Check for updates

METHOD ARTICLE

Inertial human motion capture with automatic IMU-to-segment position and orientation calibration using JointTracker and deep learning with uncertainties

[version 1; peer review: 2 approved with reservations]

Bertram Taetz[1,2], Gabriele Bleser-Taetz[1,2], Markus Miezal[2], Didier Stricker[3,4]

[1]IT and Engineering, IU International University of Applied Sciences, Erfurt, Thuringia, 99084, Germany
[2]sci-track GmbH, Mesenich, Rhineland-Palatinate, 56820, Germany
[3]Augmented Vision, German Research Center for Artificial Intelligence, Kaiserslautern, Rhineland-Palatinate, 67663, Germany
[4]Augmented Vision Lab, RPTU University Kaiserslautern-Landau, Kaiserslautern, Rhineland-Palatinate, 67663, Germany

v1 **First published:** 31 Mar 2026, **6**:86
https://doi.org/10.12688/openreseurope.22844.1
Latest published: 31 Mar 2026, **6**:86
https://doi.org/10.12688/openreseurope.22844.1

Abstract

This paper introduces a novel approach to self-calibrating online inertial human motion capture designed for dense IMU configurations. By combining model-based sensor fusion with deep learning, the proposed method eliminates the need for static pose or functional calibration, enabling automatic IMU-to-segment (I2S) orientation and position calibration during dynamic activities such as walking. The approach integrates the previously established JointTracker framework with a novel probabilistic deep learning module for I2S orientation prediction. The JointTracker estimates global IMU orientations and joint positions from synchronized IMU measurements using recursive filtering, while the calibration module extracts kinematic features, employs probabilistic sequence-to-sequence prediction with an ensemble network trained in a self-supervised manner, and applies active selection of I2S updates based on predicted uncertainties to ensure robust performance under unseen motions. The evaluation demonstrates that the proposed method achieves comparable accuracy to existing static pose or functional calibration methods performed under supervised conditions. Notably, it performs on par or surpasses previous self-calibration approaches in accuracy on simulated and real IMU data and is, to the best of our knowledge, the first method to estimate I2S orientations and positions on the fly. The training works on kinematic data from real human motions, no real IMU data needed, making it easily adaptable to different motion scenarios. Qualitative results on newly collected motion data with simulated gait deviations further

Open Peer Review

Approval Status ? ?

	1	2
version 1 31 Mar 2026	? view	? view

1. **Elisa Digo**, Politecnico di Torino, Turin, Italy
2. **Moein Younesi Heravi**, North Dakota State University, Fargo, USA

Any reports and responses or comments on the article can be found at the end of the article.

validate the method's practical applicability. This research advances inertial motion capture by offering a robust and flexible solution broadening the potential for deployment in diverse applications where dynamic calibration is essential.

Keywords
IMU-to-segment calibration, inertial motion capture, sequence-to-sequence prediction, self-supervision, probabilistic deep learning, ensemble network



This article is included in the Horizon Europe gateway.



This article is included in the Horizon 2020 gateway.

Corresponding author: Bertram Taetz (Bertram.Taetz@iu.org)

Author roles: Taetz B: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Bleser-Taetz G**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Miezal M**: Data Curation, Resources, Software, Writing – Review & Editing; **Stricker D**: Funding Acquisition, Project Administration, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No [826304] (Personalized Body Sensor Networks with Built-In Intelligence for Real-Time Risk Assessment and Coaching of Ageing workers, in all types of working and living environments [BIONIC]), grant agreement No [101092889] (Embodied Social Experiences in Hybrid Shared Spaces [SHARESPACE]) and grant agreement No [101058236] (Human Centered Technologies for a Safer and Greener European Construction Industry [HumanTech]).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Taetz B *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Taetz B, Bleser-Taetz G, Miezal M and Stricker D. **Inertial human motion capture with automatic IMU-to-segment position and orientation calibration using JointTracker and deep learning with uncertainties [version 1; peer review: 2 approved with reservations]** Open Research Europe 2026, **6**:86 https://doi.org/10.12688/openreseurope.22844.1

First published: 31 Mar 2026, **6**:86 https://doi.org/10.12688/openreseurope.22844.1

1. Introduction

Human motion capture (HMC) is a well-studied field with applications ranging from human-machine interaction over healthcare and sports to workplace analysis, robotics and more (Caserman, Garcia-Agundez, and Göbel 2020; Wong et al. 2015; Bleser et al. 2015; Menolotto et al. 2020). While camera-based HMC, despite great advances in the field (Zheng et al. 2023), faces inherent limitations such as occlusions caused by external factors or self-occlusions, inertial HMC (IHMC), which uses multiple inertial measurement units (IMUs) mounted on body segments using straps or clothing integration (Cereatti et al. 2024; Schepers, Giuberti, and Bellusci 2018; Xenoma Inc., n.d.; Lorenz et al. 2022), does not suffer from the line-of-sight problem and can be used location-independent. This makes it a promising approach, in particular for in-field capturing.

There are two major IHMC approaches: (1) The more classic approach, which underlies commercial systems like (Schepers, Giuberti, and Bellusci 2018; Xenoma Inc., n.d.), is based on the assumption that an IMU is attached to each body segment which should be tracked (dense IMU placement). The pose estimation then typically uses model-based sensor fusion in combination with a personalized biomechanical model (Bleser, Taetz, and Lukowicz 2018; Taetz et al. 2025). (2) More recent studies have achieved pose estimation (prediction) with a sparse set of IMUs (sparse IMU placement), in combination with deep learning (Huang et al. 2018; Yi, Zhou, and Xu 2021; Yi et al. 2022). Here, it is assumed that the missing information is compensated by a trained model. While a sparse IMU placement is preferable in terms of setup time, wearing comfort and hardware cost, the pose prediction results mainly depend on the training data used to train the model (rather than on generally applicable physical and biomechanical models), and small segments like hands and feet are typically neglected. Therefore, the classic approach is still more widespread in areas where detailed and accurate body motion capture is required, such as in biomechanics, medicine and sports (Cereatti et al. 2024).

Besides these trade-offs, there is a challenge which both IHMC approaches have in common and which is currently considered the main challenge for IHMC in biomechanical motion analysis, since it is a prerequisite for obtaining anatomically accurate joint angles and positions: the calibration of (1) the sensor poses (positions and orientations) with respect to the anatomical reference frames of the associated body segments, the so-called IMU-to-segment (I2S) calibrations, and (2) the segment lengths (Vitali and Perkins 2020). For successful and reliable in-field usage of IHMC systems, a self-calibrating approach seems crucial, in order to support ease-of-use and reduce human-induced errors due to manual procedures.

In (Taetz et al. 2025), a dense IMU placement approach with I2S position self-calibration is proposed. This is based on a recursive estimator for the simultaneous online estimation of sensor poses and joint positions of a kinematic chain. The body pose can then be illustrated by connecting the segments at the estimated joint positions, as it is common practice in the vision community (Zheng et al. 2023). The segment lengths are obtained implicitly as well. However, this approach does not address I2S orientation calibration, and as a consequence, it does not provide anatomical 3D joint orientations. Note, I2S orientation errors propagate linearly into joint-angle errors (Miezal, Taetz, and Bleser 2016).

For I2S orientation calibration, different methods are in use. Most widespread is still the static pose calibration that requires the user to take on one or more specific poses, such as N- or T-pose (Palermo et al. 2014; Liu, Zhang, and Zeng 2019). Functional calibration requires the user to perform motions around predefined axes (Favre et al. 2009; Nazarahari et al. 2019). Clearly, the accuracies of these methods depend on how precisely the predefined poses and motions are performed and can show large variances (Bouvier et al. 2015; Pacher et al. 2020). In order to reduce human-induced errors, (Seel, Raisch, and Schauer 2014; Müller et al. 2017; Taetz, Bleser, and Miezal 2016; Laidig et al. 2019) exploit kinematic constraints (assuming a dense IMU placement) and numerical optimization in combination with arbitrary motions to achieve I2S orientation self-calibration. However, these methods tend to be limited to restricted joints (with one or two degrees of freedom) and require extensive motions in the available degrees of freedom, such as circling around the considered joints, in order to converge (Pacher et al. 2020).

Another promising approach to I2S orientation self-calibration uses trained models in order to predict I2S orientations from IMU measurements (Zimmermann, Taetz, and Bleser 2018; Zuo et al. 2025). Both contributions work on a per IMU basis, making them applicable to dense as well as sparse IMU placements, and use available motion capture datasets as well as data augmentation through simulation for model training. However, these methods can still predict incorrect calibrations in the presence of irregular motions, not present in the training dataset.

This article proposes a novel, entirely self-calibrating online IHMC method designed for dense IMU placements, which integrates two core components: (1) model-based sensor fusion for tracking kinematic chains and calibrating I2S positions as well as implicitly segment lengths, and (2) a novel deep learning architecture for predicting I2S orientations with uncertainties. This uses the foundation laid in (Taetz et al. 2025) (a recursive filter named JointTracker) and extends

it with an I2S orientation calibration module that enhances the work in (Zimmermann, Taetz, and Bleser 2018) by incorporating higher-level input features derived from JointTracker states. These include fused IMU orientations and joint directions rather than direct IMU measurements, facilitating an ensemble network architecture tailored for probabilistic sequence-to-sequence orientation prediction. Additionally, the method employs an active I2S orientation selection strategy that uses predicted uncertainties to circumvent updates during unseen motions, which are typically associated with significant uncertainty. This new approach aims at increasing the system's resilience to variations in IMU data, while achieving successful self-calibration on the fly with minimal magnetometer usage, as inherited from JointTracker.

In the following, Section 2 provides the mathematical background for this article, and it describes the task to be solved. Section 3 describes the overall method, and it details the different aspects of the I2S orientation calibration. An experimental evaluation on a lower body gait dataset is presented in Section 4. Finally, Section 5 concludes and describes the open challenges for future work.

2. Notation and task description

First, the notation, relevant geometric quantities and assumptions, based on (Taetz et al. 2025), are introduced in Section 2.1. Then, the task to be solved is described in Section 2.2.

