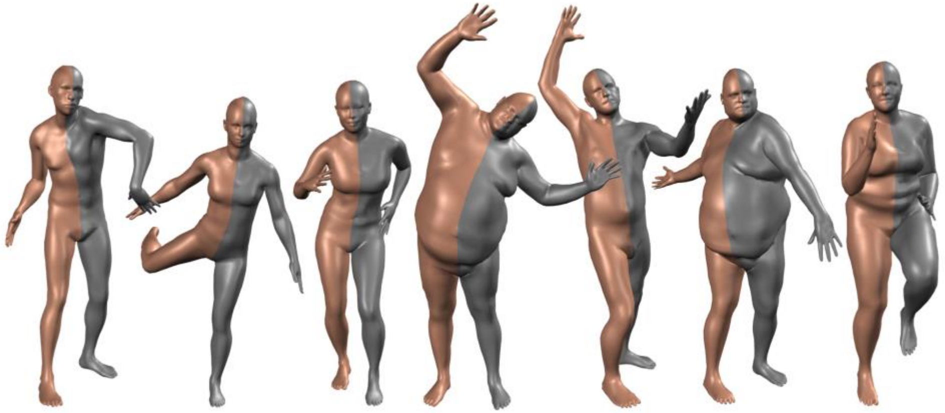
11. Keep it SMPL Automatic Estimation of 3D Human Pose and Shape from a Single Image

Thai Thanh Tuan, Young sik Yun

Create: 13th July 2021

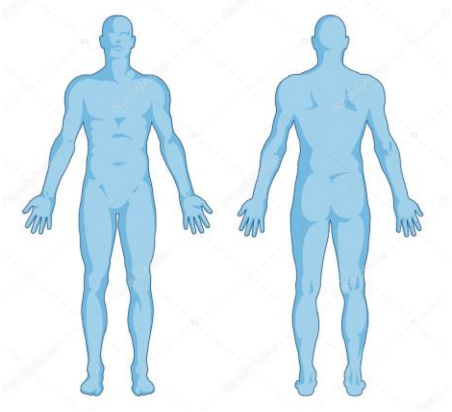


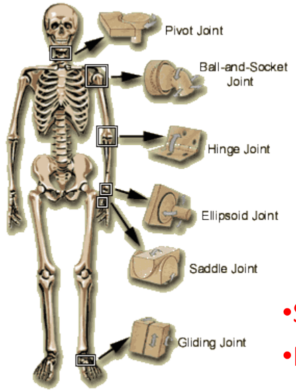
Introduction

- Project page: https://smplify.is.tuebingen.mpg.de/index.html
- Paper: http://files.is.tue.mpg.de/black/papers/BogoECCV2016.pdf
- Supplementary Video: https://youtu.be/OgX49T2Cqdo
- Conference: ECCV 2016

Introduction

- Challenges:
  - Complexity of the human body
  - Articulation
  - Occlusion
  - Clothing
  - Lighting
  - Ambiguity in inferring 3D from 2D



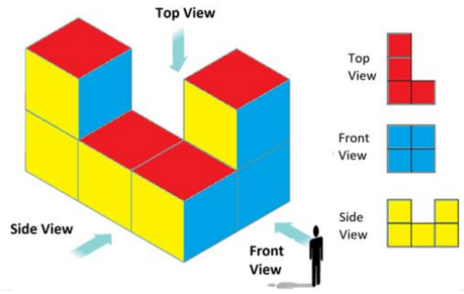


Joint/Articulation

Classified by:

•Structure (composition)

•Function (range of motion)



Challenges

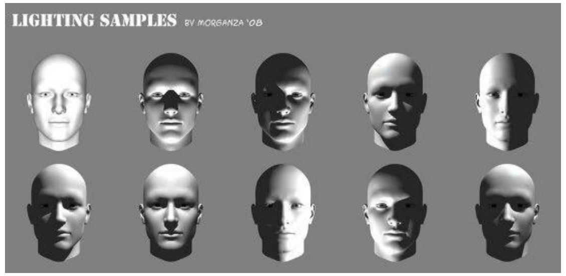


Figure 1: Heavily occluded people are better separated using *human pose* than using *bounding-box*.

