European Commission

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101092889, Topic HORIZON-CL4-2022-HUMAN-01-14

SHARESPACE

Embodied Social Experiences in Hybrid Shared Spaces

SHARE SPACE .eu

Project Reference No	**101092889**
Deliverable	D3.1. Self-calibrating ego-centric visual-inertial body tracking v1
Workpackage	WP3: task 3.2
Nature	D (Deliverable)
Dissemination Level	PU - Public
Date	31/05/2024
Status	v1.0
Editor(s)	**Bertram, Taetz (DFKI)** Markus, Miezal (DFKI)
Involved Institutions	DFKI
Document Description	This deliverable presents the first results on task 3.2

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Table 1: List of Abbreviations

Term / Abbreviation	Definition
IMU	Inertial Measurement Unit
REQ	Requirements
I2S	IMU-to-Segment

1 INTRODUCTION

The technologies for human body tracking has significantly advanced applications in fields ranging from virtual reality to health monitoring. However, achieving precise, full-body tracking with ease of use remains a complex challenge. Our lightweight demonstrator system aims to address the prevalent difficulties associated with full-body tracking and its integration into a digital environment. By combining ego-centric visual-inertial tracking with minimal setup requirements, this system seeks to offer a practical and efficient solution for both professional and casual users.

At the core of our demonstrator is an innovative approach to tracking the human body in a global coordinate system, a task often marred by challenges relating to sensor placement, calibration, and accuracy. Traditional tracking systems typically require a large number of inertial sensors that must be meticulously calibrated to specific body segments. This not only adds to the complexity but also increases the potential for inaccuracies caused by sensor mispositioning. Our system addresses these issues by reducing the number of necessary inertial sensors while maintaining high levels of accuracy in joint angle tracking.

Ease of use is a critical component of our system's design philosophy. Traditional full-body tracking systems often involve cumbersome setups that are not only time-consuming but also intrusive to the user’s experience. Our demonstrator, in contrast, emphasizes a simple and unobtrusive setup. This makes the system particularly suitable for a variety of applications, from interactive virtual environments and gaming to real-time motion analysis in sports and rehabilitation.

In essence, our lightweight demonstrator represents a step forward in the realm of human body tracking for virtual environments. By addressing the key issues of accurate full-body tracking, sensor calibration, and setup complexity, we offer a solution that is both high-performing and user-friendly. Whether utilized for enhancing user experiences in digital environments or for precise monitoring in professional settings, our system stands out as a versatile and reliable tool for full-body tracking.

2 BACKGROUND

Human body pose and motion tracking from an ego-centric perspective, where the tracking device is mounted on the body, has emerged as a significant area of research in wearable computing and immersive environments. This approach leverages the user's perspective to enhance interaction with digital systems, creating a seamless bridge between the physical and digital worlds.
The goal of this deliverable is to demonstrate the first version of our developments towards a lightweight, full-body tracking system that is easy to setup, can track joint angles of the full human body and can be used to locate the human body within a digital environment accurately.

The following requirements have a direct influence on the demonstrator and are derived from the health scenario and encompass (based on the requirements defined for the health scenario):

- **REQ1:** To ensure that the patient performs the exercises is the first priority (so the motivation to/fear of exercise -> make the available exercise sets ‘easy’, ‘doable’, ‘fun’, ‘engaging’, ‘challenging’) **–** the system needs to offer a variety of movement options at different levels of difficulty
- **REQ3:** Motion tracking sensors need to provide data allowing for computation of: amplitude of motion, speed of movement and smoothness -> spatiotemporal data with a point of origin (for calculation of displacement)

- **REQ4:** The system should support/provide all exercises usually done in a session at the hospital (lying on a bed or on the floor, standing, leaning against the wall)
- **REQ5:** Reliable tracking of the upper legs, pelvis, and spine
- **REQ6:** Motion prediction of body segments which do not have IMU sensors (info from cameras)
- **REQ7:** Fusion of the IMU information with the information of a camera image from an laptop/smartphone or the head-mounted display.
- **REQ8:** The remote patient should get the feeling of being part of the group and performs the exercises with the group
- **REQ15:** Sensors need to be attached by Velcro to strap. Clear instruction needs to be provided where to place them on the body and how to close them.
- **REQ 17:** head-mounted display (Phase I and Phase II) should be small, lightweight, and wireless and easy to connect to the processing unit (laptop or smartphone)