2.1 Notation

The following definitions are illustrated in Figure 1.

The navigation frame N is the global reference frame. Its z-axis is aligned with gravity, and its x-axis points (initially) towards local magnetic north. As stated in (Taetz et al. 2025), magnetometers are only used once during system initialization to create the initial heading alignment. Note, the global frame N is distinguished from the egocentric frame G, which is detailed at the end of this section.

We assume a kinematic chain with $N_I \geq 2$ rigid segments connected by $N_I - 1$ joints, where one IMU is mounted on each segment. A segment and its corresponding IMU share the same index, and these are collected in the set I, with $|\mathrm{I}| = N_I$, the

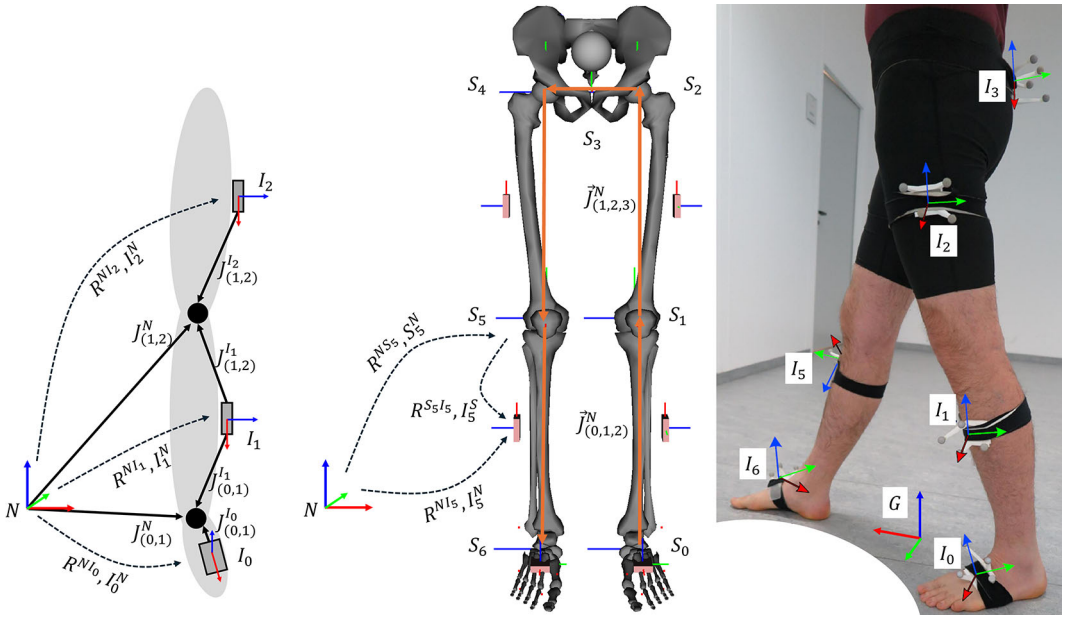

Figure 1. Left: Kinematic chain model (cutout of left leg) used in (Taetz et al. 2025) with IMUs (grey rectangles), joint centers (black circles), global and IMU-centered joint position vectors (solid black arrows) and global IMU poses (dashed black arrows). The rigid segments, on which the IMUs are mounted, are indicated as light grey ellipses. Middle: Anatomical skeleton model (Rajagopal et al. 2016) with IMU and segment coordinate systems S, where the global segment and IMU pose (w.r.t. the global frame N) and IMU-to-segment (I2S) pose are annotated for the right lower leg. Moreover, (orange) joint direction vectors are also shown, with mathematical symbols annotated exemplary for the left leg. Right: Lower body real setup with IMU coordinate frames I and egocentric reference frame G.

number of IMUs and segments. Moreover, we assume the kinematic chain to move in space (e.g. through locomotion), with position or velocity information being available at least for one segment for positional drift compensation (e.g. zero velocity detections as proposed in (Taetz et al. 2025)). Figure 1 shows a lower body configuration, which fulfills these assumptions and which is also used in the experiments.

Each IMU $i \in I$, has its own coordinate system, I_i, which is related to the navigation frame via a position (vector) $I_i^N \in R^3$ and a rotation (matrix) $R^{NI_i} \in SO(3)$. Each body segment $i \in I$, also has its own coordinate system, S_i, which is related to the navigation frame via a position $S_i^N \in R^3$ and a rotation $R^{NS_i} \in SO(3)$. The I2S orientation $R^{S_iI_i}$ and position I_i^S is the rigid transformation between IMU i and its associated segment i. Note, the I2S poses are here assumed to be rigid throughout one capturing session. This requirement could be relaxed, as further discussed in Section 5.

Using the I2S poses, the segment poses can be computed from the IMU poses for each segment and IMU $i \in I$ via

$$R^{NS_i} = R^{NI_i}\left(R^{S_iI_i}\right)^T \tag{1}$$

$$S_i^N = I_i^N - R^{NS_i} I_i^S. \tag{2}$$

Following (Taetz et al. 2025), the position of a joint connecting two neighboring segments with rigidly attached IMUs I_i and I_j is denoted in each IMU coordinate frame as $J_{(i,j)}^{I_i}$ and $J_{(i,j)}^{I_j}$ with $(i,j) \in J$, the set of all joint index pairs. We call a joint center, which is given relative to an adjacent IMU coordinate frame, an IMU-centered joint center. Global joint centers $J_{(i,j)}^N$ can be computed from the respective IMU-centered joint centers and IMU poses as

$$J_{(i,j)}^N = I_i^N + R^{NI_i} J_{(i,j)}^{I_i} \; or \tag{3}$$

$$J_{(i,j)}^N = I_j^N + R^{NI_j} J_{(i,j)}^{I_i}, \tag{4}$$

where both vectors can be averaged to obtain one result.

Note that, given the above definitions, I2S positions and associated IMU-centered joint positions are rigidly coupled via the I2S orientations (indices left out for brevity)

$$I^S = -R^{SI} J^I. \tag{5}$$

Here, J^I denotes the relevant adjacent joint to the considered IMU I, and, without loss of generality, a constant offset between the joint center J and the origin of the segment coordinate frame S is omitted.

Figure 1 (middle) also shows global joint directions, i.e. vectors (in frame N) pointing from one joint center to the next one in the kinematic chain, denoted $\vec{J}_{(i,j,k)}^N$, with $(i,j,k) \in J^2$, the set of all neighboring joint pairs when moving in a given direction through the kinematic chain. The joint directions are calculated from the respective global joint positions for each joint pair $(i,j,k) \in J^2$ as

$$\vec{J}_{(i,j,k)}^N := J_{(j,k)}^N - J_{(i,j)}^N. \tag{6}$$

The normalized global joint direction of joint pair $(i,j,k) \in J^2$ is denoted

$$\hat{J}_{(i,j,k)}^N := \frac{\vec{J}_{(i,j,k)}^N}{\|\vec{J}_{(i,j,k)}^N\|}. \tag{7}$$

In addition to N, we also use an egocentric reference frame $G := G_t$, which moves with the person in terms of position and heading direction and whose z-axis is also aligned with gravity (cf. Section 3.2). Note that this frame is therefore time-dependent (time index left out for brevity).

Besides rotation matrices R, we use a continuous six-dimensional orientation parametrization $\tilde{R} := \left[R_{*0}^T, R_{*1}^T\right]^T$, where $R_{*c}, c \in \{0,1,2\}$ denotes the cth orthonormal column vector of rotation matrix R, for I2S orientation prediction. This continuous six-dimensional representation was shown to provide superior (more accurate) results compared to the more redundant rotation matrices (as used in (Zuo et al. 2025)) or the discontinuous Euler angles (as used in (Zimmermann,

Taetz, and Bleser 2018)) in the context of neural network-based orientation regression (Zhou et al. 2019). Moreover, from this representation, a rotation matrix can be easily reconstructed by obtaining the third column via the cross product.

2.2 Task description

Overall, the proposed method should be able to provide global segment poses, i.e. positions $S^N_{i,t}$ and orientations $R^{NS_{i,t}}$, for each segment i and timestep t from the synchronized measurements of an IMU-equipped kinematic chain, with known structure according to the assumptions described in Section 2.1. The IMU measurement for IMU i at time t comprises the acceleration $y^{I_i}_{a,t} \in R^3$, angular velocity $y^{I_i}_{\omega,t} \in R^3$ and magnetic field $y^{I_i}_{m,t} \in R^3$, where the latter is only used for initialization. Reliable segment poses should be provided online, i.e. with minimal delay and only based on past and current measurements. While a short period of initial motion for self-calibration after system initialization is considered acceptable (e.g. walking a few steps), no predefined calibration poses or motions should be required in order to obtain valid results.

In (Taetz et al. 2025), given an IMU-equipped kinematic chain according to the assumptions described in Section 2.1, the global IMU orientations and the IMU-centered joint positions are estimated online for each time t from the synchronized IMU measurements of that time using a recursive filter. Note, after filter initialization, the joint position estimates require some time for convergence, (Taetz et al. 2025) reports five seconds for randomly initialized joint positions estimated from 60 Hz walking data. Note, the filter is subsequently referred to as JointTracker, as already introduced in Section 1.

Here, according to the above-described overall goal, we are interested in obtaining the global segment poses, from which then joint angles (relative segment orientations) can be derived. According to Section 2.1, this requires the I2S orientations $R^{S_iI_i}$ for each $i \in I$ to be calibrated, while the I2S positions can be derived from the above quantities, see Equation (5).

Hence, a critical task to be solved is the online prediction of reliable I2S orientations from available input features. This is done using the filter (JointTracker) estimates over suitable motion segments in combination with probabilistic deep learning, as already indicated in Section 1 and further detailed below.

3. Method

First, the overall method including the interaction between the JointTracker and a newly developed deep learning based I2S orientation calibration is described. Then, the different aspects of the I2S orientation calibration are detailed.