Introduction

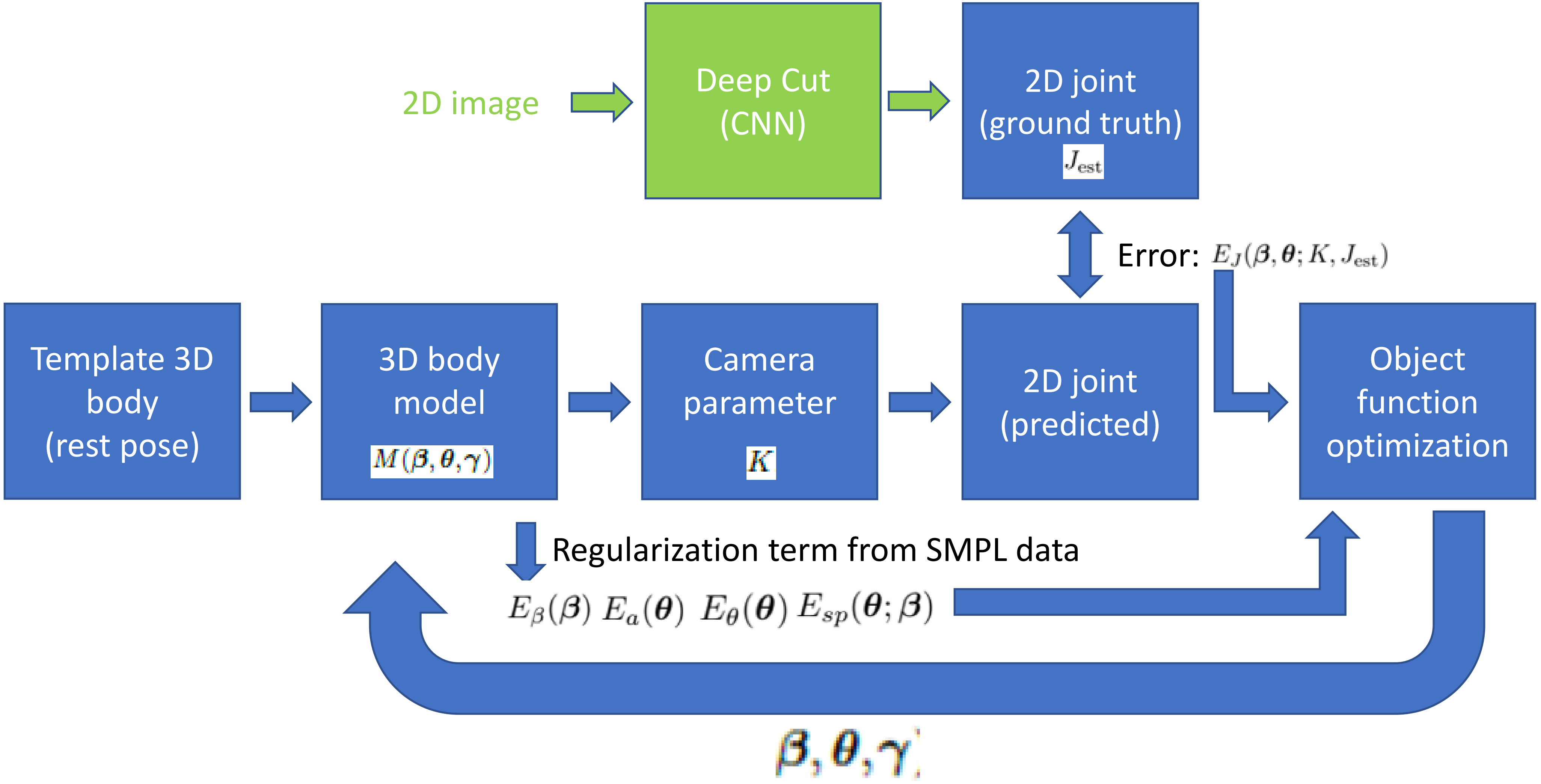
- Motivation:
  - Many applications
  - Previous works only on pose and ignore 3D human shape
- Evaluation:
  - CMU dataset [3]
  - Compare recently published methods [4, 39, 58]
  - Qualitatively
  - Quantitatively
- Contribution:
  - First fully automatic method of estimating 3D body shape and pose from 2D joints
  - An interpenetration term
  - Novel objective function that matches a 3D body model to 2D joints
  - Provide the code/2D joints/ 3D models

Is it better to add body shape information?

Related work

- 3D pose from 2D joints:
  - Have weak/non-existent models of human shape.
  - →stronger model of body shape →reduce ambiguity
- 3D pose and shape:
- Make it automatic:
  - None of the methods are automatic

Estimate 3D joint from 2D joint



Objective function

Objective function $$E(\boldsymbol{\beta}, \boldsymbol{\theta}) = E_J(\boldsymbol{\beta}, \boldsymbol{\theta}; K, J_{\text{est}}) + \lambda_\theta E_\theta(\boldsymbol{\theta}) + \lambda_a E_a(\boldsymbol{\theta}) + \lambda_{sp} E_{sp}(\boldsymbol{\theta}; \boldsymbol{\beta}) + \lambda_\beta E_\beta(\boldsymbol{\beta})$$

Joint-based data term $$E_J(\boldsymbol{\beta}, \boldsymbol{\theta}; K, J_{\text{est}}) = \sum_{\text{joint } i} w_i \rho(\Pi_K(R_\theta(J(\boldsymbol{\beta})_i)) - J_{\text{est},i})$$

Pose prior (elbows and knees) $$E_a(\boldsymbol{\theta}) = \sum_i \exp(\boldsymbol{\theta}_i)$$

Pose prior $$E_\theta(\boldsymbol{\theta}) \equiv -\log \sum_j (g_j \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{\theta,j}, \Sigma_{\theta,j})) \approx -\log(\max_j(c g_j \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{\theta,j}, \Sigma_{\theta,j})))$$

$$= \min_j \left(-\log(c g_j \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{\theta,j}, \Sigma_{\theta,j}))\right)$$

Interpenetration error term $$E_{sp}(\boldsymbol{\theta}; \boldsymbol{\beta}) = \sum_i \sum_{j \in I(i)} \exp\left(\frac{||C_i(\boldsymbol{\theta}, \boldsymbol{\beta}) - C_j(\boldsymbol{\theta}, \boldsymbol{\beta})||^2}{\sigma_i^2(\boldsymbol{\beta}) + \sigma_j^2(\boldsymbol{\beta})}\right)$$

Shape prior $$E_\beta(\boldsymbol{\beta}) = \boldsymbol{\beta}^T \Sigma_\beta^{-1} \boldsymbol{\beta}$$

