

Probabilistic Inertial Poser (ProbIP): Uncertainty-aware Human Motion Modeling from Sparse Inertial Sensors

Supplementary Material

6. Implementation details

We adopt **Mamba** as the backbone architecture. The Mamba blocks are connected via residual connections and employ RMS normalization. Throughout, we denote N as the number of sensors (ranging from 6 to 2), and x_t as the linearly projected sensor signal at time t . Given the input x_t and the previous hidden state h_{t-1} , the Mamba block produces a hidden representation h_t , which is subsequently used to compute the output latent vector v_t .

Each Mamba block's transition matrix A is initialized using the S4D method. We use a model dimension of 128, an expansion factor of 2, and a state dimension of 8 for the Mamba layers. During training, the kernel K within each Mamba layer is computed using a parallel scan method, which evaluates all values in $2 \log_2(T)$ steps for a sequence of length T . Except for the first RU-Mamba block (which uses the lowest scaling factor), sampled poses from the previous RU-Mamba block are concatenated with x_t at each layer.

The latent vector v_t is passed through two separate linear blocks to generate R_t , representing the mean rotation, and u_t , representing the uncertainty, for each joint in the reduced joint set. Following DynaIP[34], we predict a reduced set of joints (N_J) from the SMPL model and reconstruct the full-body pose accordingly. For each joint, we predict both R_t and u_t , which together define the distribution parameter F , from which we sample poses as input for the subsequent block. To predict joint velocity, we utilize the final posterior obtained by concatenating the predicted rotation and uncertainty values before scaling, rather than sampled poses for stability. The model is trained for 300 epochs using the AdamW optimizer with cosine annealing. The learning rate is scheduled between a maximum of 1×10^{-3} and a minimum of 1×10^{-6} .

In scenarios with fewer sensors, the model is further tasked with predicting joint rotations for regions without direct sensor attachments, adapting to the current configuration. For comparison, models such as DynaIP [34] and TransPose [32] are configured similarly, taking global orientation and acceleration as inputs, given that the pelvis IMU is omitted in the fewer-sensor setup. For DynaIP, we add an additional subposer to predict joints with missing sensors. As the number of sensors decreases, some subposers lack direct information based on the original model structure. In these cases, we provide all available sensor data as inputs to compensate for missing information.

7. Exponentially Scaled Normalizing Constant

Given the probability density function as:

$$p(R|F) = \frac{1}{c(F)} \exp(\text{tr}[F^T R]) \quad (13)$$

The negative log-likelihood (NLL) of the function with scaling factor s_α , given that $U, S, V^T = \text{SVD}(F)$, is calculated as follows:

$$L_{NLL}^{s_\alpha} = -\log(p(R|F)) = -\text{tr}[F^T R] + \left(1 + \frac{1}{3 * \exp(s_\alpha)}\right) * (\text{tr}[S] + \bar{c}(F)) \quad (14)$$

The exponentially scaled normalizing constant is calculated using techniques proposed in [15]. Given that $S = [s_i, s_j, s_k]$, and $B_0 = \frac{1}{2}(s_i - s_j)(1 - x)$, $B_1 = \frac{1}{2}(s_i + s_j)(1 + x)$, and $E = \exp((\min\{s_i, s_j\} + s_k)(x - 1))$

$$\bar{c}(S) = \int_{-1}^1 \frac{1}{2} \bar{I}_0[B_0] * \bar{I}_0[B_1] * E dx \quad (15)$$

We note that $D_{\bar{I}}(x) = \frac{d\bar{I}_0(x)}{dx} = \bar{I}_1(x) - \text{sgn}[x]\bar{I}_0(x)$, where \bar{I} indicates the exponentially scaled modified Bessel functions of the first kind [1]. Therefore, the first order derivative for gradient descent of $\bar{a}(S)$ is evaluated by

$$\begin{aligned} \frac{d\bar{c}(S)}{ds_i} = & \int_{-1}^1 \frac{1}{4} (1 - x) D_{\bar{I}}(B_0) \bar{I}_0[B_1] E \\ & + \frac{1}{4} (1 + x) D_{\bar{I}}(B_1) \bar{I}_0[B_0] E \\ & + \frac{1}{2} (x - 1) \bar{I}_0[B_0] \bar{I}_0[B_1] E dx \end{aligned} \quad (16)$$

$$\begin{aligned} \frac{d\bar{c}(S)}{ds_j} = & \int_{-1}^1 \frac{1}{4} (1 + x) D_{\bar{I}}(B_1) \bar{I}_0[B_0] E \\ & - \frac{1}{4} (1 - x) D_{\bar{I}}(B_0) \bar{I}_0[B_1] E \\ & + \frac{1}{2} (x - 1) \bar{I}_0[B_0] \bar{I}_0[B_1] E dx \end{aligned} \quad (17)$$

$$\frac{d\bar{c}(S)}{ds_k} = \int_{-1}^1 \frac{1}{2} (x - 1) \bar{I}_0[B_0] \bar{I}_0[B_1] E dx \quad (18)$$

