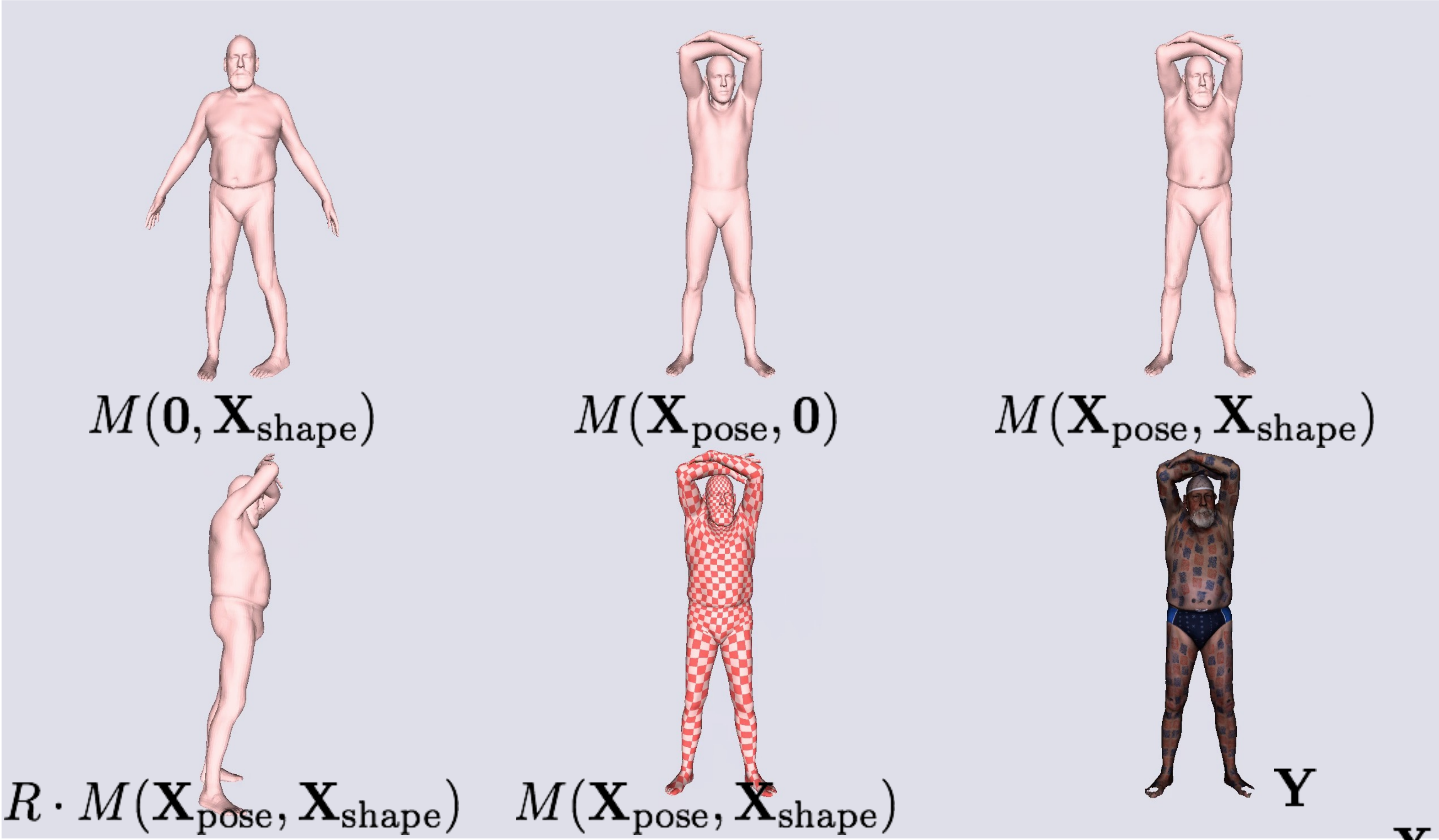
Virtual Humans – Winter 23/24

Lecture 7_1 – Fitting SMPL to IMU with Optimization

Prof. Dr.-Ing. Gerard Pons-Moll
University of Tübingen / MPI-Informatics

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

A Body Model is a function



$$\mathbf{X} = \{\mathbf{X}_{\text{pose}}, \mathbf{X}_{\text{shape}}\}$$

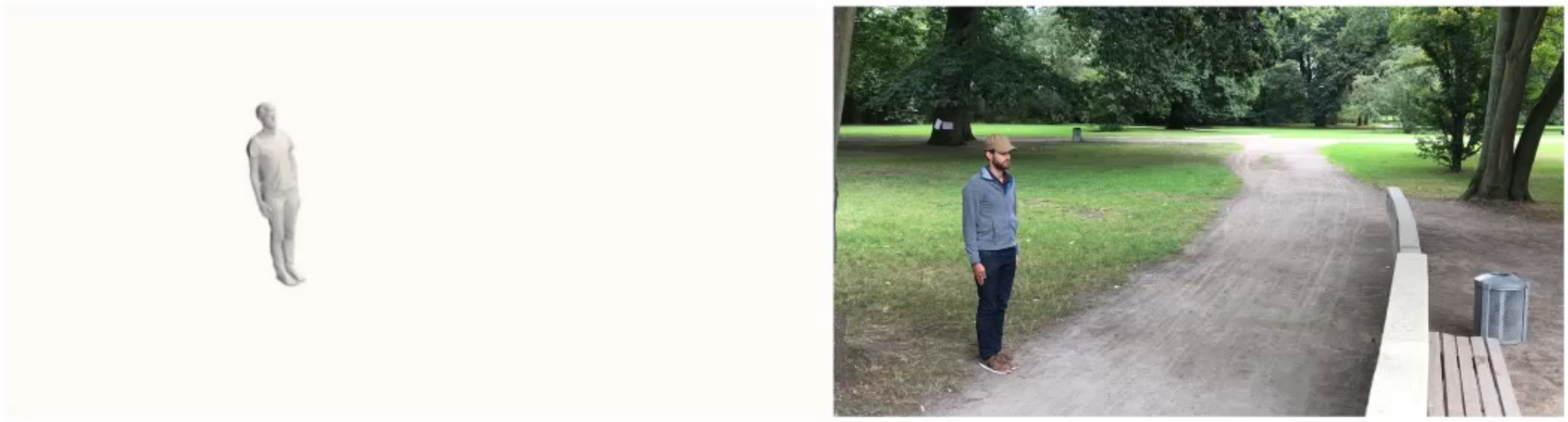
Sparse Inertial Poser

Automatic 3D Human Pose Estimation from Sparse IMUs

Timo v. Marcard[1], Bodo Rosenhahn[1], Michael J. Black[2], Gerard Pons-Moll[2]

[1]Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover
[2]MPI for Intelligent Systems, Perceiving Systems, Tübingen

3D Human Motion Capture

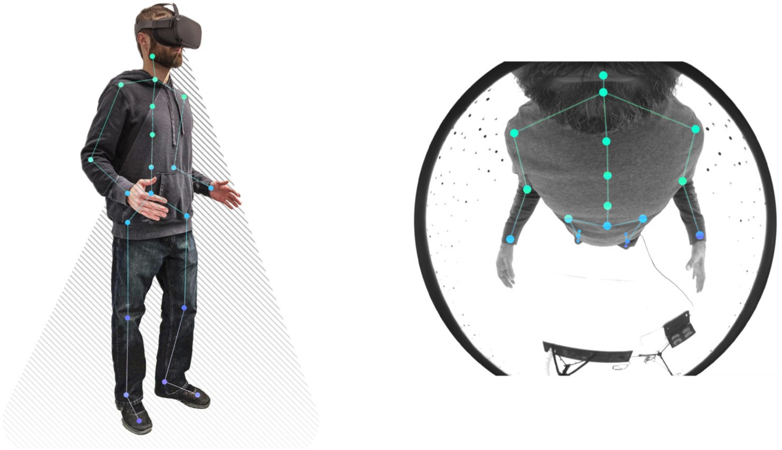


Vision-based Motion Capture

Image based such as SMPLify require external camera

- Limited recording volume
- Certain actvities can not be captured

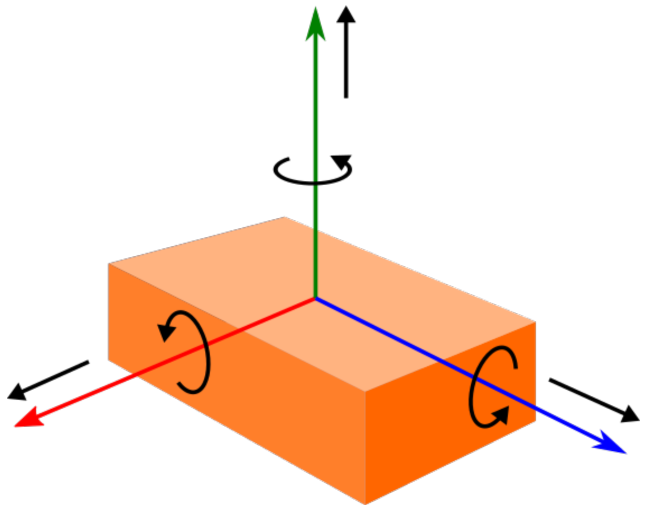
[Rhodin et al., 2016]



[Tome SelfPose et al., 2018]

IMU-based Motion Capture

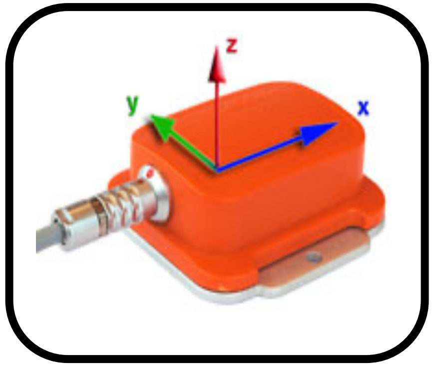
- IMU = Inertial Measurement Unit



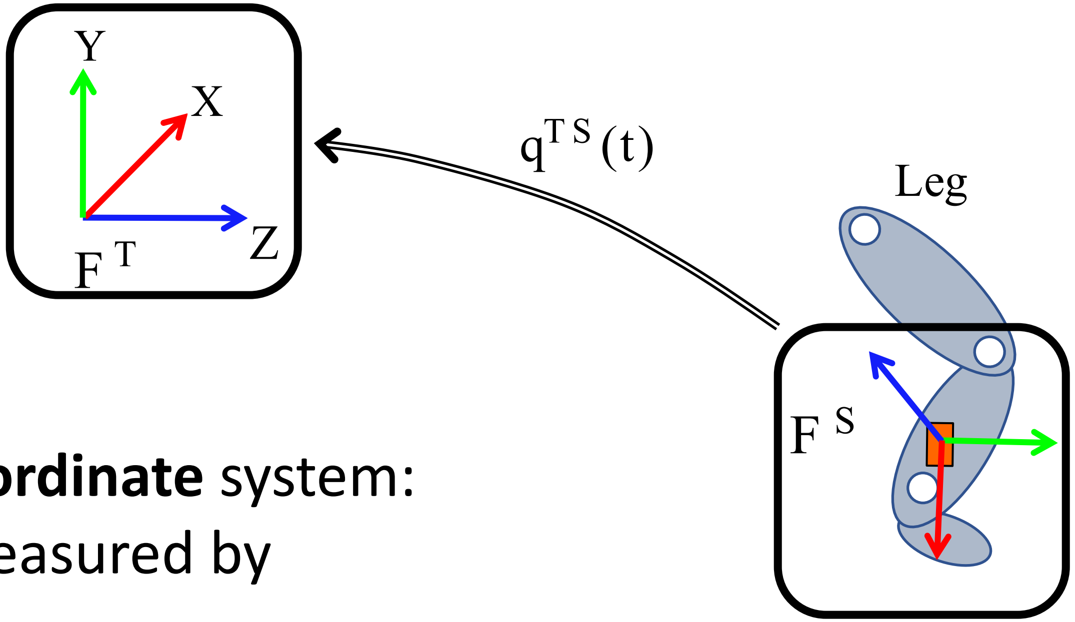
[Roetenberg et al., 2007]