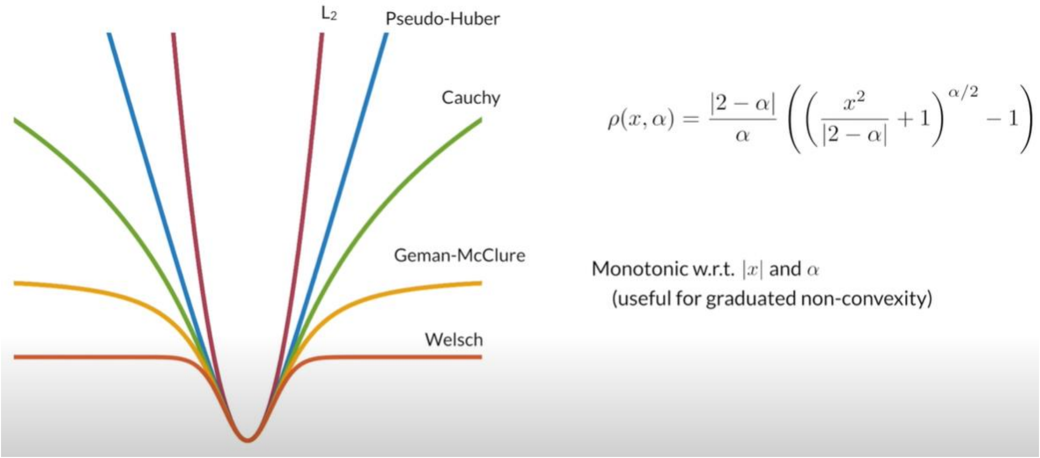
Joint-based data term

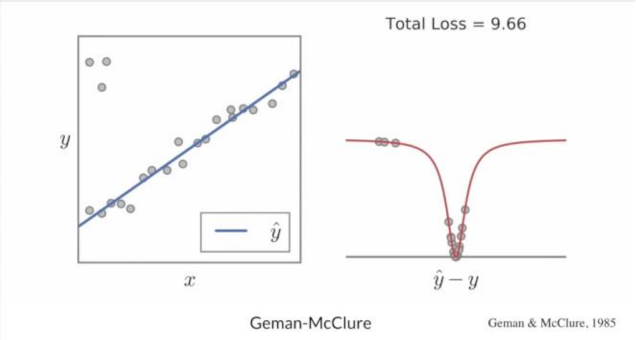
the confidence of estimated joints (Deep pose)
→occluded joints has low confidence(weight)

$$E_J(\boldsymbol{\beta}, \boldsymbol{\theta}; K, J_{\text{est}}) = \sum_{\text{joint } i} w_i \rho(\Pi_K(R_\theta(J(\boldsymbol{\beta})_i)) - J_{\text{est},i})$$

Noisy estimate → differentiable Geman-McClure penalty function

the projection from 3D to 2D induced by a camera with parameters K





https://www.youtube.com/watch?v=BmNKbnF69eY&ab_channel=JonBarron

Pose prior (elbows and knees)

- Penalizing elbows and knees that bend unnaturally

$$E_a(\boldsymbol{\theta}) = \sum_i \exp(\boldsymbol{\theta}_i)$$

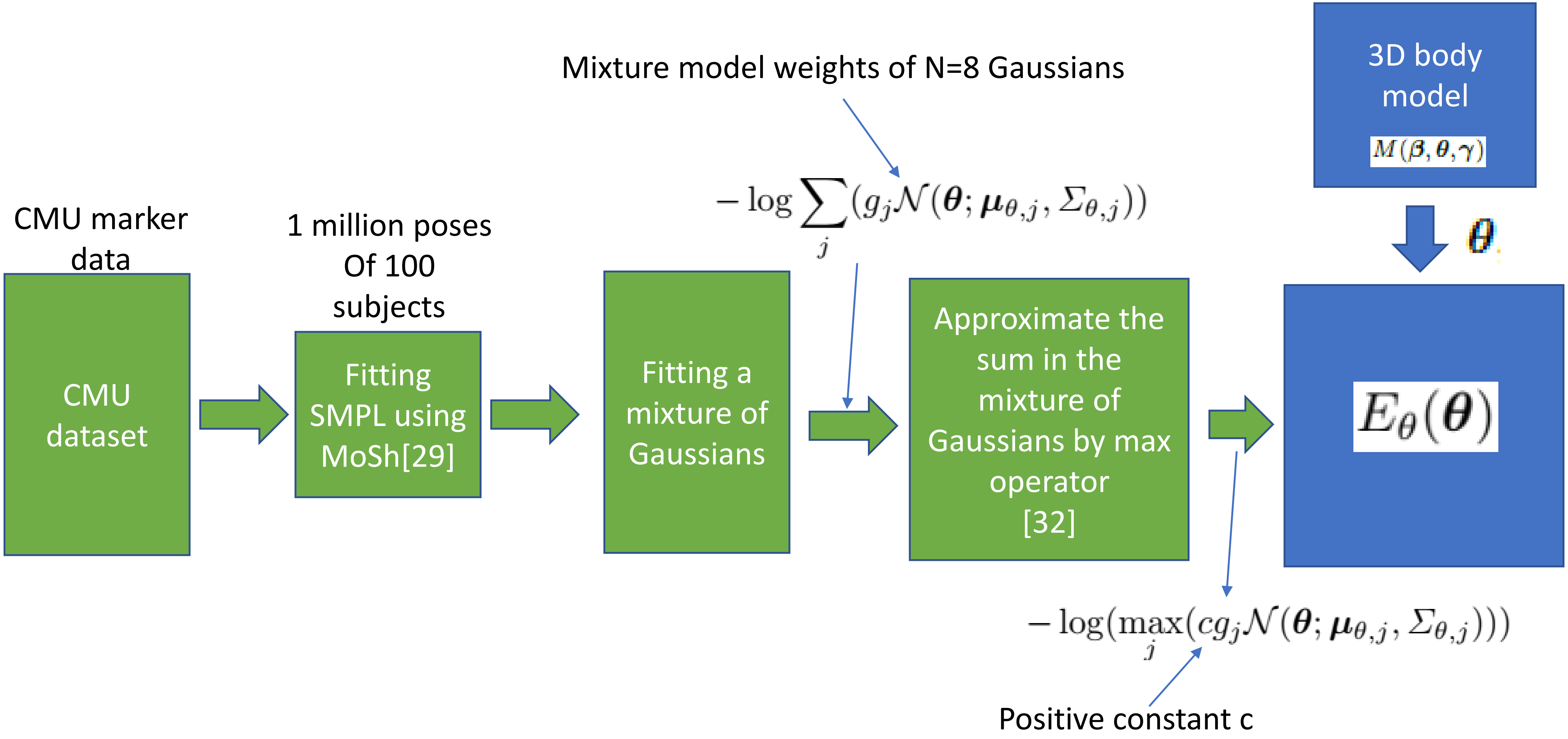
right knee, left knee, right elbow, left elbow

