Transformer IMU Calibrator: Dynamic On-body IMU Calibration for Inertial Motion Capture

CHENGXU ZUO, Xiamen University, China
JIAWEI HUANG, Xiamen University, China
XIAO JIANG, Xiamen University, China
YUAN YAO, Xiamen University, China
XIANGREN SHI, Bournemouth University, United Kingdom
RUI CAO, Xiamen University, China
XINYU YI, Tsinghua University, China
FENG XU, Tsinghua University, China
SHIHUI GUO, Xiamen University, China
YIPENG QIN, Cardiff University, United Kingdom

In this paper, we propose a novel dynamic calibration method for sparse inertial motion capture systems, which is the first to break the restrictive *absolute static assumption* in IMU calibration, i.e., the coordinate drift $R_{G'G}$ and measurement offset R_{BS} remain constant during the entire motion, thereby significantly expanding their application scenarios. Specifically, we achieve real-time estimation of $R_{G'G}$ and R_{BS} under two relaxed assumptions: i) the matrices change negligibly in a short time window; ii) the human movements/IMU readings are diverse in such a time window. Intuitively, the first assumption reduces the number of candidate matrices, and the second assumption provides diverse constraints, which greatly reduces the solution space and allows for accurate estimation of $R_{G'G}$ and R_{BS} from a short history of IMU readings in real time. To achieve this, we created synthetic datasets of paired $R_{G'G}$, R_{BS} matrices and IMU readings, and learned their mappings using a Transformer-based model. We also designed a calibration trigger based on the diversity of IMU readings to ensure that assumption ii) is met before applying our method. To our knowledge, we are the first to achieve implicit IMU calibration (i.e., seamlessly putting IMUs into use without the need for an explicit calibration process), as well as the first to enable long-term and accurate motion capture using sparse IMUs. The code and dataset are available at *https://github.com/ZuoCX1996/TIC*.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Authors' Contact Information: Chengxu Zuo, Xiamen University, Xiamen, China, zuochengxu@stu.xmu.edu.cn; Jiawei Huang, Xiamen University, Xiamen, China, 30920231154349@stu.xmu.edu.cn; Xiao Jiang, Xiamen University, Xiamen, China, ferster@stu.xmu.edu.cn; Yuan Yao, Xiamen University, Xiamen, China, furtheryao@stu.xmu.edu.cn; Xiangren Shi, Bournemouth University, Bournemouth, United Kingdom, xshi@bournemouth.ac.uk; Rui Cao, Xiamen University, Xiamen, China, mec2109494@xmu.edu.my; Xinyu Yi, Tsinghua University, School of Software and BNRist, Beijing, China, yixy20@mails.tsinghua.edu.cn; Feng Xu, Tsinghua University, School of Software and BNRist, Beijing, China, feng-xu@tsinghua.edu.cn; Shihui Guo, Xiamen University, Xiamen, China, guoshihui@xmu.edu.cn; Yipeng Qin, Cardiff University, Cardiff, United Kingdom, qiny16@cardiff.ac.uk.

ACM 1557-7368/2025/8-ART
https://doi.org/10.1145/3730937

Additional Key Words and Phrases: Inertial Motion Capture, Dynamic Calibration, Transformer

ACM Reference Format:
Chengxu Zuo, Jiawei Huang, Xiao Jiang, Yuan Yao, Xiangren Shi, Rui Cao, Xinyu Yi, Feng Xu, Shihui Guo, and Yipeng Qin. 2025. Transformer IMU Calibrator: Dynamic On-body IMU Calibration for Inertial Motion Capture. *ACM Trans. Graph.* 44, 4 (August 2025), 14 pages. https://doi.org/10.1145/3730937

1 Introduction

Motion capture with sparse inertial sensors has received increasing attention in recent years, given their advantage in high usability and reduced hardware cost [Huang et al. 2018; Jiang et al. 2022b;

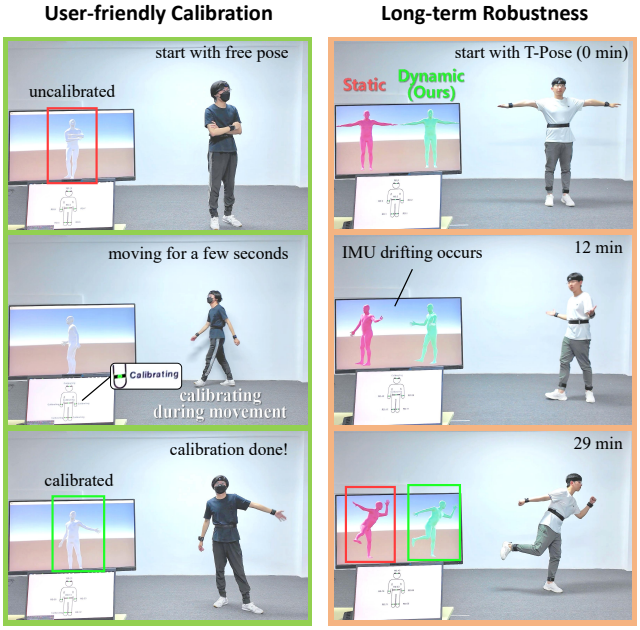

Fig. 1. Live demonstration of our *dynamic* calibration method against the conventional *static* calibration method. Our method provides a user-friendly experience (**w/o IMU heading reset and T-Pose**) and ensures long-term robustness for inertial motion capture.

Yi et al. 2022, 2021, 2024]. These systems predict full-body skeletal rotations $\boldsymbol{\theta}$ based on partial skeletal orientations $R_{GB} \in \mathbb{R}^{3\times3}$ and their associated joint endpoint acceleration $a_G \in \mathbb{R}^{3\times1}$ in the global coordinate system G: $\boldsymbol{\theta} = f(R_{GB}, a_G)$. In practice, R_{GB} and a_G are measured by Inertial Measurement Units (IMUs) readings: rotation R_{IMU} and acceleration a_{IMU}, respectively, which are defined as:

$$\begin{aligned} R_{\text{IMU}}(t) &= R_{G'G}(t) \cdot R_{GB}(t) \cdot R_{BS}(t) \\ a_{\text{IMU}}(t) &= R_{G'G}(t) \cdot a_G(t) \end{aligned} \tag{1}$$

where t denotes time; G' is G's offset version; $R_{G'G} \in \mathbb{R}^{3\times3}$ denotes the *coordinate drift* caused by factors such as magnetic field interference and/or gyroscope integration error; R_{BS} denotes the *measurement offset* caused by ambiguous wearing orientation of sensor, where B represents the bone for which the attitude is to be measured and S represents the IMU sensor. To faithfully measure R_{GB} and a_G with R_{IMU} and a_{IMU}, a *calibration* process is required to eliminate the effects of $R_{G'G}$ and R_{BS}.

Traditionally, the calibration performed with an ideal but restrictive assumption that we call:

Assumption 1 (Long-term Static Assumption). *The coordinate drift $R_{G'G}(t)$ and measurement offset $R_{BS}(t)$ remain constant during the entire motion sequence (t=0, 1, 2, ..., T), i.e.,*

- $R_{G'G}(t) = R_{G'G}(0) = I$
- $R_{BS}(t) = R_{BS}(0) = \hat{R}_{BS}(0)$

where $\hat{R}_{BS}(0)$ is estimated using a calibration pose (e.g., T-pose).

Under this assumption, IMU calibration can *only* be performed at the start of motion, and it has become common practice for users to first calibrate the IMU to estimate and remove $R_{G'G}(0)$ using tools provided by the sensor manufacturer before wearing the device, then to wear the device and perform a specific calibration pose to estimate and remove $R_{BS}(0)$, and finally to start motion capture. However, such static calibration is only a *makeshift solution* because of the restrictive nature of the underlying Assum. 1:

- $R_{G'G}(t)$ change dynamically due to magnetic field interference and cumulative error of the gyroscope;
- $R_{BS}(t)$ vary over time due to accumulated changes, such as IMU placement offsets in long-term use;
- Assuming $R_{BS}(0) = \hat{R}_{BS}(0)$ is suboptimal due to users' imperfect calibration poses [Yi et al. 2024];

These shortcomings indicate that conventional static calibration struggles to remain robust in long-term use and is sensitive to the accuracy of the calibration pose. This results in a suboptimal user experience, limiting the adoption of inertial motion capture in applications such as gaming and sports fitness.

In this paper, we address these shortcomings by proposing a novel *dynamic* on-body calibration technique that operates automatically and imperceptibly during motion capture (Fig. 1). Our key insight is that although estimating $R_{G'G}(t)$ and $R_{BS}(t)$ from $R_{\text{IMU}}(t)$ is an inherently *ill-posed problem*, a feasible solution can be achieved by reducing its solution space and imposing additional constraints. Specifically, we replace the restrictive Assum. 1 with 2 relaxed assumptions:

Assumption 2 (Short-term Static Assumption). *Let $[t_a, t_b]$ be a short time window in a motion sequence, $t_i \in [t_a, t_b]$, we assume coordinate drift and measurement offset matrices $R_{G'G}(t_i)$ and $R_{BS}(t_i)$ change negligibly in $[t_a, t_b]$, i.e.:*

- $R_{BS}(t_i) \approx R_{BS}(t_b)$;
- $R_{G'G}(t_i) \approx R_{G'G}(t_b)$.

Assumption 3 (Short-term Diversity Assumption). *Let $t_i, t_j \in [t_a, t_b]$, $i \neq j$, $R_{\text{IMU}}(t)$ is diverse in $[t_a, t_b]$, i.e.:*

- $R_{\text{IMU}}(t_i) \neq R_{\text{IMU}}(t_j)$;

Then, given a short history of IMU readings $\{R_{\text{IMU}}(t_1), ..., R_{\text{IMU}}(t_n)\}$, $t_i \in [t_a, t_b]$ $(1 \leq i \leq n)$, Assum. 2 reduces its solution space from $\{R_{G'G}(t_1), ..., R_{G'G}(t_n)\}$ and $\{R_{BS}(t_1), ..., R_{BS}(t_n)\}$ to $R_{G'G}(t_b)$ and $R_{BS}(t_b)$; and the inherent diversity among its elements serves as additional constraints (Assum. 3). Thanks to the arbitrary choices of $[t_a, t_b]$, our method can track dynamic changes of $R_{G'G}$ and R_{BS} in the entire motion sequence. To achieve this, we created two synthetic datasets of paired ($R_{G'G}$, R_{BS}) and R_{IMU} using the AMASS [Mahmood et al. 2019] and DIP [Huang et al. 2018] datasets, and learned the mapping between them using a Transformer-based model. We also designed a calibration trigger based on the diversity of IMU readings to ensure that Assum. 3 is met before applying our method. Empirically, we compare our dynamic calibration with traditional static calibration methods in bone orientation and global acceleration measurement accuracy, and apply to 6 state-of-the-art sparse inertial motion capture methods, demonstrating its superiority in user-friendliness and long-term use. In summary, our main contributions include:

- Conceptually, we propose 2 fundamental assumptions (Assums. 2 and 3) for dynamic calibration in sparse inertial motion capture, enabling accurate and on-the-fly estimation of IMU coordinate drift and measurement offset.
- Technically, we propose a practical dynamic calibration workflow including 1) a Transformer IMU Calibrator (TIC) network to estimate $R_{G'G}$ and R_{BS} with a short history of IMU orientations and accelerations; 2) a calibration trigger mechanism based on IMU rotation diversity to ensure effective use of TIC. To our knowledge, we are the *first* to achieve implicit IMU calibration, i.e., seamlessly putting IMUs into use without requiring an explicit calibration process.
- Additionally, we collected the first long-duration inertial motion capture dataset that explicitly incorporates IMU coordinate drift and measurement offset, providing a valuable resource for analyzing their characteristics.

