many more cues to use beyond silhouettes
- Assume calibration here

Applications

- Surveillance
  - Extract body shape measurements from surveillance video.
  - Gender from video (94.3% correct on our clothing dataset).
- Games
  - Avatar creation and animation.
- Fashion
  - Extract body shape measurements.
  - Virtual try-on.

Summary of approach

- Estimating human shape may be as important as estimating 3D pose
- Detailed graphics model practical for vision
  - Sometimes complexity makes things easier (or possible).
  - Better match to image evidence (e.g. torso shape).
  - Initial optimization approx 5hrs (!) per frame; down to about 2min.
- Shape under clothing
  - Combine constraints on shape constancy and skin with clothing-appropriate observation model.
- Representation
  - Supports more than “detection”; is someone tall? fat? male? old?
  - 3D model allows us to explain illumination, shading, pose variation, muscle bulging, ….
  - This allow us to extract what is constant over time (identity, shape, etc.)

Current & future work

- Monocular estimation and tracking in multiple frames with moving camera
  - Constraints on shape consistency
- Beyond silhouettes
  - Internal structure & motion
- Modeling clothing and hair
- Dynamics of soft tissue and cloth
- Evaluation of accuracy
  - Pose and shape

See Balan *et al*, ICCV ‘07

The evolution of man

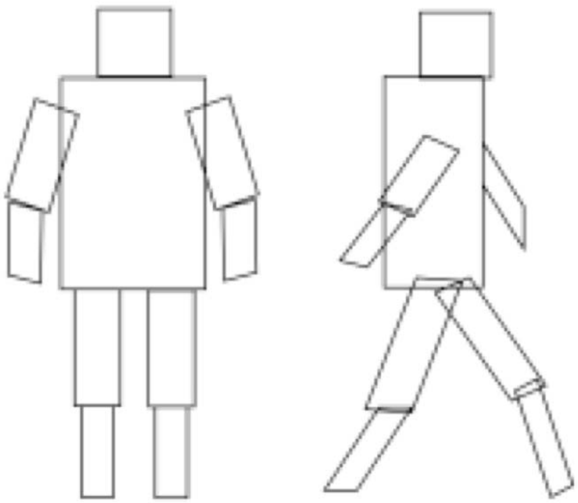
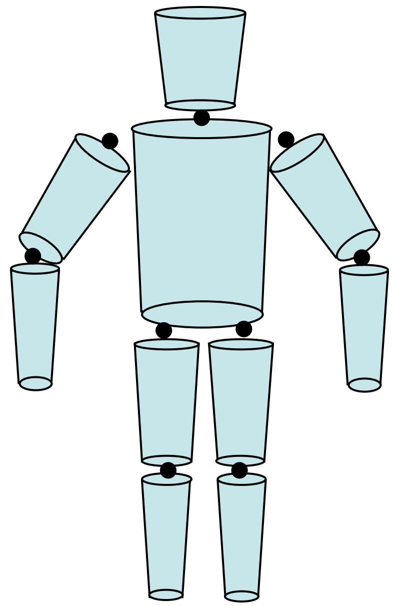


Figure 1: The cardboard person model. The limbs of a person are represented by planar patches.

1996



2000

2007

Acknowledgements

Funding Sources:

- NSF IIS-0535075, IIS-0812364
- NIH-NINDS R01 NS 50967-01
  - part of the Collaborative Research on Computational Neuroscience Program
- Intel Corporation

Collaborators: Horst Haussecker (Intel), Leonid Sigal (Toronto), James Davis (UCSC)

Thanks: Dragomir Anguelov (Google), Alex Weiss (Brown), Matthew Loper (Brown), Steven Scher (UCSC), Loretta Reiss (Brown), David Hirshbirg (Brown)