[Vlasic et al., 2007]

Inertial sensors



Inertial Measurement Unit (IMU)



Global orientation w.r.t. **global coordinate** system:

- X is magnetic north direction measured by compass
- Y is the direction of gravity measured by accelerometer

Coordinate frames involved

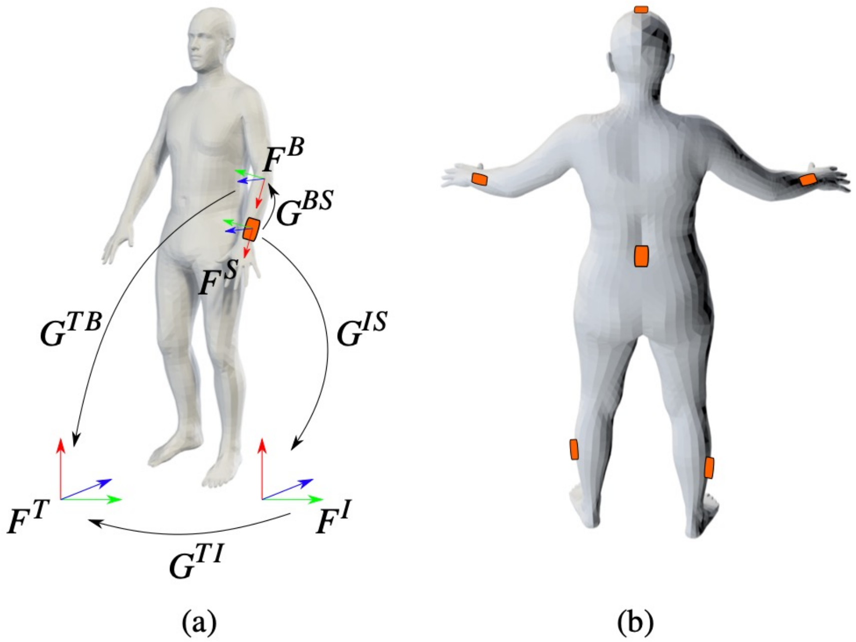
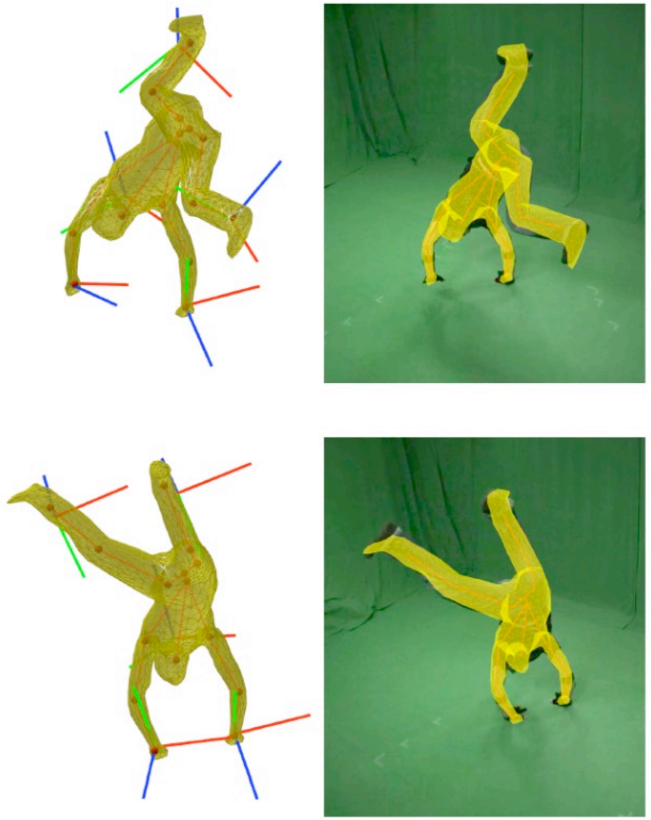
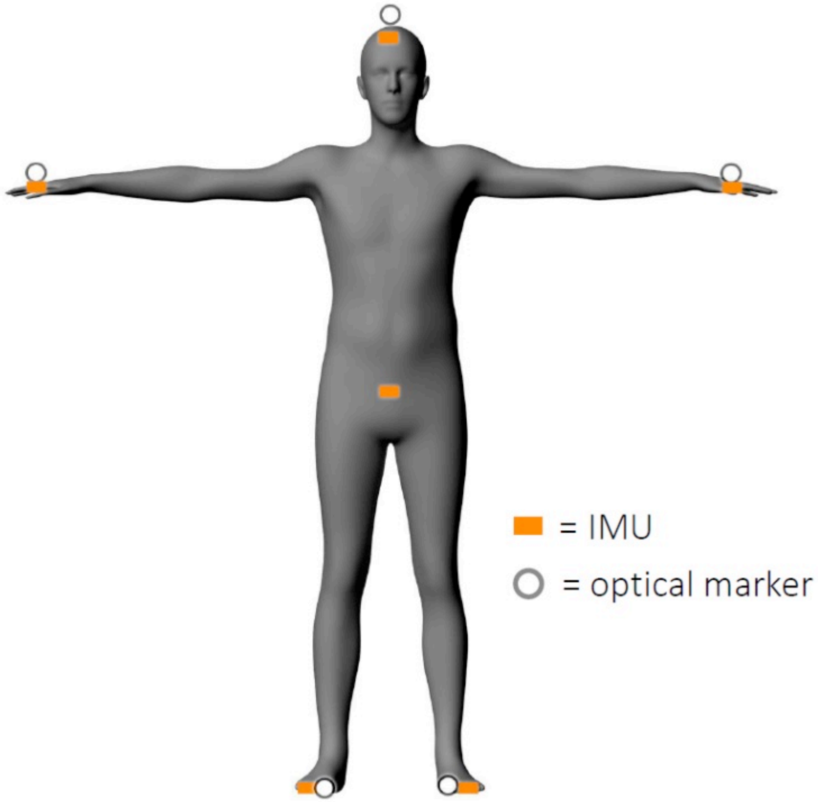


Figure 3: *(a) Coordinate frames: Global tracking coordinate frame F^G, Inertial coordinate frame F^I, Bone coordinate frame F^B and Sensor coordinate frame F^S. (b) Sensor placement at head, lower legs, wrists and back.*

Sparse IMUs + Vision



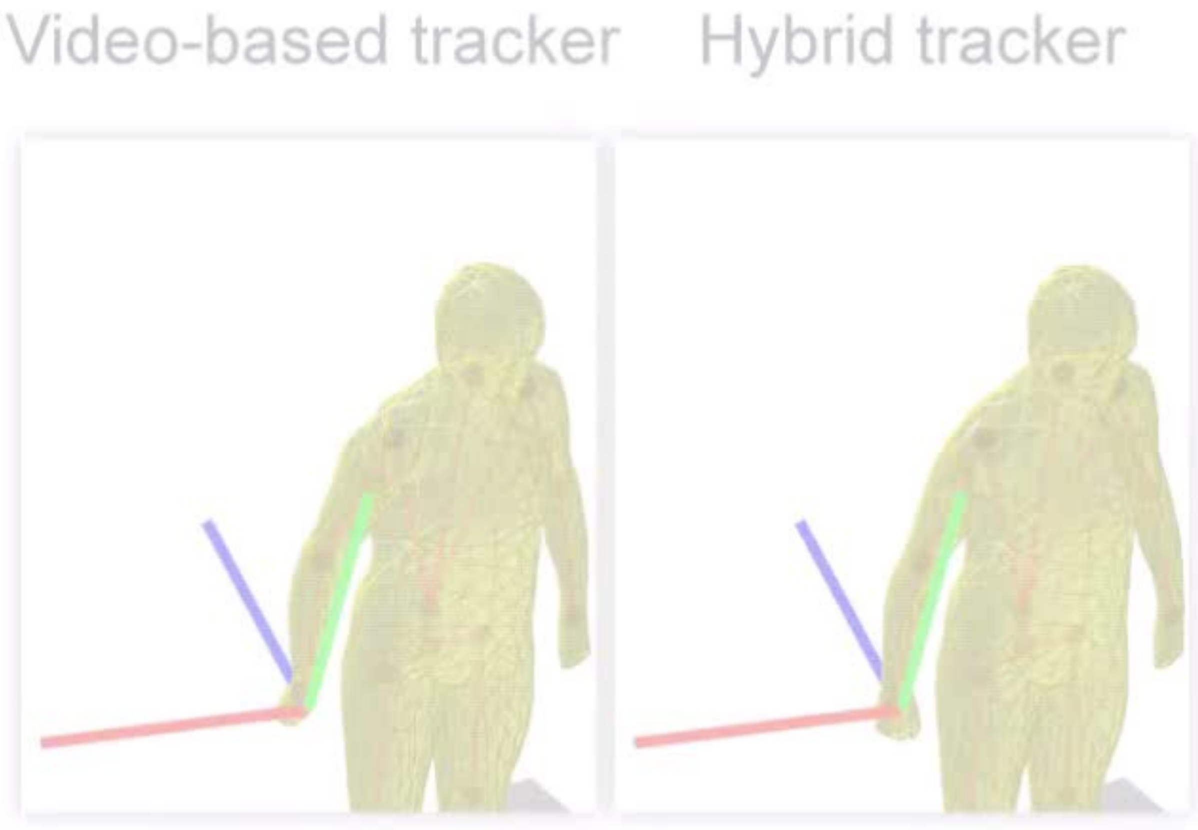
5 IMU + Video [Pons-Moll et al., 2010 & 2011]



[Andrews et al., 2016]

IMU+Video

First combination of IMU and vision for full body capture



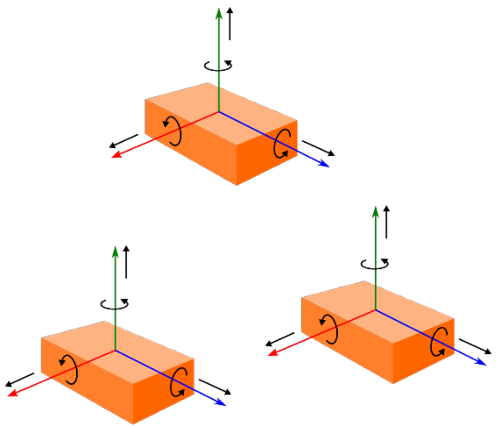
Key idea:
- Combine vision based (good localization of joints) with inertial tracking (good orientation of limbs).
- Compensate for drift with IMU

[Pons-Moll et al., 2010]

Related Work

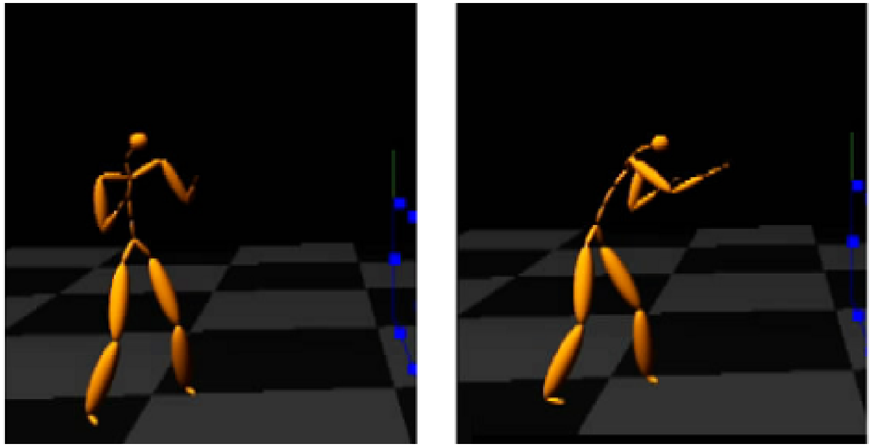
- Motion reconstruction with sparse IMUs

Sparse IMU data



[Slyper et al., 2008],
[Tautges et al., 2011],
[Schwarz et al., 2009]

