Loose Inertial Poser: Motion Capture with IMU-attached Loose-Wear Jacket

Chengxu Zuo[1] Yiming Wang[1] Lishuang Zhan[1] Shihui Guo[1*†] Xinyu Yi[2]
Feng Xu[2] Yipeng Qin[3]
[1]School of Informatics, Xiamen University, China [2]School of Software and BNRist, TsinghuaUniversity, China
[3]School of Computer Science & Informatics, Cardiff University, UK

Abstract

Existing wearable motion capture methods typically demand tight on-body fixation (often using straps) for reliable sensing, limiting their application in everyday life. In this paper, we introduce Loose Inertial Poser, a novel motion capture solution with high wearing comfortableness, by integrating four Inertial Measurement Units (IMUs) into a loose-wear jacket. Specifically, we address the challenge of scarce loose-wear IMU training data by proposing a Secondary Motion AutoEncoder (SeMo-AE) that learns to model and synthesize the effects of secondary motion between the skin and loose clothing on IMU data. SeMo-AE is leveraged to generate a diverse synthetic dataset of loose-wear IMU data to augment training for the pose estimation network and significantly improve its accuracy. For validation, we collected a dataset with various subjects and 2 wearing styles (zipped and unzipped). Experimental results demonstrate that our approach maintains high-quality real-time posture estimation even in loose-wear scenarios. Our dataset and code are available at: https://github.com/ZuoCX1996/Loose-Inertial-Poser.

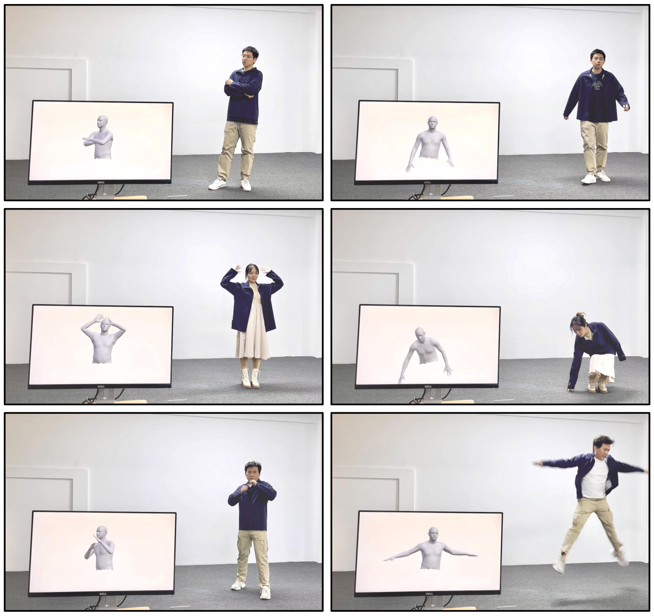


Figure 1. Loose Inertial Poser (LIP) achieves real-time upper-body motion capture through a loose-wear jacket equipped with 4 IMUs.

1. Introduction

Wearable motion capture enjoys its advantages in terms of portability, privacy friendliness and robustness against extreme lighting/occlusion compared to vision-based approaches. Recent works achieve posture estimation with a sparse number (3-6) of IMUs [13, 16, 33, 54, 55]. However, these methods still require the IMU sensors to be tightly attached to the body for stable measurement, which inevitably produces uncomfortable wearing experiences. Ideally, if motion capture can be achieved with our daily (mostly loose) clothes, the burden on users could be largely reduced, benefiting applications such as chronic disease monitoring and ubiquitous body-centric interaction.

The use of loose-wear clothes leads to the challenge that the secondary motion between clothes and human bodies degrades the quality of IMU readings for measurements of body orientation and acceleration. One direct approach to handle this data fluctuation is to train the model with a large volume of data that covers all possible secondary motions. Naively, such data can be collected by *simulating* IMU-attached clothing on existing human pose datasets. Nevertheless, in the case of loose wearing, IMU data simulation should encompass various body types and dressing styles, meaning that data simulation must cover a wide range of

This work is supported by National Natural Science Foundation of China (62072383, 61702433), the public technology service platform project of Xiamen City (No.3502Z20231043), the Fundamental Research Funds for the Central Universities (20720210044). This work is partially supported by Royal Society (IEC \NSFC \211022) *Corresponding author: Shihui Guo (guoshihui@xmu.edu.cn) †Also with Key Laboratory of Digital Protection and Intelligent Processing of Intangible Cultural Heritage of Fujian and Taiwan, Ministry of Culture and Tourism.

scenarios. Even with the use of deep learning-based clothing simulation [37] (approximately 1,000 times faster than physical simulation), it remains infeasible to simulate all wearing conditions within an acceptable timeframe.

In this paper, we introduce Loose Inertial Poser (LIP), which achieves real-time and accurate pose estimation with sparse IMUs attached to a loose-wear jacket (Fig. 1). To handle the aforementioned data scarcity challenge, we propose Secondary Motion AutoEncoder (SeMo-AE) to model and synthesize the effects of secondary motion between the skin and loose clothing on IMU data. Our SeMo-AE consists of two novel techniques. First, we propose *noise-guided latent space learning*, where we learn a latent space under the assumption that i) the IMU signal and the effects of secondary motion follow an additive relationship in the latent space and ii) the latent representation of the secondary motion (*i.e.*, noise) follows a Gaussian distribution. The key insight of our approach is that limited secondary motion samples in the simulated data pose a challenge in providing sufficient constraints for an additional generator network. Thus, it is more effective to leverage an autoencoder already well-constrained by reconstruction losses to meet an additional Gaussian prior, thereby effectively mitigating the data-hungry problem.

Second, we proposed a temporal coherence scheme to model the dependency of secondary motion in successive frames, resulting in less jittering and more realistic results. This scheme is based on our key observation that loose-wear IMU signals tend to deform smoothly over time, exhibiting local temporal coherence. After training, we concatenate our SeMo-AE with the pose estimation network to provide an unlimited supply of simulated ad-hoc data for training the network to estimate poses in loose-wear clothing.

To validate the effectiveness of our approach, we collected a real-world testing dataset covering different users and wearing styles. Extensive experimental results demonstrated that our method can effectively adapt to different wear conditions, achieving a mean joint rotation error of less than 20 degrees. Our main contributions include:

- We propose a real-time and accurate approach for human motion capture using loose-wear clothes embedded with a sparse number of IMU sensors, which guarantees a comfortable user experience.
- We propose a novel Secondary Motion AutoEncoder (SeMo-AE) network for synthesizing loose-wear IMU data. SeMo-AE models secondary motion as additive Gaussian noise in the latent space, enabling it to generate synthetic IMU data with novel secondary motions by sampling the Gaussian noise distribution using a very limited simulation dataset.
- We propose a temporal coherence scheme to model local temporal coherence of secondary motion, thereby reducing jittering and producing more realistic results.

2. Related Works

2.1. Motion Capture

Here we roughly categorized motion capture methods into vision-based and non-vision-based ones [10, 30].