Negative bending = natural → penalty goes to zero
Positive bending = unnatural → penalized high

$$E_\theta(\boldsymbol{\theta}) \equiv -\log \sum_j (g_j \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{\theta,j}, \Sigma_{\theta,j})) \approx -\log(\max_j(c g_j \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{\theta,j}, \Sigma_{\theta,j}))) \quad (4)$$

$$= \min_j \left(-\log(c g_j \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{\theta,j}, \Sigma_{\theta,j}))\right) \quad (5)$$

3D pose prior

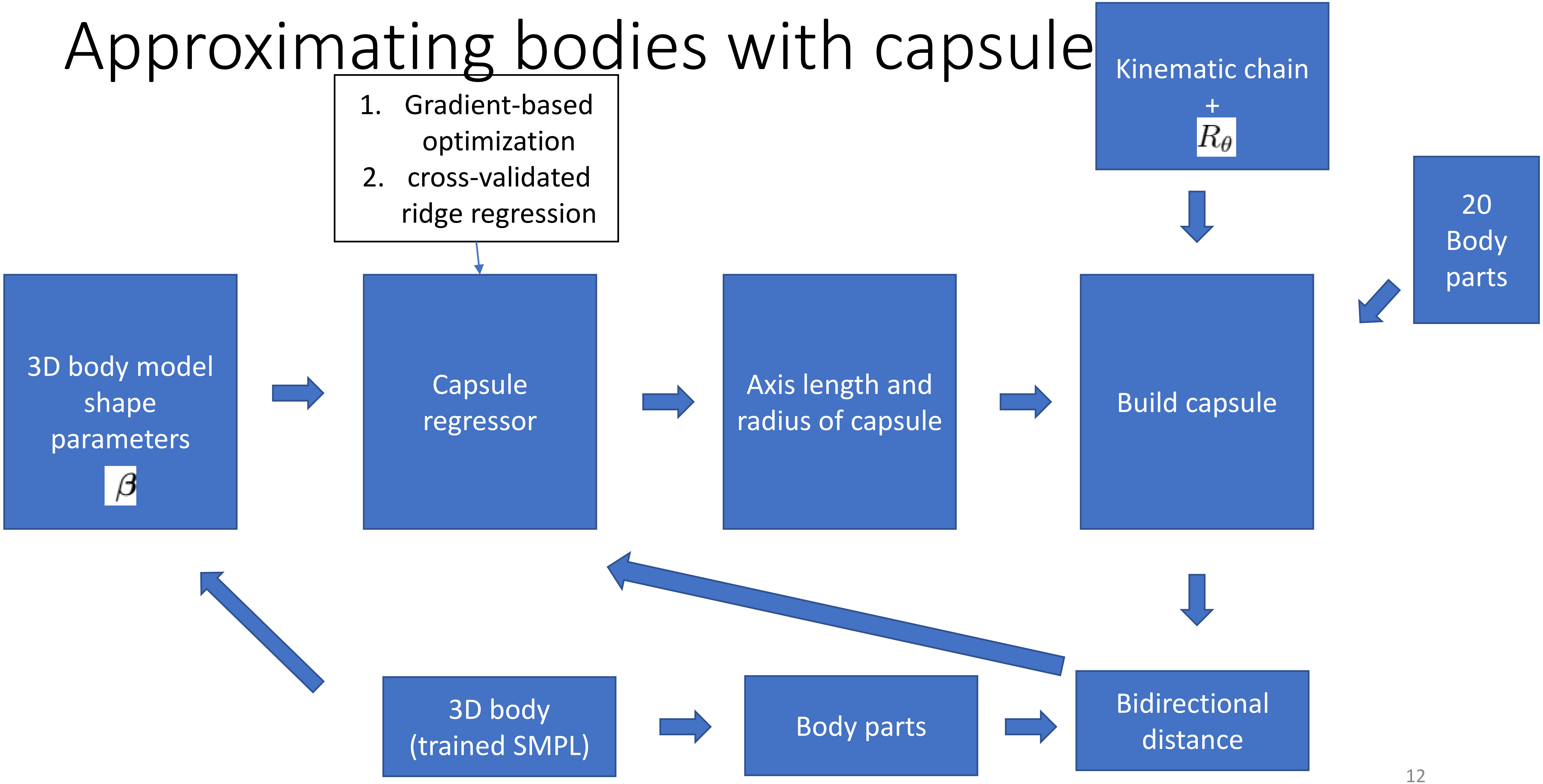


Approximating bodies with capsules

- Complex, non-convex body surfaces → hard to compute interpenetration
- ➔follow [10,50] to use proxy geometries to compute collisions

Fig. 3. Body shape approximation with capsules. Shown for two subjects. Left to right: original shape, shape approximated with capsules, capsules reposed. Yellow point clouds represent actual vertices of the model that is approximated.

Approximating bodies with capsule



Along the capsule axis and radius

Interpenetration error term

Along the capsule axis and radius $r(\boldsymbol{\beta})$

Center of sphere

Mixture of 3D Gaussians model in [47]

$$E_{sp}(\boldsymbol{\theta};\boldsymbol{\beta}) = \sum_{i} \sum_{j \in I(i)} \exp\left(\frac{||C_i(\boldsymbol{\theta},\boldsymbol{\beta}) - C_j(\boldsymbol{\theta},\boldsymbol{\beta})||^2}{\sigma_i^2(\boldsymbol{\beta}) + \sigma_j^2(\boldsymbol{\beta})} \right)$$

The spheres incomparable with i

$\sigma(\boldsymbol{\beta}) = \frac{r(\boldsymbol{\beta})}{3}$

➔ Would bias the body shape to be thin ➔Not use in optimizing shape.