8. Rejection Sampler

We incorporate a rejection sampling technique based on the method proposed by [13, 25] to generate samples from predicted distribution parameter F . However, because sampling must be conducted for each motion sequence during training, we modified previous methods to perform batch-wise sampling with relaxed acceptance criteria, allowing each rotation to be sampled iteratively. This adaptation enhances efficiency, enabling faster training sessions.

Algorithm 1 Differentiable Rejection Sampler (Fast Batch Version)

Input: $\{F_i\} \in \mathbb{R}^{3 \times 3}, i = 1, 2, \dots, N$, where N is number of independent probability parameters

Output: $\{\hat{R}_i\} \in SO(3)$ such that $\hat{R}_i \sim \mathcal{M}(F_i)$

- 1: let $b = 1.5, M = \exp\left(\frac{b-4}{2}\right) \left(\frac{4}{b}\right)^2$
- 2: $U_i, [S_{i1}, S_{i2}, S_{i3}], V_i = SVD(F_i)$
- 3: $\bar{U}_i = U_i * \text{diag}(1, 1, \det(U_i))$
- 4: $\bar{V}_i = V_i * \text{diag}(1, 1, \det(V_i))$
- 5: $\bar{S}_i = [S_{i1}, S_{i2}, S_{i3} \cdot \det(U_i) \cdot \det(V_i)]$
- 6: $\mathcal{A}_i = \text{diag}(0, 2(\bar{S}_{i2} + \bar{S}_{i3}), 2(\bar{S}_{i1} + \bar{S}_{i3}), 2(\bar{S}_{i1} + \bar{S}_{i2}))$
- 7: $\Omega_i = I_4 + \frac{2}{b}\mathcal{A}_i$
- 8: Sample $e_i \sim \mathcal{N}(0, I_4), i = 1, 2, \dots, N$
- 9: $y_i = (\Omega_i^{-1})^{\frac{1}{2}} e_i$
- 10: Propose $x_i = y_i / \|y_i\|$ such that $x_i \in S^3$
- 11: Sample $w_i \sim U[0, 1], i = 1, 2, \dots, N$
- 12: **if** $w_i < \frac{\exp(-x_i^T \mathcal{A}_i x_i)}{M(x_i^T \Omega_i x_i)^{-2}}$ **then**
- 13: $\hat{Q}_i = \text{quat_to_matrix}(x_i)$ such that $\hat{Q}_i \in SO(3)$
- 14: **else**
- 15: $\hat{Q}_i = \text{mode}(U_i \text{diag}(1, 1, \det(U_i V_i)) V_i^T)$
- 16: **end if**
- 17: **return** $\{\hat{R}_i\} = U_i \hat{Q}_i V_i^T$

This batch-based differentiable rejection sampler builds upon prior sampling algorithms by introducing a more efficient, parallelized approach for handling multiple input samples simultaneously. Unlike traditional rejection sampling methods, which process one sample at a time and repeat the sampling if conditions are not met, this algorithm operates on entire batches of samples, allowing for faster and more scalable training. The acceptance condition is modified to improve training efficiency by applying it across the batch in a single pass. If the acceptance criterion fails for any individual sample, instead of resampling, the algorithm directly returns the distribution mode for that sample, reducing computational overhead. Additionally, batch-wise operations such as quaternion to matrix transformations, matrix multiplications, and normalizations are performed collectively, streamlining the processing pipeline. This approach maintains accuracy while significantly reducing the time required for sampling, making