Vision-based Methods. Traditional optical motion capture systems utilize multiple cameras and marker points [5, 9, 22], as seen in commercial systems like OptiTrack [39]. In recent years, the advancement of deep learning has opened new possibilities for markerless motion capture [41]. Single-camera 2D/3D pose estimation methods such as HRNet [45], SMAP [59], PARE [20], ViTPose[51] and others [23, 51, 52, 58] have had remarkable progress on human pose estimation. Additionally, approaches using RGB-D data [17, 56, 63] or multiple-view images [49, 53, 57] not only enhance accuracy but also mitigate challenges posed by the absence of depth information in images.

Non-vision-based Methods. These methods often use wearable sensors to capture human movement [11]. Inertial sensors, primarily comprising accelerometers and gyroscopes, provide an effective alternative for vision-based motion capture systems [21]. Commercial inertial motion capture systems, such as Xsens [38] and Noitom [35], fix multiple Inertial Measurement Units at various joints of the body to achieve accurate pose estimation.

A key focus of current research is exploring methods to configure sparse IMUs, aiming to reduce costs and invasiveness while preserving the accuracy of motion capture. In this work, we follow this direction and use 4 IMUs for upper body motion capture.

As an emerging direction, sparse inertial motion capture reconstructs the motion information with limited sensing inputs. Researchers attempt to address this issue using statistical optimization or deep learning approaches. Marcard et al. [46] utilized data from six IMUs to reconstruct human motion. However, due to its optimization-based nature, this method requires a considerable amount of time to process the entire sequence. In contrast, deep learning methods can learn to predict the current motion state with a small number of past frames, trading off for lower latency to achieve near real-time effects. For instance, Huang and colleagues [13] achieved real-time human motion capture with sparse IMUs through bidirectional RNN, but this approach is limited to estimating human body pose, overlooking body displacement. Another work, TransPose [54], further advanced sparse IMUs motion capture by integrating multi-stage pose estimation and a blended global displacement estimation, incorporating a module for physical dynamics optimization in subsequent work [55]. Jiang et al. [16] introduced Transformer into sparse inertial motion capture, simultaneously considering human motion in non-planar scenarios. Building upon this, they accomplished the task of motion capture and generated topographic height maps for human motion trajectories. In practical application domains [4, 15, 48], Ponton and his team [40] utilized 6 six-degrees-of-freedom VR trackers, incorporating a convolutional autoencoder and a learning-based inverse kinematics adjustment component for real-time full-body pose reconstruction. Additionally, some studies have integrated sparse IMUs with other forms of information [1, 7, 34, 47, 60]. For instance, Pan et al. [36] fused signal inputs from images and sparse IMUs to obtain more robust motion capture results.

Although using a sparse set of sensors, current approaches predominantly fix IMUs with straps at specific positions. This provides relatively accurate information, but compromises user experience since the fixation introduces noticeable sensation of rigid on-body gadgets.

2.2. Motion Capture on Clothing

Instead of treating wearable sensors as attachable items, another solution explores integrating the capability of motion capture into clothing [6]. Mainstream approaches include fixing marker points or inertial sensors at key joint locations in tight-wear garments, such as the TESLASUIT [44] motion capture clothing, which integrates 14 IMUs on its tight-wear garment. Another work [2] implemented end-to-end motion capture on tight-wear garments with sparse spatio-temporally synchronized infrared depth cameras and optical markers. In the field of smart clothing, flexible fabric sensors have also garnered attention [12, 26, 42]. For instance, Chen et al. [3] utilized six flexible stretchable sensors on the elbow pad to predict joint bending angles. Liang and his team [24] crafted garments using pressure-sensitive fabric material, identifying specific body postures based on the pressure distribution applied by users to the clothing. A recent work [64] proposed an adaptive motion tracking model to address the challenge of data offset resulting from unknown displacements in flexible sensors during motion.

Although the aforementioned studies primarily use tight-wear garments to obtain accurate data, loose-wear clothing is more aligned with consumer preferences and comfort considerations in daily situations. Some recent works made progress in studying human body poses based on loose-wear clothing [8, 29, 50]. For example, Zhou et al. [61] integrated a multi-channel capacitive sensor into a jacket, utilizing a deep regressor to predict upper body joint coordinates from 16-channel fabric capacitive sensors. Lorenz et al. [28] explored mapping motion data from loose-wear to tight-wear clothing using multiple IMUs, but with limited generalization capability. While some studies have employed IMUs attached to clothing for activity recognition tasks [14, 25, 31, 43], the challenge of sparse inertial sensing on loose-wear clothing is still under-explored.

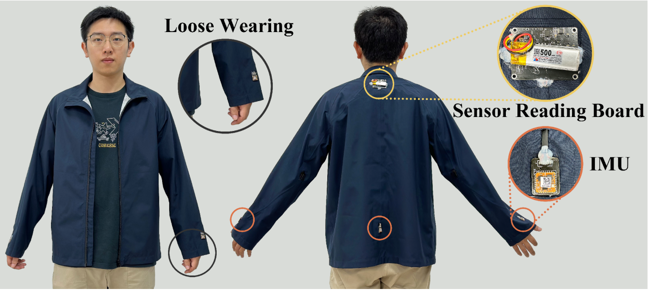


Figure 2. Prototype demonstration of the loose-wear jacket embedded with IMU sensors and the circuit board for data collection.

3. Hardware

Fig. 2 shows the prototype of the loose-wear jacket constructed for upper body motion tracking. Without loss of generality, we select the upper body to demonstrate the effectiveness of our method. With appropriate modification to our device and procedure, our approach can also be effectively applied to the scenario of lower body. The clothing is made of 100% polyester and is designed in a standard size of XL. The garment is fabricated using an unaltered pattern of a commercial outdoor jacket. The typical parameters of body girths include shoulder (52 cm), breast (106 cm), waist (110 cm) and wrist (28 cm), suitable for subject with height range (170-185 cm) and weight range (60-85 kg). The close-up view of the cuff position visually demonstrates the loose-wear feature (Fig. 2).

For upper body pose estimation, the *loose-wear* jacket is equipped with four IMU sensors. The IMU sensor is Xsens MTI-3. The IMUs are positioned on the left forearm, right forearm, back (integrated with the sensor reading board), and waist, respectively. All IMUs are connected to the sensor reading board with flexible cables, which are hot-pressed onto the garment and seamlessly integrated with the clothes. The board collects data from each IMU at a rate of 30Hz and transmits wirelessly to the computer via Bluetooth. The complete electronics system is powered by a Lithium battery with a capacity of 1000mAh (equivalently 5 hours for one charge). This multi-IMU setup allows us to capture upper body motion by fusing the orientation measurements from all four sensor locations.