3.1 Overall method

Figure 2 illustrates the overall method, with two interacting components, (1) the skeleton tracking comprising all tasks related to model-based estimation of IMU and segment poses, with the core part being the JointTracker according to (Taetz et al. 2025) (light blue box) and (2) the I2S orientation calibration comprising all tasks related to deep learning

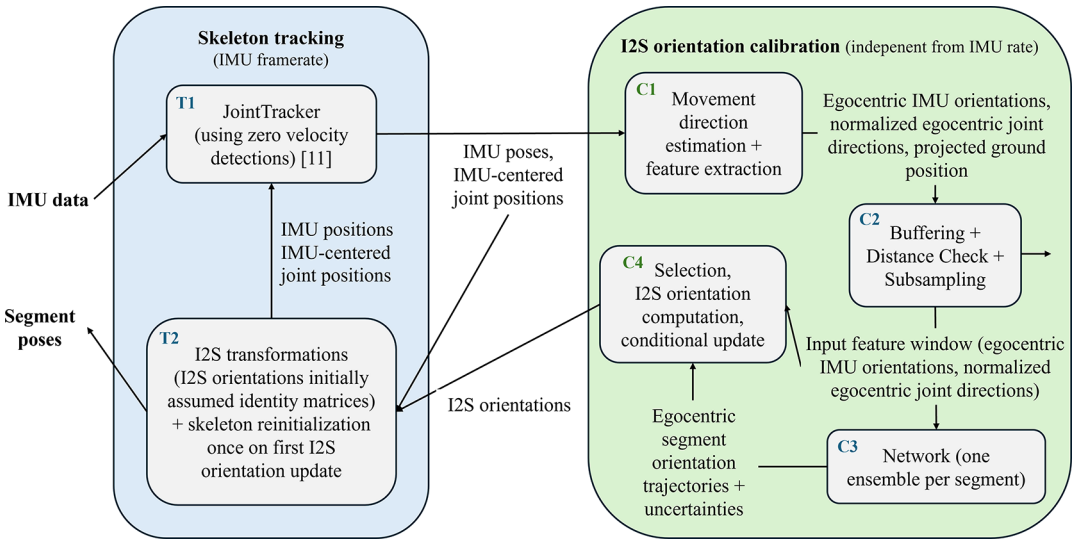

Figure 2. Illustration of the overall method.

based prediction of I2S orientations (light green box). The data exchanged by the two components is shown between the boxes (IMU poses, IMU-centered joint positions, and I2S orientations), and the inputs (IMU data) and outputs (segment poses) of the complete method are annotated on the left side of the figure.

Note, while skeleton tracking runs on IMU frame rate to deliver segment poses with minimal delay (and the currently best available I2S poses), I2S orientation calibration runs independently, on a lower rate, and it delivers I2S orientations only, if updates are available.

In the figure, the arrows indicate the data (or extracted information) flow to arrive from IMU data at segment poses. The gray boxes within each component indicate the necessary tasks, with (dark blue) labels *Ts* for skeleton tracking step *s* and (dark green) labels *Cs* for calibration step *s*. The different steps are described in the following, with the item labels corresponding to the task labels in the figure:

- ***T1*:** Synchronized IMU data at time t are processed by the JointTracker (according to (Taetz et al. 2025) with zero velocity detections and randomly initialized joint positions). This provides IMU poses for each IMU $i \in I$ and IMU-centered joint positions for each joint $(i,j) \in J$ for time t.

- ***T2*:** Initially assuming identity matrices for the I2S orientations of all segments, and updating those with I2S orientations (per segment) received from the I2S orientation calibration, segment poses are calculated from the estimated kinematic data using Equations (1), (2) and (5). Note, together with $T1$, this solves the task described in Section 2.2 after completed self-calibration. However, on the first complete I2S orientation update (i.e. once all I2S orientations have been updated at least once), the I2S orientations and IMU orientations are used together with assumed standard segment lengths (according to 1.76 meter height and the anthropometry in (Tilley and Dreyfuss 2001)) and assumed segment-centered IMU positions (IMUs sitting in the middles of the segments) to recompute the IMU positions and IMU-centered joint positions and reinitialize those states in the JointTracker. Note, this is only done once with the first updated I2S orientations to speed up convergence of the joint position estimates, since these are randomly initialized to start with.

- ***C1*:** The first task for I2S orientation calibration is to extract suitable input features for I2S orientation prediction from the kinematic data provided by $T1$. These features need to contain sufficient information for the task, but at the same time be invariant against irrelevant conditions such as global movement direction or segment lengths. The steps to achieve this are described in the following (cf. Figure 1):

  1. Global joint positions $J^N_{(i,j),t}$ are computed for each $(i,j) \in J$ by averaging the results of (3) and (4).

  2. The current pose of the egocentric reference frame G w.r.t. N is estimated from the global joint positions. The construction is based on computing a 2D projected ground position and estimating the current movement direction of the tracked kinematic chain. The details are further described in Section 3.2. Note, this subtask requires access to the kinematic data of the previous time $t-1$, i.e. a small buffer, which is omitted in the figure for better readability.

  3. All global IMU orientations $i \in I$ are transformed into the egocentric reference frame G. These are denoted egocentric IMU orientations in Figure 2, and they are represented as six-dimensional vectors $\tilde{R}^{GI_i}_t$, as introduced in Section 2.1.

  4. Normalized joint directions are calculated from the global joint positions for each joint pair $(i,j,k) \in J^2$ according to Equation (7). The resulting vectors are rotated into the egocentric reference frame G and are denoted $\tilde{J}^G_{(i,j,k),t}$ (normalized egocentric joint directions).

- ***C2*:** Since I2S orientation prediction requires some contextual motion information to reduce ambiguity, the features computed for time t (egocentric IMU orientations, normalized egocentric joint directions) are buffered and resampled to construct N_I (one per IMU) suitable input feature windows (motion segments) with T_N samples (in the experiments, 60 samples corresponding to 1 second of 60 Hz data are collected in a sliding window, each window is subsampled to 10 samples). Moreover, using the corresponding 2D projected ground positions from the previous step, a distance check is performed on a motion segment before passing this on to the next step (in the experiments, motion segments showing a traveled distance of less than 0.5 meters are omitted). This avoids predictions from uninformative input features.

- ***C3*:** This is the core of the I2S orientation calibration, where each feature window $i \in I$ provided by $C2$ with T_N samples comprising the egocentric IMU orientations of IMU i and normalized egocentric joint direction vectors of all joint pairs $(l,m,n) \in J^2$ is used to predict a motion segment of T_N egocentric segment orientations and uncertainties (variances) via sequence-to-sequence prediction from the previously trained ensemble network to obtain:

$$\mathrm{p}\left(\mathrm{y}|\tilde{R}^{GI_i}, \widehat{J}^G, X, Y\right) = N\left(\tilde{\mathrm{R}}^{GS_i}, \Sigma^{GS_i}\right). \tag{8}$$

Here, $\tilde{R}$ denotes the six-dimensional continuous orientation representation introduced in Section 2.1, (X,Y) represents the training data with inputs X and labels Y. The inputs are (as mentioned above):

- IMU orientation sequence: $\tilde{\mathrm{R}}^{GI_i} = \left[\tilde{\mathrm{R}}_1^{GI_i}, \cdots, \tilde{\mathrm{R}}_{T_N}^{GI_i}\right]$ and the corresponding
- Joint direction sequence: $\widehat{J}^G = \left[\left[\widehat{J}^G_{(l,m,n),t}\right]_{(l,m,n) \in J^2}\right]_{t=1}^{T_N}$.

The prediction (predicted normal distribution) y equals:

$$N\left(\tilde{\mathrm{R}}^{GS_i}, \Sigma^{GS_i}\right) = \left[N\left(\tilde{\mathrm{R}}_t^{GS_i}; \mu_t^{GS_i}, \Sigma_t^{GS_i}\right)\right]_{t=1}^{T_N}. \tag{9}$$

Here, Σ^{GS_i} is a six- dimensional diagonal matrix

$$\Sigma^{GS_i} = diag\left(\sigma_0^{GS_i}, \ldots, \sigma_5^{GS_i}\right), \tag{10}$$

with variances σ^{GS_i}, as predicted by the ensemble network, based on (Lakshminarayanan, Pritzel, and Blundell 2017). The network structure and training routine are detailed in Section 3.3. Note that an individual ensemble network is trained for each segment (i.e. the predictions are per segment), while the input feature vector for segment i contains IMU orientations of IMU i and of all joint directions in the kinematic chain. Within the input features, the joint directions provide information about the egocentric segment pose, which helps the network to discriminate between I2S orientation and segment orientation variations.

- ***C4*:** This task stores the best available I2S orientations and associated uncertainties (initialized with infinity) for all segments. When a new prediction is available from $C3$, per segment $i \in I$, from all T_N predicted egocentric segment orientations, the sample t with the lowest uncertainty (lowest maximum variance over orientation parameters at time t) is selected, and, together with the corresponding input egocentric IMU orientation, this is used to compute a new I2S orientation $R^{SI_i} = \left(R_t^{GS_i}\right)^T R_t^{GI_i}$. Here, the rotation matrices are reconstructed from the six-dimensional representations using the cross product. For each segment, if the current uncertainty is smaller compared to the stored one, the new I2S orientation is stored with uncertainty. New I2S orientations are provided to $T2$.

Note, for the complete method, once self-calibration has completed, in order to reduce the computational burden, the JointTracker can be replaced by the tracking method proposed in (Miezal, Taetz, and Bleser 2017) (without joint center estimation) using the calibrated I2S transformations. In the current implementation, self-calibration is considered as completed, if all IMU-centered joint position estimates have converged according to the convergence indicator proposed in (Taetz et al. 2025).

3.2 Construction of egocentric frame *G*

As mentioned in Section 2.1 and illustrated in Figure 1, the egocentric reference frame G moves with the tracked person in terms of position and heading direction, and, like N, its z-axis is aligned with gravity. Its construction is based on the global joint position estimates as derived from the JointTracker estimates.

The position g_t^N of G in N at time t, is approximated as the average global joint position $\frac{1}{N_J - 1}\sum_{(i,j) \in J} J^N_{(i,j),t}$ projected to the ground (i.e. with the z-coordinate set to zero).