2 Related Works

2.1 Inertial Motion Capture

Inertial motion capture offers advantages in portability, privacy, and resilience to challenging lighting and occlusion conditions compared to vision-based methods. Popular commercial IMU-based systems like Xsens [Paulich et al. 2018] and Noitom [Noitom 2017] use multiple wear-on IMU to capture user's motion.

Recent studies have achieved posture estimation using a sparse set (3-6) of IMUs [Huang et al. 2018; Jiang et al. 2022b; Mollyn et al. 2023; Van Wouwe et al. 2024; Yi et al. 2022, 2021; Zhang et al. 2024]. TransPose [Yi et al. 2021] enhanced sparse IMUs motion capture by

integrating multi-stage pose estimation alongside a fused global displacement estimation, integrating a module for optimizing physical dynamics in subsequent endeavors [Yi et al. 2022]. TIP [Jiang et al. 2022b] incorporated Transformer architecture into sparse inertial motion capture, thereby accounting for human motion in non-planar scenarios simultaneously. A real-time full-body posture estimation system, IMUPoser [Mollyn et al. 2023] and the follow-up research MobilePoser [Xu et al. 2024] utilize IMU data from consumer-level devices to estimate body pose and global translation. Xiao [Xiao et al. 2024] propose a hight efficiency network architecture which enabled the deployment on mobile terminals. DynaIP [Zhang et al. 2024] involves the integration of real-world motion capture data from diverse skeleton formats and introduces a novel part-based approach to enhance the robustness and accuracy of pose estimation. DiffusionPoser [Van Wouwe et al. 2024] allows immediate use of arbitrary sensor configurations and thus optimizing these configurations for specific activities. LIP [Zuo et al. 2024] presents a loose-wear jacket equipped with 4 IMUs for comfortable upper-body motion tracking.

Some emerging research explores the fusion of multiple sensor modalities to capture specific limb movements or body parts because of the high expectation to precision and flexibility, such as Ultra Inertial Poser (UIP) [Armani et al. 2024] and SmartPoser [DeVrio et al. 2023], which introduce ultra-wideband (UWB) sensors to improve accuracy. More recent studies taking the advance of AR/VR applications have explored various approaches that can further sparsify input requirements to only the upper body to estimate the entire body pose by head and hand poses [Ahuja et al. 2021; Du et al. 2023; Jiang et al. 2023, 2022a; Yang et al. 2021; Zheng et al. 2023]. To address the ongoing difficulties in achieving precise joint angle and position estimations, researchers have reconsidered the use of external [Pan et al. 2023; von Marcard et al. 2018] or body-worn [Yi et al. 2023] cameras for visual-inertial tracking. In contrast, EM-Pose [Kaufmann et al. 2021] assesses the relative 3D offsets and orientations between joints through the utilization of 6–12 custom electromagnetic (EM) field-based sensors.

In all these methods, a calibration process is typically required during system initialization to acquire accurate skeleton measurement. However, this calibration cannot effectively address the dynamic change of calibration parameters, posing a challenge in maintaining accurate pose estimation over prolonged usage.

2.2 IMU Calibration

With the miniaturization MEMS sensors, IMUs have been widely used for tasks such as navigation and attitude estimation. However, low-cost MEMS-based IMUs are usually affected by axes misalignment, bias and cross-axis sensitivities, leading to significant systematic errors in measurements [Harindranath and Arora 2024]. Traditional intrinsic calibration methods [Kim and Golnaraghi 2004; Syed et al. 2007; Titterton and Weston 2004] often require expensive high-precision equipment. [Tedaldi et al. 2014] achieve highly accurate estimation of correction parameters by placing the IMU in a set of different static positions, without the use of external equipment. [Li et al. 2012] propose a Kalman filter technique which enables the gyro triad and the accelerometer triad to calibrate each other by applying the pseudo-observations, eliminating the need for the quasi-static stays at different attitudes.

Besides the *intrinsic* calibration of the IMU, calibrating the spatial *extrinsic* parameters (relative translation and rotation) between the IMU and the object to which it is mounted through extrinsic calibration is equally important for IMU applications. In motion capture, this is commonly achieved by asking the subject to maintain a static pose, such as T-pose or N-pose, during system initialization [Choe et al. 2019; Liu et al. 2019]. Similarly, functional calibration requires the subject to perform specified movements such as hip flexion/extension and abduction/adduction [Favre et al. 2009, 2008; Nazarahari et al. 2019]. Since these methods rely on predefined postures and motions, their accuracy depends on how precisely the subject performs them. To alleviate this requirement, methods [Müller et al. 2016; Seel et al. 2014; Taetz et al. 2016] explore the use of arbitrary motion for calibration and eliminate the need for precisely executing predefined postures or motions. However, these methods require attaching IMUs to adjacent joints for calibration via kinematic constraints, making them unsuitable for sparse inertial motion capture systems.

Despite these successes, all existing calibration methods require an *explicit* calibration process during system initialization, which is less user-friendly and suffers from performance degradation over time. In contrast, our work is the *first* to achieve implicit IMU calibration (i.e., seamlessly putting IMUs into use without requiring an explicit calibration process), enabling user-friendly, long-term, and accurate inertial motion capture in a variety of real-world scenarios.

3 Problem Formulation

3.1 IMU Calibration in Inertial Motion Capture

As discussed in Sec. 1, in inertial motion capture, IMU calibration aims to estimate and remove (via multiplying inverse matrix) coordinate drift and measurement offset matrices $R_{G'G}$ and R_{BS} from Eq. 1 to obtain calibrated IMU readings $R_{\text{cali}}(t)$ and $a_{\text{cali}}(t)$:

$$\begin{aligned} R_{\text{cali}}(t) &= R^T_{G'G}(t) \cdot R_{\text{IMU}}(t) \cdot R^T_{BS} \\ &= R^T_{G'G}(t) \cdot [R_{G'G}(t) \cdot R_{GB}(t) \cdot R_{BS}(t)] \cdot R^T_{BS}(t) \\ &= R_{GB}(t) \\ a_{\text{cali}}(t) &= R^T_{G'G}(t) \cdot a_{\text{IMU}}(t) = R^T_{G'G}(t) \cdot [R_{G'G}(t) \cdot a_G(t)] \\ &= a_G(t) \end{aligned} \tag{2}$$

where rotation matrices $R^T_{G'G}(t) = R^{-1}_{G'G}(t)$ and $R^T_{BS}(t) = R^{-1}_{BS}(t)$.

3.2 Dynamic Calibration

The motivation for the proposed dynamic calibration arises from the intrinsic requirement to continually update calibration parameters during the motion capture (Fig. 2), and we define it as follows:

Egocentric yaw (ego-yaw) coordinate system. Without loss of generality, we define the global coordinate system G as one with zero roll and pitch rotations, while its yaw rotation is synchronized with user's body, called *ego-yaw coordinate system*. This definition aims to eliminate the non-solvable drift component between the drifted ego-yaw and drifted world coordinate system ($R_{W'G'}$) (please refer to the supplementary materials).

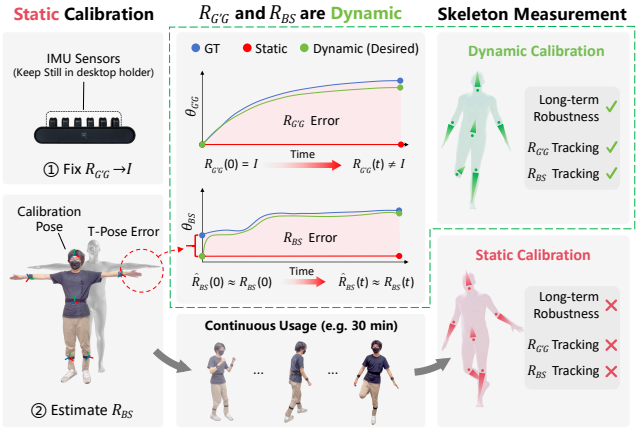

Fig. 2. Motivation of our dynamic calibration. In contrast to static calibration that suffers from T-pose errors and the $R_{G'G}$ and R_{BS} errors that increase over time, our dynamic calibration tracks changes of $R_{G'G}$ and R_{BS} during use, ensuring long-term robustness. $\theta_{G'G}, \theta_{BS}$: rotation angles.

Task Definition. According to Assum. 2 and Assum. 3, we define our dynamic calibration task as estimating:

$$\begin{aligned}\hat{R}_{G'G}(t) &= f_d(R_{\text{IMU}}(t-n+1), ..., R_{\text{IMU}}(t), \\ &\qquad a_{\text{IMU}}(t-n+1), ..., a_{\text{IMU}}(t)) \\ \hat{R}_{BS}(t) &= f_o(R_{\text{IMU}}(t-n+1), ..., R_{\text{IMU}}(t), \\ &\qquad a_{\text{IMU}}(t-n+1), ..., a_{\text{IMU}}(t))\end{aligned} \tag{3}$$

where n denotes the n-th historical IMU frame from time t; f_d and f_o are learned by:

$$\begin{aligned}f_d &= \arg\min_f \mathbb{E}_{i,t}\mathcal{L}[\hat{R}^{(i)}_{G'G}(t), R^{(i)}_{G'G}(t)] \\ f_o &= \arg\min_f \mathbb{E}_{i,t}\mathcal{L}[\hat{R}^{(i)}_{BS}(t), R^{(i)}_{BS}(t)]\end{aligned} \tag{4}$$

where $\mathcal{L}$ denotes a loss function; i is an index of IMU sensor in training dataset consisting of paired $R_{G'G}$, R_{BS}, and R_{IMU}; f is a hypothesis function (model) specified by the user.

Acceleration Auxiliary (ACCA). We introduced a_{IMU} as additional inputs to f_d and f_o because a_{IMU} is only influenced by $R_{G'G}$ and is independent of R_{BS} (Eq. 1), which helps the model distinguish $R_{G'G}$ and R_{BS}, thereby improving calibration accuracy.