Challenge 1 (Hardware) *Although significantly increasing the wearing comfort, the loose coupling between the garment and skin induces secondary motion that acts as a complex disturbance to the IMU measurements. This is in contrast to previous works that focus on tight-wear wearables [13, 16, 33, 54, 55] where the garment moves in sync with the underlying body segments, resulting in cleaner IMU signals.*

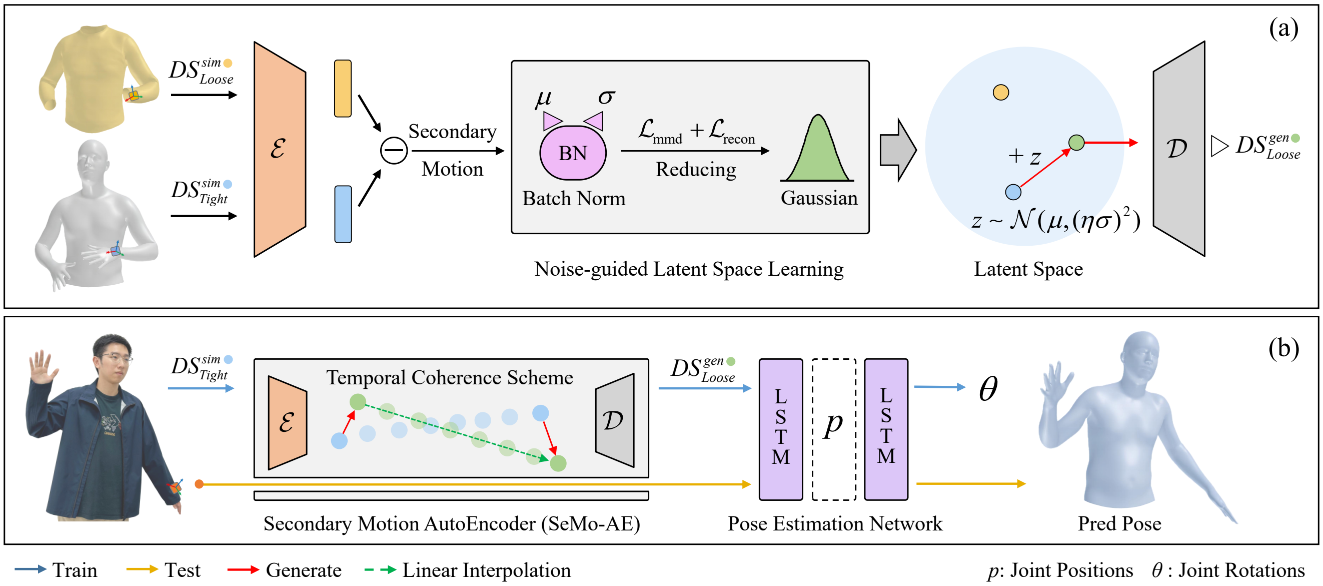


Figure 3. Overview of Loose Inertial Poser. (a) The training process of SeMo-AE and how it generates loose-wear IMU data from tight-wear IMU data. (b) The training of the pose estimation model and how it performs pose estimation. We generate diverse loose-wear IMU data using SeMo-AE and Temporal Coherence Scheme to train the pose estimation model, making it suitable for our loose-wear jacket.

4. Background and Problem Definition

Pose Estimation with IMUs. Let $\boldsymbol{\theta}(t) \in \mathbb{R}^{J\times 3}$ be the 3D rotation angles at time t defined on J joints. The in-garment IMU sensor network consists of M IMUs located at $\mathbf{s}_m, m = 1, ..., M$. Each IMU m measures accelerations $\mathbf{a}_m(t) \in \mathbb{R}^3$ and rotation $\boldsymbol{\mathcal{R}}_m(t) \in \mathbb{R}^{3\times 3}$. We employ a two-stage pipeline to learn the mapping between joint positions $\mathbf{p}(t) \in \mathbb{R}^{J\times 3}$, $\boldsymbol{\theta}(t)$ and IMU signals $\text{IMU}(t) = (\{\mathbf{a}_1(t), \mathbf{a}_2(t), ..., \mathbf{a}_m(t)\}, \{\boldsymbol{\mathcal{R}}_1(t), \boldsymbol{\mathcal{R}}_2(t), ..., \boldsymbol{\mathcal{R}}_m(t)\})$:

$$\begin{aligned} \mathbf{p}(t) &= \text{LSTM}(\text{IMU}(t), \mathcal{H}), \\ \boldsymbol{\theta}(t) &= \text{LSTM}(\mathbf{p}(t), \text{IMU}(t), \mathcal{H}), \end{aligned} \tag{1}$$

where $\mathcal{H}$ denotes the LSTM-encoded motion history, and we use $J = 10$ to represent upper body joints. Note that due to the loose-wear nature, the rotation measurement of the waist IMU cannot be directly used as the pelvis rotation. Consequently, our model incorporates a separate pelvis rotation estimate, which is different from the tight-wear situation.

Pose Estimation with loose-wear IMUs. As aforementioned, loose-wear IMUs induce secondary motion which adds disturbances $\delta(t)$ to the ideal tight-wear IMU signals $\text{IMU}(t)$, resulting in $\widehat{\text{IMU}}(t)$:

$$\widehat{\text{IMU}}(t) = f(\text{IMU}(t), \delta(t)), \tag{2}$$

where f is an unknown function; $\delta(t)$ is determined by various factors such as the user pose, body shape, and clothing fit, exhibiting complex and nonlinear dynamics with elements of randomness that are difficult to model analytically. Although challenging, both f and $\delta(t)$ can be estimated in a data-driven way using Eq. 1, under the assumption that *there is abundant motion data collected from loose-wear IMUs*. However, this assumption is difficult to satisfy due to the aforementioned various factors affecting $\delta(t)$. Thus, we have:

Challenge 2 (Software) *While deep learning provides a viable solution for pose estimation from loose-wear IMUs, it relies on the availability of abundant training data. However, the collection of diverse real-world motion capture data across user poses, body shapes, clothing fit and randomness poses a practical data scarcity challenge.*

Therefore, there is a critical need for a synthetic data generation approach capable of producing realistic and diverse loose-wear IMU signals to train pose estimation models, which can be formulated as:

$$\widehat{\text{IMU}}(t, z) = G(\text{IMU}(t), z), \tag{3}$$

where G denotes a deep generative model, z denotes the noise vectors (usually sampled from a standard normal distribution) used in G that capture the randomness of $\delta(t)$.

5. Secondary Motion AutoEncoder

5.1. Noise-guided Latent Space Learning

We propose a novel autoencoder framework for secondary motion modeling and synthetic loose-wear IMU data generation (Fig. 3), with $\mathcal{E}$ denotes its encoder, $\mathcal{D}$ denotes its decoder. Our key idea is to utilize the universal approximation capability of deep neural networks to learn a latent space in which:

- $\delta(t)$ is captured by $\mathcal{E}(\widehat{\text{IMU}}(t, z)) - \mathcal{E}(\text{IMU}(t)) = z$;
- z follows a normal distribution, *i.e*., $z \sim \mathcal{N}(\mu, \sigma^2)$;

then we can model secondary motion as Gaussian noise in the latent space. Unlike the naive approach that learns the distribution of z in a given latent space, we reverse the idea to learn a latent space where the distribution of z is predefined, thus calling it noise-guided latent space learning.

