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
A Body Model is a function

\[ M(0, X_{\text{shape}}) \]

\[ M(X_{\text{pose}}, 0) \]

\[ M(X_{\text{pose}}, X_{\text{shape}}) \]

\[ R \cdot M(X_{\text{pose}}, X_{\text{shape}}) \]

\[ M(X_{\text{pose}}, X_{\text{shape}}) \]

\[ Y \]

\[ X = \{X_{\text{pose}}, X_{\text{shape}}\} \]
Sparse Inertial Poser
Automatic 3D Human Pose Estimation from Sparse IMUs

Timo v. Marcard\textsuperscript{1}, Bodo Rosenhahn\textsuperscript{1}, Michael J. Black\textsuperscript{2}, Gerard Pons-Moll\textsuperscript{2}

\textsuperscript{1}Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover
\textsuperscript{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 activities can not be captured

[Tome SelfPose et al., 2018]

[Rhodin et al., 2016]
IMU-based Motion Capture

• IMU = Inertial Measurement Unit
Inertial sensors

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

[Refs: 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

\[ x \in \mathbb{R}^{75} \text{ pose} \]

Sparse Inertial Poser

\[
\underset{x}{\text{arg min}} \left( \right) \text{ Orientation consistency} \quad \text{Acceleration consistency} \quad \text{Anthropometric consistency}
\]
Anthropometric Consistency

**Objective**

enforce human-like poses

\[ N(\mu_x, \Sigma_x) \]

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

\[ E_{\text{anthro}}(x) = d_{\text{mahal}}(x)^2 + ||e_{\text{limits}}(x)||^2 \]
Orientation Consistency

Objective:
sensor & bone orientation consistency

How do you compute distance between orientations??
Distance metrics in SO(3)

**Angular distance**

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

$$\theta = d_\angle(S, R) = d_\angle(SR^T, I) = \| \log(SR^T) \|_2$$

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

The sign of $\omega$ must be chosen so that theta lies between 0 and $\pi$.

Geodesic distance in SO(3)
Distance metrics in SO(3)

**Chordal distance**
Frobenious norm of rotation matrix difference

\[ d_{\text{chord}}(S, R)^2 = \| S - R \|_F^2 = \| SR^\top - I \|_F^2 \]

\[ = 2(\sin^2(\theta) + (1 - \cos(\theta))^2) \]

\[ = 8 \sin^2(\theta/2) \]

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

Relation to angular distance

**Properties:**

1. \( \exp(\theta \hat{\omega}) = I + \hat{\omega} \sin(\theta) + \hat{\omega}^2 (1 - \cos(\theta)) \)
2. \( \hat{\omega} \) and \( \hat{\omega}^2 \) are orthogonal under Frobenious norm
3. \( \| \hat{\omega} \|_F^2 = \| \hat{\omega}^2 \|_F^2 = 2 \)
Distance metrics in SO(3)

Quaternion distance

\[ d_{\text{quat}}(S, R) = \min\{\|s - r\|_2, \|s + r\|_2\} \]

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

\[ d_{\text{quat}}(S, R) = \min\{\|s - r\|_2, \|s + r\|_2\} \]

Recall that quaternions \( q \) and \( -q \) represent the same rotation. This ambiguity is resolved by taking the min.

The relationship to the angular distance:

\[ e = (1, 0, 0, 0) \quad s \cdot r^{-1} = (\cos(\theta/2), \hat{v} \sin(\theta/2)) \]

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

- Inner product of two 4-vectors equals \( \cos(\alpha) \)
- Hence \( \alpha = \theta/2 \)
Distance metrics in SO(3)

Quatenable distance

\[ d_{\text{quat}}(S, R) = \min\{\|s - r\|_2, \|s + r\|_2\} \]

Recall that quaternions \( q \) and \( -q \) represent the same rotation. This ambiguity is resolved by taking the min.

The relationship to the angular distance:

\[ e = (1, 0, 0, 0) \quad s \cdot r^{-1} = (\cos(\theta/2), \hat{v} \sin(\theta/2)) \]

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

\[ \|s \cdot r^{-1} - e\|_2 = \|s - r\|_2 \]

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

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

- The distance of two unit vectors separated by \( \alpha \) is \( 2 \sin(\alpha/2) \)
Distance metrics for SO(3)

Angular distance
\[
\theta = d_\triangle(S, R) = d_\triangle(SR^T, I) = \| \log(SR^T) \|_2
\]
\[0 \leq \theta \leq \pi\]