The heading direction (initialized with a zero vector) at time t is approximated as the smoothed velocity

$$\dot{g}_t^N = \alpha \frac{g_t^N - g_{t-1}^N}{\Delta t} + (1-\alpha)\dot{g}_{t-1}^N, \tag{11}$$

where Δt is the sampling time and $\alpha = 0.5$ in the experiments. The x-axis (heading direction) of G in N is aligned with the resulting velocity vector, i.e. $x_{G,t}^N = \dot{s}_t^N$. Due to the gravity alignment, $z_G^N = [0,0,1]^T$ (independently of t). Then R_t^{NG} can be constructed as

$$R_t^{NG} = \left[x_{G,t}^N \quad z_{G,t}^N \times x_{G,t}^N \quad z_{G,t}^N \right]. \tag{12}$$

3.3 Deep learning based IMU-to-segment orientation prediction with uncertainties

After shortly summarizing the mathematical background of Deep Ensembles, first, the inference with uncertainties is described, then the training data generation, network architecture and training are detailed.

3.3.1 Deep Ensembles

To obtain reliable predictive uncertainties, we adopt Deep Ensembles (Lakshminarayanan, Pritzel, and Blundell 2017) and model the conditional predictive distribution for a target vector $y \in R^D$ given an input x as a Gaussian

$$p(y|x,X,Y) \approx N(y;\mu(x),\Sigma(x)), \tag{13}$$

where (X,Y) denotes the training set. An ensemble consists of M neural networks $f_i = \{f^m\}_{m=1}^M$ with parameters W^m, trained independently (different initializations and data shuffling). Each network outputs (1) a mean vector and (2) a diagonal (heteroscedastic) covariance:

$$[\mu_m(x), s_m(x)] = f_{W^m}^m(x), m = 1, \ldots, M, \tag{14}$$

with $\mu_m(x) \in R^D$ and $s_m(x) \in R^D$ the log-variances. The predicted diagonal covariance of network m is

$$\Sigma_m(x) = diag(exp(s_m(x))), \tag{15}$$

which represents aleatoric uncertainty (input-dependent noise) (Kendall and Gal 2017).

Training objective. Each ensemble member is trained by minimizing the negative log-likelihood under a Gaussian noise model (heteroscedastic regression). For training pairs $\{(x_j, y_j)\}_{j=1}^N$, the loss for member m is

$$L_{ENS}(W^m) = \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{2} (y_j - \mu_m(x_j))^\top \Sigma_m(x_j)^{-1} (y_j - \mu_m(x_j)) + \frac{1}{2} log\left(det(\Sigma_m(x_j))\right) \right) + const. \tag{16}$$

Since $\Sigma_m(x)$ is diagonal, this corresponds to a weighted squared error plus a regularizer on the predicted variances, and the aleatoric uncertainty is learned implicitly (no uncertainty labels required) (Kendall and Gal 2017).

Predictive mean and variance (Gaussian mixture moments). At test time, the ensemble defines a mixture of Gaussians. We approximate it by a single Gaussian via moment matching. The predictive mean for an input x_s is

$$\mu_{ENS}(x_s) = \frac{1}{M} \sum_{m=1}^M \mu_m(x_s), \tag{17}$$

and the predictive covariance is

$$\Sigma_{ENS}(x_s) = \frac{1}{M} \sum_{m=1}^M \Sigma_m(x_s) + \frac{1}{M} \sum_{m=1}^M \mu_m(x_s) \mu_m(x_s)^\top - \mu_{ENS}(x_s) \mu_{ENS}(x_s)^\top. \tag{18}$$

Thus, Σ_{ENS} captures both average aleatoric uncertainty and epistemic uncertainty via the disagreement between ensemble members (Lakshminarayanan, Pritzel, and Blundell 2017). In our implementation we retain only the diagonal entries of Σ_{ENS}, consistent with the diagonal output parameterization.

3.3.2 I2S prediction per motion segment

As already described for task $C2$ in Section 3.1, orientation prediction requires contextual motion information to reduce ambiguity. Therefore, the per-timestep features computed at time t (egocentric IMU orientations and normalized egocentric joint directions) are buffered into sliding windows and resampled to construct, for each IMU/segment $i \in I$, an input feature window (motion segment) of length T_N. In the experiments, 60 samples (corresponding to 1 second at 60 Hz) are collected and then subsampled to 10 samples per window. In addition, using the 2D projected ground positions from the previous step, we apply a distance check and omit windows with traveled distance below 0.5 m, which avoids predictions from uninformative motion.

Per segment ensemble predictor. We train an individual Deep Ensemble for each segment $i \in I$. The resulting ensemble predictor f_i maps the windowed input sequences to a distribution over a sequence of egocentric segment orientations, with the symbols being introduced already with Equations (8) to (10):

$$N\left(\tilde{\mathrm{R}}^{GS_i}, \Sigma^{GS_i}\right) = \mathrm{f}_i\left(\tilde{\mathrm{R}}^{GI_i}, \hat{J}^{G}\right). \tag{19}$$

The output is a per-time step Gaussian distribution

$$N\left(\tilde{\mathrm{R}}^{GS_i}, \Sigma^{GS_i}\right) = \left\{N\left(\tilde{R}_t^{GS_i}; \mu_t^{GS_i}, \Sigma_t^{GS_i}\right)\right\}_{t=1}^{T_N}, \tag{20}$$

where $\mu_t^{GS_i} \in R^6$ and $\Sigma_t^{GS_i} \in R^{6\times 6}$ is diagonal. Both are obtained by deep ensemble moment matching, introduced in Equations (17) to (18) applied at each timestep t.

Why sequence-to-sequence instead of direct I2S prediction. Rather than predicting a single I2S rotation directly from the full window (as in direct approaches (Zimmermann, Taetz, and Bleser 2018; Zuo et al. 2025)), we cast the task as sequence-to-sequence prediction and estimate an egocentric segment orientation distribution for each timestep. Empirically, direct window-to-one-rotation regression yielded higher predictive variances and less reliable predictions, especially for motions underrepresented in the training set. In contrast, the per-timestep distributions typically provide more homogeneous uncertainty behavior and enable robust selection of informative timesteps for I2S extraction. The most confident I2S rotation is then extracted from the predicted distribution as described for task $C4$ in Section 3.1.

3.3.3 Network Architecture

Each ensemble member uses the same sequence-to-sequence architecture as depicted in Figure 3. The design reflects that the input contains

1. a time-varying component (the motion of the kinematic chain) and
2. an approximately time-constant component (the I2S rotation).

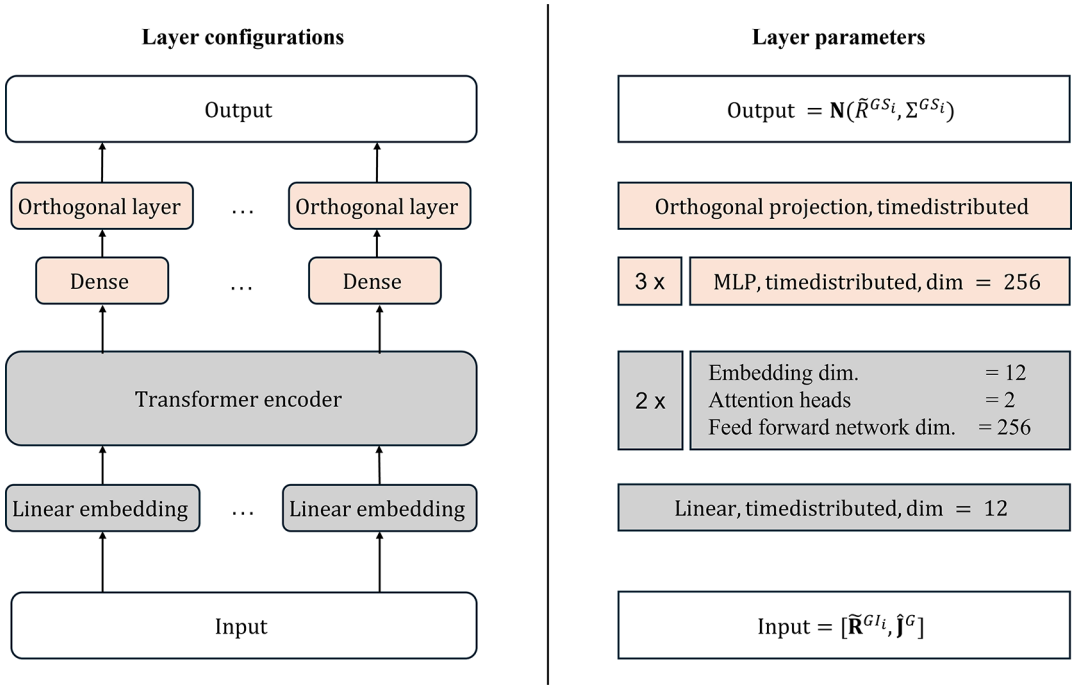

Figure 3. Network architecture with the layer configuration (left) and the corresponding parameters (right). The orange-colored boxes model the time-constant parts, and the gray colored boxes model the time-varying parts.

Concretely, for each timestep, the input features are first linearly projected into an embedding space. We add sinusoidal positional encodings and process the sequence using a stack of Transformer encoder blocks (Vaswani et al. 2017) to model temporal dependencies. The Transformer outputs are passed through a time-distributed Multilayer Perceptron (MLP) head (three fully connected layers) that produces, at each timestep, (1) the mean of the 6D rotation representation and (2) the corresponding log-variances. Finally, we apply an orthonormalization layer as in (Zhou et al. 2019) to ensure valid 6D rotation outputs (two orthonormal 3-vectors), consistent with the representation used in task $C3$.

For each segment/IMU $i \in I$ we use an ensemble of $M = 5$ independently trained networks and compute the per-timestep predictive mean and covariance via deep ensemble moment matching, see (17, 18) or (Lakshminarayanan, Pritzel, and Blundell 2017).