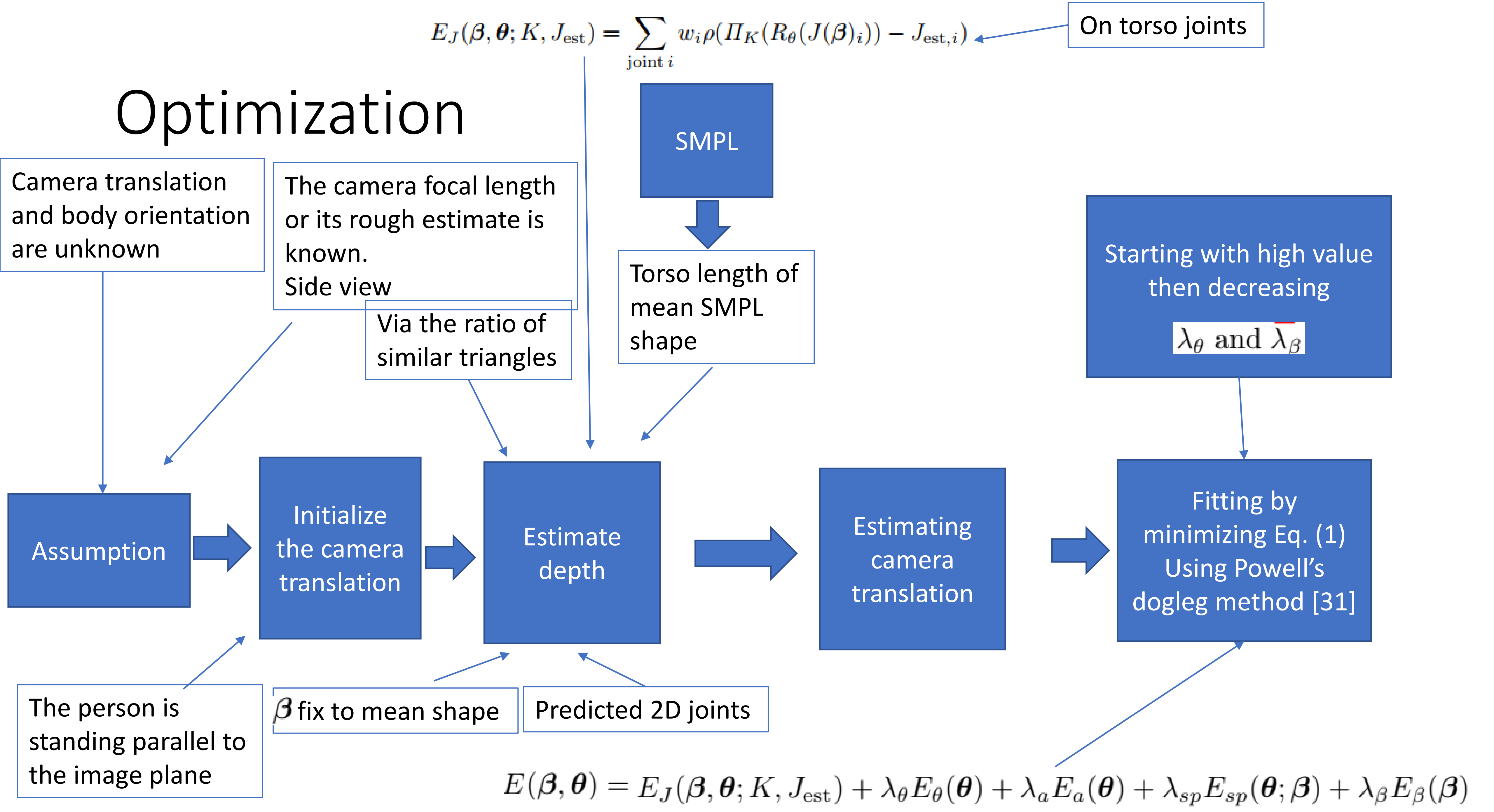
Shape prior

$$E_\beta(\boldsymbol{\beta}) = \boldsymbol{\beta}^T \Sigma_\beta^{-1} \boldsymbol{\beta}$$

A diagonal matrix with the squared singular values estimated via Principal Component Analysis from the shapes in the SMPL training set.