4 Method

Our dynamic calibration includes 1) *TIC Network* for $R_{G'G}$ and R_{BS} estimation and 2) *Calibration Trigger via Rotation Diversity* (*RD*). Fig. 3 illustrates the rationale behind this design: our calibration trigger mimics the human ability to evaluate whether a given short motion sequence contains sufficient information to determine its naturalness, while our TIC network mimics the human ability to infer the original motion through calibration. Furthermore, we filter out unreliable results based on *RD* to satisfy the diversity requirement in Assum. 3. Please see Fig. 4 for an intuitive illustration of our dynamic calibration workflow.

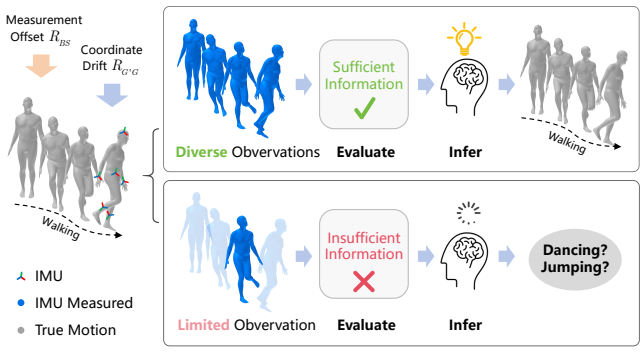

Fig. 3. The rationale of our dynamic calibration. The $R_{G'G}$ and R_{BS} can lead to unnatural dynamics in the captured human motion (e.g., movements that cannot maintain balance), and humans can evaluate it (our calibration trigger) and infer the original motion (our calibration).

4.1 TIC Network

In this work, we instantiate the hypothesis function f (Eq. 4) with our Transformer IMU Calibrator (TIC) network.

Network Architecture. Considering the dependence of the task on temporal information (IMU sequence), we opt to use a Transformer comprising two components:

- Encoder (E): A standard Transformer encoder consisting of 3 Transformer encoder blocks for feature extraction;
- Transformer-Pooling-Mapping (TPM) module (x2): Each consists of a Transformer encoder block, a temporal average pooling layer, and a linear mapping layer; One for $R_{G'G}$ (TPM_d), the other one for R_{BS} (TPM_o).

Thus, we instantiate (Eqs. 3 and 4) with our TIC as:

$$\begin{aligned}f_d(\cdot) &= \text{TPM}_d(\text{E}(\cdot)) \\ f_o(\cdot) &= \text{TPM}_o(\text{E}(\cdot))\end{aligned} \tag{5}$$

Model Training. We instantiate $\mathcal{L}$ in Eq. 4 with an MSE loss and define the calibration loss $\mathcal{L}_{\text{cali}}$ as:

$$\begin{aligned}\mathcal{L}_{\text{cali}} =& ||f_d(\mathbf{R}^{1\rightarrow n}_{\text{IMU}}, \mathbf{a}^{1\rightarrow n}_{\text{IMU}}) - R_{G'G}(n)||^2_2 \\ &+||f_o(\mathbf{R}^{1\rightarrow n}_{\text{IMU}}, \mathbf{a}^{1\rightarrow n}_{\text{IMU}}) - R_{BS}(n)||^2_2\end{aligned} \tag{6}$$

where $\mathbf{R}^{1\rightarrow n}_{\text{IMU}}$ and $\mathbf{a}^{1\rightarrow n}_{\text{IMU}}$ are n frames of input IMU orientation and acceleration readings, respectively.

4.2 Calibration Trigger via Rotation Diversity

As discussed in Sec. 1, Assum. 3 should be met to ensure reliable $R_{G'G}$ and R_{BS} estimation. To achieve this, we propose a calibration trigger technique based on IMU rotation diversity. Rotation diversity is quantified by counting covered grid points in the discretized Euler angle space. We discretize the continuous Euler angle space ($\theta_x \in [-180, 180]$, $\theta_y \in [-90, 90]$, $\theta_z \in [-180, 180]$) at 15° intervals to form a $24 \times 12 \times 24$ discrete space S. The rotation diversity *RD* of an IMU sequence can be calculated using Algorithm 1.

4.3 Dynamic Calibration in Motion Capture

Based on the TIC network and aforementioned rotation diversity, we have incorporated dynamic calibration into the existing inertial

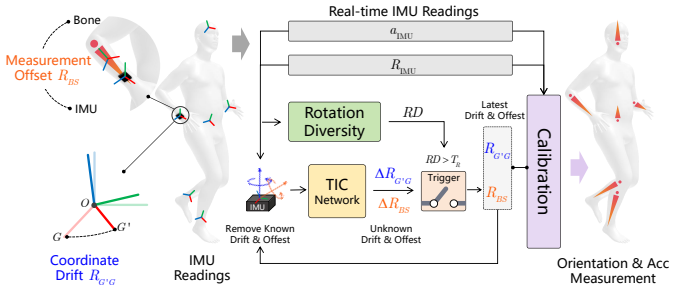

Fig. 4. Our dynamic calibration workflow. With real-time IMU inputs, the TIC network operates at fixed short time intervals (e.g., every 2 seconds) to track the dynamic changes of $R_{G'G}$ and R_{BS}. The updating of calibration paremeters are controlled by a Rotation Diversity (RD) based trigger, ensuring only reliable results that meet Assum 3 are used.

ALGORITHM 1: IMU Rotation Diversity

Data: A sequence of IMU Rotation $\{R_{\text{IMU}}(1), ..., R_{\text{IMU}}(n)\}$.
Result: Rotation Diversity RD.
Initialize discrete Euler angle space $\mathbf{S} \in \mathbb{R}^{24\times12\times24}$ with $\mathbf{S}_{i,j,k} = 0$ $\forall i, j, k$;
for $t = 1$ ***to*** n **do**
  $i, j, k \leftarrow$ Calculate the coordinates of $R_{\text{IMU}}(t)$ in S;
  $\mathbf{S}_{i,j,k} \leftarrow \mathbf{S}_{i,j,k} + 1$;
end
$RD \leftarrow \text{Count}(\mathbf{S}_{i,j,k} > 0)$;

motion capture system (Alg. 2). It is worth noting that for each IMU in the system, we independently calculate RD, and if the calibration is triggered, the TIC network will calculate $R_{G'G}$ and R_{BS} for all IMUs, but only calibrate those whose RD exceeds the threshold T_R.

5 Experiments

5.1 Training Data Synthesis

Uncalibrated IMU data and their corresponding $R_{G'G}(t)$ and $R_{BS}(t)$ data are necessary to train the TIC network. However, the cost of collecting such dataset is prohibitive because $R_{G'G}(t)$ and $R_{BS}(t)$ are vary randomly during device usage, which cannot be manually controlled to collect sufficient samples.

To address this challenge, we adopted a data synthesis approach to obtain the required training data. Firstly, we need well-calibrated IMU data $\mathcal{D}^{\text{cali}}_{\text{IMU}}$. Similarly to works in inertial motion capture, the $\mathcal{D}^{\text{cali}}_{\text{IMU}}$ we used includes both synthetic data based on AMASS [Mahmood et al. 2019] and real-world data from the DIP [Huang et al. 2018] dataset, which helps the TIC network adapt to the characteristics of the real IMU signal. Then, we simulated uncalibrated IMU data based on Assum. 2. Specifically, for each IMU data sequence (256 frames at 30Hz, 8.53s) in $\mathcal{D}_{\text{cali}}$, we use random Euler angle transformations to create $R_{G'G}$ and R_{BS} and apply them to each frame in a sequence. These Euler angles were extensively sampled from a uniform distribution within fixed intervals (see supplementary materials) to cover all possible cases.

It is worth noting that R_{IMU} and a_{IMU} for each batch of data are synthesized on demand during model training. Compared to using a pre-synthesized dataset of fixed/limited size, this can provide a more diverse set of $R_{G'G}$ and R_{BS} samples.

ALGORITHM 2: Dynamic Calibration in Motion Capture

Data: Data index $i = 1$, Uncalibrated IMU orientation and acceleration readings $\{R_{\text{IMU}}(1), R_{\text{IMU}}(2), ...\}$, $\{a_{\text{IMU}}(1), a_{\text{IMU}}(2), ...\}$, Data Buffer B_n with maximum length $= n$, Trained TIC Network TIC, Rotation diversity threshold T_R. Timing signal at intervals of t seconds S_t.
Result: Calibrated IMU readings $\hat{R}_{\text{IMU}}$, $\hat{a}_{\text{IMU}}$.
Initialize $R_{G'G}$ and R_{BS} with identity matrix I;
while *True* **do**
  $\hat{R}_{\text{IMU}}(i), \hat{a}_{\text{IMU}}(i) \leftarrow$ Calibrate $R_{\text{IMU}}(i), a_{\text{IMU}}(i)$ using $R_{G'G}$ and R_{BS} (Eq. 2);
  B_n.append($R^{(i)}_{\text{IMU}}$, $a^{(i)}_{\text{IMU}}$);
  if $|B_n| == n$ *and* $S_t == True$ **then**
    $\mathbf{R}^{1\to n}_{\text{IMU}}, \mathbf{a}^{1\to n}_{\text{IMU}} \leftarrow B_n$;
    // Remove the known drift and offset.
    $\mathbf{R}^{1\to n}_{\text{IMU}}, \mathbf{a}^{1\to n}_{\text{IMU}} \leftarrow R^T_{G'G} \cdot \mathbf{R}^{1\to n}_{\text{IMU}} \cdot R^T_{BS}, R^T_{G'G} \cdot \mathbf{a}^{1\to n}_{\text{IMU}}$;
    // Estimating the unknown changes.
    $\Delta R_{G'G}, \Delta R_{BS} \leftarrow TIC(\mathbf{R}^{1\to n}_{\text{IMU}}, \mathbf{a}^{1\to n}_{\text{IMU}})$;
    $RD \leftarrow$ RotationDiversity(B_n);
    // Update drift and offset matrices.
    if $RD > T_R$ **then**
      $R_{G'G} \leftarrow R_{G'G} \cdot \Delta R_{G'G}$;
      $R_{BS} \leftarrow \Delta R_{BS} \cdot R_{BS}$;
    end
    B_n.clear();
  end
  i $\leftarrow$ i + 1;
end

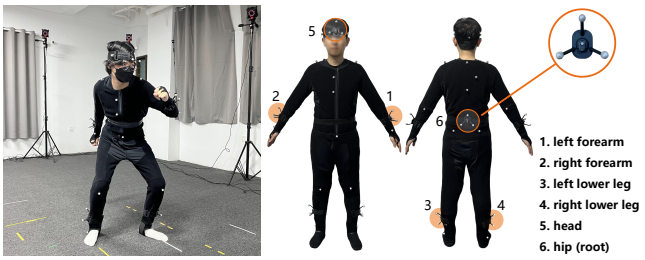

Fig. 5. Our data collection system. The absolute IMU orientation and acceleration are tracked by optical markers (orange circles). Body motions (skeleton orientation, global translation) and raw IMU readings are synchronously collected.