To reach the stated goal and address the given requirements, ego-centric visual-inertial tracking was utilized, i.e., a combination of HMD-integrated cameras and inertial sensors mounted on body segments. This approach solves the following technical challenges and matches the requirements as follows:

- Fused tracking of full-body in one joint coordinate system of visual (HMD) and inertial tracking system with position tracking and possible localization relative to a starting point for usage in a virtual environment **(REQ8, REQ7, REQ6, REQ4, REQ3, REQ1)**
- Reduced number of inertial sensors for full-body tracking (5-7 sensors + 1 HMD) **(REQ 17)**
- Self-calibration of joint positions and segment length estimates/ease of setup **(REQ 15)**
- Accurate joint angle estimation of the lower body with robustness to drift and magnetic disturbances **(REQ 5)**

Related work:

One of the primary challenges addressed in our system is full-body tracking using a reduced number of inertial sensors while maintaining accurate and drift-free lower-body joint angle estimates. Full-body inertial motion capture typically involves a dense array of sensors distributed across various body segments (Roetenberg et al., 2009) to capture detailed joint angles reliably. One remaining problem is the sensor-to-segment calibration, which typically requires an exact pose or motion to be performed (Di Raimondo et al., 2022; Ekdahl et al., 2023). Usually, between 11 to 18 sensors must be mounted on the body for full-body tracking, which renders the setup cumbersome and intrusive, limiting practical applications and user comfort.

The issue of calibrating inertial sensors to body segments is a critical factor in ensuring accurate tracking. Self-calibration techniques or calibration-free approaches have gained prominence in recent research (Laidig et al., 2022; McGrath & Stirling, 2020, 2022; Taetz et al., 2016; Zimmermann et al., 2018), aiming to automate the alignment process, thereby reducing the dependency on manual calibration procedures which are prone to human error.

Another direction is concerned with inertial full-body tracking from a reduced amount of inertial sensors (Huang et al., 2018; Yi et al., n.d., 2022). These approaches try to estimate the full-body pose from 5 or 6 sensors mounted at the extremities and close to the body center. The main challenge in these approaches are the increased ambiguity of the measurements. This is typically addressed via a recurrent neural network approach, sophisticated body model and physically relevant detections and modeling. However, increased drift or failure to capture subtle movements renders these approaches often unreliable for long-term pure inertial tracking.

There are some recent approaches to combine a sparse amount of inertial sensors in the above mentioned setting, i.e. that are sparsely distributed over the body segments, and combined with an external camera (Pan et al., 2023) of an HMD (Kim & Lee, 2022).

In contrast, our approach reduces the number of required sensors while maintaining accuracy, with reduced drift, due to a dense sensor network mounting on the lower-body segments, making the system mobile, practical and reliable for everyday usage.

Therefore, our proposed system aims to balance the requirements of high joint angle and positional tracking accuracy and unobtrusive setup, by utilizing a synergistic approach combining ego-centric visual upper-body tracking from a HMD with inertial sensor-based lower-body tracking with self-calibrating joint position estimation to obtain anatomically correct kinematic estimates of the full human body. The approach utilizes advanced calibration and drift mitigation techniques to ensure that the system can be deployed easily and comfortably, targeting seamless interaction within virtual or augmented environments.

3 APPROACH

Fused visual-inertial tracking

The proposed approach consists of a head-mounted display (Meta Quest Pro[1]) and seven inertial sensors (Xsens AWINDA[2]) that stream data via Wi-Fi to a Desktop PC, as shown in Figure 1 on the left-hand side. The latter is used to fuse the data into a consistent kinematic skeleton, as shown in Figure 1 on the right-hand side.