3D motion



[Liu et al., 2011]

Our Approach: Sparse Inertial Poser

Analysis-by-Synthesis

6 IMUs

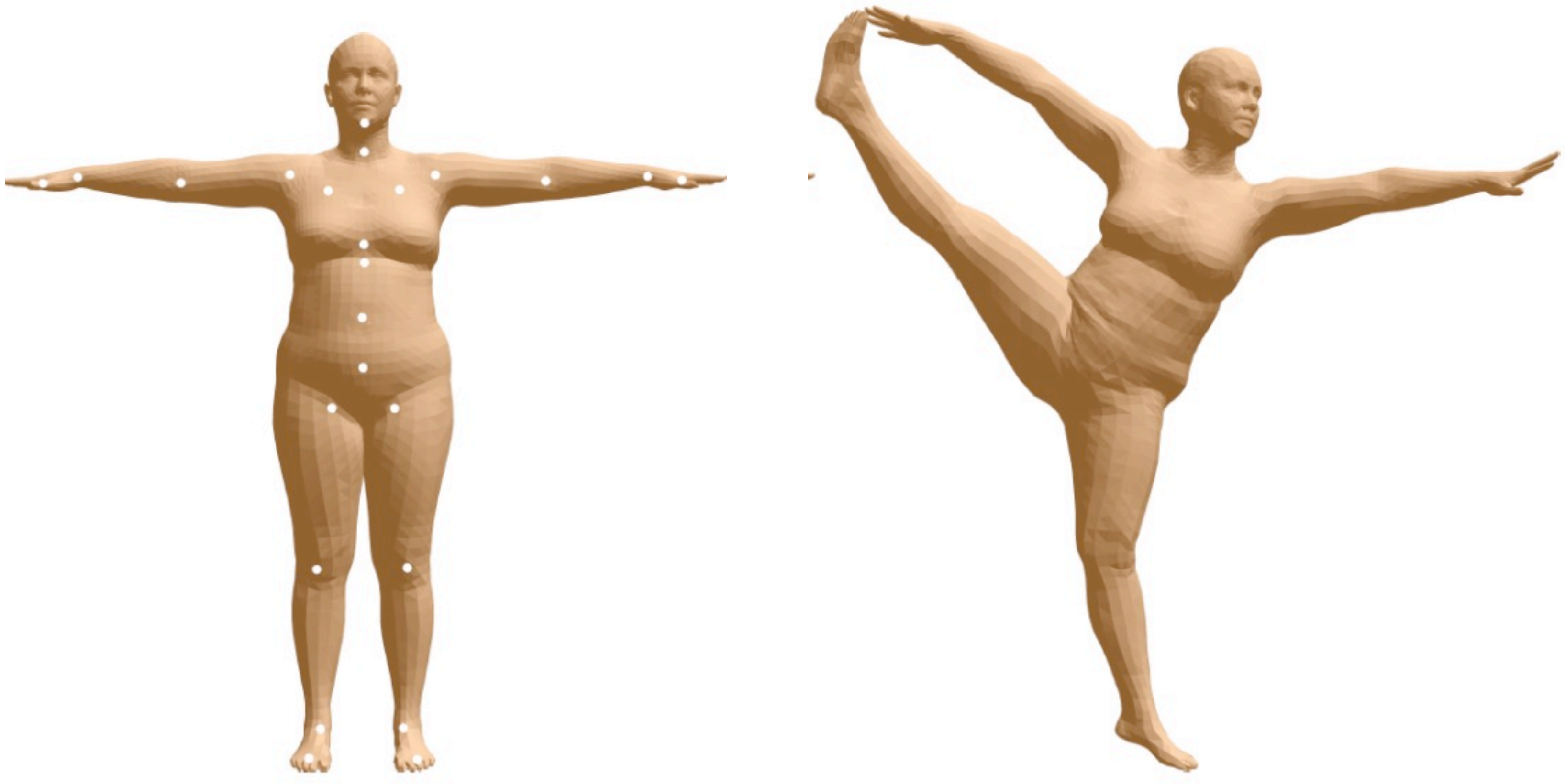
- $a \in \mathbb{R}^3$ acceleration
- $R \in SO(3)$ orientation

Sparse Inertial Poser

SMPL body model[1]

23 ball joints

$x \in \mathbb{R}^{75}$ pose



[1] SMPL: A skinned multi-person linear model, Loper et al., Siggraph Asia, 2015

Sparse Inertial Poser

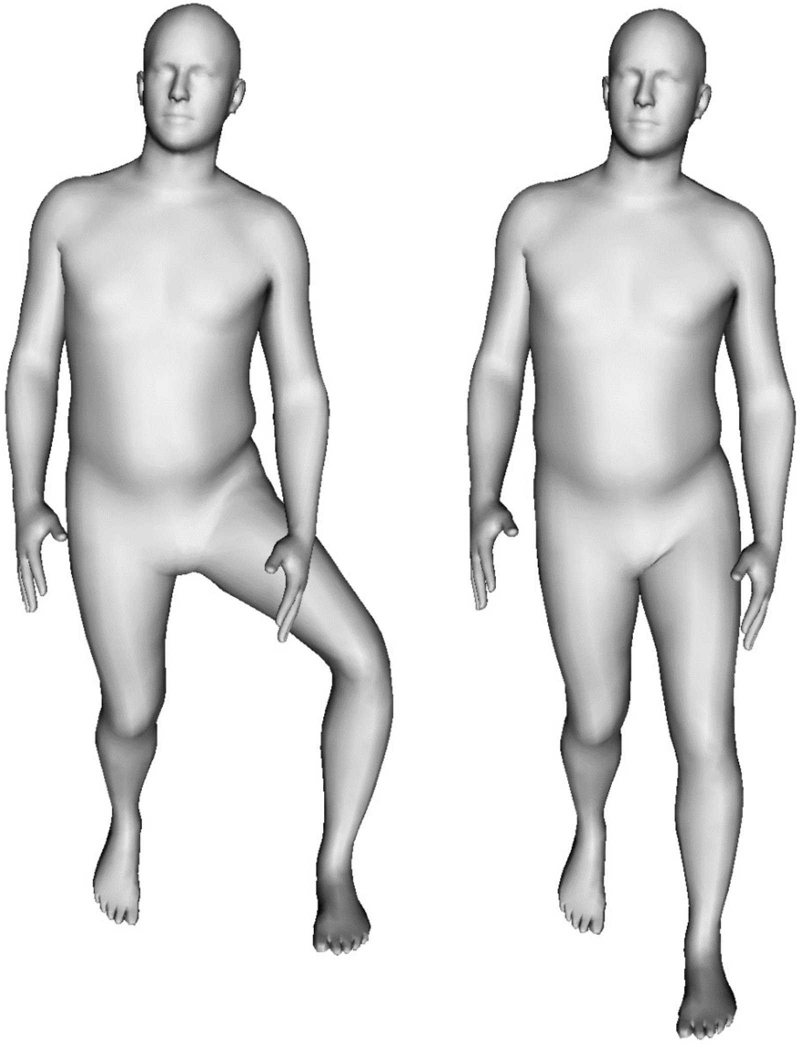
$$\arg\min_x \Big(\qquad , \qquad , \qquad \Big)$$

Orientation consistency

Acceleration consistency

Anthropometric consistency

Anthropometric Consistency



Objective

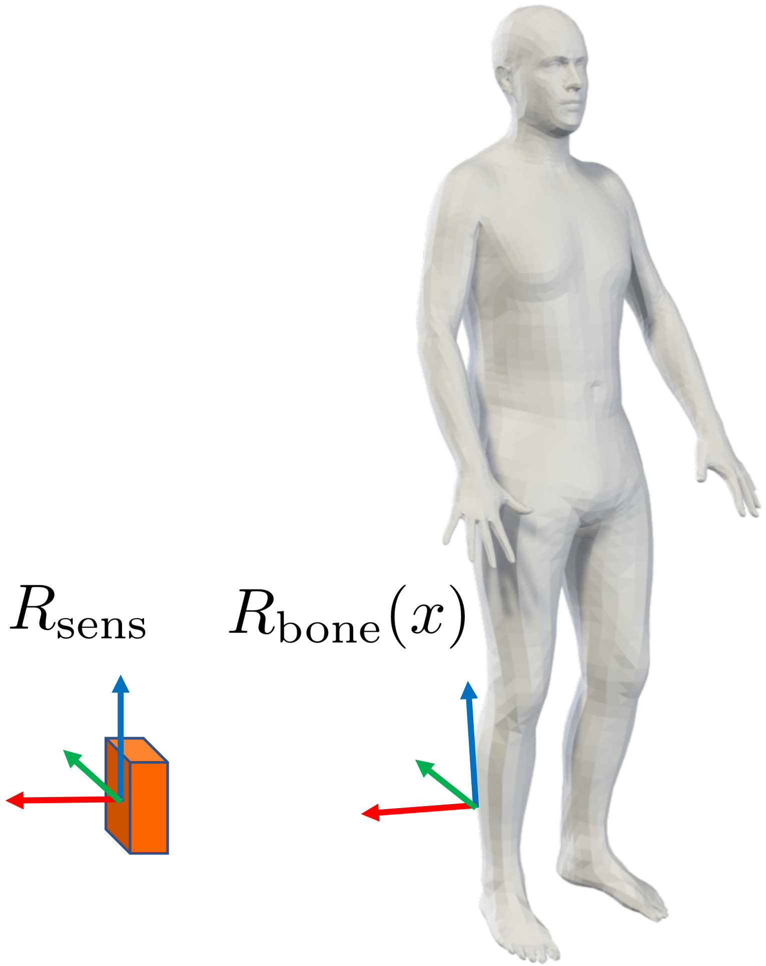
enforce human-like poses

$$\mathcal{N}(\mu_x, \Sigma_x)$$

$$d_{\text{mahal}}(x) = \sqrt{(x-\mu_x)^T \Sigma_x^{-1} (x-\mu_x)}$$

$$E_{\text{anthro}}(x) = d_{\text{mahal}}(x)^2 + ||e_{\text{limits}}(x)||^2$$

Orientation Consistency



Objective:

sensor & bone orientation consistency

How do you compute distance between orientations??

Distance metrics in SO(3)

Angular distance

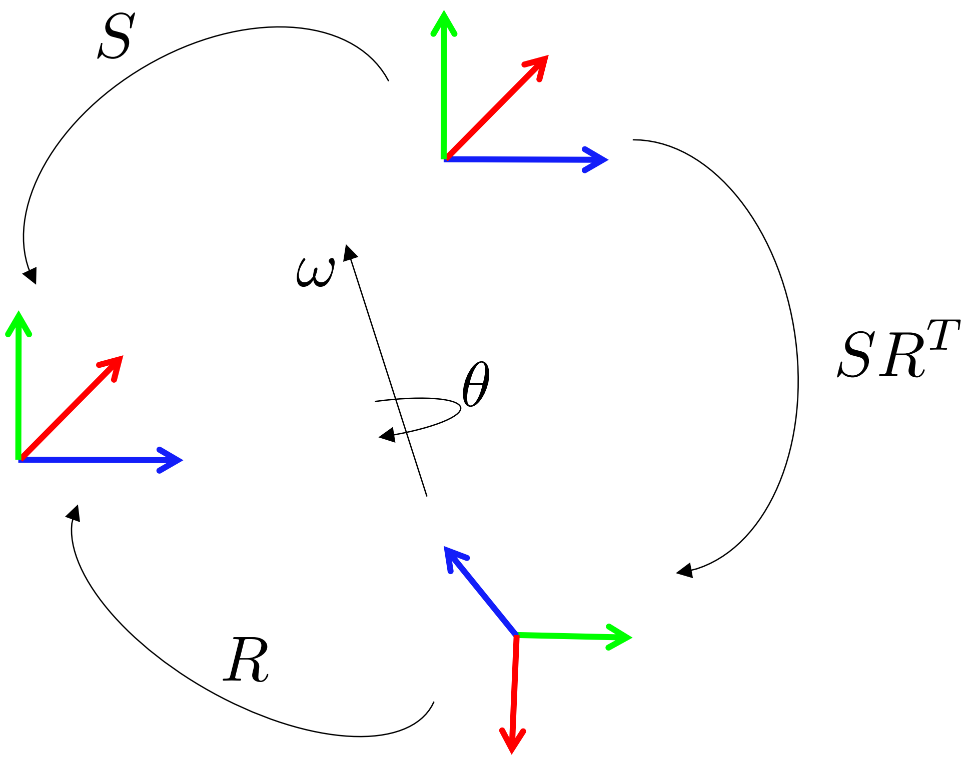
The anglular distance of two rotations R and S is the angle of SR^T:

$$\theta = d_\angle(\mathtt{S}, \mathtt{R}) = d_\angle(\mathtt{SR}^\top, \mathtt{I}) = \|\log(\mathtt{SR}^\top)\|_2$$

$$0 \leq \theta \leq \pi$$

The sign of ω must be chosen so that theta lies between 0 and pi.