Training. Given training data $\text{IMU}(t)$ and $\widehat{\text{IMU}}(t)$, we feed them into the encoder $\mathcal{E}$, and implement our noise-guided latent space learning with the reparameterization method [19] and a Maximum Mean Discrepancy (MMD) loss $\mathcal{L}_{\text{mmd}}$ [27], which measures the distributional distance between batch normalized $\mathcal{E}(\widehat{\text{IMU}}(t, z)) - \mathcal{E}(\text{IMU}(t))$ and a standard normal distribution:

$$\mathcal{L}_{\text{mmd}} = \text{MMD}(\text{BN}(\mathcal{E}(\widehat{\text{IMU}}(t, z)) - \mathcal{E}(\text{IMU}(t))), \quad (4)$$

where BN denotes a batch normalization layer. Our motivation of employing batch normalization differs from its conventional use in training optimization. In our case, batch normalization addresses the challenge that directly imposing $\mathcal{E}(\widehat{\text{IMU}}(t, z)) - \mathcal{E}(\text{IMU}(t))$ on a specific parameterized Gaussian distribution (e.g., each dimension with $\mu = 0$, $\sigma = 1$) could otherwise constrain the solution space and impede the model from converging to potentially better solutions. Therefore, we incorporate batch normalization into $\mathcal{L}_{\text{mmd}}$, ensuring that any parameterized Gaussian distribution of $\mathcal{E}(\widehat{\text{IMU}}(t, z)) - \mathcal{E}(\text{IMU}(t))$ minimizes $\mathcal{L}_{\text{mmd}}$.

Together with the standard MSE reconstruction losses used in autoencoders:

$$\begin{aligned}\mathcal{L}_{\text{recon}} = \frac{1}{n}\sum_{t=1}^{n} & ||\mathcal{D}(\mathcal{E}(\text{IMU}(t))) - \text{IMU}(t)||_2^2 \\ & + ||\mathcal{D}(\mathcal{E}(\widehat{\text{IMU}}(t))) - \widehat{\text{IMU}}(t)||_2^2,\end{aligned} \quad (5)$$

we train our model with the overall loss function:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{mmd}}. \quad (6)$$

Inference. After training, we can generate diverse synthetic loose-wear IMU data by:

$$\widehat{\text{IMU}}(t, z) = \mathcal{D}(\mathcal{E}(\text{IMU}(t)) + z), \quad (7)$$

where $z \sim \mathcal{N}(\mu, \sigma^2)$, μ and σ are obtained from the batch normalization layer.

Data Extrapolation. To enable generating synthetic data with increased levels of secondary motion, we introduce a controllable looseness parameter η ($\eta > 1$) into z (Eq. 7):

$$z^* \sim \mathcal{N}(\mu, (\eta\sigma)^2). \quad (8)$$

We use $\eta = 2$ during inference of SeMo-AE. Empirically, we show that such extrapolated data improves the training of our pose estimation model (see supplementary material).

5.2. Temporal Coherence Scheme

Although effective, Eq. 7 neglects the temporal coherence of secondary motions between successive frames in a motion sequence. To address this issue, we propose a Temporal Coherence Scheme as a heuristic to linearly interpolate the noise z to $z(t)$ and have:

$$z(t) = (1 - \frac{t}{n}) \cdot z(1) + \frac{t}{n} \cdot z(n), \quad (9)$$

where $z(1), z(n) \sim \mathcal{N}(\mu, (\eta\sigma)^2)$, n is a user-specified hyper-parameter controlling the length of the interpolation over the motion sequence. Intuitively, larger n produces smoother synthetic IMU signals. Accordingly, we have:

$$\widehat{\text{IMU}}(t, z(t)) = \mathcal{D}(\mathcal{E}(\text{IMU}(t)) + z(t)). \quad (10)$$

It is worth noting that when $n = 2$, Eq. 10 will degenerate to Eq. 7. We use $n = 128$ in our work, based on the autocorrelation analysis of the noise vector sequence z (see supplementary material).

6. Experiments

6.1. Dataset and Metrics

Simulation Dataset. This dataset, consisting of paired loose-wear and tight-wear simulated IMU data DS^{sim}_{Loose} and DS^{sim}_{Tight}, is used to train the proposed SeMo-AE:

- DS^{sim}_{Loose}: First, we utilized the TailorNet [37] to simulate the corresponding clothing models required for various poses within the AMASS dataset [54], which consists of over 9 million frames. TailorNet can rapidly simulate topologically consistent clothing given SMPL pose and body shape, meeting the requirements for IMU data simulation. We selected the *Shirt* model from the ones provided by the authors, which is most similar to our loose jacket. Additionally, we configured the SMPL model's physique as *Tall Thin* to ensure that greater space between the clothing and the skin. This configuration results in a more accurate representation of a loosely worn scenario. *Overall, this single simulation required approximately 4 days to complete using an NVIDIA RTX 4080 graphics card.* Then, to simulate an IMU, we selected 4 nearby vertices on the simulated clothing based on their positions

We omit z as it denotes the training data without the synthesis process in our jacket to describe 2 axis directions, then calculated the third axis direction through cross-products to simulate orientation measurement. Additionally, we used the geometric center of these four vertices as the position of the IMU for acceleration simulation.

- DS^{sim}_{Tight}: We simulate tight-wearing IMU signals by directly placing the virtual IMUs, using the method described above, on the human body mesh obtained from the AMASS dataset [54]. Notably, we adapted the joint and mesh vertex settings to match our upper body IMU setup (left forearm, right forearm, back and waist).

Synthetic Dataset (On-demand). We synthesize novel loose-wear IMU data, denoted as DS^{gen}_{Loose}, using the proposed SeMo-AE on-demand to provide training data for the pose estimation network.

Testing Dataset (Real). We recruited a total of five individuals with varying body shapes and collected a real-world dataset, denoted as DS^{real}_{Loose}, for evaluation using the loose-wear jacket integrated with four IMUs. All participants were informed of the experiment purpose and signed the consent agreement for participation. Their body size fits the designed garment. More information on participants' body characteristics can be found in the supplementary materials. Each participant was instructed to perform data collection in two different wearing styles: zipped and unzipped. During data collection, participants were asked to perform five predefined actions, including walking, running, jumping, boxing, and ping-pong, as well as five times of free-form movements. Each action lasted for one minute. For ground truth pose, we used the Perception Neuron 3 system [35] to capture upper body poses using 11 tight-wear IMUs. Overall, we collected 212,496 frames of 30 fps data, with a total duration of about 2 hours.