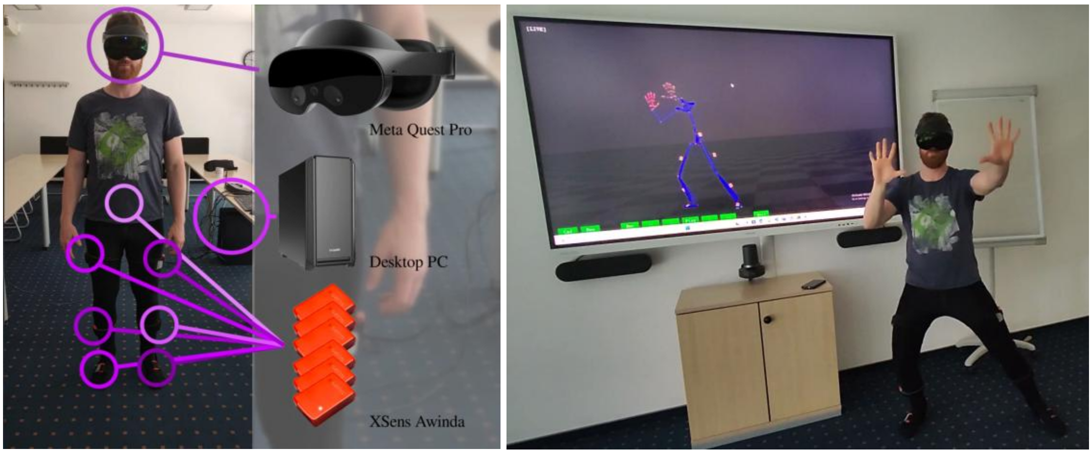

Figure 1 - Overview of visual-inertial full-body tracking system

[1] https://www.meta.com/de/quest/quest-pro/
[2] https://www.movella.com/products/wearables/xsens-mtw-awinda