Geodesic distrance in SO(3)



Distance metrics in SO(3)

Chordal distance

Frobenious norm of rotation matrix difference

$$
\begin{aligned}
d_{\text{chord}}(\mathtt{S},\mathtt{R})^2 &= \|\mathtt{S}-\mathtt{R}\|_{\mathrm{F}}^2 = \|\mathtt{S}\mathtt{R}^\top - \mathtt{I}\|_{\mathrm{F}}^2 \\
&= 2(\sin^2(\theta) + (1-\cos(\theta))^2) \\
&= 8\sin^2(\theta/2)
\end{aligned}
$$

$$
d_{\text{chord}}(\mathtt{S},\mathtt{R}) = 2\sqrt{2}\,\sin(\theta/2)
$$

Relation to angular distance

Properties:

- 1) $\exp(\theta\widehat{\omega}) = I + \widehat{\omega}\sin(\theta) + \widehat{\omega}^2(1-\cos(\theta))$
- 2) $\widehat{\omega}$ and $\widehat{\omega}^2$ are orthogonal under Frobenious norm
- 3) $\|\widehat{\omega}\|_F^2 = \|\widehat{\omega}^2\|_F^2 = 2$

Distance metrics in SO(3)

Quaternion distance

$$d_{\text{quat}}(\mathtt{S}, \mathtt{R}) = \min\{\|\mathbf{s} - \mathbf{r}\|_2, \|\mathbf{s} + \mathbf{r}\|_2\}$$

Recall that quaternions q and –q represent the same rotation. This ambiguity is resolved by taking the min.

Distance metrics in SO(3)

Quaternion distance

$$d_{\text{quat}}(\mathtt{S},\mathtt{R}) = \min\{\|\mathbf{s}-\mathbf{r}\|_2, \|\mathbf{s}+\mathbf{r}\|_2\}$$

Recall that quaternions q and –q represent the same rotation. This ambiguity is resolved by taking the min.

The relationship to the angular distance:

$$\mathbf{e} = (1,0,0,0) \qquad \mathbf{s}\cdot\mathbf{r}^{-1} = (\cos(\theta/2), \hat{\mathbf{v}}\sin(\theta/2))$$

$$\langle \mathbf{e}, \mathbf{s}\cdot\mathbf{r}^{-1}\rangle = \cos(\theta/2)$$

- Inner product of two 4-vectors equals $\cos(\alpha)$
- Hence $\alpha = \theta/2$

Distance metrics in SO(3)

Quaternion distance

$$d_{\text{quat}}(\mathtt{S}, \mathtt{R}) = \min\{\|\mathbf{s} - \mathbf{r}\|_2, \|\mathbf{s} + \mathbf{r}\|_2\}$$

Recall that quaternions q and –q represent the same rotation. This ambiguity is resolved by taking the min.

The relationship to the angular distance:

$$\mathbf{e} = (1, 0, 0, 0) \qquad \mathbf{s} \cdot \mathbf{r}^{-1} = (\cos(\theta/2), \hat{\mathbf{v}}\sin(\theta/2))$$

$$\left\langle \mathbf{e}, \mathbf{s} \cdot \mathbf{r}^{-1} \right\rangle = \cos(\theta/2)$$

- Inner product of two 4-vectors equals $\cos(\alpha)$
- Hence $\alpha = \theta/2$

$$\|\mathbf{s} \cdot \mathbf{r}^{-1} - \mathbf{e}\|_2 = \|\mathbf{s} - \mathbf{r}\|_2$$

- The distance of two unit vectors separated by α is $2\sin(\alpha/2)$

$$d_{\text{quat}}(\mathtt{S}, \mathtt{R}) = 2\sin(\alpha/2) = 2\sin(\theta/4)$$

Distance metrics for SO(3)

Angular distance

$$\theta = \quad d_\angle(\mathtt{S},\mathtt{R}) = d_\angle(\mathtt{S}\mathtt{R}^\top,\mathtt{I}) = \|\log(\mathtt{S}\mathtt{R}^\top)\|_2$$

$$0 \le \theta \le \pi$$

Chordal distance

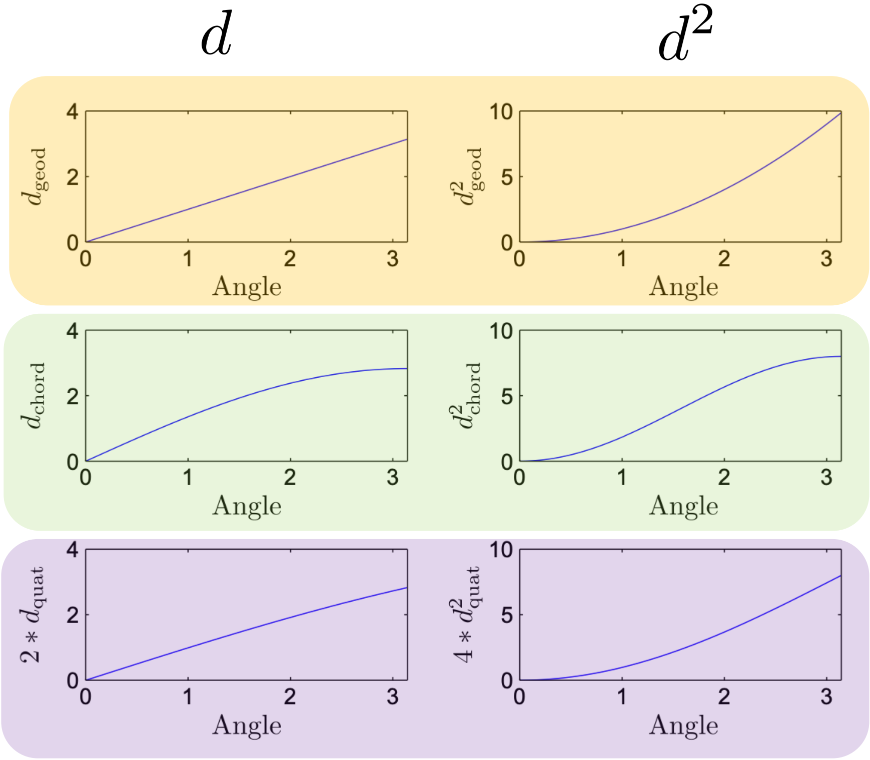
$$d_{\text{chord}}(\mathtt{S},\mathtt{R})^2 = \|\mathtt{S}-\mathtt{R}\|_{\text{F}}^2 = \|\mathtt{S}\mathtt{R}^\top - \mathtt{I}\|_{\text{F}}^2$$

$$d_{\text{chord}}(\mathtt{S},\mathtt{R}) = 2\sqrt{2}\,\sin(\theta/2)$$

Quaternion distance

$$d_{\text{quat}}(\mathtt{S},\mathtt{R}) = \min\{\|\mathbf{s}-\mathbf{r}\|_2, \|\mathbf{s}+\mathbf{r}\|_2\}$$

$$d_{\text{quat}}(\mathtt{S},\mathtt{R}) = 2\sin(\alpha/2) = 2\sin(\theta/4)$$



Distance in angle-axis space

- Euclidean distance between corresponding scaled axis angles of log(R) and log(S). This metric is **not continuous**!!
- If log(R) is taken to be the smallest length vector, rotations about angles near π about opposite axes are not close to each other under this metric (but they are in the angular distance metric)

See Hartley et al. IJCV'13

Distance in angle-axis space

- Euclidean distance between corresponding scaled axis angles of log(R) and log(S). This metric is **not continuous**!!
- If log(R) is taken to be the smallest length vector, rotations about angles near π about opposite axes are not close to each other under this metric (but they are in the angular distance metric)
- **Solution** take the min over all choices of vectors

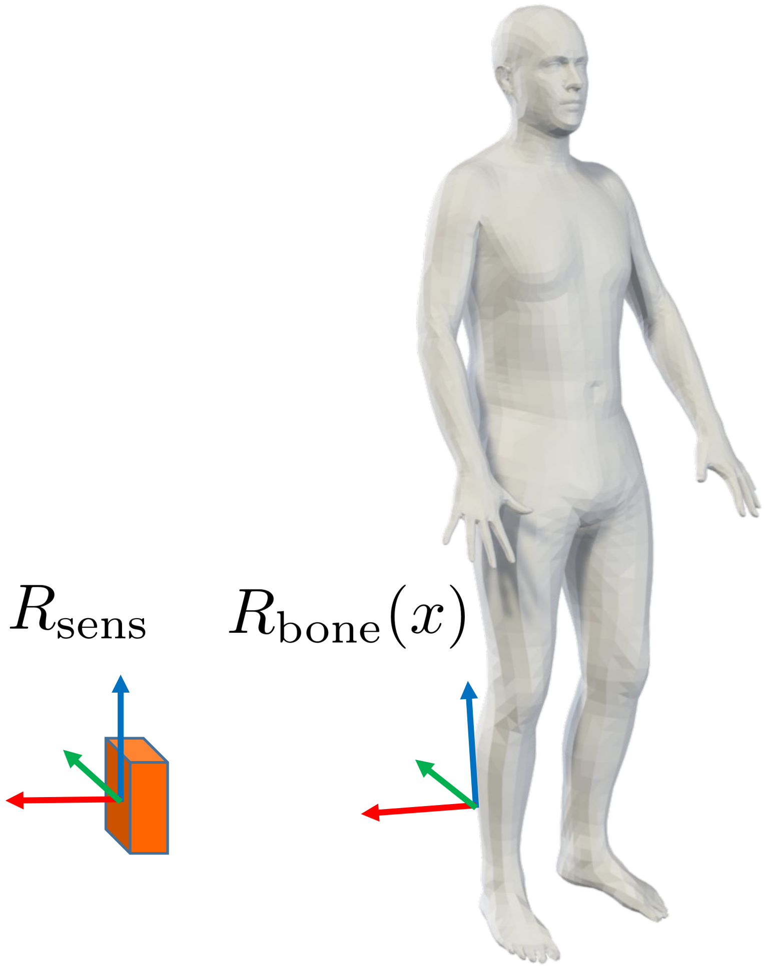
$$d_{\log}(\mathtt{S}, \mathtt{R}) = \min \|\mathbf{v}_r - \mathbf{v}_s\|_2$$

where the minimum is taken over all choices of vectors $\mathbf{v}_r$ and $\mathbf{v}_s$ such that $\exp[\mathbf{v}_r]_\times = \mathtt{R}$ and $\exp[\mathbf{v}_s]_\times = \mathtt{S}$.

Many regressors to human pose use this metric, but do not take the min, which can be a problem!!