5.2 Test Data Collection

We recruited 5 volunteers (3 male, 2 female) to collect the real-world dataset. All these volunteers have experience with using inertial motion capture devices and are familiar with the process of Static Calibration. Volunteers were asked to wear both optical motion capture suits and 6 IMUs simultaneously, placed on the left forearm, right forearm, left lower leg, right lower leg, head, and hips (see Fig. 5). All 6 IMUs are integrated with 3 optical markers, allowing the absolute orientation and acceleration of IMUs to be captured.

Table 1. Dataset Summary. We collected sufficient real samples to validate the performance of the proposed dynamic calibration. n_{seq}: The number of IMU data sample sequences in the dataset (256 frames, 30 Hz).

Dataset	Purpose	$\mathcal{D}_{\text{cali}}$	$R_{G'G}$ & R_{BS}	n_{seq}
DS_{AMS}	Train	Synthesis	Synthesis	1.83M
DS_{DIP}	Train	Real	Synthesis	114k
DS_{TIC}	Test	Real	Real	1.04M

We collect continuous 60-minute data at 60fps for each volunteer, which is sufficient to capture drift signals of IMUs and ensure volunteers do not overexert themselves. In each session, volunteers sequentially perform stretching, walking/running/jumping, table tennis, aerobics, boxing, and free activity, each lasting 10 minutes. IMU data are recorded using the NOITOM Axis Lab system and calibrated through Static Calibration before system start-up. Human body pose, absolute IMU orientation and acceleration are obtained using the NOKOV motion capture system.

Let $R_{G'S}(t)$ be the IMU orientation reading, $R_{GB}(t)$ be the captured skeleton orientation, $R_{GS}(t)$ be the absolute IMU orientation, then $R_{G'G}(t)$ and $R_{BS}(t)$ can be obtained by the following formulas:

$$\begin{aligned} R_{G'G}(t) &= R_{G'S}(t) \cdot R_{GS}^{T}(t) \\ R_{BS}(t) &= R_{GB}^{T}(t) \cdot R_{GS}(t) \end{aligned} \tag{7}$$

5.3 Metrics of Calibration

As shown in Eq. 1, the purpose of calibration is to obtain accurate measurements of joint orientation and global acceleration. We defined two metrics to evaluate the accuracy of calibration:

- **Orientation Measurement Error** (OME). The angular error between the calibrated IMU orientation and the ground-truth skeletal orientation in the ego-yaw coordinate system.
- **Acceleration Measurement Error** (AME). The Euclidean distance between the calibrated IMU acceleration and the ground-truth acceleration (captured by NOKOV system in our work) in the ego-yaw coordinate system.

5.4 TIC v.s. Static Calibration

To demonstrate the advantages of applying our dynamic calibration TIC in inertial motion capture, we applied different calibration strategies on DS_{TIC} and created two different datasets:

- w TIC: Static calibration at system initialization, and apply TIC during usage;
- w/o TIC: Static calibration only at system initialization.

Table 2 shows the OME and AME on these two datasets, indicating that dynamic calibration based on TIC provides more accurate skeleton motion measurements. It can be observed that the skeletal measurement error obtained using Static Calibration for the root node (hip) is lower. This is because in the ego-yaw coordinate frame, the yaw rotation is defined by the yaw rotation of the root node, so its pose measurement error is not affected by IMU drifting.

With dynamic updating of $R_{G'G}$ and R_{BS}, TIC can significantly improve the robustness of inertial motion capture systems during long-term use. We used six state-of-the-art methods, DIP [Huang et al. 2018], Transpose [Yi et al. 2021], TIP [Jiang et al. 2022b], PIP [Yi et al. 2022], DynaIP[Zhang et al. 2024] and PNP [Yi et al. 2024], to test the effectiveness of TIC. The error metrics used include:

Table 2. Static Calibration vs. our TIC (dynamic) on DS_{TIC}.

Joint	OME (°) ↓		AME (m/s^2) ↓	
	w TIC	w/o TIC	w TIC	w/o TIC
left forearm	**21.40**	63.81	**1.53**	3.94
right forearm	**20.56**	68.57	**1.90**	4.77
left lower leg	**13.79**	47.87	**1.54**	2.17
right lower leg	**12.93**	55.23	**1.42**	2.21
head	**16.56**	57.37	**0.62**	1.44
hip	**6.00**	7.03	**0.80**	0.81
Avg	**15.20**	49.98	**1.30**	2.56

Table 3. Performance of SOTA inertial motion capture methods with / without our dynamic calibration during long-term usage (evaluated on DS_{TIC}).

Method	Pose error metrics with / without dynamic calibration			
	Angular (°)	Positional (cm)	SIP (°)	Mesh (cm)
DIP	**19.31**/37.18	**9.36**/13.25	**20.89**/32.38	**10.97**/17.22
TransPose	**17.90**/36.88	**8.43**/12.72	**18.98**/32.06	**10.33**/16.11
TIP	**16.50**/37.56	**7.28**/13.41	**17.28**/33.05	**8.92**/16.86
PIP	**16.21**/32.39	**7.53**/12.02	**15.77**/29.86	**9.30**/14.84
DynaIP	**16.71**/35.16	**7.35**/12.14	**16.56**/30.14	**9.29**/15.02
PNP	**15.52**/30.60	**7.20**/12.68	**14.18**/25.13	**8.84**/15.26

- Angular Error. The global rotation error of all joints;
- Positional Error. The joint position error of all joints;
- SIP Error. The global rotation error of hips and shoulders;
- Mesh Error. The vertex error of the posed SMPL meshes.

Both metrics were calculated with the root joint (Hip) aligned.

Table 3 shows that pose estimation metrics with TIC are significantly lower than those without TIC. This implies that unreliable skeleton measurements caused by the calibration parameters changing severely affect pose estimation, and TIC effectively addresses this issue. In particular, we found that PNP achieved the optimal result, which is attributed to the PNP includes calibration errors simulation caused by a small volume of $R_{G'G}$ and R_{BS} in the training data, thus achieve better adaption to imperfect calibration.

In Fig. 6 we visualize both OME and the corresponding predicted poses at different time points. The comparison demonstrates that, with only static calibration, the OME gradually increases as a result of the change of calibration parameters, leading to incorrect pose estimation. In contrast, thanks to the tracking of calibration parameters change, our dynamic calibration ensures accurate skeleton orientation and acceleration measurement over long durations, thereby maintaining robust pose estimation.

5.5 Ablation Study

To assess the necessity of the rotation diversity (*RD*) trigger and the Acceleration Auxiliary (ACCA), we recorded the error metrics of dynamic calibration after their removal. The results presented in Table 4 justify our claim that the effectiveness of using TIC relies on Assum. 3, whose fulfillment is ensured by the rotation diversity (*RD*) trigger (Case 1 vs. case 2). Furthermore, as expected, the ACCA

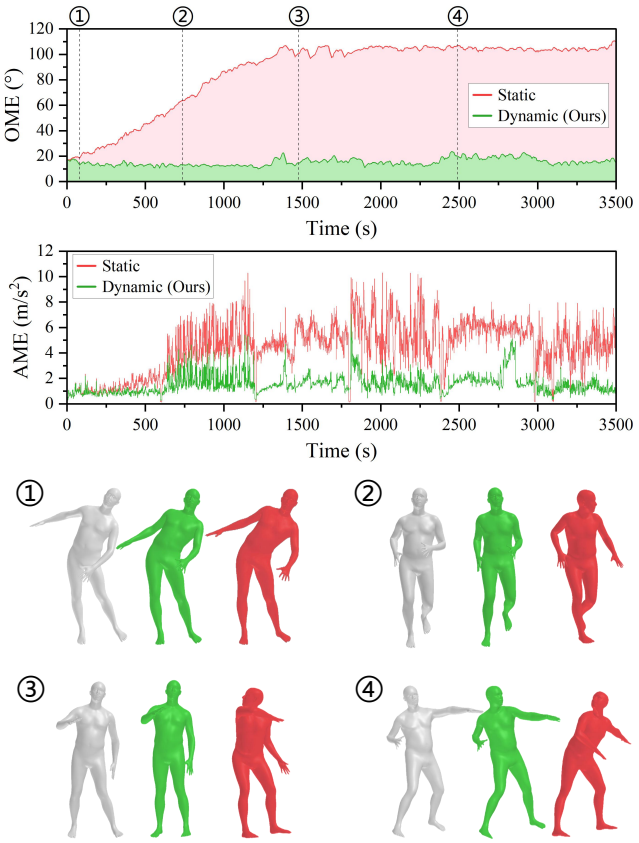

Fig. 6. Top: Static Calibration (PNP) vs. our Dynamic Calibration (TIC) on joint orientation measurement error over an extended period. Bottom: visualization of predicted poses at the four time points highlighted in Top.

has a significant impact on the accuracy of $R_{G'G}$. The removal of ACCA resulted in an increase of 85.4% in the $R_{G'G}$ error (Case 1 vs. Case 3), which consequently led to a rise in AME.

Additionally, Case 5 presents the ablation results of DS_{DIP} finetuning, demonstrating the benefits of real-world IMU data in TIC network training.

Fig. 7 provides a more intuitive illustration of the impact of *RD* Trigger and ACCA. It can be observed that after removing the *RD* Trigger, there is a significant error and oscillation in $R_{G'G}$ tracking under low *RD* conditions. This is due to the unreliable outputs not being filtered and applied to $R_{G'G}$ updating. On the other hand, after removing ACCA, even with *RD* Trigger, the tracking accuracy of $R_{G'G}$ still suffers a noticeable decline. This reflects the important role of ACCA in accurate $R_{G'G}$ estimation.

5.6 Error Analysis

In this section, we analyze the errors in dynamic calibration. As shown in Table 5, the R_{BS} error is significantly higher than $R_{G'G}$, particularly for the IMUs located on the left and right forearms, which leads to higher OME. For further investigation, we visualized the R_{BS} error and positional error (calculated via the predicted human pose) for two forearm joints over the first 200 seconds of

Table 4. Ablation Study on Rotation Diversity (*RD*) trigger, the acceleration auxiliary (ACCA) and DS_{DIP} finetuning. $R_{G'G}/R_{BS}$ Err: Regression error of $R_{G'G}/R_{BS}$ (°).