3.3.4 Data preparation and network training

Our training methodology employs a self-supervised approach requiring only a dataset of global segment kinematics, specifically global segment orientations and joint positions, as input data. The kinematic chain definition used must remain consistent between the training phase and subsequent inference operations. In our experimental setup, we utilized the lower body kinematics from the musculoskeletal model described in (Rajagopal et al. 2016), with $N_I = 7$ segments/IMUs.

Data format. During the training process the input consists of motion sequences of fixed length T_N, comprising an egocentric 6D IMU orientation $\tilde{R}_t^{GI} \in R^6$ and normalized egocentric joint direction vectors $\hat{J}_t^G \in R^{(N_I-2)\times 3}$ for each time step. When concatenated and stacked, the input for a single motion sequence is represented as $\left[\tilde{R}_t^{GI}, \hat{J}_t^G\right]_{t=1}^{T_N} \in R^{T_N \times 3N_I}$. As already mentioned for task $C2$ in Section 3.1, in our experiments we set $T_N = 10$, corresponding to sub-sampled data from one second of kinematic recordings at 60 Hz by a factor of 6, effectively simulating data recording at 10 Hz. Each motion sequence had an overlap of 9 samples, equivalent to online buffering at 10 Hz.

Training setup. All experiments were conducted on an AMD Ryzen 9 9950X processor with an NVIDIA RTX 5090 GPU. The network ensemble was implemented in Tensorflow 2.21.0 and CUDA 12.8. The training batch size was configured to 2000 with an initial learning rate of 10^{-4} utilizing AdamW optimzer with default parameters. The model was trained for 5 epochs, with evaluation performed after each epoch to select the best-performing model as the final output.

Algorithm 1 outlines the process for generating training data from a single segment sequence. The algorithm operates in a sliding window fashion over the kinematic data input. Notably, in Experiment 2, we extended our approach through simple transfer learning by fine-tuning the best-performing model on new simulated datasets (cf. Section 4.2).

Algorithm 1. Training data generation.

Input: Data sequence of global segment kinematics (global segment orientations, global joint positions) with N samples (timesteps).

Output: Training data sets T_i with inputs (motion segments with T_N egocentric orientations of IMU i and T_N egocentric normalized joint direction sequences) and corresponding outputs (motion segments with T_N egocentric orientations of segment i) for each segment $i \in I$.

$t \leftarrow 0$

$T_i \leftarrow$ empty, for all $i \in I$

while t < N − T_N do

    Collect a window of T_N global segment orientations and joint positions $(t \ldots t + T_N - 1)$.

    Construct egocentric frame G from the joint positions according to Section 3.2.

    for each sample in the window do

        Compute egocentric segment orientations for each segment $i \in I$.

Compute egocentric normalized joint direction vectors according to $C1(4)$ in Section 3.1 and (7).
end for
if distance check fails ($C2$ in Section 3.1) **then**
$t \leftarrow t+1$
continue
end if
for each segment i **do**
Create an output motion segment with T_N egocentric orientations for segment i.
function ***SIMULATE I2S ORIENTATIONS***
Generate K uniform random samples on a unit sphere and construct the respective normals as 3-vectors ($K = 1000$ in the experiments).
for each random sample **do**
Generate D I2S orientations by aligning the virtual IMU coordinate system's z-axis (z_i^I) with the sampled normal and adding D rotations around the normal, each by an angle of $\frac{360}{D}$ degrees ($D = 8$ in the experiments).
end for
Reject new orientation samples, which are too close to a previous sample (5° in the experiments).
return Simulated I2S orientations ($K \cdot D$)
end function
for each simulated I2S orientation **do**
Compute the T_N egocentric IMU orientations for (virtual) IMU i according to (1).
Create an input motion segment from the egocentric IMU orientations and the above constructed egocentric normalized joint direction vectors. Note, while the egocentric IMU orientations are different for each simulated I2S orientation, the normalized egocentric joint directions are the same, representing the body pose. This enables better discrimination between I2S orientation variation and motion as compared to relying purely on the egocentric IMU orientations alone. Note that we do not include direct IMU sensor information as input features (in contrast to (Zimmermann, Taetz, and Bleser 2018), (Zuo et al. 2025)) to allow for training data generation from any kinematic chain sequences.
Add the input and output motion segments to T_i.
end for
end for
$t = t+1$
end while
return T_i for all $i \in I$.

4. Evaluation

This section presents an experimental evaluation of the proposed method on the TUK-6-minute-walking dataset (Teufl et al. 2023), the lower body human gait dataset also used in (Taetz et al. 2025), and on newly collected lower body IMU data. The capturing setup of the TUK-6-minute-walking dataset is illustrated in Figure 1 (right). The dataset contains synchronized lower body IMU and optical reference data from 25 young, healthy test persons (13 females) and two sessions. The hardware setup consisted of twelve OptiTrack Prime 13 cameras (NaturalPoint, Inc., Corvallis, OR, USA) and seven XSens MTw Awinda (Xsens Technologies BV, Enschede, The Netherlands) IMUs.

We provide three levels of evaluation:

- **Experiment 1:** Stochastic cross validation of the I2S orientation calibration (light green box on the right side of Figure 2) using the optical data (global segment kinematics) from the gait dataset (Teufl et al. 2023). For this, we used the first 2000 samples of the first recording session per person. We randomly selected one person for testing and trained a Deep Ensemble on the other 24 persons (recordings) as described in Section 3.3. We tested the I2S orientation calibration (steps $C1$ to $C4$ in Figure 2) on 100 randomly generated I2S orientations (corresponding IMU orientation trajectories) for the test person. The complete test was performed three times (for three

randomly selected test persons). This is an isolated quantitative evaluation of the I2S orientation calibration. The results are detailed in Section 4.1.

- **Experiment 2:** Quantitative evaluation of the self-calibration capability of the proposed method as illustrated in Figure 2 (but excluding $T2$). For this, the Deep Ensemble was trained on all 25 recordings of the gait dataset (segment kinematics from the optical data) as described in Section 3.3, and the inference was based on the IMU and joint kinematics as estimated by the JointTracker (processing the real IMU data). The JointTracker was used with the settings described in (Taetz et al. 2025). Per recorded session (first 2000 samples), the most certain I2S orientation (as selected by the I2S orientation calibration) was compared to the reference I2S orientation obtained from the optical data (according to the actual IMU mounting of the person). This experiment is also a sim-to-real test, since the training was purely based on simulated IMU data, while the inference was based on kinematics estimated from real IMU data. The results are presented in Section 4.2. Moreover, we also illustrate the uncertainty-based selection ($C4$ in Figure 2) for one exemplary test person and segment in this section.

- **Experiment 3:** Qualitative evaluation of the overall method being applied to newly collected lower body gait data with unseen motions via putting on the pants while recording and simulated gait deviations. The results (including handover from self-calibration to pure tracking) are presented in Section 4.3.

The quantitative evaluations focus on I2S orientation errors (differences) $R_{err} := R_{ref} R_{pred}^T$, where R_{ref} is a reference orientation and R_{pred} is the predicted orientation. As evaluation metric, we use means and standard deviations over complete angle differences $\phi(R_{err}) = \arccos\left(\frac{\mathrm{trace}(R_{err})-1}{2}\right)$ (Miezal, Taetz, and Bleser 2016).

The focus on I2S orientations in this article is for the following reasons: (1) Joint position and segment length estimates (which are the basis for I2S positions) are already evaluated in (Taetz et al. 2025), and (2) I2S orientation errors linearly propagate into joint-angle errors and are the dominant source of error for anatomical joint angle tracking (Miezal, Taetz, and Bleser 2016).

4.1 Experiment 1: Stochastic cross-validation of I2S orientation calibration (quantitative)

The results of the stochastic cross-validation in terms of angle differences between the selected and the ground truth I2S orientation for each segment are shown in Table 1. The overall mean errors range between 1.87 degrees for RightUpperLeg and 6.06 degrees for LeftUpperLeg, with standard deviations (SD) ranging from 0.50 degrees for Pelvis to 2.58 degrees for LeftFoot. Overall, the mean errors and standard deviations are lower compared to those reported based on static and functional calibration methods. For instance, the review in (Pacher et al. 2020) reports mean angle errors between 4.3 and 17 degrees for the lower body segments, with standard errors (not standard deviations) ranging between 1.7 and 5.17 degrees. Among the self-calibration approaches, (Zimmermann, Taetz, and Bleser 2018) (Section 4.3.2, dense IMU placement) reports a mean angle error of 21.35 degrees over all lower body segments as best results, and (Zuo et al. 2025) (sparse IMU placement) reports angle errors between 7.76 and 14.33 degrees for the lower body, which are both higher. Note that the results in Table 1 are purely based on the optical data as initial plausibility test though.

4.2 Experiment 2: Evaluation on real IMU data (quantitative)

The results of the quantitative evaluation of the overall self-calibration method as illustrated in Figure 2 (but excluding $T2$) in terms of angle differences between the selected and the ground truth I2S orientation for each segment on real IMU data are shown in Tables 2 and 3.

For the results in Table 2, the Deep Ensemble was trained on all 25 recordings of the gait dataset using global segment kinematics from the optical data as described in Algorithm 1, and the inference was based on the IMU and joint kinematics as estimated by the JointTracker (processing the real IMU data). In Table 2, the angle differences are clearly higher compared to those in Table 1, due to the processing of real IMU data through the JointTracker. The overall mean errors range between 8.36 degrees for Pelvis and 19.56 degrees for LeftUpperLeg, with standard deviations (SD) ranging from 5.23 degrees for LeftLowerLeg to 11.66 degrees for LeftUpperLeg. Overall, these mean errors are slightly higher compared to those reported based on static and functional calibration methods (mean angle errors between 4.3 and 17 degrees for the lower body segments according to (Pacher et al. 2020). However, the mean over all segments (13.20 degrees) is still considerably lower compared to the best mean over all lower body segments reported in (Zimmermann, Taetz, and Bleser 2018) for the dense IMU placement (21.35 degrees). Moreover, the means for Pelvis (8.36 degrees), LeftLowerLeg (10.77 degrees) and RightLowerLeg (14.12 degrees) are comparable to the mean angle errors reported in (Zuo et al. 2025) for the sparse IMU placement (7.76 degrees for hip (Pelvis), 14.33 degrees for left lower leg, 14.32 degree for right lower leg). Note, none of these report standard deviations.