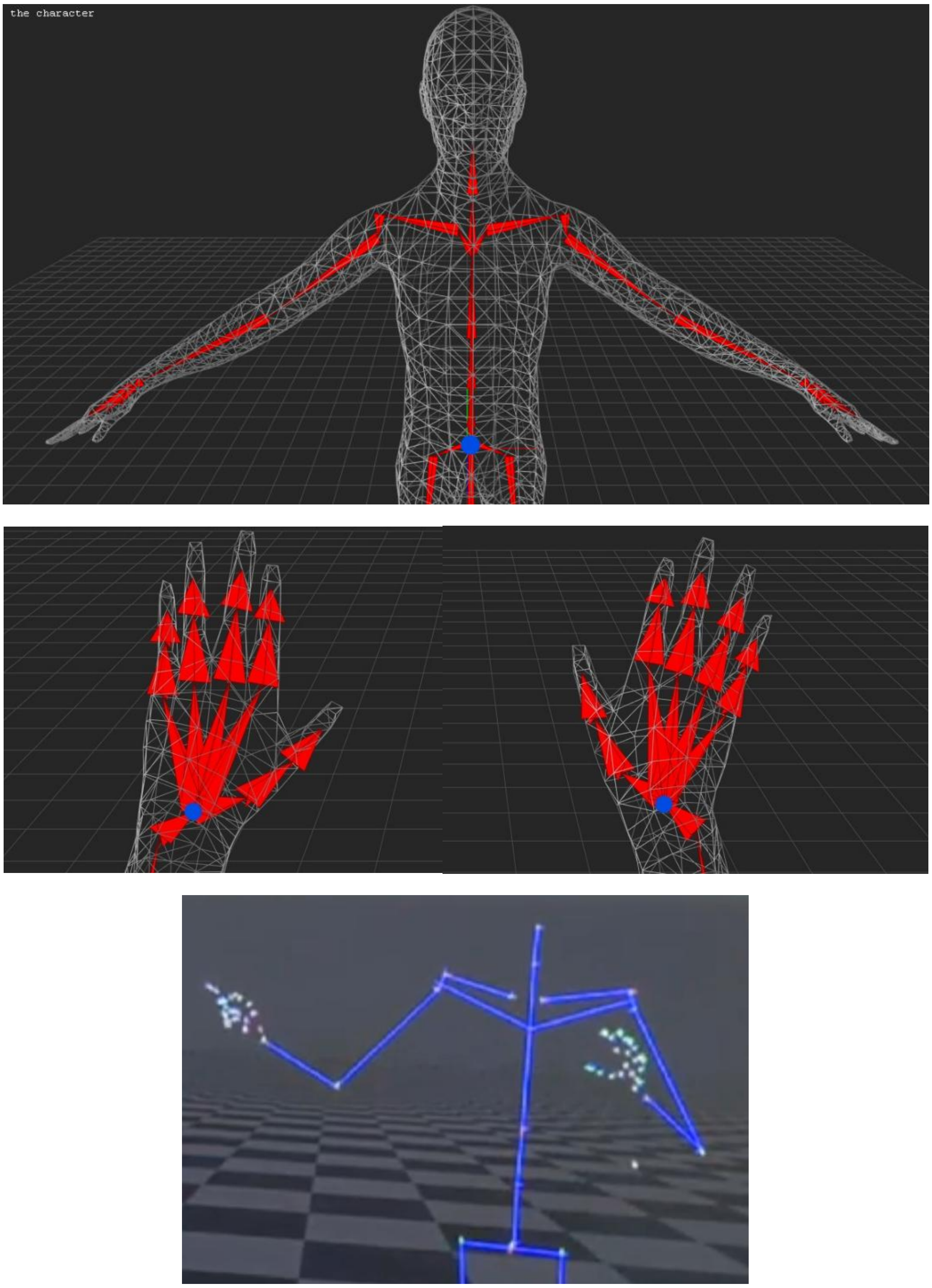

Figure 2: Segments of the upper body (red lines) and root joint (blue dot) in the upper image, segments, and root joint of the left and right hand (middle image)[3], and the visualization of both together in our visualizer, lower image.

Figure 2 shows the upper body skeleton. The SDK description describes all joint names[4].

Figure 3 shows the lower body skeleton from the inertial motion tracking. Note, the current demonstrator is based on the IMU-based QuatTracker and the three-step calibrationas described in

[3] https://developer.oculus.com/documentation/native/android/move-ref-body-joints/
[4] https://developer.oculus.com/documentation/native/android/move-ref-body-joints/

detail in (Miezal, 2021). However, a novel self-calibrating approach, not yet integrated, is being developed, as described below.

Self-calibrating inertial tracking (JointTracker)

In this section we consider inertial lower body tracking using the lower body configuration illustrated in Figure 3, with 7 inertial sensors (I_i^N) and six joints ($J_{i,j}^N$). In the figure, the joints are connected via blue lines and an IMU-to-Segment calibration is not given. Inertial motion tracking typically exploits a fixed biomechanical model with a prior calibration step where the orientation of the inertial sensor is estimated with respect to the corresponding segment of the biomechanical model (Ekdahl et al., 2023). We investigated a calibration-free real-time estimation approach that estimates the IMU pose (orientation and position) in a navigation frame and anatomical joint positions adapted (self-calibrated) to the person wearing the inertial sensor network. This is performed via an online recursive estimation of the IMU states and the joint position as further state variables that are estimated alongside the states of the IMU pose; the method, including evaluations, is published in (Taetz et al., 2024).

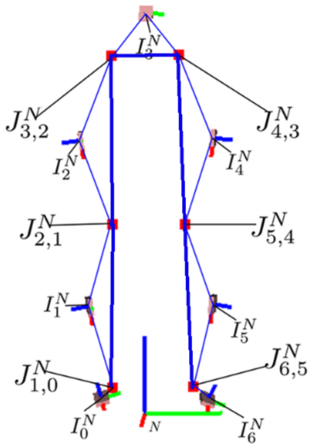

Figure 3: Illustration of the kinematic model of the lower body in the navigation frame (N) with segments (blue lines), joint positions (red dots) and inertial sensors (small boxes) with their coordinate systems.

Important features of this approach that support the requirements are:

The approach:

- works in real-time and is calibration-free, i.e. works directly online on a data stream of inertial data from a body sensor network, without exact placement of the sensors on the segment or calibration poses or functional movements being required
- allows for segment length estimation of all segments with two joints to personalize the segment length to a specific person wearing the inertial sensor network.

4 RESULTS

A video of the integrated demonstrator is shown here:
https://www.youtube.com/watch?v=S9xITW0cD1Y

JointTracker results

The following plots show the results of estimating segment length based on the above-mentioned JoinTracker states. The segment length is computed as the Euclidean distance between the two joints flanking each segment (see Figure 3). For further explanations, see (Taetz et al., 2024).
Segment length estimation based on joint position estimates from forward walking motion (dataset: 6-minute walking test[5]). The segment length converges after about 3-4 seconds except on the pelvis segment, which has less motion, and the convergence takes about twice as long as can be observed in the plots shown in Figure 4.