sensor configurations	GA(^o)	LA(^o)	legsLA Err(^o)	backLA(^o)
DiffusionPoser (h, p, w_l, w_r, l_l, l_r)	14.4	13.0	14.0	10.9
ProbIP (h, p, w_l, w_r, l_l, l_r)	10.6	11.6	10.5	10.8
DiffusionPoser (h, w_l, w_r, l_l, l_r)	17.7	-	17.2	12.4
ProbIP (h, w_l, w_r, l_l, l_r)	11.7	12.7	12.0	11.6
DiffusionPoser (w_l, w_r, l_l, l_r)	21.6	-	20.4	15.1
ProbIP (w_l, w_r, l_l, l_r)	13.7	13.6	11.9	13.6
DiffusionPoser (h, p, w_l, w_r)	18.0	-	19.2	11.4
ProbIP (h, p, w_l, w_r)	11.5	12.2	12.2	11.2
DiffusionPoser (h, w_l, w_r)	24.0	-	21.0	14.7
ProbIP (h, w_l, w_r)	13.6	13.3	13.6	11.9
DiffusionPoser (plv, w_l, w_r)	22.5	-	22.9	13.1
ProbIP (plv, w_l, w_r)	13.3	13.0	13.3	12.7
DiffusionPoser (w_l, w_r)	27.5	-	21.7	16.9
ProbIP (w_l, w_r)	16.1	13.8	14.5	12.7

Table 6. Performance comparison under various sparse IMU configurations on the TotalCapture Real dataset. GA: Global Angular Error; LA: Local Angular Error. Sensor abbreviations: h (head), p (pelvis), w_l/w_r (left/right wrist), l_l/l_r (left/right leg).

it well-suited for large-scale or high-dimensional applications where sampling efficiency is crucial.

9. Additional Performance Comparison

We further analyze the inference time and jitter of our model in comparison to state-of-the-art (SOTA) models. The Tab. 7 presents inference time comparisons under the condition of purely iterative inference on incoming signals, excluding bottlenecks caused by Bluetooth communication with sensors and rendering processes.

Due to the incorporation of singular value decomposition and distribution modeling, our model exhibits relatively higher inference time and jitter compared to PIP and DynaIP. However, it achieves lower jitter and faster inference time compared to the Transpose model. Importantly, our model still maintains strong performance, supporting real-time visualization at 30Hz. Further optimization, such as simplifying the projection and transformation of distributions within a single time frame, would be a valuable direction for future work.

10. Additional Performance Comparison on sparse IMU configuration with DiffusionPoser

As shown in Tab. 6, we perform additional comparisons with DiffusionPoser[30] under both our proposed sensor configurations and the optimal configurations reported in DiffusionPoser to ensure a comprehensive evaluation. Some local error metrics are omitted due to the unavailability of the original model implementation. Our model consistently outperforms DiffusionPoser across all configurations.

Metric	Transpose	PIP	DynaIP	ProbIP (ours)
Jitter (DIP)	14.6	2.4	5.1	10.7
Inference Time (DIP)	0.0197	0.0104	0.0112	0.0171

Table 7. Performance comparison of the proposed model and state-of-the-art (SOTA) methods in terms of jitter ($10^2 m/s^3$) and inference time (s).

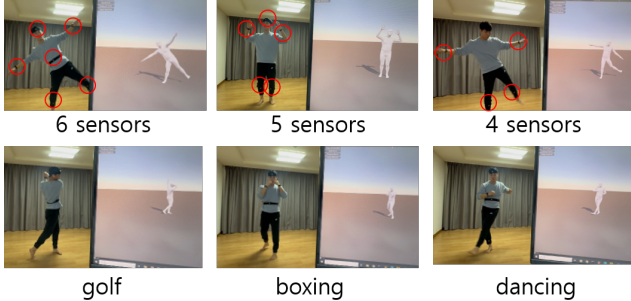


Figure 7. Real-time human motion inference using ProbIP with 6, 5, and 4 IMU sensors

11. Additional Analysis

In this section, we present additional experimental results for our proposed model, ProbIP, which generates human motion from varying numbers of sparse IMU sensors in real time while effectively accounting for motion uncertainty, as shown in Fig. 7. Additionally, the supplementary videos demonstrate various motions, including golf swings, boxing, and dancing, which involve rapid and simultaneous movements across multiple joints.

Additional Study on the Effect of Progressive Distribution Narrowing. We further analyze how layers with different scaling factors contribute to the model’s output by visualizing changes in average uncertainty values for different body parts across a sequence of motions. Fig. 8 illustrates the average singular values for the lower body, upper body, and main body in layers with varying scaling factors. For motions primarily involving upper body movement with stable lower body positioning, the average singular values for the lower and main body remain high in the final layer (the layer with the highest scaling factor), while the upper body singular values stay low to capture diverse plausible hand motions. In contrast, for full-body movements, such as jumping, where all body parts move synchronously, the average singular values fluctuate across layers.