Table 1. Experiment 1: Means and standard deviations (SD) for angle differences for all segments.

Segment	Mean [*deg*]	SD [*deg*]
LeftFoot	4.0814	2.5757
LeftLowerLeg	5.5963	2.0005
LeftUpperLeg	6.0642	1.5963
Pelvis	2.4838	0.4953
RightFoot	5.7076	1.6810
RightLowerLeg	3.7262	0.8989
RightUpperLeg	1.8727	0.7080
All	4.2189	1.4222

Table 2. Experiment 2: Means and standard deviations (SD) for angle differences for all segments.

Segment	Mean [*deg*]	SD [*deg*]
LeftFoot	16.7662	7.6509
LeftLowerLeg	10.7734	5.2277
LeftUpperLeg	19.5629	11.6582
Pelvis	8.3571	5.7656
RightFoot	10.0735	5.4510
RightLowerLeg	14.1207	9.5661
RightUpperLeg	12.7640	9.3954
All	13.2025	7.8164

Table 3. Experiment 2 augmented: Means and standard deviations (SD) for angle differences for all segments.

Segment	Mean [*deg*]	SD [*deg*]
LeftFoot	12.6134	5.5802
LeftLowerLeg	8.6141	4.6973
LeftUpperLeg	13.3848	8.5642
Pelvis	12.6882	8.2499
RightFoot	14.8571	8.4247
RightLowerLeg	11.5967	5.1679
RightUpperLeg	9.7000	5.8950
All	11.9220	6.6542

For the results in Table 3, the trained Deep Ensemble was finetuned with augmented training data, better reflecting the influence of the JointTracker in the input features (during training). For this, Algorithm 1 was slightly extended to use an additional Input (line 1) for creating the input motion segments (line 28). This additional input consists of global segment kinematics as obtained from the JointTracker, when processing IMU data simulated from the optical segment kinematics using identity matrices as I2S orientations. IMU data simulation was done as in (Miezal, Taetz, and Bleser 2016). Note that this augmentation doesn't use additional data recordings nor real IMU data or mountings. The whole training procedure is still self-supervised, i.e. purely based on simulated IMU data and mountings.

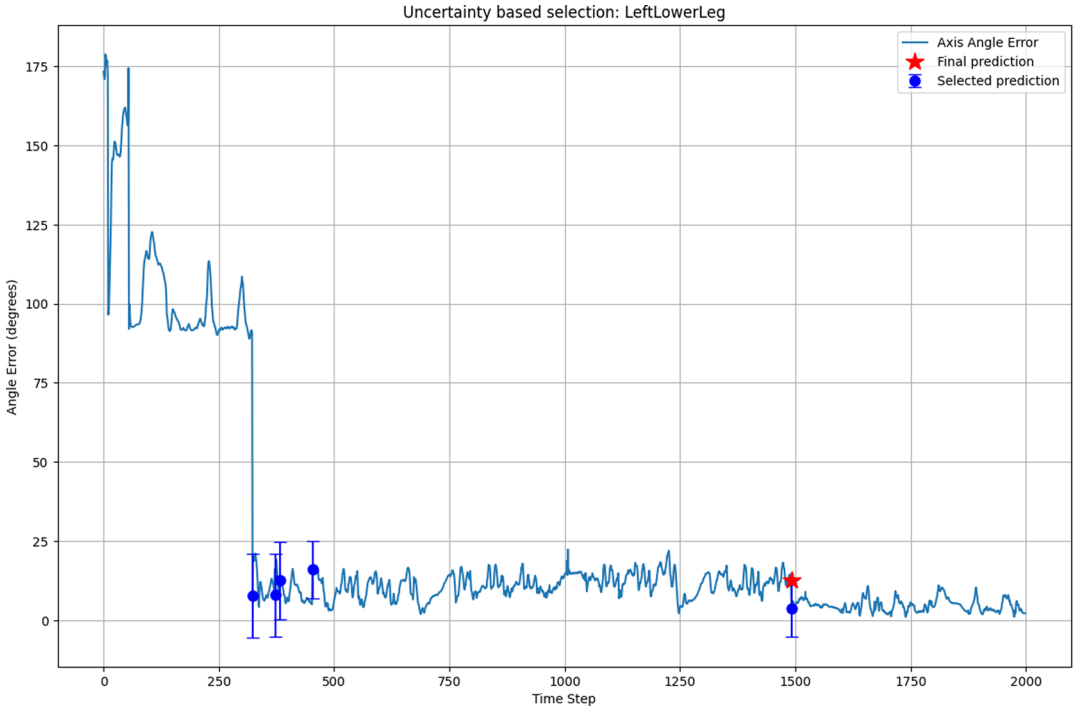

Figure 4. Uncertainty based selection exemplified for one test person and LeftLowerLeg.

This augmentation improved the overall results (reduced mean and standard deviation over all segments), and, importantly, it resulted in lower maximum mean errors. In Table 3, the overall mean errors range between 8.61 degrees for LeftLowerLeg and 14.86 degrees for RightFoot. These mean errors are comparable to those reported based on static and functional calibration methods (highest mean error is lower, see above), lower compared to (Zimmermann, Taetz, and Bleser 2018) for the dense IMU placement and comparable to (Zuo et al. 2025) for the sparse IMU placement.

Figure 4 illustrates the uncertainty based I2S selection for one test person and one segment (LeftLowerLeg). The blue line shows the angle difference (Axis Angle Error) between the currently assumed and the ground truth I2S orientation per timestep. Each I2S orientation update ($C4$ in Figure 2) is shown by a blue dot with indicated prediction uncertainty. In the beginning, the I2S orientation is assumed to be an identity matrix until the first selection. Therefore, the angle error drops at the first selection. The first selection happens at the first time $C4$ is executed, i.e. the first time enough data has been collected in the buffer via $C1$ to $C3$. The figure then shows four additional I2S orientation updates (selections) due to smaller prediction uncertainties. The finally selected I2S orientation is marked with a red star and is indeed at a low error level. This demonstrates the I2S orientation selection procedure.

Inference Time. The model achieves an average inference time of 3.53 milliseconds per segment on a single AMD Ryzen 9 9950X CPU, including computations for mean and variance as well as uncertainty-based selection. It is worth noting that all motion segments can be processed independently and simultaneously, enabling real-time inference.

4.3 Experiment 3: Overall method (qualitative)

The complete system as described in Section 3.1 was implemented in C++ for online usage with Xenoma's e-skin MEVA pants (Xenoma Inc., n.d.). For qualitative evaluation, two data sequences were collected, starting with a person putting on the IMU pants and then starting to walk, once normally (sequence 1) and once with a simulated gait deviation, a circumduction of the right hip (sequence 2). The test person gave her written informed consent to publish her details and collected data. The images in Figure 5 show single frames from sequence 2: an external view of the person and a screenshot from the skeleton visualization with overlaid IMUs and convergence indicators (bubbles around the IMUs). The images are from three phases:

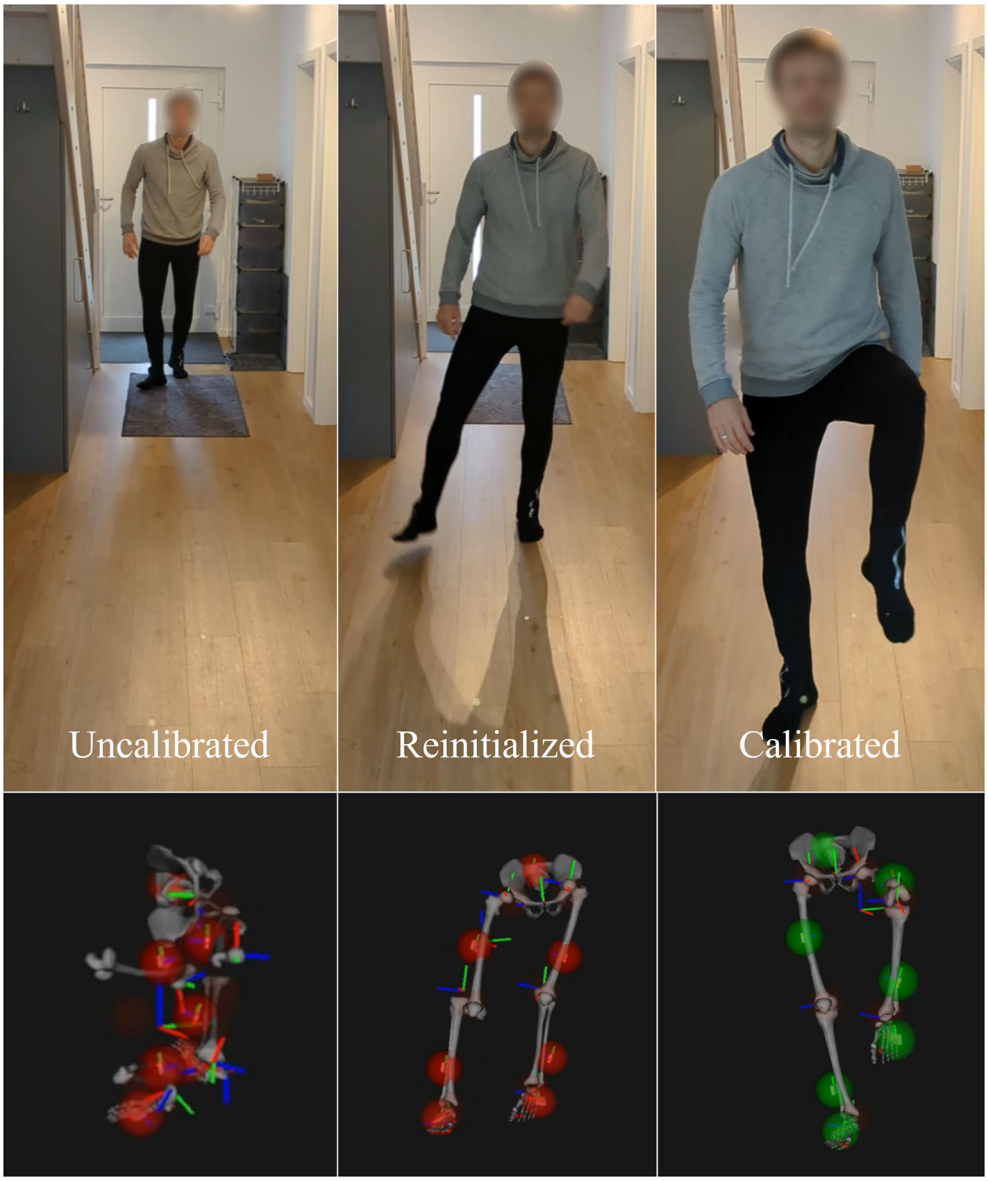

Figure 5. Experiment 3: Qualitative results from the overall system. The underlying data sequence and the external video were recorded by the authors. The screenshots are from the graphical user interface of the C++ implemented tracking system.