Case	*RD*	ACCA	OME	AME	$R_{G'G}$ Err	R_{BS} Err
1	+	+	15.20	1.30	9.18	15.28
2	-	+	15.48	1.32	9.27	15.56
3	+	-	16.35	1.85	15.22	16.53
4	-	-	16.86	1.91	16.21	16.90
5	w/o DS_{DIP}		16.45	1.30	9.21	16.38

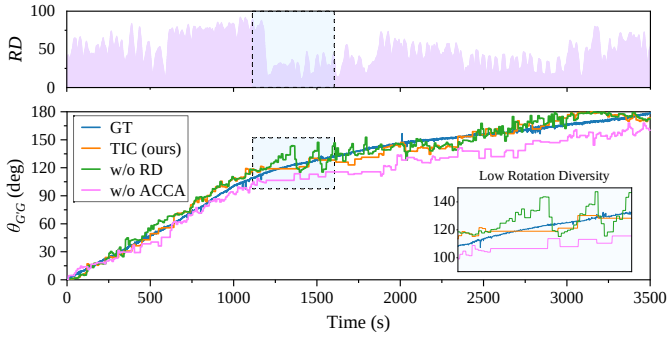

Fig. 7. Qualitative evaluation on of Rotation Diversity (*RD*) Trigger and ACCA in $R_{G'G}$ tracking (subject 1, right lower leg IMU in DS_{TIC}). $\theta_{G'G}$: angle of coordinate drift $R_{G'G}$ in degree; GT: ground truth coordinate drift in ego-yaw frame.

Table 5. Error of $R_{G'G}$ and R_{BS} and corresponding OME in DS_{TIC}.

IMU location	$R_{G'G}$ Err (°)	R_{BS} Err (°)	OME (°)
left forearm	10.49	21.41	21.40
right forearm	15.44	17.33	20.56
left lower leg	6.66	14.33	13.79
right lower leg	10.19	14.32	12.93
head	9.94	16.53	16.56
hip	2.34	7.76	6.00

motion capture. As shown in Fig. 8, our dynamic calibration reduces both R_{BS} error and positional error caused by non-standard calibration poses. Notably, compared to static calibration, the reduction in positional error using dynamic calibration is significantly greater (55.64% vs. 40.01%), ensuring accurate joint position estimation. This indicates that our dynamic calibration more effectively tracks the rotation components affecting joint position in R_{BS}, while disregarding part of rotation components unrelated to joint position (e.g., axial rotation of the forearm).

5.7 Evaluation on Global Translation

Our dynamic calibration significantly enhances the accuracy of translation estimation over extended periods. As shown in Table 6, the translation error using dynamic calibration (TIC) in the ego-yaw

Fig. 8. Qualitative evaluation on R_{BS} error. The visualized R_{BS} error and positional error are the average values of left and right forearm. The examples are from subject 5 in DS_{TIC}.

Table 6. Translation error under different coordinate frame with / without our dynamic calibration TIC. The translation is captured using PIP [Yi et al. 2022]. 1/2/5/10s: duration of time window. Ego-Yaw: fit the translation to the ego-yaw coordinate system of the first sample within the time window.

Coordinate System	TIC	Translation Error (cm)			
		1s	2s	5s	10s
Ego-Yaw	+	**8.38**	**12.46**	**22.24**	**36.43**
	-	15.31	23.16	34.06	44.96
SMPL	+	18.03	26.79	38.54	49.50
	-	14.09	19.84	29.97	41.69

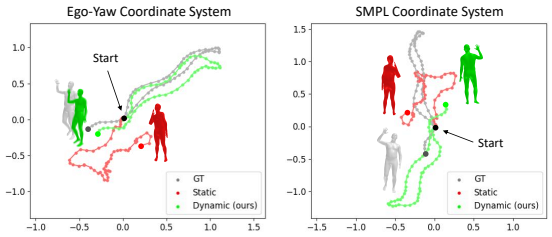

Fig. 9. Visualization of global translation tracking on DS_{TIC} in the SMPL frame and ego-yaw frame, unit: m.

coordinate system is markedly lower than when it is not used. This improvement is attributed to the fact that translation estimation relies on reliable pose estimation results for forward kinematics calculations [Yi et al. 2022, 2021, 2024], which TIC effectively ensures. However, since TIC can only solve IMU drift within the ego yaw coordinate system, it cannot correct the drift outside the ego yaw coordinate system (such as the drift of the root IMU). Consequently, for translation in a fixed global coordinate system (e.g., the SMPL coordinate system), accuracy of translation may not be guarantee (see Fig. 9).

5.8 Evaluation on Consumer-grade IMU

As shown in Table 7, we also verified our dynamic calibration TIC on the IMUPoser dataset [Mollyn et al. 2023], in which data were collected from 5 consumer-grade IMUs integrated in smartphones, smartwatches, and earbuds. These IMUs were located on head, wrists and front pockets of pants, with a sensor layout different from that in our work. The results demonstrated that TIC significantly reduced OME in the IMUPoser dataset, which highlights the potential of TIC for consumer-grade inertial motion capture systems.

Table 7. Results on the IMUPoser Dataset.

IMU Location	OME (°)	
	with TIC	without TIC
left wrist	**20.74**	29.42
right wrist	**21.95**	33.11
left front Pocket	**16.93**	19.68
right front Pocket (ego-yaw)	10.64	**9.18**
head	**24.51**	26.66
Avg	**18.95**	23.61

6 Limitations

Since our dynamic calibration is based on Assum. 2 and Assum. 3, it may fail under the following conditions: i) large and sudden changes in $R_{G'G}$ and R_{BS} (e.g., extreme environmental changes or sensor hardware inconsistencies); ii) low-activity scenarios, such as office work, watching TV, etc, where the rotation diversity may not meet the requirement for effective calibration; iii) we only consider the coordinate drift in the ego-yaw frame and do not support the correction of global yaw drift (Fig. 9); iv) irregular motions may lead to incorrect calibration (Fig. 10).

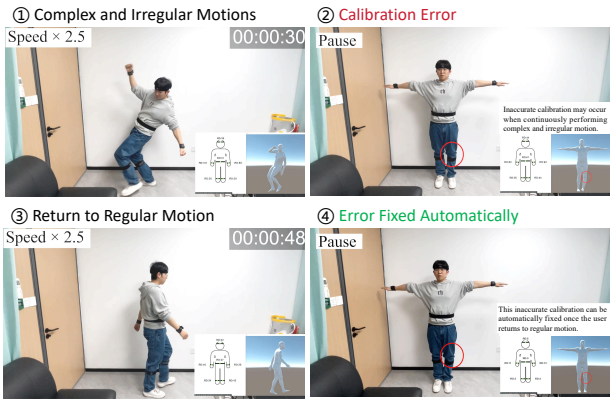

Fig. 10. Evaluation on complex and irregular motions in the demo video. We observed that such scenarios could lead to inaccurate calibration, as these types of motions are underrepresented in existing motion datasets. Fortunately, our approach can automatically recover and correct the calibration once the user resumes regular motion.

7 Conclusion

This work presented a novel dynamic calibration method for sparse inertial motion capture systems that breaks the restrictive absolute static assumption in traditional IMU calibration. Our key innovations include: i) real-time calibration parameters estimation under two relaxed assumptions that they change negligibly over short windows and human movements provide diverse IMU readings in that window, and ii) creating a Transformer-based model trained on synthetic data to learn the mapping from IMU readings to calibration parameters. Our work represents the first implicit IMU calibration technique that integrates calibration into regular usage without requiring an explicit calibration process. The promising results demonstrate the significant potential of our dynamic calibration framework in extending the capture duration and expanding the applications of inertial motion capture.

Acknowledgments

This work is supported by National Natural Science Foundation of China (62472364, 62072383), the Public Technology Service Platform Project of Xiamen City (No.3502Z20231043), Xiaomi Young Talents Program / Xiaomi Foundation and the Fundamental Research Funds for the Central Universities (20720240058). This work is partially supported by Royal Society (IEC \NSFC \211022). Shihui Guo is the corresponding author.

References

Karan Ahuja, Eyal Ofek, Mar Gonzalez-Franco, Christian Holz, and Andrew D. Wilson. 2021. CoolMoves: User Motion Accentuation in Virtual Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2 (2021).

Rayan Armani, Changlin Qian, Jiaxi Jiang, and Christian Holz. 2024. Ultra Inertial Poser: Scalable Motion Capture and Tracking from Sparse Inertial Sensors and Ultra-Wideband Ranging. In *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery.

Namchol Choe, Hongyu Zhao, Sen Qiu, and Yongguk So. 2019. A sensor-to-segment calibration method for motion capture system based on low cost MIMU. *Measurement* 131 (2019), 490–500.

Nathan DeVrio, Vimal Mollyn, and Chris Harrison. 2023. Smartposer: Arm pose estimation with a smartphone and smartwatch using uwb and imu data. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–11.

Yuming Du, Robin Kips, Albert Pumarola, Sebastian Starke, Ali Thabet, and Artsiom Sanakoyeu. 2023. Avatars Grow Legs: Generating Smooth Human Motion From Sparse Tracking Inputs With Diffusion Model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 481–490.

Julien Favre, Rachid Aissaoui, Brigitte M Jolles, Jacques A de Guise, and Kamiar Aminian. 2009. Functional calibration procedure for 3D knee joint angle description using inertial sensors. *Journal of biomechanics* 42, 14 (2009), 2330–2335.

Julien Favre, BM Jolles, Rachid Aissaoui, and K Aminian. 2008. Ambulatory measurement of 3D knee joint angle. *Journal of biomechanics* 41, 5 (2008), 1029–1035.

Aparna Harindranath and Manish Arora. 2024. A systematic review of user - conducted calibration methods for MEMS-based IMUs. *Measurement* 225 (2024), 114001.

Yinghao Huang, Manuel Kaufmann, Emre Aksan, Michael J Black, Otmar Hilliges, and Gerard Pons-Moll. 2018. Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–15.

Jiaxi Jiang, Paul Streli, Manuel Meier, and Christian Holz. 2023. EgoPoser: Robust Real-Time Ego-Body Pose Estimation in Large Scenes. arXiv:2308.06493 [cs.CV]

Jiaxi Jiang, Paul Streli, Huajian Qiu, Andreas Fender, Larissa Laich, Patrick Snape, and Christian Holz. 2022a. AvatarPoser: Articulated Full-Body Pose Tracking from Sparse Motion Sensing. In *Proceedings of European Conference on Computer Vision*. Springer.

Yifeng Jiang, Yuting Ye, Deepak Gopinath, Jungdam Won, Alexander W Winkler, and C Karen Liu. 2022b. Transformer Inertial Poser: Real-time human motion reconstruction from sparse IMUs with simultaneous terrain generation. In *SIGGRAPH Asia 2022 Conference Papers*. 1–9.

Manuel Kaufmann, Yi Zhao, Chengcheng Tang, Lingling Tao, Christopher Twigg, Jie Song, Robert Wang, and Otmar Hilliges. 2021. EM-POSE: 3D Human Pose Estimation From Sparse Electromagnetic Trackers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 11510–11520.