See Hartley et al. IJCV'13

Orientation Consistency

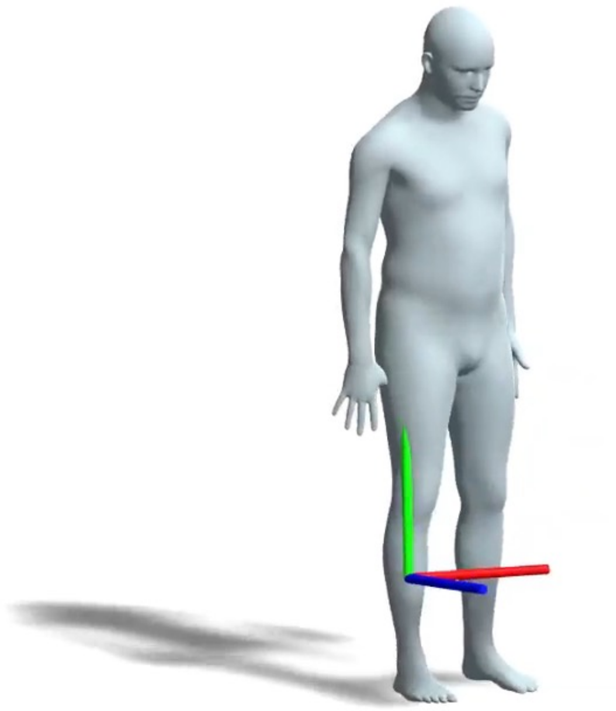


Objective

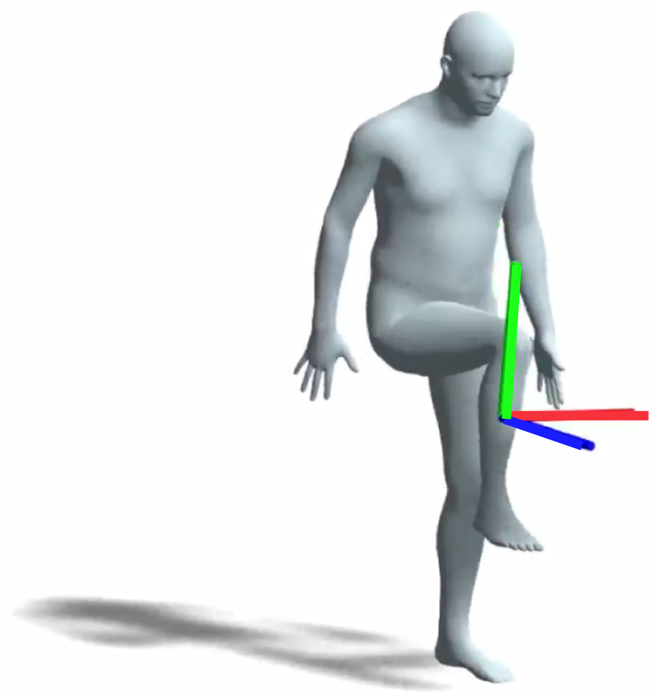
sensor & bone orientation consistency

$$e_{\text{ori}}(x, R_{\text{sens}}) = \log\left(R_{\text{bone}}(x)\,(R_{\text{sens}})^{-1}\right)$$

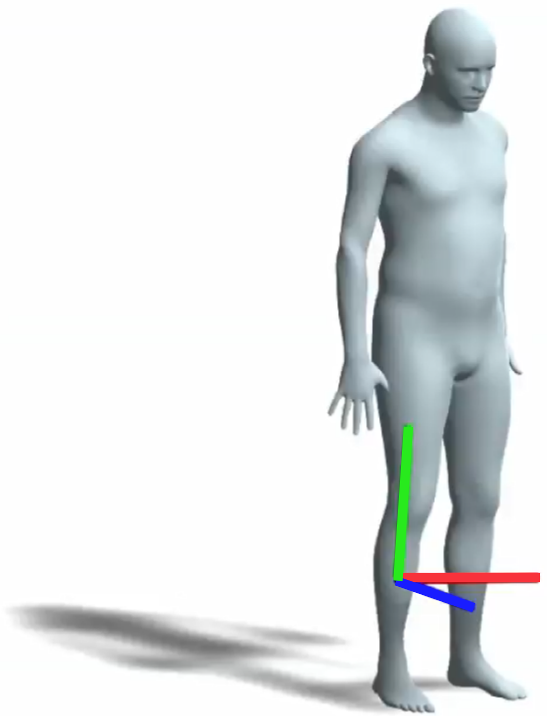
Underconstrained Pose Space



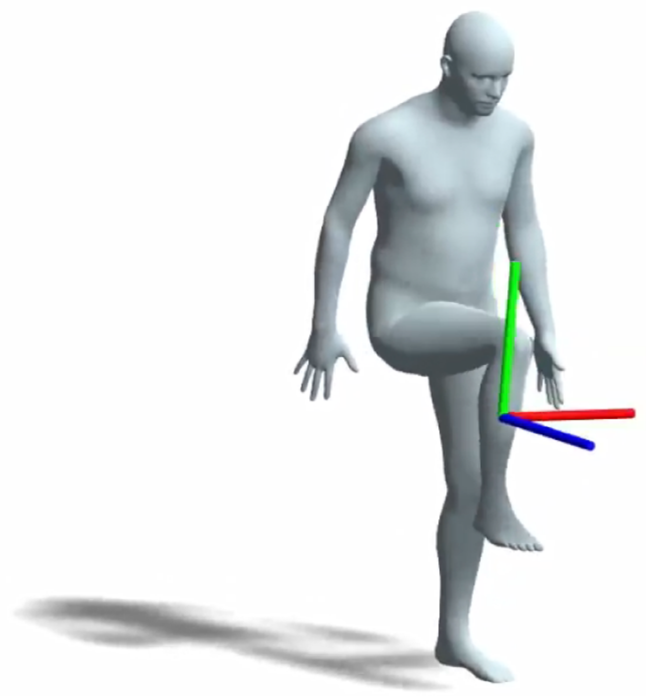
Underconstrained Pose Space



pose
different

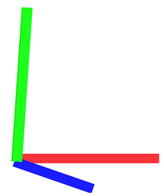


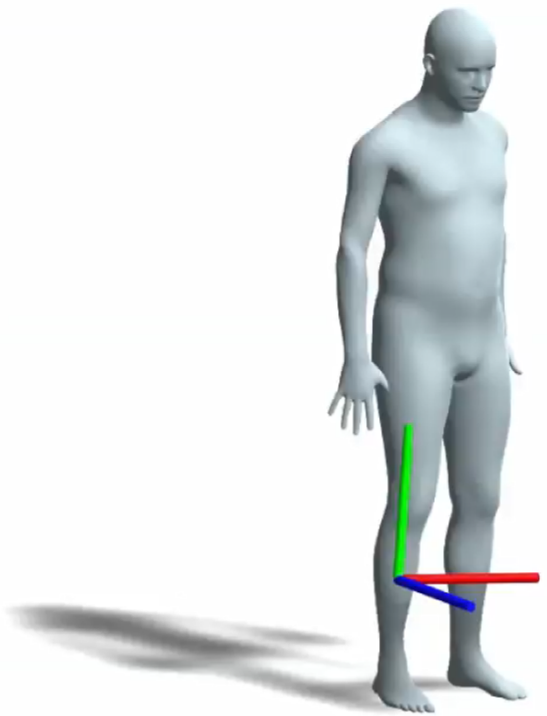
Underconstrained Pose Space



pose
different

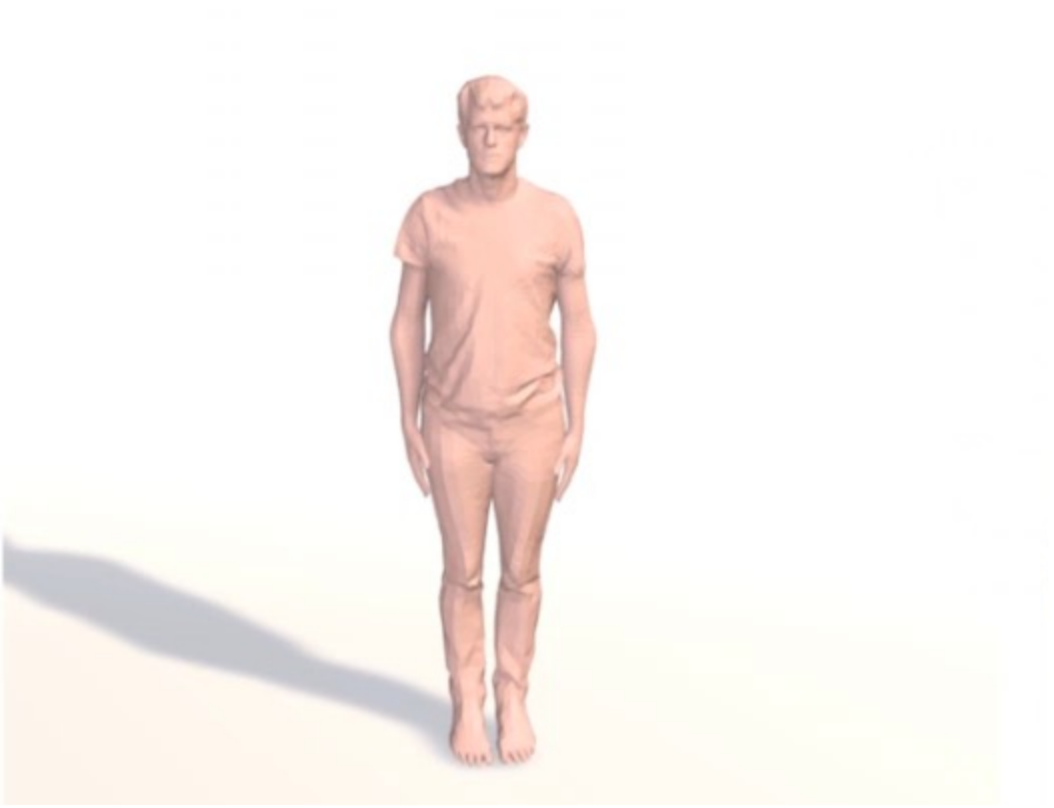
orientation
measurement
nearly identical



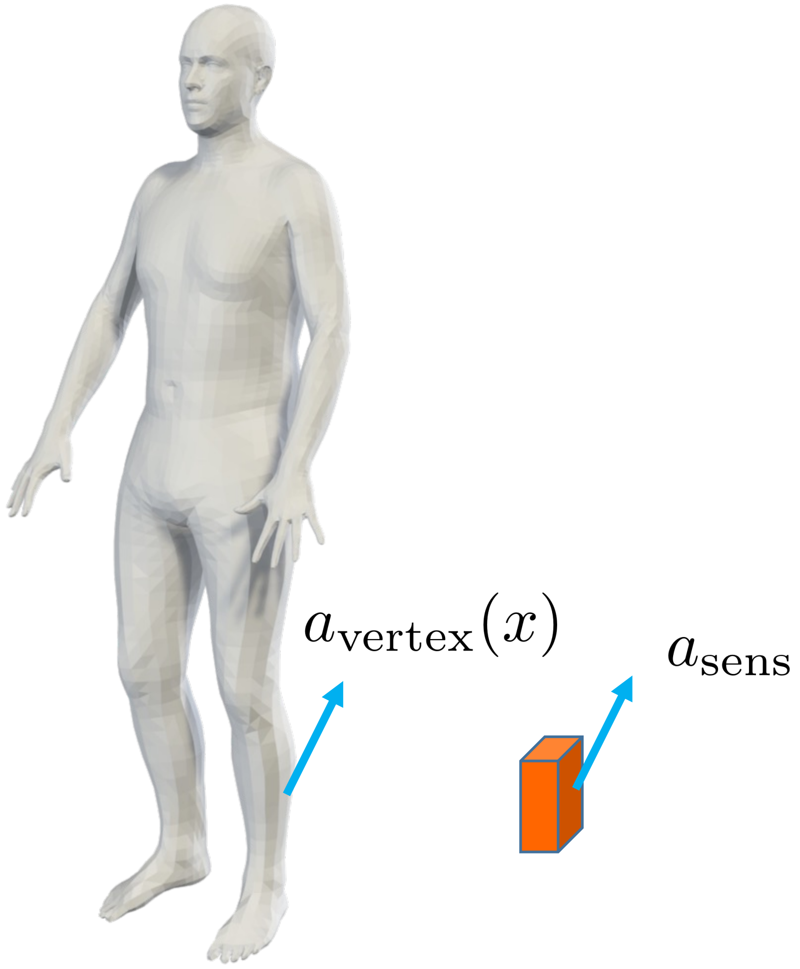


Orientation Consistency

Sparse Orientation Poser (SOP)