[5] https://zenodo.org/records/10253111

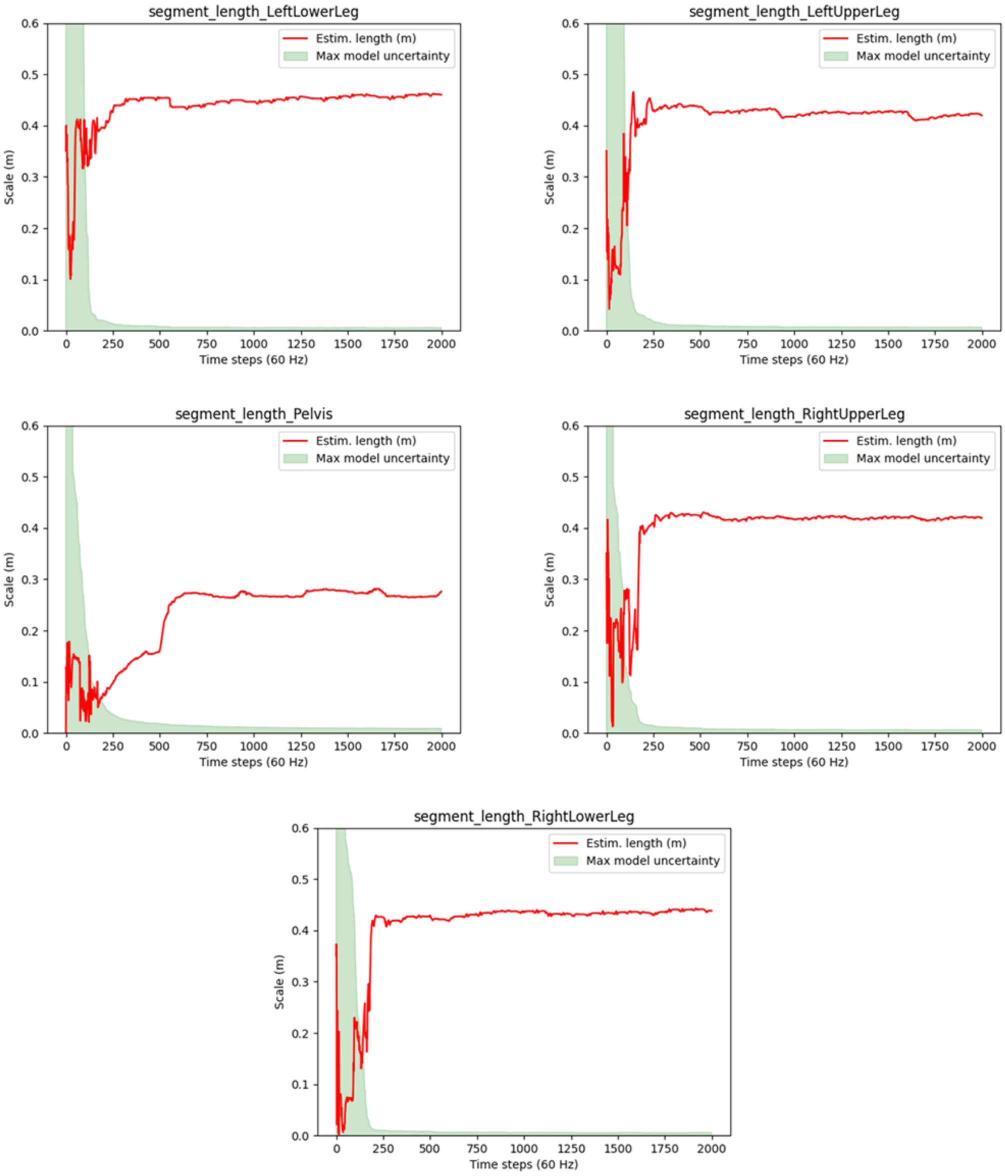

Figure 4: The plot shows the online segment length estimation and the timeline of the different lower body segments. The green area displays an uncertainty measure from the Bayesian recursive estimator, revealing the uncertainty of the joint position estimate.

5 LIMITATIONS AND NEXT STEPS

Intensive user testing for robustness, validity and usability needs to be conducted.

Furthermore, the synchronization and registration of one or multiple external camera(s) with the ego-centric visual inertial setup could allow to further reduce the number of required inertial sensors (Pan et al., 2023).

The JointTracker requires a minimum amount of motion in all degrees of freedom to estimate the joint positions accurately. Its applicability within the application scenarios needs to be further investigated.