Anthony Kim and MF Golnaraghi. 2004. Initial calibration of an inertial measurement unit using an optical position tracking system. In *PLANS 2004. Position Location and Navigation Symposium (IEEE Cat. No. 04CH37556)*. IEEE, 96–101.

You Li, Xiaoji Niu, Quan Zhang, Hongping Zhang, and Chuang Shi. 2012. An in situ hand calibration method using a pseudo-observation scheme for low-end inertial measurement units. *Measurement Science and Technology* 23, 10 (2012), 105104.

Yu-Tao Liu, Yong-An Zhang, and Ming Zeng. 2019. Sensor to segment calibration for magnetic and inertial sensor based motion capture systems. *Measurement* 142 (2019), 1–9.

Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. 2019. AMASS: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF international conference on computer vision*. 5442–5451.

Christian Mandery, Ömer Terlemez, Martin Do, Nikolaus Vahrenkamp, and Tamim Asfour. 2015. The KIT whole-body human motion database. In *2015 International Conference on Advanced Robotics (ICAR)*. IEEE, 329–336.

Vimal Mollyn, Riku Arakawa, Mayank Goel, Chris Harrison, and Karan Ahuja. 2023. IMUPoser: Full-Body Pose Estimation using IMUs in Phones, Watches, and Earbuds. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–12.

Philipp Müller, Marc-André Bégin, Thomas Schauer, and Thomas Seel. 2016. Alignment-free, self-calibrating elbow angles measurement using inertial sensors. *IEEE journal of biomedical and health informatics* 21, 2 (2016), 312–319.

Milad Nazarahari, Alireza Noamani, Niloufar Ahmadian, and Hossein Rouhani. 2019. Sensor-to-body calibration procedure for clinical motion analysis of lower limb using magnetic and inertial measurement units. *Journal of Biomechanics* 85 (2019), 224–229.

L Noitom. 2017. *Perception neuron*. https://www.noitom.com/

Shaohua Pan, Qi Ma, Xinyu Yi, Weifeng Hu, Xiong Wang, Xingkang Zhou, Jijunnan Li, and Feng Xu. 2023. Fusing Monocular Images and Sparse IMU Signals for Real-time Human Motion Capture. *arXiv preprint arXiv:2309.00310* (2023).

Monique Paulich, Martin Schepers, Nina Rudigkeit, and Giovanni Bellusci. 2018. Xsens MTw Awinda: Miniature wireless inertial-magnetic motion tracker for highly accurate 3D kinematic applications. *Xsens: Enschede, The Netherlands* (2018), 1–9.

Abhinanda R Punnakkal, Arjun Chandrasekaran, Nikos Athanasiou, Alejandra Quiros-Ramirez, and Michael J Black. 2021. BABEL: Bodies, action and behavior with english labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 722–731.

Thomas Seel, Jorg Raisch, and Thomas Schauer. 2014. IMU-based joint angle measurement for gait analysis. *Sensors* 14, 4 (2014), 6891–6909.

Zainab F Syed, Prizanka Aggarwal, Christopher Goodall, Xiaoji Niu, and Naser El-Sheimy. 2007. A new multi-position calibration method for MEMS inertial navigation systems. *Measurement science and technology* 18, 7 (2007), 1897.

Bertram Taetz, Gabriele Bleser, and Markus Miezal. 2016. Towards self-calibrating inertial body motion capture. In *2016 19th International Conference on Information Fusion (FUSION)*. IEEE, 1751–1759.

David Tedaldi, Alberto Pretto, and Emanuele Menegatti. 2014. A robust and easy to implement method for IMU calibration without external equipments. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3042–3049.

David Titterton and John L Weston. 2004. *Strapdown inertial navigation technology*. Vol. 17. IET.

Matthew Trumble, Andrew Gilbert, Charles Malleson, Adrian Hilton, and John P Collomosse. 2017. Total capture: 3D human pose estimation fusing video and inertial sensors.. In *BMVC*, Vol. 2. London, UK, 1–13.

Tom Van Wouwe, Seunghwan Lee, Antoine Falisse, Scott Delp, and C Karen Liu. 2024. DiffusionPoser: Real-time Human Motion Reconstruction From Arbitrary Sparse Sensors Using Autoregressive Diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2513–2523.

Timo von Marcard, Roberto Henschel, Michael J. Black, Bodo Rosenhahn, and Gerard Pons-Moll. 2018. Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Xuan Xiao, Jianjian Wang, Pingfa Feng, Ao Gong, Xiangyu Zhang, and Jianfu Zhang. 2024. Fast Human Motion reconstruction from sparse inertial measurement units considering the human shape. *Nature Communications* 15, 1 (2024), 2423.

Vasco Xu, Chenfeng Gao, Henry Hoffmann, and Karan Ahuja. 2024. MobilePoser: Real-Time Full-Body Pose Estimation and 3D Human Translation from IMUs in Mobile Consumer Devices. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–11.

Dongseok Yang, Doyeon Kim, and Sung-Hee Lee. 2021. LoBSTr: Real-time Lower-body Pose Prediction from Sparse Upper-body Tracking Signals. *Computer Graphics Forum* 40, 2 (2021), 265–275.

Xinyu Yi, Yuxiao Zhou, Marc Habermann, Vladislav Golyanik, Shaohua Pan, Christian Theobalt, and Feng Xu. 2023. EgoLocate: Real-time Motion Capture, Localization, and Mapping with Sparse Body-mounted Sensors. *ACM Trans. Graph.* 42, 4 (2023).

Xinyu Yi, Yuxiao Zhou, Marc Habermann, Soshi Shimada, Vladislav Golyanik, Christian Theobalt, and Feng Xu. 2022. Physical inertial poser (pip): Physics-aware real-time human motion tracking from sparse inertial sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13167–13178.
Xinyu Yi, Yuxiao Zhou, and Feng Xu. 2021. Transpose: Real-time 3d human translation and pose estimation with six inertial sensors. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–13.
Xinyu Yi, Yuxiao Zhou, and Feng Xu. 2024. Physical Non-inertial Poser (PNP): Modeling Non-inertial Effects in Sparse-inertial Human Motion Capture. In *SIGGRAPH 2024 Conference Papers*.
Yu Zhang, Songpengcheng Xia, Lei Chu, Jiarui Yang, Qi Wu, and Ling Pei. 2024. Dynamic Inertial Poser (DynaIP): Part-Based Motion Dynamics Learning for Enhanced Human Pose Estimation with Sparse Inertial Sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1889–1899.
Xiaozheng Zheng, Zhuo Su, Chao Wen, Zhou Xue, and Xiaojie Jin. 2023. Realistic Full-Body Tracking from Sparse Observations via Joint-Level Modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 14678–14688.
Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. 2019. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5745–5753.
Chengxu Zuo, Yiming Wang, Lishuang Zhan, Shihui Guo, Xinyu Yi, Feng Xu, and Yipeng Qin. 2024. Loose Inertial Poser: Motion Capture with IMU-attached Loose-Wear Jacket. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2209–2219.

A Drift & Offset Simulation

Orientation. We simulate the impact of coordinate drift and measurement offset on orientation measurement based on Assum.2 and Eq.2 in the main paper. As shown in Table 8, the $R_{G'G}$ and R_{BS} used are obtained by randomly sampling from a uniform distribution within a specific range. For R_{BS}, we set a distribution range of 45 degrees equally for rotations around the x, y, and z axes, as R_{BS} is influenced by calibration pose errors and can occur around any rotation axis. For $R_{G'G}$, we distinguish two cases: 1) The $R_{G'G}$ for the non-root IMU is primarily set to rotate around the yaw (the y-axis in the SMPL frame), aligning with the characteristics of actual drifting; 2) The yaw rotation of the $R_{G'G}$ for the root IMU is set to 0, as the root IMU measures the root joint of the human body, and its yaw rotation always be 0 in the ego-yaw coordinate system, i.e.,

$$Yaw(R_{G'G}^{(root)} \cdot R_{GB_{root}}) = Yaw(R_{GB_{root}}) = 0 \tag{8}$$

where G represents the ego-yaw coordinate system, G' represents the drifted G, and B_{root} refers to the root bone, $R_{G'G}^{(root)}$ is the drifting applied to the root IMU. To ensure that the formula holds, the yaw rotation of $R_{G'G}^{(root)}$ must be 0.

Table 8. Distribution of random $R_{G'G}(t)$ and $R_{BS}(t)$ samples. $U(a,b)$: a uniform distribution between a and b (inclusive of a and b).

Rotation Matrix	Distribution		
	Θ_x	Θ_y	Θ_z
R_{BS}	$U(-45,45)$	$U(-45,45)$	$U(-45,45)$
$R_{G'G}$ (root)	$U(-20,20)$	$U(0,0)$	$U(-20,20)$
$R_{G'G}$ (non-root)	$U(-20,20)$	$U(-60,60)$	$U(-20,20)$

Hardware Level Acceleration. Here we illustrate our Hardware Level Acceleration simulation under coordinate drift. The global acceleration measurement of a real IMU is a **corrected value** obtained by manually removing the influence of Gravitational Acceleration (GA). The hardware level acceleration $\tilde{a}_G$ can be expressed as follows:

$$\begin{aligned} \tilde{a}_G &= R_{GS} \cdot (a_S - g_S) + g_{\text{bias}} \\ &= a_G - g_G + g_{\text{bias}} \end{aligned} \tag{9}$$

where a_S and g_S are the linear and gravitational acceleration in the sensor coordinate system, respectively. The g_{bias} is typically set to accurate gravitational acceleration in the global coordinate system G. In the hardware level, the coordinate drift $R_{G'G}$ can lead to inaccurate GA removal, which we refer to as **GA leakage** in accelerations measurement (Fig. 11). Based on Eq.2, the influence of coordinate drift on $\tilde{a}_G$ can be modeled as follow:

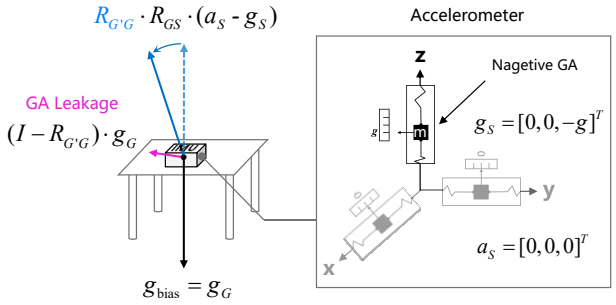

Fig. 11. Illustration of gravitational acceleration (GA) leakage in hardware level acceleration measurement. The accelerometer inherently includes a negative GA measurement. The $R_{G'G}$ can lead to inaccurate GA removal.

$$\begin{aligned} \tilde{a}_{G'} &= R_{G'G} \cdot (a_G - g_G) + g_{\text{bias}} \\ &= R_{G'G} \cdot a_G - R_{G'G} \cdot g_G + g_G \\ &= R_{G'G} \cdot a_G + (I - R_{G'G}) \cdot g_G \end{aligned} \tag{10}$$

The term $(I - R_{GG}) \cdot g_G$ represents the GA leakage. During the simulation, we set g_{bias} to $[0, -9.80665, 0]^T$, which accurately represents the gravitational acceleration in the ego-yaw coordinate system of the SMPL body.