1. Uncalibrated: Segments have arbitrary orientations and are not attached at the joints, bubbles are red.

2. After skeleton reinitialization ($T2$ in Figure 2): Segments are roughly aligned and roughly attached at the joints, bubbles are still red (uncalibrated).

3. Calibrated, after handover to pure kinematics estimation (without joint center estimation): Segments are visually well aligned and attached at the joints, bubbles are green, indicating that self-calibration has completed.

For this specific sequence, three to four gait cycles corresponding to about 6 seconds were required until self-calibration was considered completed. A video showing both sequences is available under https://youtu.be/o1uQaR5wvYA. Note that, even with simulated gait deviations, which were not available in the training dataset, the I2S selection resulted in a

visually good calibration. Also, the person putting on the pants at the beginning of the first sequence did not disturb the calibration process.

5. Conclusion and open challenges

This article proposes and evaluates the enhancement of JointTracker, a previously published online sensor fusion method for global IMU orientation and IMU-centered joint position estimation in kinematic chains, with a newly developed probabilistic deep learning based I2S orientation prediction to obtain biomechanically interpretable joint kinematics.

Together, this creates a self-calibrating online inertial human motion capture method for a dense IMU placement, which does not depend on the user accurately performing pre-defined poses or movements for I2S calibration, but just to walk a few steps. This is an important step towards a plug-and-play system.

The per segment I2S orientation calibration is based on an ensemble network for sequence-to-sequence prediction which is trained purely on simulated IMU mountings and data. Hence, the training routine only uses recorded joint kinematics as they are available in various online accessible datasets (e.g. AMASS). It does not require real IMU data or IMU mountings for training. This is thanks to the fact that only higher-level input features (fused IMU orientations and joint directions) are used instead of IMU measurements, as e.g. in (Zimmermann, Taetz, and Bleser 2018; Zuo et al. 2025). Moreover, the ensemble network architecture enables probabilistic prediction and, based on this, an active I2S orientation selection using the predicted uncertainties. This can be used to avoid updates in the presence of unseen motions, which typically cause large uncertainties.

The proposed method is evaluated quantitatively on an existing lower body gait dataset, being trained purely on simulated IMU data from this dataset. After confirming the proper functioning of the I2S orientation calibration in isolation (experiment 1), testing the complete method on real IMU data (experiment 2) shows comparable results to those reported for supervised static pose or functional calibration methods, while the upper error bound is even lower. Moreover, it outperforms a previous self-calibration approach for a dense IMU placement (Zimmermann, Taetz, and Bleser 2018) and provides comparable results to a previous recent self-calibration approach for a sparse IMU placement (Zuo et al. 2025) (which provides results only for lower legs and pelvis), though both previous works used higher amounts of training data, i.e. CMU (CMU Graphics Lab, n.d.) in (Zimmermann, Taetz, and Bleser 2018) and AMASS (Mahmood et al. 2019) in (Zuo et al. 2025), and mixed simulated with real IMU data for training in order for the network to adapt to the IMU data characteristics (cf.(Zuo et al. 2025), Section 5.1). Furthermore, a qualitative evaluation also shows visually good results when applying the overall method to a newly collected lower body gait data with simulated gait deviations (experiment 3), thus supporting the positive impact of the active selection. Consequently, the proposed method extend upon earlier self-calibration works (Zimmermann, Taetz, and Bleser 2018; Zuo et al. 2025) through improved feature integration, new network architecture, predictive accuracy and estimation robustness via uncertainty based I2S selection.

As is visible in Figure 4, the uncertainty-based selection does not necessarily select the I2S orientation with the smallest error. Therefore, a deeper investigation of the predicted uncertainties and the selection criterion, based on a larger training dataset with a greater variety of movements, will be part of our future work. Moreover, while the current evaluation focuses on a static I2S calibration throughout a recording session, future work will focus on extending this to a dynamic recalibration, e.g. after an IMU has shifted. Finally, while the evaluation focused on the lower body, the proposed method is not limited to this and can be extended to work with the full body as well.

Software availability

Source code for data visualization and inspection available from: https://github.com/BTaetz/Visualizer-SkeletalMotionHDF5

License: MIT

Ethics and consent

The TUK-6-minute-walking dataset (Teufl et al. 2023) study was approved by the Ethics Committee of the Department of Social Sciences at Rheinland-Pfälzische Technische Universität (RPTU), formerly known as Technische Universität Kaiserslautern (TUK) (ethics application: “Studie 1: Validierung eines mobilen Inertial Sensor Systems gegen ein optisches Infrarotkamerasystem an jungen, gesunden Probanden, Studie 2: Testung der Reliabilität eines mobilen Inertial Sensor Systems”, date of approval 5.12.2017), and meets the criteria of the declaration of Helsinki. After receiving all relevant study information, the participants signed a written informed consent to the study including a permission to publish the data.

Data availability

Underlying data

Zenodo: Underlying data for 'Inertial human motion capture with automatic IMU-to-segment position and orientation calibration using JointTracker and deep learning with uncertainties', https://doi.org/10.5281/zenodo.10253111 (Teufl et al. 2023) (TUK-6-minute-walking dataset)

This project contains the following underlying data:

- Description.pdf
- BVH 6 min walking test.zip
- HDF5 6 min walking test.zip

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0)

References

Bleser G, Damen D, Behera A, *et al.*: **Cognitive Learning, Monitoring and Assistance of Industrial Workflows Using Egocentric Sensor Networks.** *PLoS One.* 2015; **10**(6): e0127769.
Publisher Full Text

Bleser G, Taetz B, Lukowicz P: **Human Motion Capturing and Activity Recognition Using Wearable Sensor Networks.** *Developing Support Technologies: Integrating Multiple Perspectives to Create Assistance That People Really Want.* Karafillidis A, Weidner R, editors. Cham: Springer International Publishing; 2018; 191–206.
Publisher Full Text

Bouvier B, Duprey S, Claudon L, *et al.*: **Upper Limb Kinematics Using Inertial and Magnetic Sensors: Comparison of Sensor-to-Segment Calibrations.** *Sensors.* 2015; **15**(8): 18813–18833.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Caserman P, Garcia-Agundez A, Göbel S: **A Survey of Full-Body Motion Reconstruction in Immersive Virtual Reality Applications.** *IEEE Trans. Vis. Comput. Graph.* 2020; **26**(10): 3089–3108.
PubMed Abstract | **Publisher Full Text**

Cereatti A, Gurchiek R, Mündermann A, *et al.*: **ISB Recommendations on the Definition, Estimation, and Reporting of Joint Kinematics in Human Motion Analysis Applications Using Wearable Inertial Measurement Technology.** *J. Biomech.* 2024; **173**: 112225.
Publisher Full Text

CMU Graphics Lab: **CMU Graphics Lab Motion Capture Database.** n.d.
Reference Source

Favre J, Aissaoui R, Jolles BM, *et al.*: **Functional Calibration Procedure for 3D Knee Joint Angle Description Using Inertial Sensors.** *J. Biomech.* 2009; **42**(14): 2330–2335.
Publisher Full Text

Huang Y, Kaufmann M, Aksan E, *et al.*: **Deep Inertial Poser: Learning to Reconstruct Human Pose from Sparse Inertial Measurements in Real Time.** *ACM Trans. Graph.* 2018; **37**(6).
Publisher Full Text

Kendall A, Gal Y; **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?.** *NIPS.* Guyon I, von Luxburg U, Bengio S, *et al.*, editors. 2017; 5574–84.

Laidig D, Lehmann D, Bégin M-A, *et al.*: **Magnetometer-Free Realtime Inertial Motion Tracking by Exploitation of Kinematic Constraints in 2-DoF Joints.** *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* 2019; 1233–38.
Publisher Full Text

Lakshminarayanan B, Pritzel A, Blundell C: **Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles.** *Advances in Neural Information Processing Systems.* Guyon I, Von Luxburg U, Bengio S, *et al.*, editors. Curran Associates, Inc; 2017; **30**.