A notable observation is the inverse relationship in uncertainty between the early and later layers. In the later layers, higher singular values are associated with narrow lower body motions, indicating increased certainty. In contrast, in the early layers, lower singular values allow for a wider range of motion. This suggests that the model modulates

uncertainty across layers, adaptively controlling motion dynamics in response to different scaling factors.

Additional Ablation Study on the Non-Deterministic Approach with Varying Sensor Counts. We compare the performance of the model with non-deterministic and deterministic outputs across multiple datasets, as shown in Tab. 8, using different numbers of sensors. The deterministic model shares the same structure as ProbIP but outputs a rotation matrix for each target joint instead of probabilistic distribution parameters F . The results demonstrate that ProbIP achieves the best overall performance across most test sets, regardless of sensor count. With fewer sensors attached to the body, our model maintains stable performance in predicting human motions, highlighting its robustness in sparse sensor configurations.

12. Additional Qualitative Results

Additional Qualitative Comparison of ProbIP Outputs with Different Numbers of Sensors. In Fig. 9, we compare the model’s performance with sensor counts ranging from 2 to 6 (from left to right). By modeling motion uncertainty, ProbIP achieves strong performance with various sensor configurations, providing optimal results with six sensors and comparable performance with only four or five sensors.

One notable observation is the effect of including a head sensor. In the four-sensor setup (with sensors on the wrists and legs), body bending motions are occasionally not well captured. In contrast, in the three-sensor setup (with sensors on the wrists and head), body bending motions are more accurately represented. In the two-sensor configuration (with sensors only on the wrists), general motion is recognized well, but detailed actions, such as body twisting and bending, are less accurately reflected, leading to a tendency to default to more generalized motion.

Additional Qualitative Comparison of Network Structure. Fig. 10 presents a qualitative comparison between our model, ProbIP, and state-of-the-art models, including TransPose [32], PIP [33], and DynaIP [34]. This visualization showcases model predictions on the DIP-IMU test set, highlighting the performance of ProbIP relative to other models. Fig. 10 presents a real-time performance comparison against state-of-the-art (SOTA) models, along with representative frames of reconstructed full-body motion under extremely sparse IMU configurations. Our model effectively recovers a wide range of dynamic and complex human motions using only 6 or even 3, 2 IMUs, while maintaining low-latency inference suitable for real-time applications.

The proposed model, ProbIP, benefits from a non-deterministic approach and a broader range distribution, allowing it to capture more detailed motion than previous models. Due to uncertainty introduced by sparse sensor

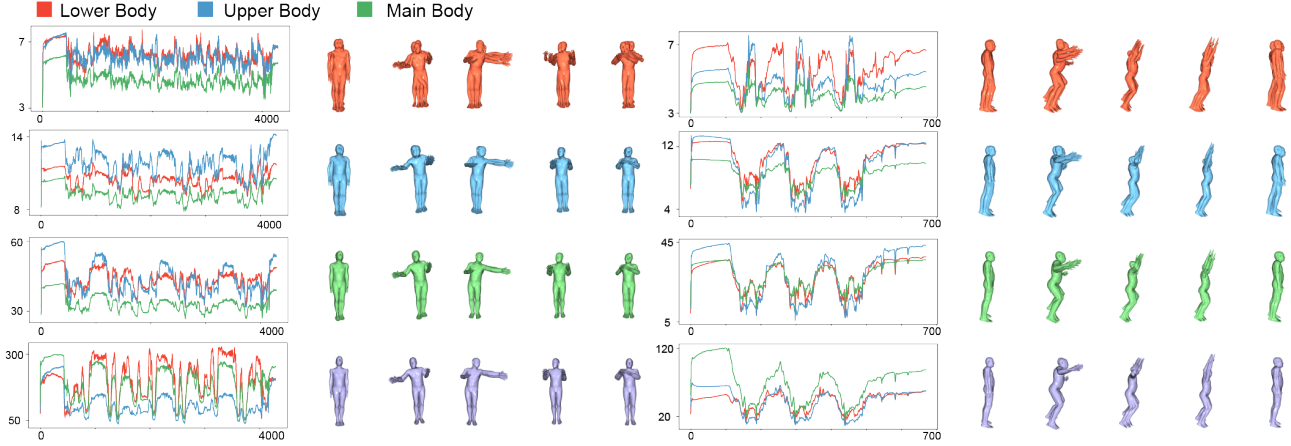


Figure 8. Qualitative results showcasing the average singular values for various body parts. From top to bottom, the figure presents the output of RU-Mamba block’s average uncertainty values, beginning with layer 1 (scaling factor 2) up to layer 4 (scaling factor 8), alongside corresponding sample poses. The red curve indicates singular values for lower body joints, including the upper leg joint. The blue curve represents upper body joints, including both shoulders and upper arms. Finally, the green curve illustrates the main body joints, including the spine.