Acceleration Consistency



Objective

sensor & vertex acceleration consistency

$$e_{\text{acc}}(x, a_{\text{sens}}) = a_{\text{vertex}}(x) - a_{\text{sens}}$$

$$\hat{\mathbf{a}}_t^G = \frac{\mathbf{p}_{t-1}^G - 2\mathbf{p}_t^G + \mathbf{p}_{t+1}^G}{dt^2}$$

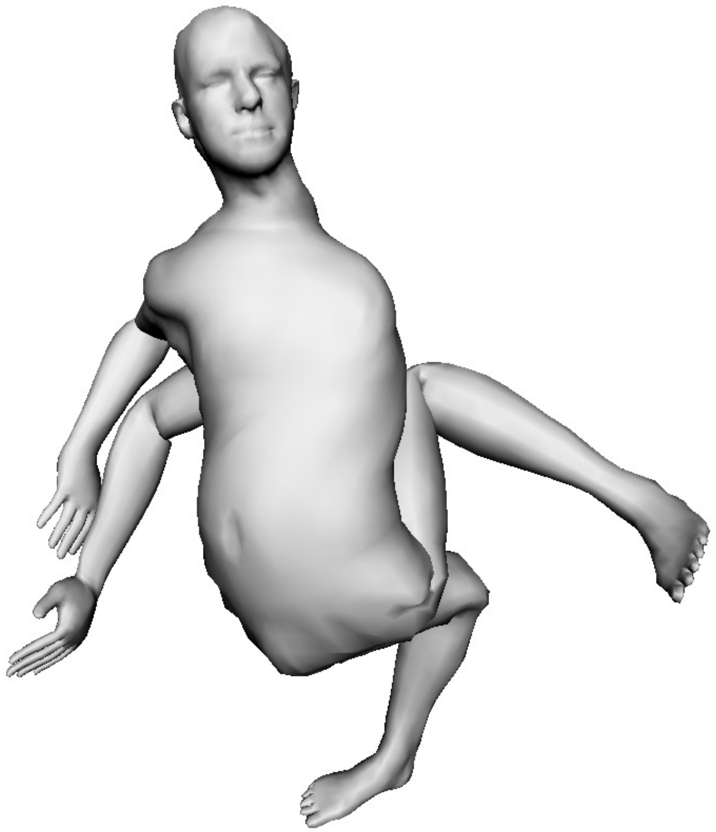
1) Use finite differences

$$\mathbf{a}_t^G = \mathbf{R}_t^{GS}\mathbf{a}_t^S - \mathbf{g}^G.$$

1) Transform from local to global
2) Subtract gravity

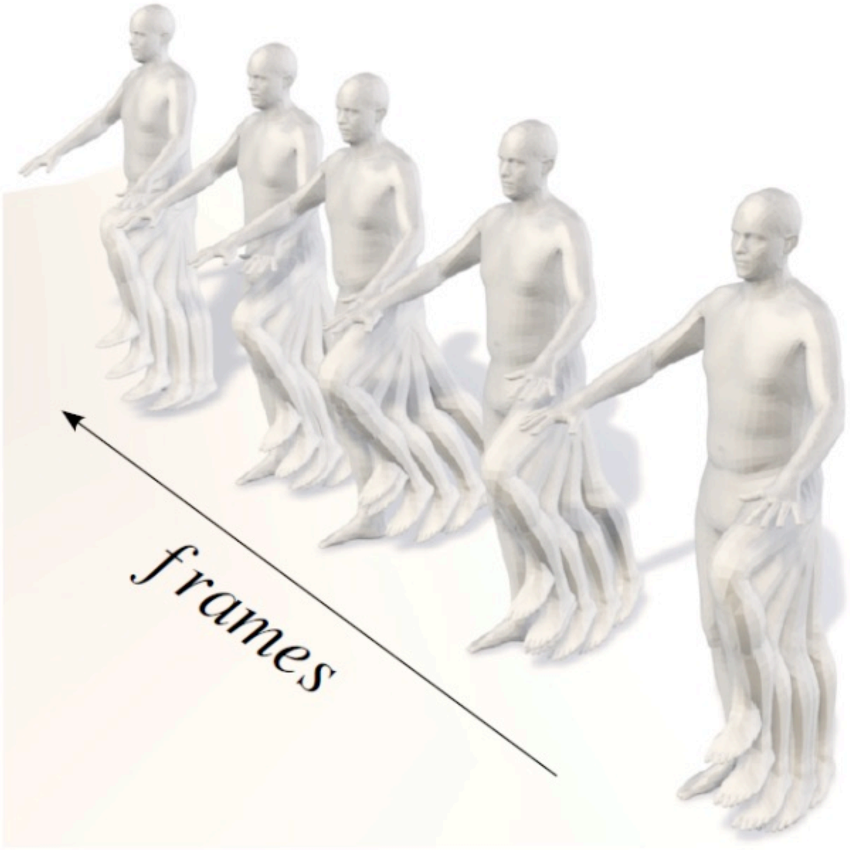
Acceleration Consistency

Sparse Acceleration Poser

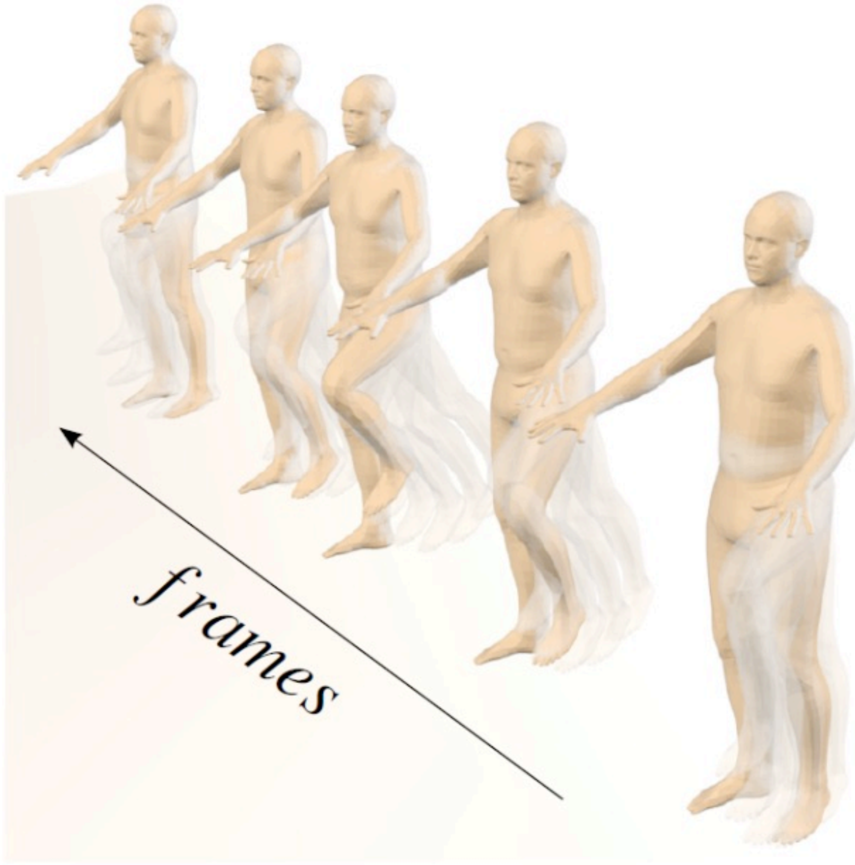


Key Observation I

Orientation only



Orientation + Acceleration



Key Observation II

Statistical body model (SMPL)

anthropometric constraints
realistic motion synthesis

Multi-Frame Optimization

$$x^*_{1:T} = \arg\min_{x_{1:T}} E_{\mathrm{motion}}(x_{1:T}, R_{1:T}, a_{1:T}) \qquad x_{1:T} \in \mathbb{R}^{75T}$$

$$E_{\mathrm{motion}}(x_{1:T}, R_{1:T}, a_{1:T}) =$$

$$w_{\mathrm{ori}} \cdot \sum_{t=1}^{T} \sum_{n=1}^{6} ||e_{\mathrm{ori}}(x_t, R_t^n)||^2$$ Orientation consistency

$$+w_{\mathrm{acc}} \cdot \sum_{t=1}^{T} \sum_{n=1}^{6} ||e_{\mathrm{acc}}(x_t, a_t^n)||^2$$ Acceleration consistency

$$+w_{\mathrm{anthro}} \cdot \sum_{t=1}^{T} E_{\mathrm{anthro}}(x_t)$$ Anthropometric consistency

Optimization

$$\mathbf{e}(\mathbf{x}, \delta\mathbf{x}) \approx \mathbf{e}(\mathbf{x}) + \mathbf{J}\delta\mathbf{x}.$$

To minimize $\mathbf{e}^T\mathbf{e}$, linearize the vector of residuals with Jacobian matrix

For example, the acceleration residuals linearized take the form:

$$\mathbf{e}_{acc}(t, \delta\mathbf{x}) \approx \mathbf{e}_{acc}(t) + \begin{bmatrix} \mathbf{J}_{p(\mathbf{x}_{t-1})} & -2\mathbf{J}_{p(\mathbf{x}_t)} & \mathbf{J}_{p(\mathbf{x}_{t+1})} \end{bmatrix} \begin{bmatrix} \delta\mathbf{x}_{t-1} \\ \delta\mathbf{x}_t \\ \delta\mathbf{x}_{t+1} \end{bmatrix}$$

The Jacobian matrix above, maps increments in parameter space to increments in vertex position where the sensor is placed

Batch Optimization over frames

Question: If the error residual for 1 frame is N, and the number of pose parameters is P, how large is the Jacobian for the residuals for all frames?

Batch Optimization over frames

Question: If the error residual for 1 frame is N, and the number of pose parameters is P, how large is the Jacobian for the residuals for all frames?

Expensive in general!! → Exploit the **block diagonal** structure

$$\begin{bmatrix} \ddots & & & & \\ & \mathbf{J}_{t-1} & & & \\ & & \mathbf{J}_t & & \\ & & & \mathbf{J}_{t+1} & \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ \delta\mathbf{x}_{t-1} \\ \delta\mathbf{x}_t \\ \delta\mathbf{x}_{t+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \mathbf{e}(t-1) \\ \mathbf{e}(t) \\ \mathbf{e}(t+1) \\ \vdots \end{bmatrix}$$

Orientation+anthopometric residual equations

$$\begin{bmatrix} \ddots & \ddots & & & \\ \ddots & -2\mathbf{J}_{t-1} & \mathbf{J}_t & & \\ & \mathbf{J}_{t-1} & -2\mathbf{J}_t & \mathbf{J}_{t+1} & \\ & & \mathbf{J}_t & -2\mathbf{J}_{t+1} & \ddots \\ & & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ \delta\mathbf{x}_{t-1} \\ \delta\mathbf{x}_t \\ \delta\mathbf{x}_{t+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \mathbf{e}_{\text{acc}}(t-1) \\ \mathbf{e}_{\text{acc}}(t) \\ \mathbf{e}_{\text{acc}}(t+1) \\ \vdots \end{bmatrix}$$

Acceleration term residual equations

Results

Results

Evaluation

Sparse Orientation Poser vs. Sparse Inertial Poser

Quantitative Evaluation

TNT15 dataset [1]