Liu Y-T, Zhang Y-A, Zeng M: **Sensor to Segment Calibration for Magnetic and Inertial Sensor Based Motion Capture Systems.** *Measurement.* 2019; **142**: 1–9.
Publisher Full Text

Lorenz M, Bleser G, Akiyama T, *et al.*: **Towards Artefact Aware Human Motion Capture Using Inertial Sensors Integrated into Loose Clothing.** *2022 International Conference on Robotics and Automation (ICRA).* 2022; 1682–88.
Publisher Full Text

Mahmood N, Ghorbani N, Troje NF, *et al.*: **AMASS: Archive of Motion Capture As Surface Shapes.** *2019 IEEE/CVF International Conference on Computer Vision (ICCV).* Seoul, Korea (South): IEEE; 2019; 5441–5450.
Publisher Full Text

Menolotto M, Komaris D-S, Tedesco S, *et al.*: **Motion Capture Technology in Industrial Applications: A Systematic Review.** *Sensors.* 2020; **20**(19).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Miezal M, Taetz B, Bleser G: **On Inertial Body Tracking in the Presence of Model Calibration Errors.** *Sensors.* 2016; **16**(July): 1132.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Miezal M, Taetz B, Bleser G: **Real-Time Inertial Lower Body Kinematics and Ground Contact Estimation at Anatomical Foot Points for Agile Human Locomotion.** *2017 IEEE International Conference on Robotics and Automation (ICRA).* 2017; 3256–63.
Publisher Full Text

Müller P, Bégin M-A, Schauer T, *et al.*: **Alignment-Free, Self-Calibrating Elbow Angles Measurement Using Inertial Sensors.** *IEEE J. Biomed. Health Inform.* 2017; **21**(2): 312–319.
PubMed Abstract | **Publisher Full Text**

Nazarahari M, Noamani A, Ahmadian N, *et al.*: **Sensor-to-Body Calibration Procedure for Clinical Motion Analysis of Lower Limb Using Magnetic and Inertial Measurement Units.** *J. Biomech.* 2019; **85**: 224–229.
PubMed Abstract | **Publisher Full Text**

Pacher L, Chatellier C, Vauzelle R, *et al.*: **Sensor-to-Segment Calibration Methodologies for Lower-Body Kinematic Analysis with Inertial Sensors: A Systematic Review.** *Sensors.* 2020; **20**(11): 3322.
Publisher Full Text

Palermo E, Rossi S, Marini F, *et al.*: **Experimental Evaluation of Accuracy and Repeatability of a Novel Body-to-Sensor Calibration Procedure for Inertial Sensor-Based Gait Analysis.** *Measurement.* 2014; **52**: 145–155.
Publisher Full Text

Rajagopal A, Dembia CL, DeMers MS, *et al.*: **Full-Body Musculoskeletal Model for Muscle-Driven Simulation of Human Gait.** *IEEE Trans. Biomed. Eng.* 2016; **63**(10): 2068–2079.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schepers M, Giuberti M, Bellusci G: *Xsens MVN: Consistent Tracking of Human Motion Using Inertial Sensing.* Xsens Technologies B.V.; 2018.
Publisher Full Text

Seel T, Raisch J, Schauer T: **IMU-Based Joint Angle Measurement for Gait Analysis.** *Sensors (Basel, Switzerland).* 2014; **14**(April): 6891–6909.
Publisher Full Text

Taetz B, Bleser G, Miezal M: **Towards Self-Calibrating Inertial Body Motion Capture.** *19th International Conference on Information Fusion, FUSION 2016, Heidelberg, Germany, July 5–8, 2016.* IEEE; 2016; 1751–59.
Reference Source

Taetz B, Lorenz M, Miezal M, *et al.*: **JointTracker: Real-Time Inertial Kinematic Chain Tracking with Joint Position Estimation [Version 2; Peer Review: 3 Approved].** *Open Res Europe.* 2025; **4**(33).
Publisher Full Text

Teufl W, Miezal M, Taetz B, *et al.*: TUK-6-Minute-Walking Dataset. [Dataset]. *Zenodo.* 2023.
Publisher Full Text

Tilley AR, Dreyfuss H: *The Measure of Man and Woman: Human Factors in Design, Revised Edition.* Wiley; 2001.

Vaswani A, Shazeer N, Parmar N, *et al.*: **Attention Is All You Need.** *Advances in Neural Information Processing Systems.* Guyon I, Von Luxburg U, Bengio S, *et al.*, editors. Curran Associates, Inc.; 2017; **30**.
Reference Source

Vitali RV, Perkins NC: **Determining Anatomical Frames via Inertial Motion Capture: A Survey of Methods.** *J. Biomech.* 2020; **106**(June): 109832.
PubMed Abstract | **Publisher Full Text**

Wong C, Zhang Z-Q, Lo B, *et al.*: **Wearable Sensing for Solid Biomechanics: A Review.** *IEEE Sensors J.* 2015; **15**(5): 2747–2760.
Publisher Full Text

Xenoma Inc: **E-Skin MEVA.** n.d.
Reference Source

Yi X, Zhou Y, Habermann M, *et al.*: **Physical Inertial Poser (PIP): Physics-Aware Real-Time Human Motion Tracking from Sparse Inertial Sensors.** *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2022; 13157–68.
Publisher Full Text

Yi X, Zhou Y, Xu F: **TransPose: Real-Time 3D Human Translation and Pose Estimation with Six Inertial Sensors.** *ACM Trans. Graph.* 2021; **40**(4).
Publisher Full Text

Zheng C, Wu W, Chen C, *et al.*: **Deep Learning-Based Human Pose Estimation: A Survey.** *ACM Comput. Surv.* 2023; **56**(1).
Publisher Full Text

Zhou Y, Barnes C, Lu J, *et al.*: **On the Continuity of Rotation Representations in Neural Networks.** *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2019; 5738–46.
Publisher Full Text

Zimmermann T, Taetz B, Bleser G: **IMU-to-Segment Assignment and Orientation Alignment for the Lower Body Using Deep Learning.** *Sensors.* 2018; **18**(1).
Publisher Full Text

Zuo C, Huang J, Jiang X, *et al.*: **Transformer IMU Calibrator: Dynamic on-Body IMU Calibration for Inertial Motion Capture.** *ACM Trans. Graph.* 2025; **44**(4).
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 03 June 2026

https://doi.org/10.21956/openreseurope.24703.r74434

© **2026 Heravi M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Moein Younesi Heravi
North Dakota State University, Fargo, North Dakota, USA

The authors present a self-calibrating, online inertial human motion capture method that integrates a JointTracker filter with a probabilistic deep learning module to actively calibrate IMU-to-segment (I2S) poses during dynamic movements.

- The abstract is currently dense and mixes methodological details with results. It would benefit from a clearer structural separation between the problem statement, the novelty of the method, and the quantitative findings.

- Section 2 involves complex coordinate frames that require a summary table for readability. Figure 2 also needs clearer node labels (e.g., explaining T1 vs. C1) and better visual distinctions for the data flow.

- The Deep Ensembles formulation in Section 3.3.1 requires clearer mapping between the general variables and the specific biomechanical quantities used in the study.

- The results tables report values with an unnecessarily high number of decimal places (e.g., 4.0814 degrees), which implies unrealistic precision and should be rounded appropriately.

- The manuscript currently lacks a dedicated discussion section, as Section 5 serves primarily as a conclusion. Expanding on limitations, underlying assumptions, and interpretations of performance differences would greatly improve the paper.

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Partly

Are sufficient details provided to allow replication of the method development and its use

by others?
Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Artificial Intelligence, human activity intensity estimation using kinematic features, vision-based frameworks, and explainable AI.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 22 April 2026

https://doi.org/10.21956/openreseurope.24703.r71835

© 2026 Digo E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Elisa Digo
Politecnico di Torino, Turin, Italy

This manuscript presents a novel approach to inertial human motion capture, combining model-based sensor fusion with probabilistic deep learning for online IMU-to-segment calibration. The work is relevant, and has the potential for significant impact in the field, particularly in applications requiring robust and user-independent calibration procedures.
Overall, the paper is methodologically sound. However, several aspects of the manuscript should be improved to enhance clarity, readability, and the strength of the scientific contribution.

- **Abstract**. The abstract is dense and mixes methodological details, contributions, and results without a clear structure. It can be improved clearly separating the problem statement, the novelty of the proposed contribution, the proposed method, and then the quantitative results. Additionally, the lack of some numerical performance indicators weakens the impact of the paper.
- **Section 2**.
  - Given the heavy use of multiple reference frames and symbols (e.g., N, G, I, S), the readability of Section 2 would benefit from a concise schematic summary (e.g., a table or diagram) that clearly defines all coordinate systems and their relationships. This

would help the reader navigate the notation more efficiently.
    - The term “task description” in Section 2.2 is potentially misleading. In biomechanics, “task” typically refers to the performed movement (e.g., gait), whereas here it describes the computational estimation problem. A clearer wording (e.g., “Problem formulation”) would improve readability and avoid ambiguity.
- **Section 3** presents a well-structured and visually appealing overview of the proposed method (Figure 2).
    - However, the clarity of the figure and its connection with the text could be improved. In particular, the meaning of the labels (T1, T2, C1–C4) is not immediately intuitive. Although it can be inferred from the text that “T” refers to tracking steps and “C” to calibration steps, this distinction is not explicitly stated and should be clarified, ideally in the figure or its caption. Moreover, the data flow is somewhat ambiguous. Variables such as “IMU poses” and “IMU-centered joint positions” play multiple roles: they are outputs of T1, inputs to subsequent tracking steps (e.g., T2), and at the same time inputs to the calibration module. This multi-level usage is not clearly represented in the figure, which may lead to confusion about the actual information flow. A clearer distinction (for example with differentiated arrows) between internal data flow within the tracking block, data exchanged between modules, and feedback signals would improve the interpretability of the pipeline.
    - Subsection 3.3.1 provides a mathematical description of the Deep Ensemble framework. However, some components of the formulation are not fully explained or clearly connected to the specific application. Providing a clearer mapping between the general variables (e.g., x, y) and the quantities used in this work would improve readability, especially for readers less familiar with deep learning methodologies.
    - Figure 3 is clear and well organized; however, it is not entirely self-contained and requires the reader to refer to the text for full understanding.
- **Section 4**.
    - The structure could be improved. Currently, the description of the experimental setup, the presentation of results, and their discussion are interspersed within Section 4, which affects readability and makes it difficult to clearly distinguish between observations and interpretation. A clearer separation between experimental design, results, and discussion would strengthen the manuscript.
    - It might be helpful to complement the quantitative evaluation with additional visual examples, particularly for Experiments 1 and 2, to provide a more intuitive understanding of the results.
    - The reported values use an unnecessarily high number of decimal places, which may give a misleading impression of precision. A more appropriate level of rounding is recommended.
- **Section 5** mainly serves as a conclusion rather than a dedicated discussion. Expanding the discussion of the results (e.g., limitations, assumptions, and interpretation of performance differences) in a more structured way would further enhance the paper.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Partly

Are sufficient details provided to allow replication of the method development and its use by others?
Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Applied Mechanics, Biomechanics, IMUs

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.