Metrics. Following TransPose [54], we measure the accuracy of pose estimation using the following 2 metrics: *1) angular error*, which measures the mean rotation error of 10 upper body joints in degrees; *2) positional error*, which measures the mean Euclidean distance error of 12 upper body joint endpoints in centimeters. Note that in practice, we only calculate the positional error of 11 joint endpoints as the position of the pelvis joint is kept at $[0, 0, 0]$.

6.2. Training Details

IMU Data Format. We utilize the acceleration and rotation measurements of IMU as the input for all models. Following TransPose [54], we applied normalization to the IMU data, transforming the acceleration and rotation measurements of the left, right, and back IMUs into relative values to the waist IMU. To facilitate the training of the pose estimation network, we convert the IMU rotation readings into their corresponding 6D representation [62]. As a result, we define IMU data as $I \in \mathbb{R}^{36}$, which is obtained by concatenating three-axis accelerations and 6D rotation of 4 IMUs, *i.e.*, $(3 + 6) \times 4 = 36$.

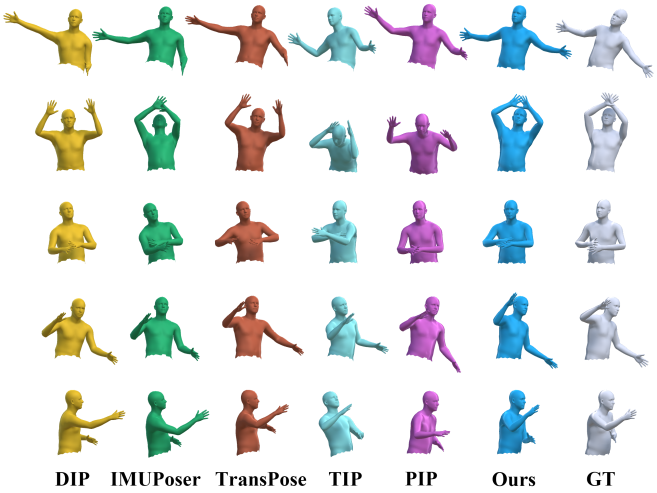


Figure 4. Qualitative results on the testing dataset DS^{real}_{Loose}. Our approach mitigates the disturbance in IMU data caused by secondary motion, achieving accurate motion capture.

Training Settings. All our experiments run on a computer with an Intel(R) Core(TM) i7-13700KF CPU and an NVIDIA RTX 4080 GPU. The model is implemented using PyTorch 1.12.1 with CUDA 11.3.

- *Training SeMo-AE.* The SeMo-AE was trained with paired DS^{sim}_{Loose} and DS^{sim}_{Tight}, utilizing a batch size of 512. We employed the Adam [18] optimizer with a learning rate of $lr = 1 \times 10^{-3}$ during training.
- *Training Pose Estimation Network.* The pose estimation network was trained with DS^{gen}_{Loose}, utilizing a batch size of 256. We employed the Adam optimizer with a learning rate of $lr = 5 \times 10^{-4}$ during training.

Please refer to the supplementary material for the detailed SeMo-AE and pose estimation network structure.

6.3. Comparison with SOTA Methods

Quantitative Results. We compared our method with state-of-the-art (SOTA) methods in sparse inertial motion capture, including DIP [13], IMUPoser [33], TransPose [54], TIP [16], and PIP [55] on the testing dataset DS^{real}_{Loose}. Since these SOTA methods are designed for tight-wear IMUs with different sensor numbers and installation positions, we followed the official code provided by the authors and retrained the models using DS^{sim}_{Loose}, and reported the angular and positional errors on DS^{real}_{Loose}. As shown in Table 1 and Table 2, SOTA methods are much less effective for motion capture using loose-wear IMUs due to the low diversity of DS^{sim}_{Loose}. In addition, we observed that PIP, the best-performing SOTA method, produces abnormal

Table 1. Experimental results of angular error (unit: degree) on five participants. The s1 indicates the subject with ID=1.

ID	DIP	IMUPoser	TransPose	TIP	PIP	Ours
s1	25.63 ± 6.06	25.90 ± 6.33	24.60 ± 6.31	25.50 ± 6.93	23.34 ± 6.02	**19.91 ± 5.86**
s2	25.16 ± 5.44	23.19 ± 5.07	22.99 ± 5.37	21.78 ± 6.35	19.97 ± 3.89	**18.10 ± 4.55**
s3	30.31 ± 7.20	29.87 ± 7.50	28.51 ± 6.92	29.42 ± 8.53	26.35 ± 6.89	**23.72 ± 7.54**
s4	26.20 ± 7.78	26.09 ± 7.49	26.75 ± 7.65	25.63 ± 8.42	22.05 ± 6.68	**19.62 ± 7.98**
s5	24.62 ± 5.99	23.20 ± 5.76	23.17 ± 6.13	22.64 ± 6.99	19.78 ± 5.01	**17.81 ± 5.84**
zipped	25.20 ± 6.54	24.59 ± 6.40	24.63 ± 6.44	24.05 ± 7.54	21.62 ± 6.08	**18.74 ± 6.28**
unzipped	27.57 ± 6.92	26.67 ± 7.29	25.73 ± 7.19	25.87 ± 8.23	22.95 ± 6.39	**20.98 ± 6.93**
all	26.38 ± 6.85	25.63 ± 6.94	25.18 ± 6.85	24.97 ± 7.94	22.28 ± 6.28	**19.83 ± 6.79**

Table 2. Experimental results of positional error (unit: cm) on five participants. The s1 indicates the subject with ID=1.

ID	DIP	IMUPoser	TransPose	TIP	PIP	Ours
s1	16.85 ± 6.84	17.75 ± 6.57	17.69 ± 6.26	16.47 ± 6.47	17.01 ± 7.68	**10.23 ± 5.12**
s2	15.19 ± 5.88	15.58 ± 5.16	15.50 ± 5.54	14.36 ± 6.31	13.83 ± 6.70	**9.08 ± 4.80**
s3	16.95 ± 6.63	18.63 ± 6.38	16.72 ± 6.26	16.17 ± 6.95	15.33 ± 6.62	**11.21 ± 6.17**
s4	20.10 ± 7.99	20.67 ± 7.39	20.40 ± 7.36	19.06 ± 7.60	19.62 ± 9.33	**12.67 ± 7.87**
s5	14.99 ± 6.39	16.45 ± 5.34	14.96 ± 5.78	15.19 ± 7.09	13.23 ± 6.76	**9.86 ± 6.41**
zipped	16.91 ± 7.00	17.99 ± 6.58	17.58 ± 6.40	16.63 ± 7.38	16.61 ± 8.22	**10.19 ± 6.08**
unzipped	16.61 ± 6.99	17.56 ± 6.29	16.41 ± 6.63	15.78 ± 6.71	14.87 ± 7.25	**10.98 ± 6.56**
all	16.77 ± 7.00	17.78 ± 6.44	17.00 ± 6.54	16.20 ± 7.07	15.74 ± 7.80	**10.58 ± 6.24**

joint position estimates, impeding the proper functioning of its physics-aware motion optimizer. In contrast, our method outperforms all SOTA methods including PIP, demonstrating a clear advantage in robustness to secondary motion.