B IMU Drifting Analysis

A direct observation of IMU drifting is demonstrated in the supplementary video (Fig 12). All six IMUs are placed in a compact holder, which ideally should maintain the IMUs to be fixed at the same orientation. A noticeable drift is visualized when monitoring the IMU orientation in real time. All six IMUs demonstrate different extents of drifting.

We further verified the IMU drifting when the sensors are placed on human body. This is achieved by tracking the IMUs, each with 3 optical markers. This provides the absolute orientation of IMUs, which allows the analysis of IMU drifting. As shown in Fig. 13, $R_{G'G}(t)$ is primarily the y-axis (vertical direction) rotation, while the rotations around the x-axis and z-axis (in the horizontal direction) are smaller. This is because the stable gravity direction can provide a reliable reference for the normal to the horizontal plane, thereby correcting the horizontal drift. However, the vertical rotation relies solely on the magnetometer signal for correction, which is highly

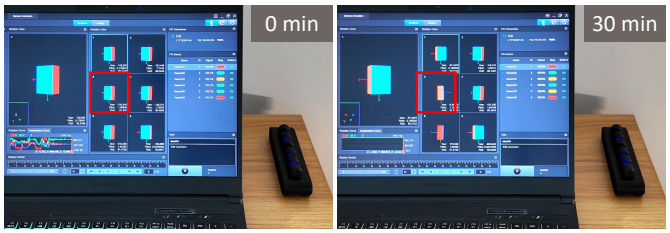

Fig. 12. Visualization of IMU drifting in the scenario of placing them in a desktop holder. In the 30-minute observation, we recorded a maximum yaw drifting of 0.07 deg/s.

susceptible to interference from metal objects, making severe IMU drift more likely to occur.

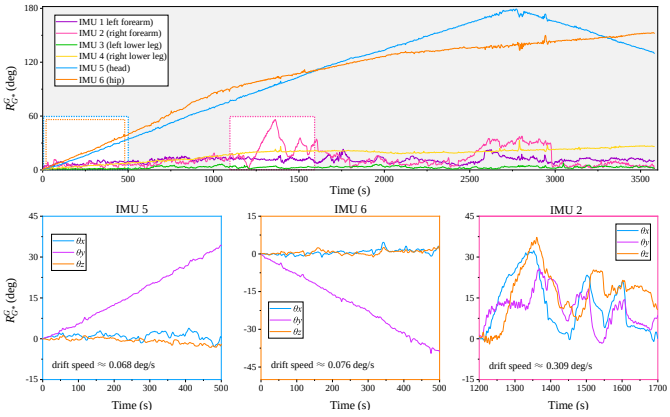

Fig. 13. Visualization of IMU drifting in real-world dataset (s1). We use SMPL frame as global frame.

Meanwhile, we also observed mixed-axis drift in the IMU on the right hand (IMU 2) over short periods, occurring during table tennis movements. This is because during this action, the right hand remains in motion for an extended period, making it difficult to obtain a stable gravity acceleration direction for correcting drift along the x-axis and z-axis.

C Calibration Pose Error Analysis

Static calibration estimates R_{BS} through specific calibration pose (e.g. T-Pose). The drawback of this method is that in real-world usage, the user's T-Pose is difficult to achieve completely accurately (see Fig. 14), leading to errors in estimating R_{BS}, thus affecting the accuracy of pose estimation.

Table 9 shows the errors in 30 T-Pose instances collected in our real-world dataset, indicating significant errors in the left and right forearms, which were well covered by the distribution of R_{BS} used in our data synthesis (±45° for the x, y and z axes).

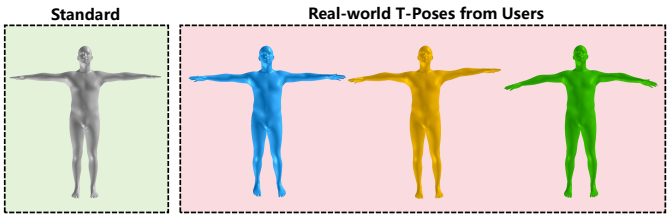

Fig. 14. T-Pose error in real-world usage.

Table 9. T-pose err of six joints.

Joint	T-Pose Err (deg)
left forearm	32.86 ± 11.43
right forearm	29.12 ± 9.35
left lower leg	11.74 ± 4.64
right lower leg	13.23 ± 4.60
head	5.01 ± 2.57
hip	2.78 ± 2.19

D Rationale of ego-yaw coordinate system

We define the ego-yaw coordinate system as the global coordinate system G in dynamic calibration due to the exist of unresolved components in the drifted foundational world coordinate systems W' (e.g. ENU or SMPL). As illustrated in Fig. 15, the overall coordinate drift $R_{W'G}$ can be decomposed into two parts: $R_{W'G'}$, corresponding to yaw rotation between drifted ego-yaw and drifted world coordinate system, which is yaw-only rotation that cannot not result in any unnatural motion; thus, it cannot be perceived and resolved through observation. In contrast, the remaining $R_{G'G}$ directly impacts the rotation of human joints, leading to unnatural movements. Therefore, we set $R_{G'G}$ as the target of dynamic calibration.

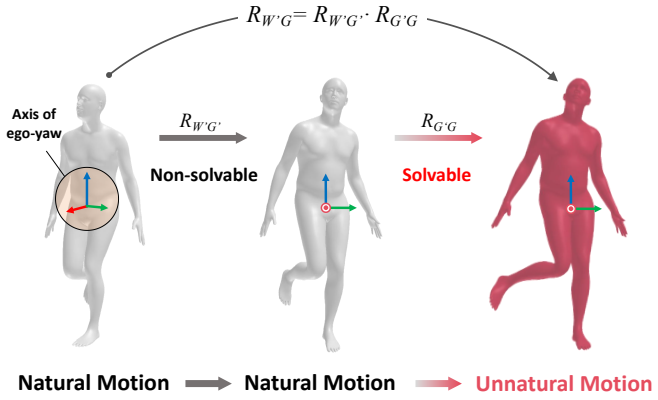

Fig. 15. Decompose of the coordinate drift based on the ego-yaw coordinate system.

E Hyperparameters Selection

E.1 T_R in Calibration Trigger

To ensure accurate calibration parameter estimation, setting a high T_R is a safe strategy. However, since everyday activities do not always maintain high RD, this could lead to calibration not being triggered in a timely manner to track changes in calibration parameters. Therefore, the T_R setting needs to balance calibration accuracy and trigger frequency.

To obtain the reference data for the T_R settings, we conducted experiments on the KIT [Mandery et al. 2015] subset of the AMASS dataset. The KIT dataset contains a large number of everyday actions, and we recorded the average rotation diversity of each consecutive $\hat{R}_{IMU}$ sequence sample per 512 frames (60Hz), as well as the OME of TIC on its corresponding 2000 randomly synthesized samples.

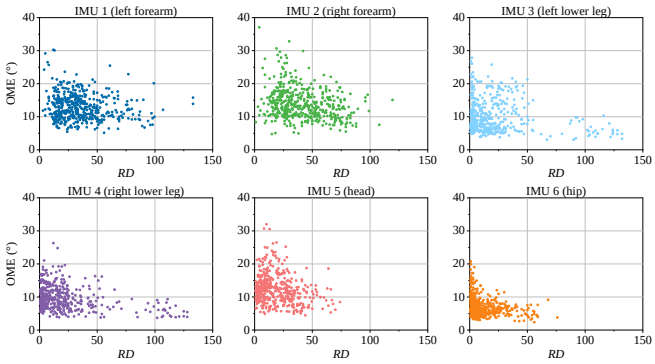

Fig. 16. The impact of rotation diversity on OME.

As shown in Fig. 16, we can observe that for each joint, lower RD tends to lead to higher OME. , demonstrating the importance of satisfying Assum.3 for accurate calibration parameter estimation. We also noticed that different joints have different T_R requirements.

In Table. 10, we present the final T_R we used, along with three indexes that guide the T_R settings for each joint: 1) $\overline{RD}$, the average RD across all motion samples; 2) $\overline{RD}_{<10}$, the average RD of samples where OME < 10° across all motion samples; and 3) s_{RD}, the RD sensitivity, defined as $s_{RD} = \overline{RD}_{<10}/\overline{RD}$, which represents the sensitivity of OME to RD. For joints with a higher s_{RD}, we tend to select T_R values above $\overline{RD}$ to ensure accuracy; conversely, for joints with a lower s_{RD}, we prefer T_R values below $\overline{RD}$ to ensure trigger frequency.

Table 10. T_R setting in TIC. $\overline{RD}$ and $\overline{RD}_{<10}$ are calculated from samples in Fig. 16.

Joint	$\overline{RD}$	$\overline{RD}_{<10}$	s_{RD}	T_R
left forearm	35.59	38.04	1.07	30
right forearm	41.85	56.99	1.36	50
left lower leg	27.46	30.96	1.13	30
right lower leg	27.01	32.20	1.19	30
head	22.09	28.40	1.29	25
hip	17.03	18.50	1.08	15

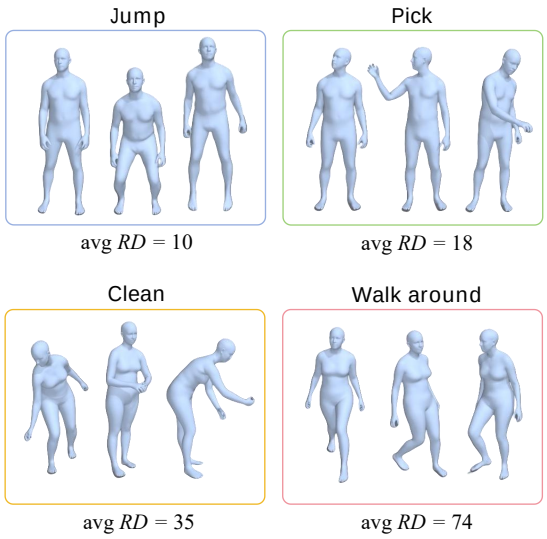

Fig. 17. Four examples of daily activities from KIT dataset. Motion labels and visualization are provided by BABEL dataset [Punnakkal et al. 2021].

Table 11. TIC performance under different running time interval t.

t (sec)	Avg OME (deg)	Avg AME (m/s^2)
0.5	15.20	1.30
1	15.20	1.30
2	15.39	1.31
5	15.42	1.31
10	15.54	1.31
20	15.85	1.34

Fig 17 shows the average RD for different daily activities. We found that activities involving body movement (such as cleaning, walking around) tend to cause changes in body's facing detection, resulting in an average RD greater than 30. Thus, our T_R settings (avg=30) is appropriate and ensures that dynamic calibration can be triggered in time during daily use.