Note that the shape coefficients ***β*** are zero-mean by construction

Optimization

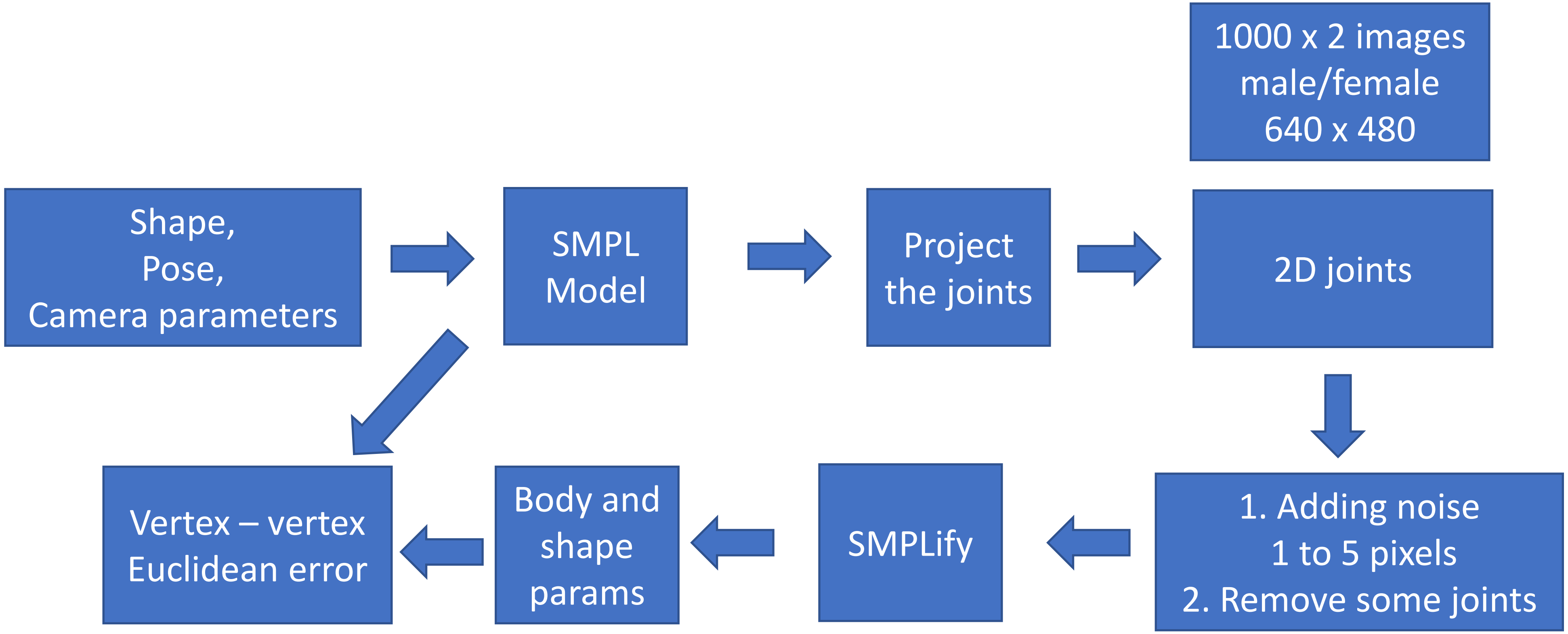


Evaluation

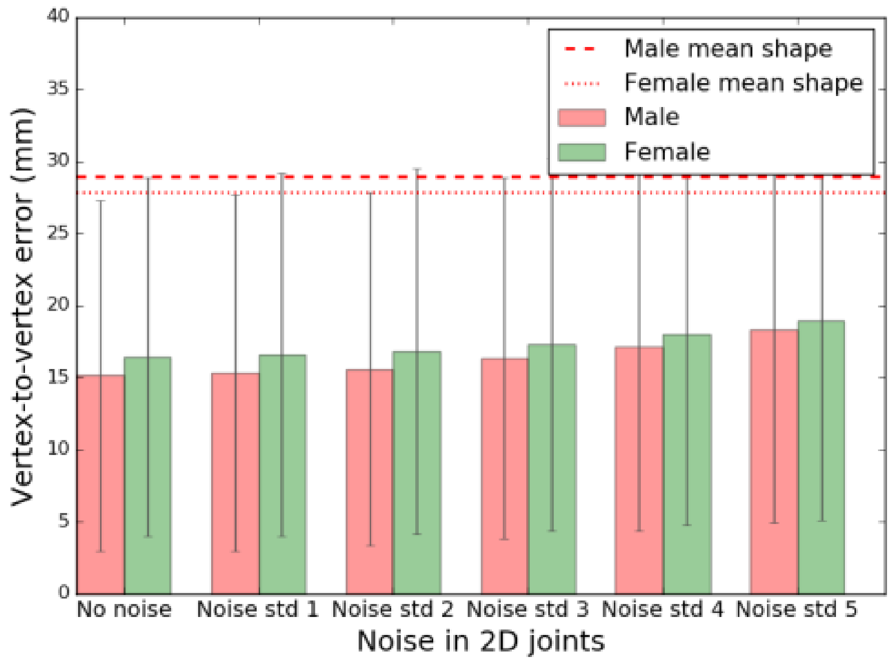
- Evaluate the accuracy of both 3D pose and 3D shape
- Datasets:
    - HumanEva-I[41]
    - Human3.6M [18]
    - Leeds Sparts Dataset (LSP)
    - Synthetic data
- Use 10 body shape coefficients

Quantitative evaluation: Synthetic data

- Synthetic bodies from the SMPL shape and pose



Quantitative evaluation



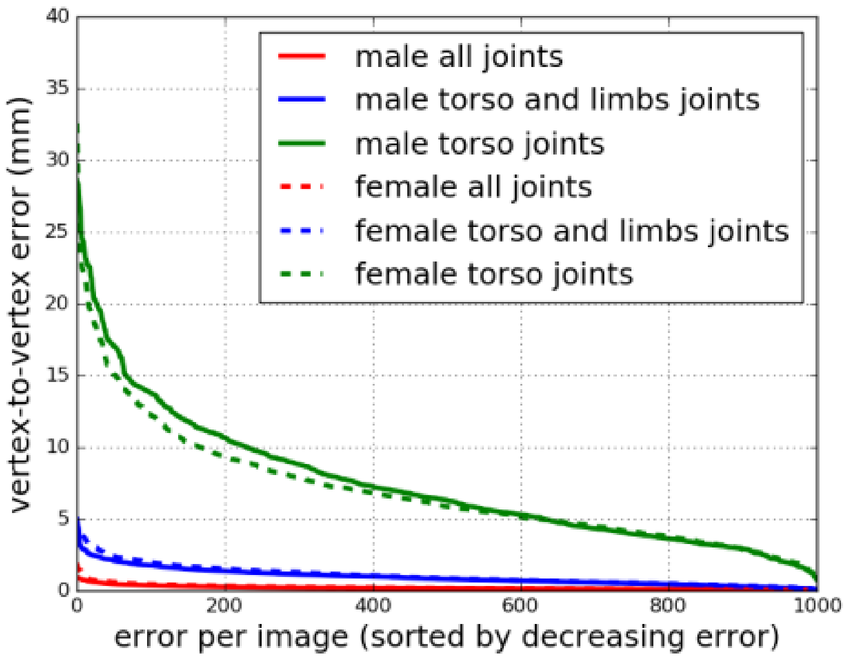
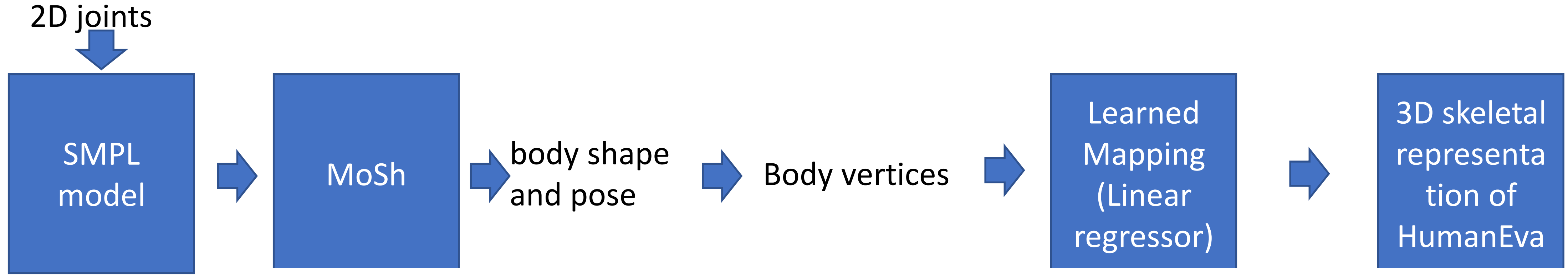