Qualitative Results. In Fig. 4, we qualitatively compare our method with SOTA ones w.r.t. the ground truth on five distinct actions. It can be observed that secondary motions introduce significant errors in the results of SOTA methods, especially for the first two actions involving large movement amplitudes. In contrast, our approach maintains stable and accurate motion capture across all actions, demonstrating clear superiority in handling secondary motion.

6.4. Ablation Study

Generator	Err Ang (deg)	Err Pos (cm)
SeMo-AE	19.83 ± 6.79	10.58 ± 6.24
CGAN	25.06 ± 7.18	16.47 ± 7.24
None	24.18 ± 7.38	18.61 ± 6.53

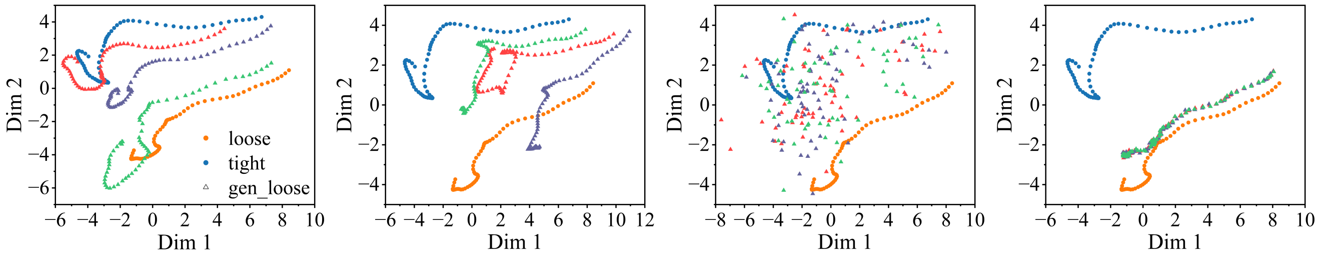
Table 3. Comparison of SeMo-AE and cGAN.

Quantitative Results. Since our loose-wear IMU data synthesis is essentially a conditional generation task (Eq. 2, IMU(t) is the condition), we compare our SeMo-AE with conditional GAN (cGAN) [32]. To facilitate a fair comparison, we train cGAN using the same data (D^{sim}_{Tight} as its condition and D^{sim}_{Loose} as its output) and latent noise distribution as our SeMo-AE. As Table 3 shows, the loose-wear IMU data generated by cGAN provides little improvement over using no synthetic data ("None", training directly on DS^{sim}_{Loose}), demonstrating much inferior performance compared to SeMo-AE.

Case	NLSL	TCS	Err Ang (deg)	Err Pos (cm)
1	+	+	19.83 ± 6.79	10.58 ± 6.24
2	+	-	21.46 ± 7.19	12.11 ± 7.23
3	-	+	21.54 ± 6.52	11.91 ± 6.33
4	-	-	21.71 ± 7.01	12.14 ± 6.99

Table 4. Ablation study on Noise-guided Latent Space Learning (NLSL) and Temporal Coherence Scheme (TCS) in our SeMo-AE.

In more details, we show the effectiveness of the proposed noise-guided latent space learning and the Temporal Coherence Scheme in Table 4. Note that in cases 3 and 4, we obtain μ and σ of the additive Gaussian noise in the latent space by the statistics of $\mathcal{E}(\widehat{\mathrm{IMU}}) - \mathcal{E}(\mathrm{IMU})$ over the entire dataset. The results demonstrate that both of the two proposed techniques are necessary to achieve the best performance.



(a) SeMo-AE (b) SeMo-AE w/o NLSL (c) SeMo-AE w/o TCS (d) cGAN

Figure 5. Qualitative results of ablation study. For each case, we visualize two data sequences from DS^{sim}_{Loose} ("loose" in the figure) and DS^{sim}_{Tight} ("tight" in the figure), respectively, along with three random samples of DS^{gen}_{Loose} ("gen_loose" in the figure), using PCA. NLSL: Noise-guided Latent Space Learning; TCS: Temporal Coherence Scheme.

Qualitative Results. As Fig. 5 shows, i) removing the proposed noise-guided latent space learning (NLSL) results in abnormal spatial patterns in the synthesized loose-wear IMU data DS^{gen}_{Loose}; ii) removing the proposed Temporal Coherence Scheme (TCS) transforms DS^{gen}_{Loose} into a noise-like signal, significantly different from DS^{sim}_{Loose}; iii) the DS^{gen}_{Loose} generated by cGAN exhibits minimal diversity, indicating that it ignores the input noise and fails to model the effects of secondary motion.

6.5. Live Demo

We have implemented a real-time pose estimation visualization system using Python and Unity. Once the user has dresses up the clothes as common, a single T-Pose calibration (5 seconds) is performed to initiate the system. During operation, the system continuously receives real-time IMU data, processes it using calibration and normalization methods similar to TransPose [54], feeds it into the pose estimation model, and displays the real-time motion capture results.

Through the live demo, we can observe that our system seamlessly combines the convenience and comfort of loose clothes with accurate motion capture capabilities. It maintains stability even during vigorous activities such as fast running and jumping, demonstrating its reliability and suitability for a wide range of applications. Please see our supplementary video for demonstration.

7. Limitations and Future Work

Limitations. *On the hardware side.* Due to cost constraints, we have currently produced only one garment. Although we have conducted comparisons with different users, the impact of clothing sizes and fabric materials on accuracy remains to be explored. Beside, the circuits integrated into the jacket and the sensors have not been waterproofed, rendering the jacket unable to undergo regular washing like conventional clothing. Additionally, as our jacket is worn loosely, the twisting of the arms causes minimal changes in IMU readings, making it difficult to accurately measure rotation in this degree of freedom. *On the software side.* While SeMo-AE significantly enhances the richness of loose-wear IMU data, due to the limited wearing styles simulation in it's training data, the generated data can not comprehensively cover different ways of wearing. Consequently, this limitation results in increased pose estimation errors, particularly evident in the unzipped wearing.

Future Work. Our work explores the use of loose clothes as the vehicle for the purpose of motion capture. This alignment with ordinary clothes in people's daily life prioritizes user comfort over other factors such as accuracy. It is worth exploring collecting a significantly large database for continuous (24x7) human motion on a large scale (expected over 100 participants). More efforts are needed to investigate the human motion pattern considering different variations in terms of body sizes and garment types. This could enable future applications such as long-term health monitoring on the population level.

8. Conclusion

In this paper, we proposed Loose Inertial Poser (LIP), a novel network to achieve real-time, accurate pose estimation using only sparse inertial sensors on loose clothing. Our key innovation is the Secondary Motion AutoEncoder (SeMo-AE), which can synthesize realistic IMU data exhibiting diverse secondary motion effects from limited simulation data. SeMo-AE employs two novel techniques: i) noise-guided latent space learning that models secondary motion as additive Gaussian noise in the latent space, and ii) temporal coherence modeling that captures the smoothness of secondary motion over time. We show SeMo-AE's effectiveness by training a pose estimation network on SeMo-AE's simulated IMU data. Extensive experiments show that our method adapts effectively across varying body shapes and motions, significantly outperforming state-of-the-art with under 20 degrees of mean joint rotation error.

