Imperial College London

Rotational Adjoint Methods for Learning-Free 3D Human Pose Estimation from IMU Data

Caterina Buizza Yiannis Demiris

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Motivation

Human pose tracking has been heavily reliant on camera systems, though Inertial Measurement Units (IMUs) can provide an alternative. Until recently, IMU systems have imposed constraints on motion due to the bulky and sometimes wired systems worn using body suits or similar uncomfortable clothing. This is changing however, with wireless sensors becoming increasingly cheap and reliable.

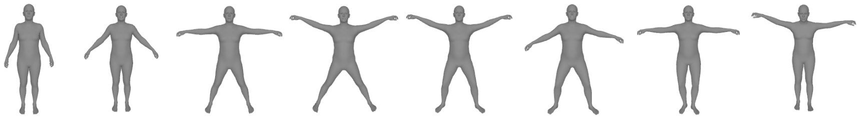
Method

Many existing pose estimation methods optimize a cost function based on several error terms.

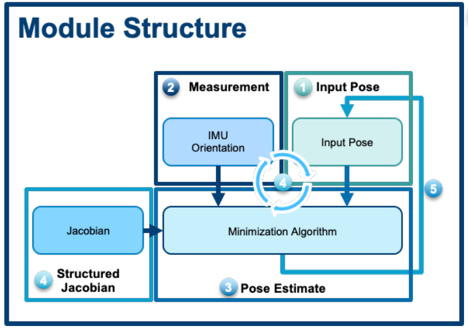
We propose to use a learning-free method, which uses only orientation measurements from a single time instance to recover human pose.

$$J(\boldsymbol{\theta}) = \frac{1}{2}(\mathbf{e}^p)^T\mathbf{P}^{-1}\mathbf{e}^p + \frac{1}{2}(\mathbf{e}^o)^T\mathbf{R}^{-1}\mathbf{e}^o$$

The cost function is calculated using two terms. First, we calculate the difference between the estimated orientations $\mathbf{R}_j$ and the orientations obtained from the sensor at the 10 sensor joints.



1. An initial guess is known from either starting pose (x_0) or previous time step (x_{t-1})
2. A set of orientation measurements is obtained from the 10 IMUs



5. Once the total error is below a tolerance, the current pose (x_t) is outputted by the algorithm

3. Both values are fed into the minimization algorithm, which will find the current pose (x_t)
4. During minimization, the Jacobian is used by the algorithm to calculate the total error change an adjustment to the estimated pose would make

Total Capture Dataset, by sequence											
Method	**N_S**	**N_C**	**S1/FS3**	**S2/FS1**	**S2/RM3**	**S3/FS1**	**S3/FS3**	**S4/FS3**	**S5/A3**	**S5/FS1**	**Mean**	**FPS**
[11] HQ	13	8	0.20	0.09	0.09	0.14	0.16	0.14	0.13	0.14	0.14	15
[11] HQ	6	8	0.28	0.16	0.15	0.23	0.27	0.22	0.21	0.21	0.22	15
[11] HS	6	4	0.32	0.19	0.19	0.28	0.34	0.26	0.25	0.26	0.26	30
Ours	10	0	0.34	0.26	0.29	0.31	0.34	0.28	0.31	0.24	0.30	30

TABLE II

Results

The method is evaluated on the TNT15 dataset, as seen in Table II. This dataset does not have any MoCap data: performance is evaluated by holding out 4 sensors and comparing the estimated orientation and position of these joints with those measured by the sensors.

We also compare our method to the results presented in [1], in which multi-video views are also used.

In addition, we have performed an ablation study with our method, where we show the effect of adding commonly used error terms to our orientation-only method. We have followed the methods for these additional terms as presented in [2].

Method	**N_S**	**Elbows**	**Hands**	**Knees**	**Ankles**	**U Back**	**L Back**	**Total**	**FPS**
Our Dataset									
Ours + acc	10	0.56	0.73	0.51	0.38	0.22	0.31	0.52	21
Ours + limits	10	0.38	0.78	0.34	0.47	0.12	0.21	0.45	25
Ours + prior	10	0.35	0.61	0.32	0.43	0.09	0.11	0.42	8
Ours + prior	6	0.76	1.22	0.51	0.73	0.19	0.39	0.63	8
Ours	10	0.46	0.78	0.31	0.43	0.11	0.09	0.36	30
Total Capture Dataset									
Ours	10	0.52	0.63	0.42	0.46	0.09	0.21	0.38	30

TABLE III

For more information, please feel free to contact us:
Email: caterina.buizza11@imperial.ac.uk
Website: www.imperial.ac.uk/personal-robotics/

References: [1] C. Malleson et al., *"Real-time full-body motion capture from video and imus,"* 3DV 2017 [2] T. von Marcard, et al., *"Sparse inertial poser: Automatic 3d human pose estimation from sparse imus,"* Computer Graphics Forum 2017