Fig. 4. Evaluation on synthetic data. Left: Mean vertex-to-vertex Euclidean error between the estimated and true shape in a canonical pose, when Gaussian noise is added to 2D joints. Dashed and dotted lines represent the error obtained by guessing the mean shape for males and females, respectively. Right: Error between estimated and true shape when considering only a subset of joints during fitting.

Quantitative evaluation: Real Data

- HumanEva dataset: Predict 3D pose from 2D joints.



	Walking			Boxing			Mean	Median
Method:	S1	S2	S3	S1	S2	S3		
Akhter & Black [4]	186.1	197.8	209.4	165.5	196.5	208.4	194.4	171.2
Ramakrishna et al. [39]	161.8	182.0	188.6	151.0	170.4	158.3	168.4	145.9
Zhou et al. [58]	100.0	98.89	123.1	112.5	118.6	110.0	110.0	98.9
SMPLify	**73.3**	**59.0**	**99.4**	**82.1**	**79.2**	**87.2**	**79.9**	**61.9**

Table 1. HumanEva-I results. 3D joint errors in mm.

Ablation study: 3D pose from 2D joints

Uses a single Gaussian

Interpenetration term does not have a significant impact on the 3D joint error. However, qualitatively, makes difference in more complex datasets

	Walking			Boxing			Mean	Median
Method:	S1	S2	S3	S1	S2	S3		
$E_\beta + E_J + E_{\theta'}$	98.4	79.6	117.8	105.9	98.5	122.5	104.1	82.3
$E_\beta + E_J + E_{\theta'} + E_{sp}$	97.9	79.4	116.0	105.8	98.5	122.3	103.7	82.3
SMPLify	**73.3**	**59.0**	**99.4**	**82.1**	**79.2**	**87.2**	**79.9**	**61.9**

Table 2. HumanEva-I ablation study. 3D joint errors in mm. The first row drops the interpenetration term and replaces the pose prior with a uni-modal prior. The second row keeps the uni-modal pose prior but adds the interpenetration penalty. The third row shows the proposed SMPLify model.

Effect of Interpenetration error term

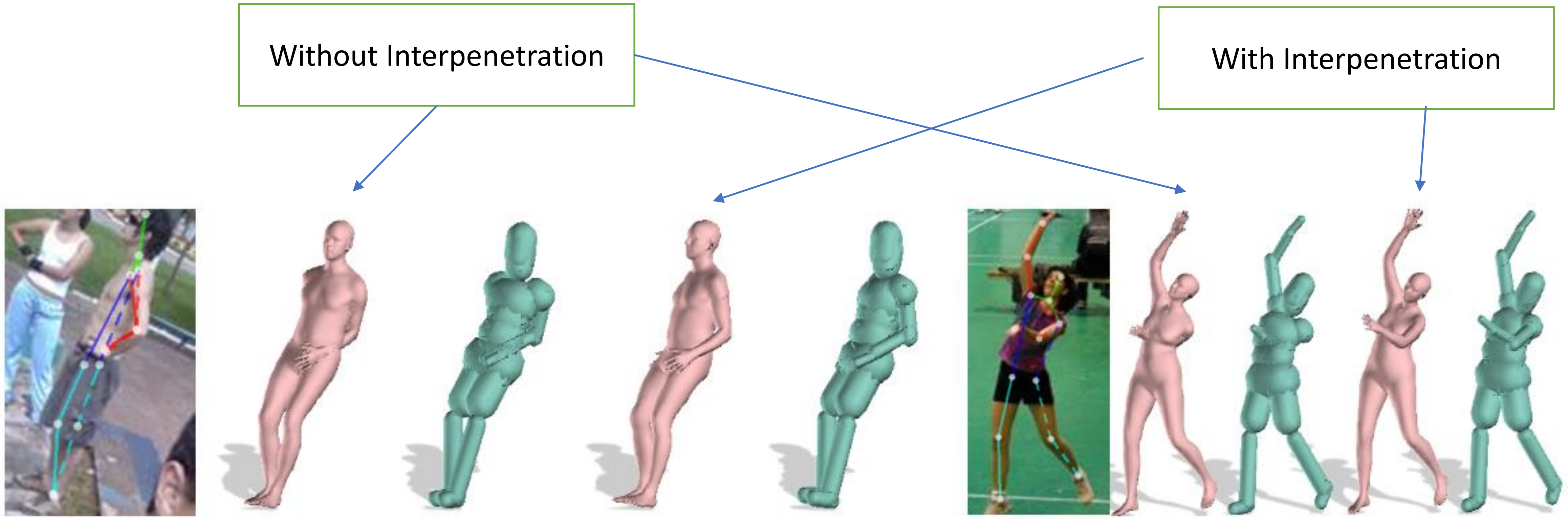


Fig. 5. Interpenetration error term. Examples where the interpenetration term avoids unnatural poses. For each example we show, from left to right, CNN estimated joints, and the result of the optimization *without* and *with* interpenetration error term.