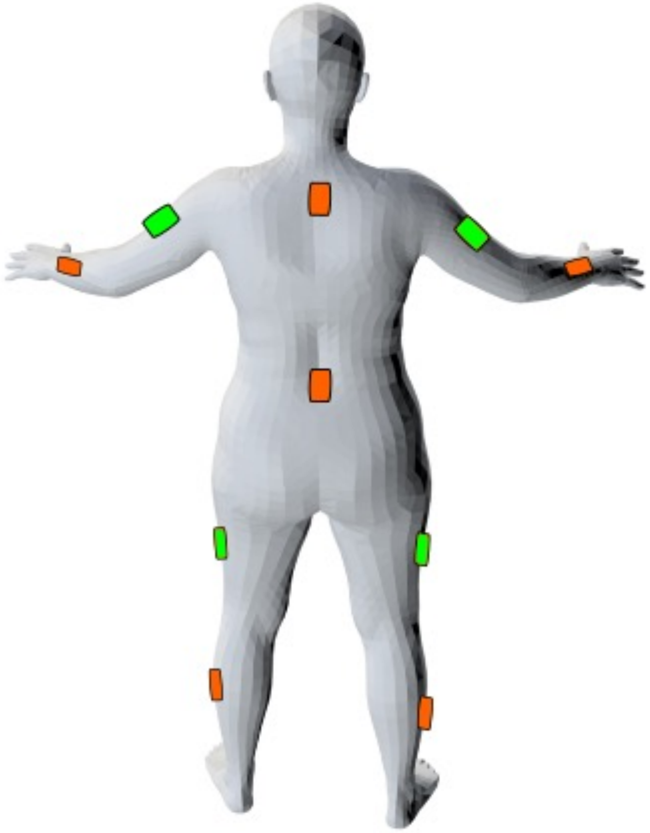
4 actors

5 activities

8 synchronized RGB-cameras

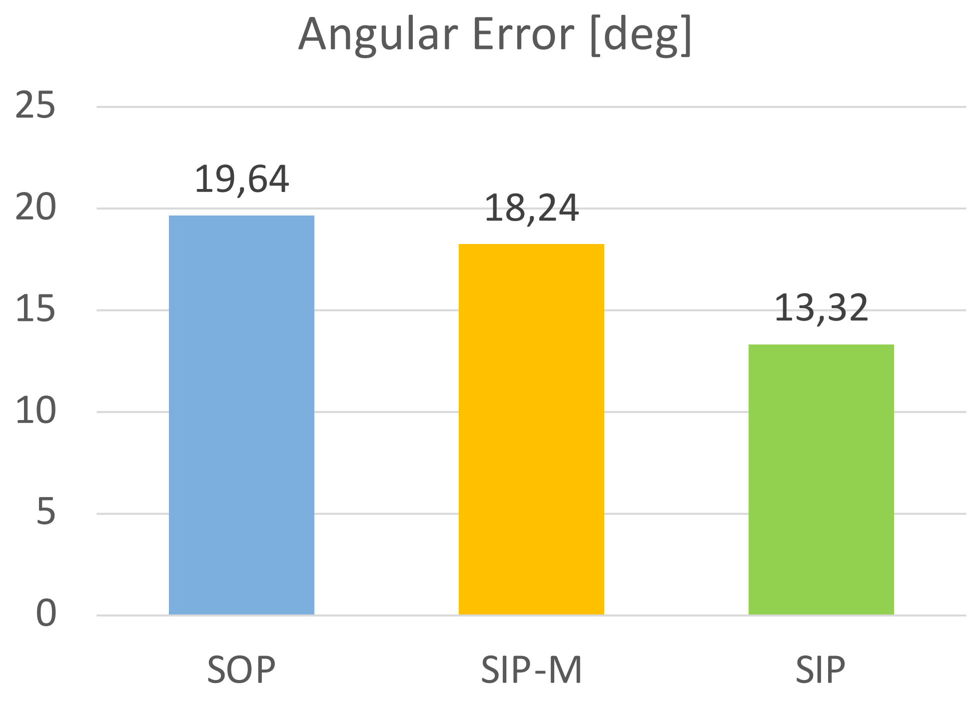
10 IMUs

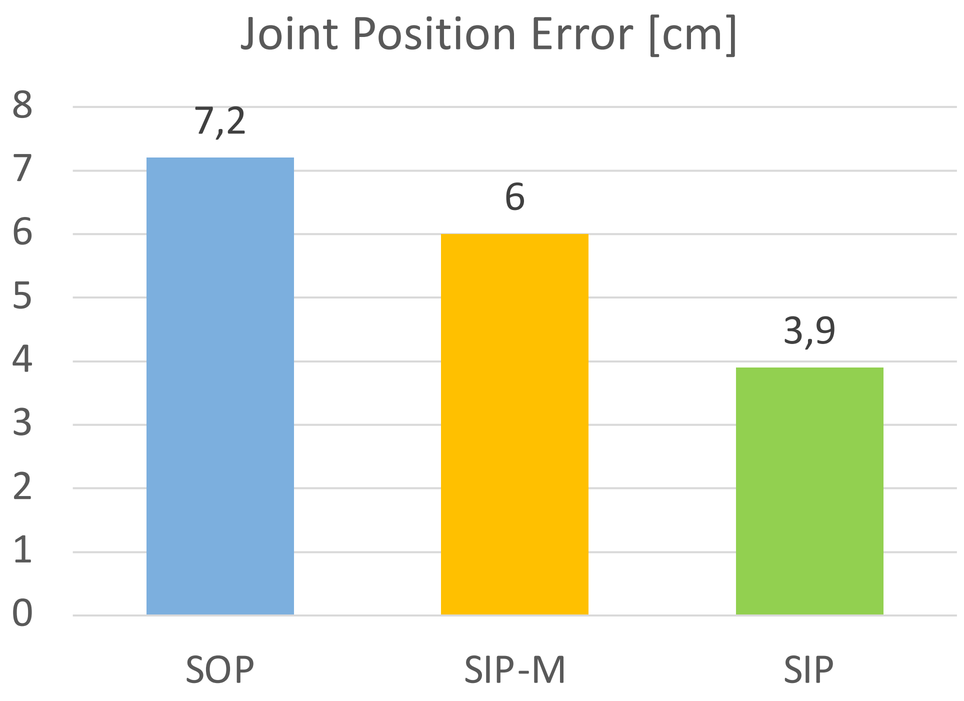
6 IMUs for tracking, 4 IMUs for validation



[1] Human pose estimation from video and IMUs, v. Marcard et al., TPAMI, 2016

Quantitative Evaluation





Limitations & Future Work

Hand and feet not tracked

Drift in global translation

Requires a laser scan

Offline approach

Conclusions

Sparse Inertial Poser

works with only 6 IMUs

reconstructs arbitrary motions

enables motion tracking in the wild

Conclusions

Sparse Inertial Poser

works with only 6 IMUs

reconstructs arbitrary motions

enables motion tracking in the wild

Computation Times

1000 frames sequence

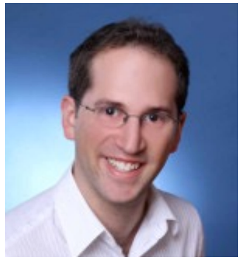
20 Levenberg-Marquardt iterations

Intel Core i7 3.5 GHz CPU

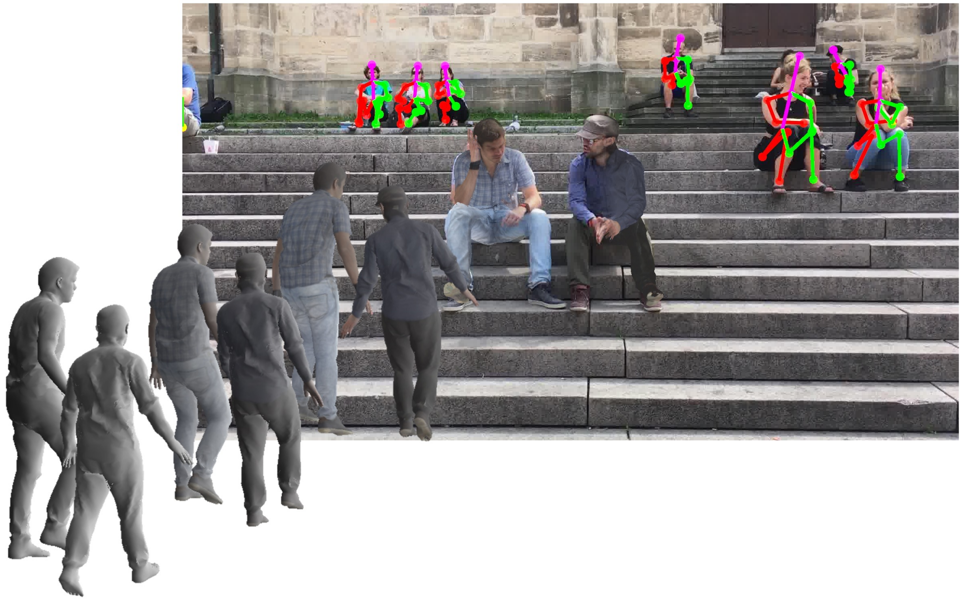
Single-core MATLAB code

Overall computation time	**7.5 min**
Model update	14.4s/iteration
Setting up Jacobians	3.3s/iteration
Solving for an update-step	1.5s/iteration

Recovering Accurate 3D Human Pose in the Wild Using IMUs and a Moving Camera



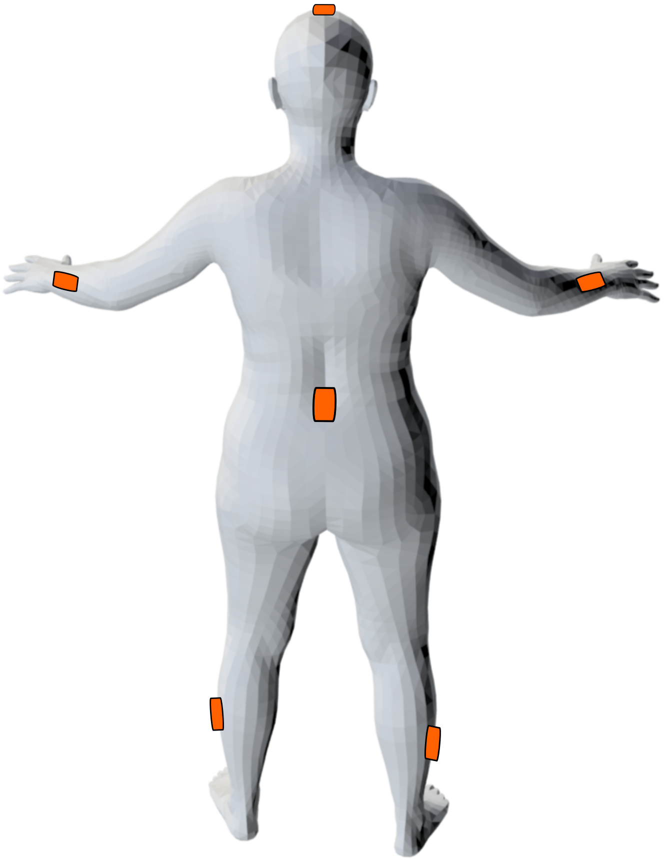
T. von Marcard R. Henschel M. Black B. Rosenhahn G. Pons-Moll