E.2 Time intervals t of timing signal in dynamic calibration

During the inertial motion capture process, we use a timing signal S_t with an interval of $t$to run the TIC network. Because in general, the changing of the calibration parameters is slow (e.g., coordinate drift of 0.1 deg/s), allowing for updates at a lower frequency to avoid unnecessary computational cost. We tested the calibration accuracy under different t settings. As shown in the table 11, when $t \leq 2$ seconds, there is no significant difference between OME and AME. However, as t increases, OME begins to rise, and a notable decline in performance is observed in both OME and AME at $t = 20$ seconds. Consequently, we have chosen $t = 1$ seconds to achieve a trade-off between performance and computational cost.

E.3 Resolution of discretized Euler angle space

The ideal discretized Euler angle space should use the lowest possible split step to achieve higher resolution and accurately represent rotation diversity. However, since the Euler angle space is three-dimensional, too high a resolution would greatly increase the computational complexity of *RD* calculation. Therefore, we use a 15-degree split step for discretization, which ensures uniform segmentation with an acceptable *RD* computation complexity (takes 2.5ms to process 256 frames data on an NVDIA RTX 4080 GPU).

F TIC Network

F.1 Network Architecture

The transformer encoder blocks in Encoder (E) of TIC network and TPM module use the same network architecture, with details shown in Table 12.

Table 12. Network details of transformer encoder blocks in Encoder (E) of TIC network and TPM module.

Param	Value
Embedding Dim	256
Attention Heads	8
FFN Size	512

F.2 Choice of Architecture

Our TIC network features three key designs as follows:

- Transformer-backbone: The Transformer architecture was chosen for its advantages in sequence modeling.
- Encoder-only: The predictions of $\Delta R_{G'G}$ and ΔR_{BS} are fixed-size outputs rather than sequences.
- Temporal Average Pooling: This design supports variable input lengths, allowing our TIC network to adapt to different motion speeds or sampling rates without retraining.

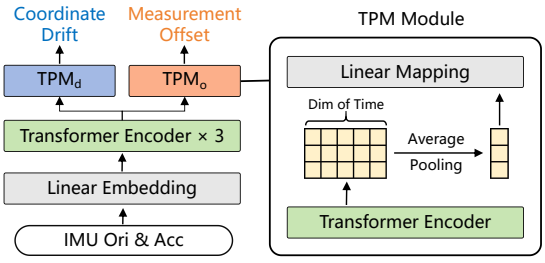

Fig. 18. Architecture of TIC network. The network uses three stacked Transformer Encoders to extract features from IMU orientation and acceleration sequences, and utilizes two TPM (Transformer Encoder-Pooling-Mapping) modules to estimate coordinate drift and measurement bias, respectively.

Table 13. Reproducibility analysis of TIC Network with 10 independent model training. Min/Max OME: 14.96/15.79; Min/Max AME: 1.30/1.32.

Joint	OME (deg)	AME (m/s^2)
left forearm	21.35 ± 0.95	1.52 ± 0.03
right forearm	20.52 ± 1.06	1.90 ± 0.05
left lower leg	13.63 ± 0.37	1.54 ± 0.01
right lower leg	12.92 ± 0.39	1.42 ± 0.01
head	16.51 ± 0.48	0.63 ± 0.02
hip	6.04 ± 0.44	0.81 ± 0.01
Avg	15.17 ± 0.23	1.30 ± 0.01

G Training Details

Data Format. In model training, the input of the TIC Network are sequences of concatenated acceleration $a_{\text{IMU}} \in \mathbb{R}^{n\times(6\times3)}$ and orientation (rotation matrices) $R_{\text{IMU}} \in \mathbb{R}^{n\times(6\times3\times3)}$ from 6 IMUs, where $n = 256$, indicating the length of the sequence. The a_{IMU} were divide by 30. The Ground Truth $R_{G'G} \in \mathbb{R}^{n\times(6\times6)}$, $R_{BS} \in \mathbb{R}^{n\times(6\times6)}$ were converted into 6D representations [Zhou et al. 2019], .

Training Settings. All experiments were conducted on a computer equipped with an Intel(R) Core(TM) i7-13700KF CPU and an NVIDIA RTX 4080 GPU. The TIC network was implemented using PyTorch 1.12.1 with CUDA 11.3. The training batch size was set to 128, learning rate to 0.001, and the Adam optimizer with default parameters was used.

The training process includes two steps: 1) *Pre-training.* The model is trained for 10 epochs using DS_{AMS}. 2) *Fine-tuning.* The model is trained for 3 epoch using the concatenated datasets DS_{AMS} and DS_{DIP}.

H Deployment

H.1 Re-calibration Cost

With an NVIDIA RTX 4080 GPU, the re-calibration requires only 4.5ms (2.5ms for RD computation, 2ms for TIC network inference), yielding an FPS of 222.2, which fully meets the real-time requirement.

H.2 Calibration Time w.r.t Initial Calibration Error

As Table 14 shows, we selected 21 samples with significant initial calibration errors from DS_{TIC} (Avg OME > 15°) and recorded the time costs to complete the calibration. It can be observed that average time costs grow with initial calibration errors. Nevertheless, the min time costs (last column) indicate that the calibration can be significantly accelerated by performing high-RD movements to enable rapid triggering and accurate $R_{G'G}$ and R_{BS} estimation.

I Evaluation on Xsens IMU

We manually selected a data segment with a high OME (8.86°) from the Total Capture dataset [Trumble et al. 2017] (collected with Xsens IMUs) and applied TIC for dynamic calibration. As shown in Table 15, although there is a slight increase in OME for the right forearm, head, and waist, the overall OME still decreased as expected (8.86°→7.92°). Fig. 19 illustrates the process of OME reduction of the right lower

Table 14. Statistic of Calibration Time w.r.t Initial Calibration Error.

Initial OME (deg)	Avg Time Cost (s)	Min Time Cost (s)
15-30	20.65±23.24	7.26 (Avg *RD*=19.17)
30-60	40.05±21.92	16.48 (Avg *RD*=24.50)
60-100	128.91±93.01	10.28 (Avg *RD*=28.16)

Table 15. Evaluation on Xsens IMUs. The data are sampled from Total Capture Dataset

Joint	OME (deg)	
	with TIC	without TIC
left forearm	**8.02**	10.93
right forearm	7.67	**7.37**
left lower leg	**13.79**	14.19
right lower leg	**9.13**	12.71
head	5.58	**4.79**
hip	3.34	**3.15**
Avg	**7.92**	8.86

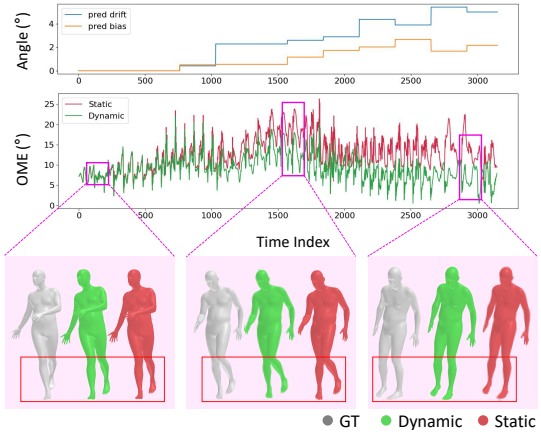

Fig. 19. An example of our dynamic calibration on Xsens IMUs. We visualize the orientation measurement error (OME) along with the predicted calibration parameters, which include the rotation angle of $R_{G'G}$ and R_{BS} of the right lower leg.

leg using our dynamic calibration. After the motion begins, the OME decreases with the updates of the coordinate drift $R_{G'G}$ and measurement bias R_{BS}. This demonstrates that our TIC does not depend on specific IMU device and has the potential for application in different types of inertial motion capture systems.

J Reproducibility Test

The use of randomly generated samples $R_{G'G}$ and R_{BS} in TIC Network training introduces additional uncertainty into the model training process, making reproducibility analysis essential. We performed 10 independent trainings of the TIC Network under a unified training setup and recorded the average OME and AME. As shown in Table 13, the performance variations across the 10 independent model trainings were minimal, indicating that the achieved metrics (OME: 15.20, AME: 1.30) are highly reproducible.

K Choice of $\Delta R_{G'G}$ & ΔR_{BS} v.s. $R_{G'G}$ & R_{BS}

We use the differences $\Delta R_{G'G}$ & ΔR_{BS} as they yield smaller output variation than $R_{G'G}$ & R_{BS}, thereby facilitating calibration accuracy as shown in Table 16.

Table 16. Comparison of differences update and absolute value update.

Method	OME	AME	$R_{G'G}$ Err	R_{BS} Err
$\Delta R_{G'G}$ & ΔR_{BS}	15.20	1.30	8.41	15.79
$R_{G'G}$ & R_{BS}	17.34	1.40	8.79	17.08

L TIC v.s. End-to-End Regression

The proposed TIC ensures accurate measurement of skeletal orientation and global acceleration through dynamic calibration parameter updates. However, a naive implementation of dynamic calibration is to use the TIC network to directly estimate the already calibrated data, known as End-to-End (End2End) Regression. To compare the effectiveness of these two approaches, we replaced the TPM module of the TIC Network with a DNN, used to estimate the calibrated orientation and acceleration. The modified model employed the same training settings as the TIC Network and was tested on real datasets.

Table 17. Competition of TIC and End-to-End regression.

Joint	OME (deg)		AME (m/s^2)	
	TIC	End2End	TIC	End2End
left forearm	**21.40**	24.68	**1.53**	1.91
right forearm	**20.56**	26.42	**1.90**	2.84
left lower leg	**13.79**	14.69	**1.54**	2.11
right lower leg	**12.93**	15.66	**1.42**	2.00
head	**16.56**	16.92	**0.62**	1.25
hip	**6.00**	6.40	**0.80**	1.16
Avg	**15.20**	17.46	**1.30**	1.88

As shown in Table 17, the OME and AME of End2End Regression are noticeably higher than those of TIC. This is because: **1) Cannot guarantee the satisfaction of ASSUM 3.** End2End Regression must operate in sync with motion capture at the same frame rate to provide real-time calibration, even when ASSUM 3 is not met. **2) Lack of prior regularization.** IMU calibration is a well-formulated process based on calibration parameters ($R_{G'G}$ and R_{BS}), and End2End Regression cannot incorporate such prior knowledge, leading to poor generalization . For example, acceleration calibration under arbitrary human motion can be accomplished via known $R_{G'G}$ and the acceleration measurement modeling (eq. 10), but using End2End Regression would require the training data to include all possible human motions, which is challenging to satisfy.