Quantitative evaluation: Human3.6M dataset

	Directions	Discussion	Eating	Greeting	Phoning	Photo	Posing	Purchases	Sit
Akhter & Black [4]	199b.2	177.6	161.8	197.8	176.2	186.5	195.4	167.3	160.7
Ramakrishna et al. [39]	137.4	149.3	141.6	154.3	157.7	158.9	141.8	158.1	168.6
Zhou et al. [58]	99.7	95.8	87.9	116.8	108.3	107.3	93.5	95.3	109.1
SMPLify	**62.0**	**60.2**	**67.8**	**76.5**	**92.1**	**77.0**	**73.0**	**75.3**	**100.3**

	SitDown	Smoking	Waiting	WalkDog	Walk	WalkTogether	Mean	Median
Akhter & Black [4]	173.7	177.8	181.9	176.2	198.6	192.7	181.1	158.1
Ramakrishna et al. [39]	175.6	160.4	161.7	150.0	174.8	150.2	157.3	136.8
Zhou et al. [58]	137.5	106.0	102.2	106.5	110.4	115.2	106.7	90.0
SMPLify	**137.3**	**83.4**	**77.3**	**79.7**	**86.8**	**81.7**	**82.3**	**69.3**

Table 3. Human 3.6M. 3D joint errors in mm.

Qualitative evaluation

- On Leeds Sports Dataset

Fig. 6. Leeds Sports Dataset. Each sub-image shows the original image with the 2D joints fit by the CNN. To the right of that is our estimated 3D pose and shape and the model seen from another view. The top row shows examples using the gender-neutral body model; the bottom row show fits using the gender-specific models.

Qualitative evaluation: failure cases

- Reason:
  - Failure of DeepCut on 2D joints prediction
  - Depth ambiguities

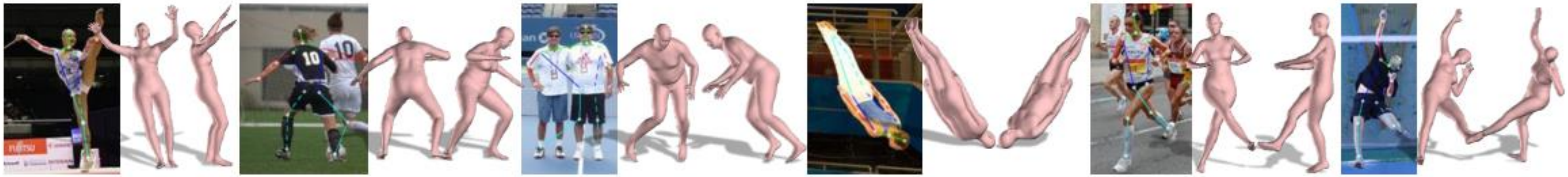


Fig. 7. LSP Failure cases. Some representative failure cases: misplaced limbs, limbs matched with the limbs of other people, depth ambiguities.

Fig. 8. Qualitative comparison. From top to bottom: Input image. Akhter & Black [4]. Ramakrishna et al. [39]. Zhou et al. [58]. SMPLify.

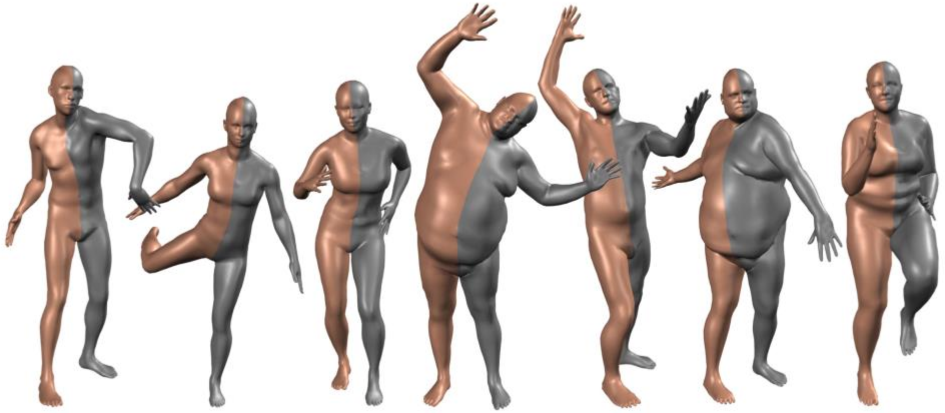
Conclusion

- SMPLIfy: a fully automated method for estimating 3D body shape and pose from 2D jojints in a single images.
  - Use SMPL body model
  - Minimizing the error between the projected joints of the model and the estimated 2D joints.

Expressive Body Capture: 3D Hands, Face, and Body from a Single Image

Thai Thanh Tuan

Create: 13th July 2021



Introduction

- Code: https://github.com/vchoutas/smplify-x
- Project page: https://smpl-x.is.tue.mpg.de/
- Paper: Expressive Body Capture: 3D Hands, Face, and Body from a Single Image
- Conference: CVPR 2019
- Paper: https://ps.is.tuebingen.mpg.de/uploads_file/attachment/attachment/497/SMPL-X.pdf
- Supplementary: https://ps.is.tuebingen.mpg.de/uploads_file/attachment/attachment/497/SMPL-X.pdf