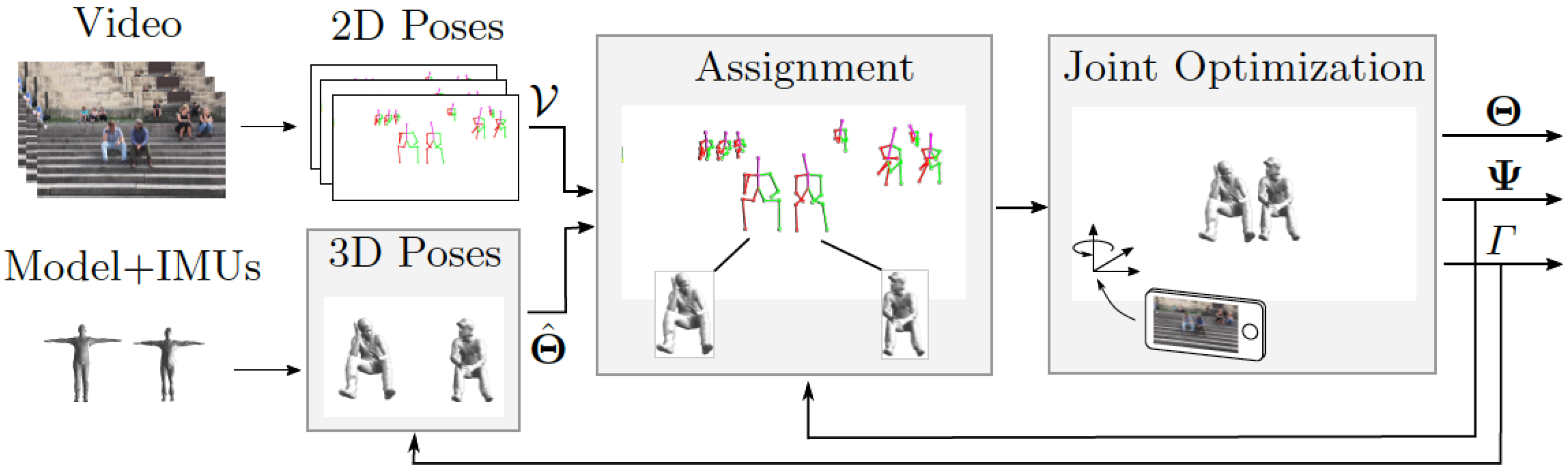
ECCV'18

3DPW: 3D Poses in the Wild

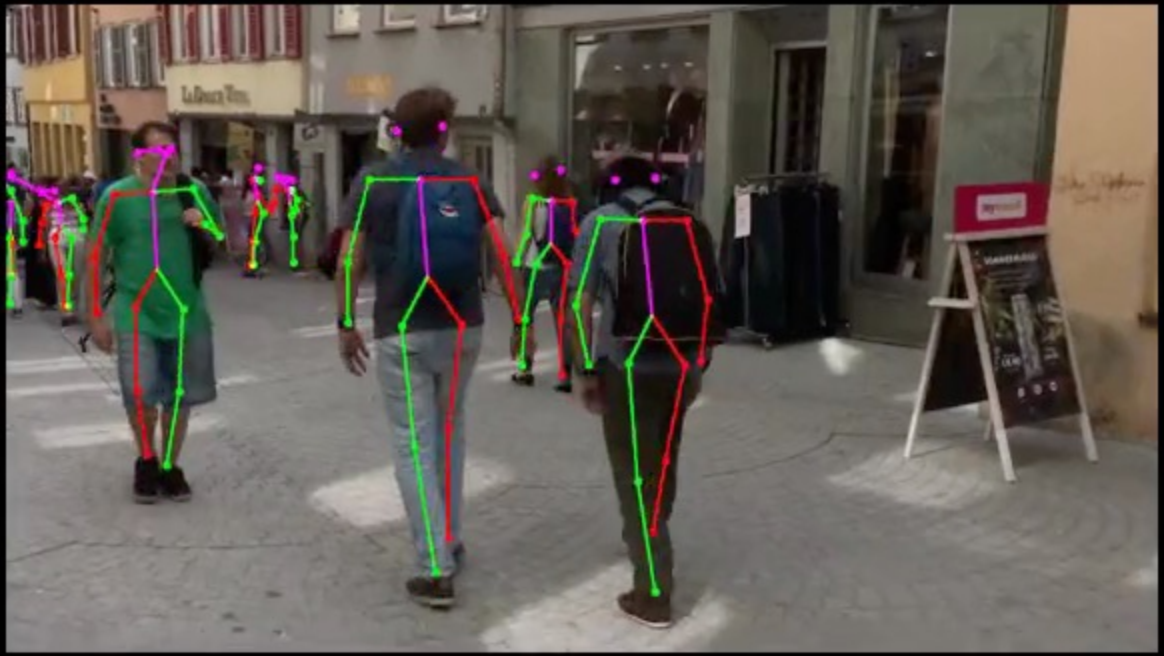
+



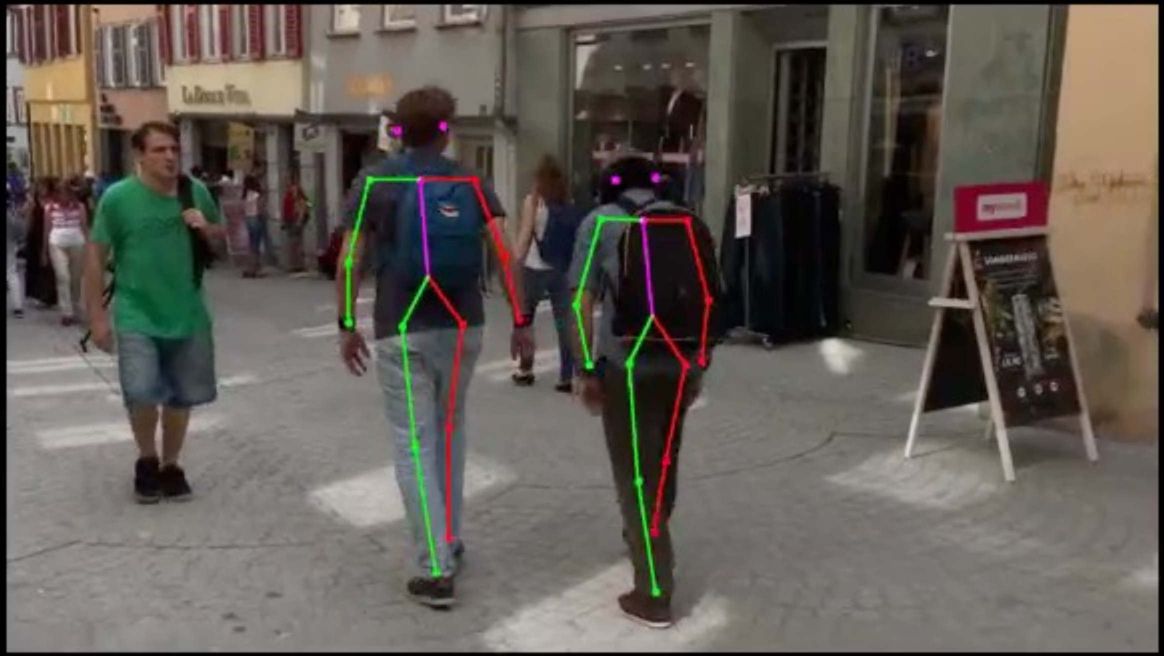
A single moving camera and IMUs on the person



Person Identification

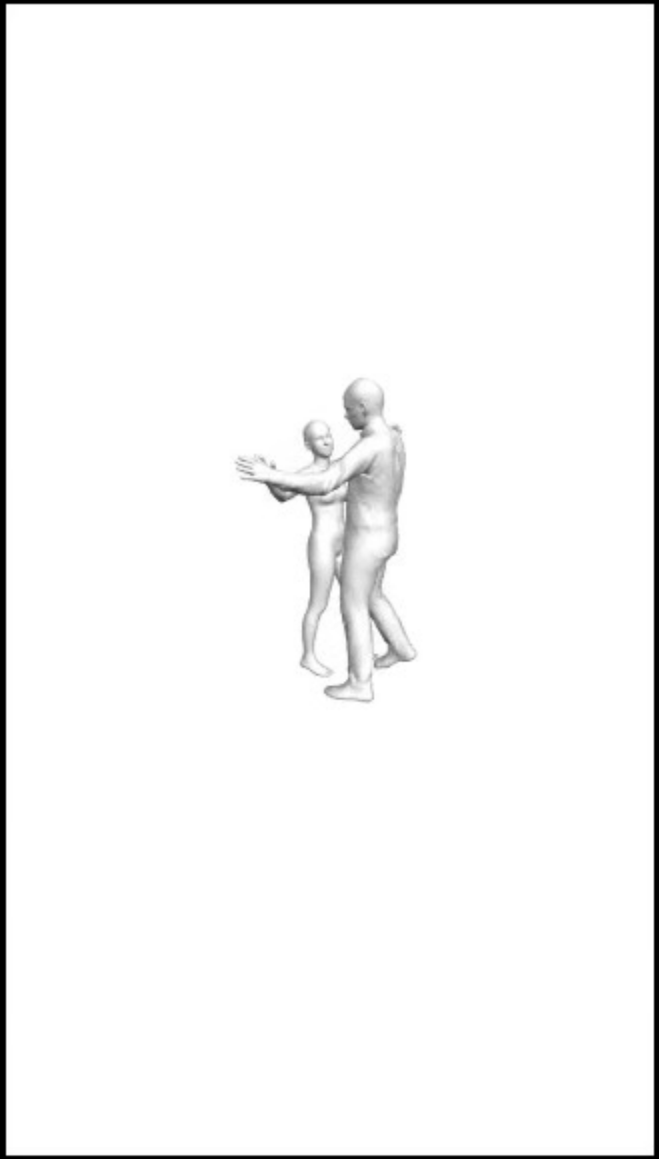


All 2D Poses



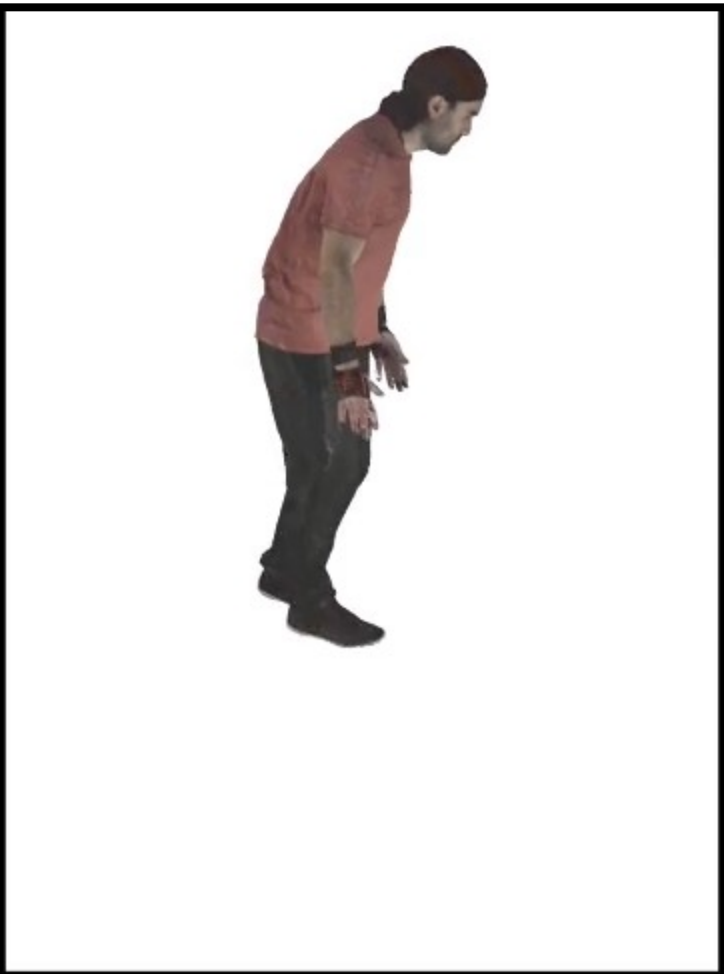
Assigned 2D Poses

3D Pose Estimation



Full dataset available:

http://virtualhumans.mpi-inf.mpg.de/3DPW/



3DPW

- 60 video sequences.
- 2D pose annotations.
- 3D reference poses.
- Camera poses for every frame in the sequences.
- 3D body scans and 3D people models (re-poseable and re-shapeable). Each sequence contains its corresponding models.
- 18 3D models in different clothing variations.

More Information

- Supplementary Video: https://www.youtube.com/watch?v=3x9dimY7o-o

- More papers on IMU-based tracking:
- https://virtualhumans.mpi-inf.mpg.de/topics/human-motion-from-wearables.html

Slide credits

- Slides on distance metrics based on Hartley et al. IJCV'13
- Slides based SIP (von Marcard et al. EG'17) and 3DPW (von Marcard et al. ECCV'18) papers, thanks to Timo von Marcard