	DIP IMU Test		CIP		Natural Motion		Andy		UNIPD	
	SIP Err(°)	Ang Err(°)	SIP Err(°)	Ang Err(°)	SIP Err(°)	Ang Err(°)	SIP Err(°)	Ang Err(°)	SIP Err(°)	Ang Err(°)
Probip (w/o prob) (sensor2)	18.25	13.82	18.77	13.84	39.50	21.20	15.09	12.44	13.39	8.53
Probip (sensor2)	17.71	12.49	18.62	13.15	31.28	17.98	15.84	11.83	11.51	6.92
Probip (w/o prob) (sensor3)	16.33	10.57	16.02	10.18	33.11	15.96	11.40	7.80	10.44	6.27
Probip (sensor3)	16.76	10.71	13.62	8.71	33.88	15.95	11.21	7.70	10.78	6.01
Probip (w/o prob) (sensor4)	14.56	8.42	13.83	7.28	32.55	11.59	10.30	5.57	9.99	4.80
Probip (sensor4)	13.84	8.18	13.19	7.04	31.07	11.42	10.70	5.50	9.17	4.40
Probip (w/o prob) (sensor5)	14.11	7.22	12.92	5.69	27.10	8.83	8.96	3.57	8.28	3.54
Probip (sensor5)	13.63	7.01	12.52	5.51	25.22	8.81	9.15	3.66	8.59	3.50
Probip (w/o prob) (sensor6)	14.37	6.74	13.98	5.35	26.15	8.78	9.61	3.55	9.98	3.63
Probip (sensor6)	13.65	6.57	11.65	4.56	21.62	7.54	8.45	3.00	8.05	2.97

Table 8. Additional ablation studies comparing deterministic (w/o prob) and non-deterministic approaches of the proposed ProbIP model across different numbers of sensors. The deterministic models maintain the same layer structure as the original ProbIP model but predict joint rotations directly, bypassing the distribution parameter prediction.

placement, current pose predictions influence later frame predictions. This means that when detailed motions are inaccurately predicted, they can lead to more significant errors in subsequent frames. However, the early layers in ProbIP are designed to accommodate a wide range of possible poses, enabling the model to recover even when initial motions are inaccurate, thereby generating refined and detailed motion sequences.

In contrast, models like DynaIP and PIP, which regulate rapid motion using velocity regressors and physical constraints, tend to produce more stable motions, especially in long-term, stationary datasets such as the Natural Motion dataset. This stabilization can be beneficial for sequences with minimal motion variations, like sitting poses, but may lack the flexibility offered by ProbIP for dynamic and detailed motion capture.

Additional Qualitative Results on Samples Generated from Different Layers.

This visualization in Fig. 11 provides a deeper qualitative analysis of sample poses generated across different layers in a single forward pass along the temporal axis. On the left side, ProbIP demonstrates a broad range of motion possibilities in the early layers, capturing a wide distribution of potential poses. As the body initiates movement and turns, the model generates a distribution that includes not only the current motion but also the forward and backward leg movements within the sample poses.

In later frames, while the model narrows the prediction space to focus on more specific motions, it still passes a broad range of possible motions along the temporal axis, allowing future frames to consider multiple plausible poses. On the right side, as the legs and arms begin to shake,