Chordal distance
\[
d_{\text{chord}}(S, R)^2 = \| S - R \|_F^2 = \| SR^T - I \|_F^2
\]
\[
d_{\text{chord}}(S, R) = 2\sqrt{2} \sin(\theta/2)
\]

Quaternions distance
\[
d_{\text{quat}}(S, R) = \min\{\| s - r \|_2, \| s + r \|_2\}
\]
\[
d_{\text{quat}}(S, R) = 2\sin(\alpha/2) = 2\sin(\theta/4)
\]
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 $\pi$ 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 \( \pi \) 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

\[
d_{log}(S, R) = \min ||v_r - v_s||_2
\]

where the minimum is taken over all choices of vectors \( v_r \) and \( v_s \) such that \( \exp[v_r]_\times = R \) and \( \exp[v_s]_\times = S \).

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
Underconstrained Pose Space

pose different

orientation measurement nearly identical
Orientation Consistency

Sparse Orientation Poser (SOP)
Acceleration Consistency

Objective

sensor & vertex acceleration consistency

\[ e_{\text{acc}}(\mathbf{x}, a_{\text{sens}}) = a_{\text{vertex}}(\mathbf{x}) - a_{\text{sens}} \]

1) Use finite differences

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

2) Subtract gravity

\[ a_t^G = R_t^{GS} a_t^S - g^G \]

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

\[ x^*_1:T = \arg \min_{x_1:T} E_{\text{motion}}(x_1:T, R_1:T, a_1:T) \quad x_1:T \in \mathbb{R}^{75T} \]

\( E_{\text{motion}}(x_1:T, R_1:T, a_1:T) = \)

\[
= w_{\text{ori}} \cdot \sum_{t=1}^{T} \sum_{n=1}^{6} \| e_{\text{ori}}(x_t, R^n_t) \|^2 + w_{\text{acc}} \cdot \sum_{t=1}^{T} \sum_{n=1}^{6} \| e_{\text{acc}}(x_t, a^n_t) \|^2 + w_{\text{anthro}} \cdot \sum_{t=1}^{T} E_{\text{anthro}}(x_t)
\]

Orientation consistency

Acceleration consistency

Anthropometric consistency
Optimization

\[ e(x, \delta x) \approx e(x) + J \delta x. \]

To minimize \( e^T e \), linearize the vector of residuals with Jacobian matrix

For example, the acceleration residuals linearized take the form:

\[ e_{acc}(t, \delta x) \approx e_{acc}(t) + \left[ J_{p(x_{t-1})} \quad -2J_{p(x_t)} \quad J_{p(x_{t+1})} \right] \begin{bmatrix} \delta x_{t-1} \\ \delta x_t \\ \delta x_{t+1} \end{bmatrix} \]

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!! $\rightarrow$ Exploit the **block diagonal** structure

$$
\begin{bmatrix}
J_{t-1} & J_t & J_{t+1} \\
\delta x_{t-1} & \delta x_t & \delta x_{t+1}
\end{bmatrix}
\begin{bmatrix}
\vdots \\
\vdots
\end{bmatrix}
\begin{bmatrix}
e(t - 1) \\
e(t) \\
e(t + 1)
\end{bmatrix}
= 
\begin{bmatrix}
\vdots \\
\vdots
\end{bmatrix}
$$

Orientation+anthropometric residual equations

$$
\begin{bmatrix}
\vdots \\
\vdots
\end{bmatrix}
\begin{bmatrix}
-2J_{t-1} & J_t & J_{t+1} \\
J_{t-1} & -2J_t & J_{t+1} \\
J_t & -2J_{t+1}
\end{bmatrix}
\begin{bmatrix}
\delta x_{t-1} \\
\delta x_t \\
\delta x_{t+1}
\end{bmatrix}
= 
\begin{bmatrix}
\vdots \\
\vdots
\end{bmatrix}
$$

Acceleration term residual equations
Results
Results

[Image of a person writing on a board and a statue with the word "European"]

[Handwritten text: "European"]
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

Quantitative Evaluation

Angular Error [deg]

- SOP: 19.64
- SIP-M: 18.24
- SIP: 13.32

Joint Position Error [cm]

- SOP: 7.2
- SIP-M: 6.0
- SIP: 3.9
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

<table>
<thead>
<tr>
<th>Component</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall computation time</td>
<td>7.5 min</td>
</tr>
<tr>
<td>Model update</td>
<td>14.4 s/iteration</td>
</tr>
<tr>
<td>Setting up Jacobians</td>
<td>3.3 s/iteration</td>
</tr>
<tr>
<td>Solving for an update-step</td>
<td>1.5 s/iteration</td>
</tr>
</tbody>
</table>
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