6 REFERENCES

Di Raimondo, G., Vanwanseele, B., van der Have, A., Emmerzaal, J., Willems, M., Killen, B. A., & Jonkers, I. (2022). Inertial Sensor-to-Segment Calibration for Accurate 3D Joint Angle Calculation for Use in OpenSim. *Sensors*, *22*(9), 3259. https://doi.org/10.3390/s22093259

Ekdahl, M., Loewen, A., Erdman, A., Sahin, S., & Ulman, S. (2023). Inertial Measurement Unit Sensor-to-Segment Calibration Comparison for Sport-Specific Motion Analysis. *Sensors*, *23*(18), 7987. https://doi.org/10.3390/s23187987

Huang, Y., Kaufmann, M., Aksan, E., Black, M. J., Hilliges, O., & Pons-Moll, G. (2018). Deep Inertial Poser: Learning to Reconstruct Human Pose from Sparse Inertial Measurements in Real Time. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, *37*, 185:1-185:15.

Keller, M., Werling, K., Shin, S., Delp, S., Pujades, S., Liu, C. K., & Black, M. J. (2023). From Skin to Skeleton: Towards Biomechanically Accurate 3D Digital Humans. *ACM Transactions on Graphics*, *42*(6), 1–12. https://doi.org/10.1145/3618381

Kim, M., & Lee, S. (2022). Fusion Poser: 3D Human Pose Estimation Using Sparse IMUs and Head Trackers in Real Time. *Sensors*, *22*(13), 4846. https://doi.org/10.3390/s22134846

Laidig, D., Weygers, I., & Seel, T. (2022). Self-Calibrating Magnetometer-Free Inertial Motion Tracking of 2-DoF Joints. *Sensors*, *22*, 9850. https://doi.org/10.3390/s22249850

McGrath, T., & Stirling, L. (2020). Body-Worn IMU Human Skeletal Pose Estimation Using a Factor Graph-Based Optimization Framework. *Sensors*, *20*(23). https://doi.org/10.3390/s20236887

McGrath, T., & Stirling, L. (2022). Body-Worn IMU-Based Human Hip and Knee Kinematics Estimation during Treadmill Walking. *Sensors*, *22*(7), 2544. https://doi.org/10.3390/s22072544

Miezal, M. (2021). *Models, methods and error source investigation for real-time Kalman filter based inertial human body tracking*. Verlag Dr. Hut.

Pan, S., Ma, Q., Yi, X., Hu, W., Wang, X., Zhou, X., Li, J., & Xu, F. (2023). Fusing Monocular Images and Sparse IMU Signals for Real-time Human Motion Capture. *SIGGRAPH Asia 2023 Conference Papers*, 1–11.

Roetenberg, D., Luinge, H., & Slycke, P. (2009). *Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors*.

Taetz, B., Bleser, G., & Miezal, M. (2016). Towards self-calibrating inertial body motion capture. *19th International Conference on Information Fusion, FUSION 2016, Heidelberg, Germany, July 5-8, 2016*, 1751–1759. https://ieeexplore.ieee.org/document/7528095/

Taetz, B., Lorenz, M., Miezal, M., Stricker, D., & Bleser-Taetz, G. (2024). JointTracker: Real-time inertial kinematic chain tracking with joint position estimation [version 1; peer review: 1 approved with reservations]. *Open Research Europe*, *4*(33). https://doi.org/10.12688/openreseurope.16939.1

Yi, X., Zhou, Y., Habermann, M., Shimada, S., Golyanik, V., Theobalt, C., & Xu, F. (2022, June). Physical Inertial Poser (PIP): Physics-aware Real-time Human Motion Tracking from Sparse Inertial Sensors. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yi, X., Zhou, Y., & Xu, F. (n.d.). *TransPose: Real-time 3D Human Translation and Pose Estimation with Six Inertial Sensors*. *40*(4).

Zhang, H., Tian, Y., Zhang, Y., Li, M., An, L., Sun, Z., & Liu, Y. (2023). PyMAF-X: Towards Well-aligned Full-body Model Regression from Monocular Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–16. https://doi.org/10.1109/TPAMI.2023.3271691

Zimmermann, T., Taetz, B., & Bleser, G. (2018). IMU-to-Segment Assignment and Orientation Alignment for the Lower Body Using Deep Learning. *Sensors*, *18*(1). https://doi.org/10.3390/s18010302