References

[1] Yiming Bao, Xu Zhao, and Dahong Qian. Fusepose: Imu-vision sensor fusion in kinematic space for parametric human pose estimation. *IEEE Transactions on Multimedia*, 2022. 3

[2] Anargyros Chatzitofis, Dimitrios Zarpalas, Petros Daras, and Stefanos Kollias. Democap: Low-cost marker-based motion capture. *International Journal of Computer Vision*, 129(12): 3338–3366, 2021. 3

[3] Xiaowei Chen, Xiao Jiang, Jiawei Fang, Shihui Guo, Juncong Lin, Minghong Liao, Guoliang Luo, and Hongbo Fu. Dispad: Flexible on-body displacement of fabric sensors for robust joint-motion tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7 (1):1–27, 2023. 3

[4] Yuming Du, Robin Kips, Albert Pumarola, Sebastian Starke, Ali Thabet, and Artsiom Sanakoyeu. Avatars grow legs: Generating smooth human motion from sparse tracking inputs with diffusion model. In *CVPR*, 2023. 3

[5] Patric Eichelberger, Matteo Ferraro, Ursina Minder, Trevor Denton, Angela Blasimann, Fabian Krause, and Heiner Baur. Analysis of accuracy in optical motion capture–a protocol for laboratory setup evaluation. *Journal of biomechanics*, 49 (10):2085–2088, 2016. 2

[6] Maryam Adnan Fadhil and Waleed F Shareef. Human activity tracking using wearable sensors in loose-fitting clothes: survey. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 21(6):1382–1390, 2023. 3

[7] Andrew Gilbert, Matthew Trumble, Charles Malleson, Adrian Hilton, and John Collomosse. Fusing visual and inertial sensors with semantics for 3d human pose estimation. *International Journal of Computer Vision*, 127:381–397, 2019. 3

[8] Sam Gleadhill, Daniel James, and James Lee. Validating temporal motion kinematics from clothing attached inertial sensors. In *Proceedings*, page 304. MDPI, 2018. 3

[9] Gutemberg Guerra-Filho. Optical motion capture: Theory and implementation. *RITA*, 12(2):61–90, 2005. 2

[10] Lorna Herda, Pascal Fua, Ralf Plankers, Ronan Boulic, and Daniel Thalmann. Skeleton-based motion capture for robust reconstruction of human motion. In *Proceedings Computer Animation 2000*, pages 77–83. IEEE, 2000. 2

[11] S Zohreh Homayounfar and Trisha L Andrew. Wearable sensors for monitoring human motion: a review on mechanisms, materials, and challenges. *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, 25(1):9–24, 2020. 2

[12] Sufeng Hu, Miaoding Dai, Tianyun Dong, and Tao Liu. A textile sensor for long durations of human motion capture. *Sensors*, 19(10):2369, 2019. 3

[13] Yinghao Huang, Manuel Kaufmann, Emre Aksan, Michael J Black, Otmar Hilliges, and Gerard Pons-Moll. Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics (TOG)*, 37(6):1–15, 2018. 1, 2, 3, 6

[14] Udeni Jayasinghe, William S Harwin, and Faustina Hwang. Comparing clothing-mounted sensors with wearable sensors for movement analysis and activity classification. *Sensors*, 20(1):82, 2019. 3

[15] Jiaxi Jiang, Paul Streli, Huajian Qiu, Andreas Fender, Larissa Laich, Patrick Snape, and Christian Holz. Avatarposer: Articulated full-body pose tracking from sparse motion sensing. In *Proceedings of European Conference on Computer Vision*. Springer, 2022. 3

[16] Yifeng Jiang, Yuting Ye, Deepak Gopinath, Jungdam Won, Alexander W Winkler, and C Karen Liu. Transformer inertial poser: Real-time human motion reconstruction from sparse imus with simultaneous terrain generation. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022. 1, 2, 3, 6

[17] Wadim Kehl, Fausto Milletari, Federico Tombari, Slobodan Ilic, and Nassir Navab. Deep learning of local rgb-d patches for 3d object detection and 6d pose estimation. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 205–220. Springer, 2016. 2

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[19] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 5

[20] Muhammed Kocabas, Chun-Hao P Huang, Otmar Hilliges, and Michael J Black. Pare: Part attention regressor for 3d human body estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11127–11137, 2021. 2

[21] Manon Kok, Jeroen D Hol, and Thomas B Schön. An optimization-based approach to human body motion capture using inertial sensors. *IFAC Proceedings Volumes*, 47(3):79–85, 2014. 2

[22] Kazutaka Kurihara, Shin'ichiro Hoshino, Katsu Yamane, and Yoshihiko Nakamura. Optical motion capture system with pan-tilt camera tracking and real time data processing. In *Proceedings 2002 IEEE international conference on robotics and automation (Cat. No. 02CH37292)*, pages 1241–1248. IEEE, 2002. 2

[23] Wenhao Li, Hong Liu, Hao Tang, Pichao Wang, and Luc Van Gool. Mhformer: Multi-hypothesis transformer for 3d human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13147–13156, 2022. 2

[24] Zhen Liang, Dongquan Zhang, Guanghua Xu, Fangting Xie, Hao Guo, Jingyuan Cheng, et al. Smart garment: A long-term feasible, whole-body textile pressure sensing system. *IEEE Sensors Journal*, 2023. 3

[25] Qi Lin, Shuhua Peng, Yuezhong Wu, Jun Liu, Hong Jia, Wen Hu, Mahbub Hassan, Aruna Seneviratne, and Chun H Wang. Subject-adaptive loose-fitting smart garment platform for human activity recognition. *ACM Transactions on Sensor Networks*, 19(4):1–23, 2023. 3

[26] Ruibo Liu, Qijia Shao, Siqi Wang, Christina Ru, Devin Balkcom, and Xia Zhou. Reconstructing human joint motion with computational fabrics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(1): 1–26, 2019. 3

[27] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation net-

works. In *International conference on machine learning*, pages 97–105. PMLR, 2015. 5

[28] Michael Lorenz, Gabriele Bleser, Takayuki Akiyama, Takehiro Niikura, Didier Stricker, and Bertram Taetz. Towards artefact aware human motion capture using inertial sensors integrated into loose clothing. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 1682–1688. IEEE, 2022. 3

[29] Timothy McGrath and Leia Stirling. Body-worn imu human skeletal pose estimation using a factor graph-based optimization framework. *Sensors*, 20(23):6887, 2020. 3

[30] Matteo Menolotto, Dimitrios-Sokratis Komaris, Salvatore Tedesco, Brendan O'Flynn, and Michael Walsh. Motion capture technology in industrial applications: A systematic review. *Sensors*, 20(19):5687, 2020. 2

[31] Brendan Michael and Matthew Howard. Activity recognition with wearable sensors on loose clothing. *Plos one*, 12(10): e0184642, 2017. 3

[32] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 7

[33] Vimal Mollyn, Riku Arakawa, Mayank Goel, Chris Harrison, and Karan Ahuja. 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*, pages 1–12, 2023. 1, 3, 6

[34] Md Moniruzzaman, Zhaozheng Yin, Md Sanzid Bin Hossain, Hwan Choi, and Zhishan Guo. Wearable motion capture: Reconstructing and predicting 3d human poses from wearable sensors. *IEEE Journal of Biomedical and Health Informatics*, 2023. 3

[35] L Noitom. Perception neuron, 2017. 2, 6

[36] Shaohua Pan, Qi Ma, Xinyu Yi, Weifeng Hu, Xiong Wang, Xingkang Zhou, Jijunnan Li, and Feng Xu. Fusing monocular images and sparse imu signals for real-time human motion capture. *arXiv preprint arXiv:2309.00310*, 2023. 3

[37] Chaitanya Patel, Zhouyingcheng Liao, and Gerard Pons-Moll. Tailornet: Predicting clothing in 3d as a function of human pose, shape and garment style. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7365–7375, 2020. 2, 5

[38] Monique Paulich, Martin Schepers, Nina Rudigkeit, and Giovanni Bellusci. Xsens mtw awinda: Miniature wireless inertial-magnetic motion tracker for highly accurate 3d kinematic applications. *Xsens: Enschede, The Netherlands*, pages 1–9, 2018. 2

[39] Natural Point. Optitrack. *Natural Point, Inc*, 2011. 2

[40] Jose Luis Ponton, Haoran Yun, Andreas Aristidou, Carlos Andujar, and Nuria Pelechano. Sparseposer: Real-time full-body motion reconstruction from sparse data. *ACM Transactions on Graphics*, 2023. 3

[41] M Rahul. Review on motion capture technology. *Global journal of computer science and technology*, 18(1):22–26, 2018. 2

[42] Amanda Jean Redhouse. *Joint Angle Estimation Method for Wearable Human Motion Capture*. PhD thesis, Virginia Tech, 2021. 3

[43] Tianchen Shen, Irene Di Giulio, and Matthew Howard. A probabilistic model of human activity recognition with loose clothing. *Sensors*, 23(10):4669, 2023. 3

[44] Tesla Suit. Teslasuit developer kit, 2020. 3

[45] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5693–5703, 2019. 2

[46] Timo Von Marcard, Bodo Rosenhahn, Michael J Black, and Gerard Pons-Moll. Sparse inertial poser: Automatic 3d human pose estimation from sparse imus. In *Computer graphics forum*, pages 349–360. Wiley Online Library, 2017. 2

[47] Timo Von Marcard, Roberto Henschel, Michael J Black, Bodo Rosenhahn, and Gerard Pons-Moll. Recovering accurate 3d human pose in the wild using imus and a moving camera. In *Proceedings of the European conference on computer vision (ECCV)*, pages 601–617, 2018. 3

[48] Alexander Winkler, Jungdam Won, and Yuting Ye. Questsim: Human motion tracking from sparse sensors with simulated avatars. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–8, 2022. 3

[49] Size Wu, Sheng Jin, Wentao Liu, Lei Bai, Chen Qian, Dong Liu, and Wanli Ouyang. Graph-based 3d multi-person pose estimation using multi-view images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11148–11157, 2021. 2

[50] Xuesu Xiao and Shuayb Zarar. Machine learning for placement-insensitive inertial motion capture. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6716–6721. IEEE, 2018. 3

[51] Yufei Xu, Jing Zhang, Qiming Zhang, and Dacheng Tao. Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in Neural Information Processing Systems*, 35:38571–38584, 2022. 2

[52] Sen Yang, Zhibin Quan, Mu Nie, and Wankou Yang. Transpose: Keypoint localization via transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11802–11812, 2021. 2

[53] Hang Ye, Wentao Zhu, Chunyu Wang, Rujie Wu, and Yizhou Wang. Faster voxelpose: Real-time 3d human pose estimation by orthographic projection. In *European Conference on Computer Vision*, pages 142–159. Springer, 2022. 2

[54] Xinyu Yi, Yuxiao Zhou, and Feng Xu. Transpose: Real-time 3d human translation and pose estimation with six inertial sensors. *ACM Transactions on Graphics (TOG)*, 40(4):1–13, 2021. 1, 2, 3, 5, 6, 8

[55] Xinyu Yi, Yuxiao Zhou, Marc Habermann, Soshi Shimada, Vladislav Golyanik, Christian Theobalt, and Feng Xu. 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*, pages 13167–13178, 2022. 1, 2, 3, 6

[56] Tao Yu, Zerong Zheng, Kaiwen Guo, Pengpeng Liu, Qionghai Dai, and Yebin Liu. Function4d: Real-time human volumetric capture from very sparse consumer rgbd sensors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5746–5756, 2021. 2

[57] Jianfeng Zhang, Yujun Cai, Shuicheng Yan, Jiashi Feng, et al. Direct multi-view multi-person 3d pose estimation. *Advances in Neural Information Processing Systems*, 34: 13153–13164, 2021. 2

[58] Qitao Zhao, Ce Zheng, Mengyuan Liu, and Chen Chen. A single 2d pose with context is worth hundreds for 3d human pose estimation. *arXiv preprint arXiv:2311.03312*, 2023. 2

[59] Jianan Zhen, Qi Fang, Jiaming Sun, Wentao Liu, Wei Jiang, Hujun Bao, and Xiaowei Zhou. Smap: Single-shot multi-person absolute 3d pose estimation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*, pages 550–566. Springer, 2020. 2

[60] Zerong Zheng, Tao Yu, Hao Li, Kaiwen Guo, Qionghai Dai, Lu Fang, and Yebin Liu. Hybridfusion: Real-time performance capture using a single depth sensor and sparse imus. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 384–400, 2018. 3

[61] Bo Zhou, Daniel Geissler, Marc Faulhaber, Clara Elisabeth Gleiss, Esther Friederike Zahn, Lala Shakti Swarup Ray, David Gamarra, Vitor Fortes Rey, Sungho Suh, Sizhen Bian, et al. Mocapose: Motion capturing with textile-integrated capacitive sensors in loose-fitting smart garments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(1):1–40, 2023. 3

[62] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5745–5753, 2019. 6

[63] Christian Zimmermann, Tim Welschehold, Christian Dornhege, Wolfram Burgard, and Thomas Brox. 3d human pose estimation in rgbd images for robotic task learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1986–1992. IEEE, 2018. 2

[64] Chengxu Zuo, Jiawei Fang, Shihui Guo, and Yipeng Qin. Self-adaptive motion tracking against on-body displacement of flexible sensors. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 3