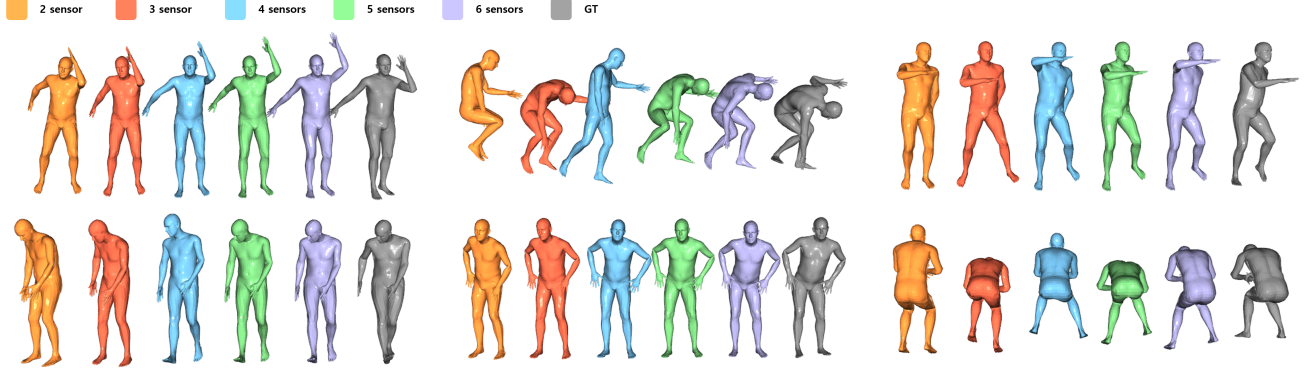


Figure 9. Additional qualitative results for different numbers of sensors with ProbIP

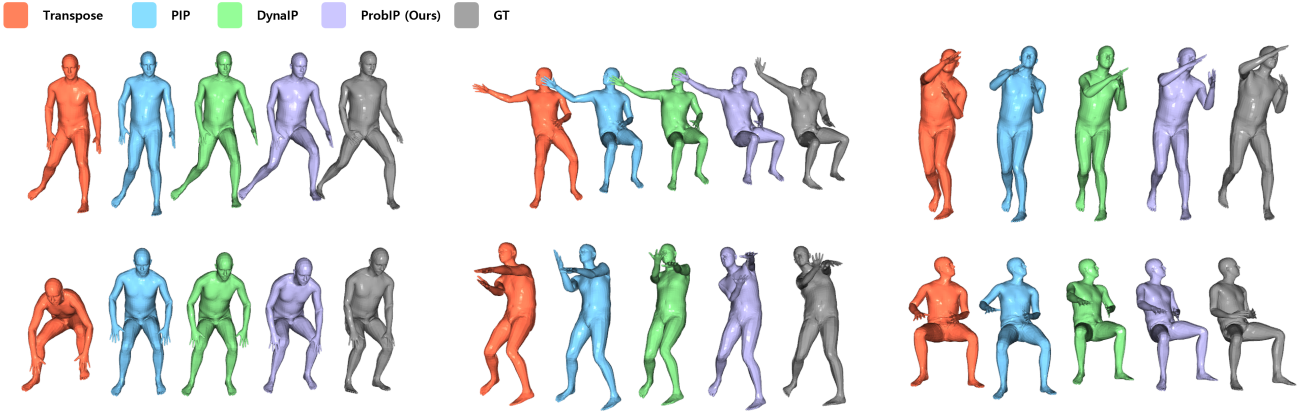


Figure 10. Additional qualitative results on DIP-IMU with previous SOTAs

ProbIP captures a wide range of motions at various scaling factors. Due to the sparse placement of IMU sensors and the inherent noise in acceleration measurements, rapid body movements often result in multiple plausible poses. ProbIP effectively captures this uncertainty, accommodating the variability in motion predictions across different layers.

13. Discussion and Future Works

To the best of our knowledge, this is the first approach to model motion in a wearable-based framework that incorporates multi-level distribution passing along the temporal axis. This innovative structure allows the model to naturally learn the uncertainties inherent to sparse sensor placements, capturing fine details of motion even when fewer sensors are attached, thus demonstrating robust performance in low-sensor setups.

However, unlike previous methods that employ post-processing or enforce physical constraints to stabilize motion, our model exhibits higher SIP errors, primarily in

long-term, stationary motions. Integrating these prior stabilization techniques, such as post-processing or physically constrained regulation, could further enhance our model’s stability. Thus, combining these strategies with our approach may yield a more robust framework capable of handling both dynamic and stable motion scenarios effectively.

It is worth noting that while our model demonstrates strong performance, the computational cost per time step is relatively higher than previous LSTM-based models. This increased cost is due to the singular value decomposition (SVD) and sampling required to generate the predicted distribution parameters during training. Future work could focus on optimizing these components to reduce the computation time and make the model more efficient. Additionally, we empirically found that an end scaling factor of 8 achieves better performance than other values. Currently, the intermediate scaling factors are linearly scaled from 2 to 8. Further optimization of both the scaling factors and the number of distribution narrowing steps could reduce computational complexity while preserving performance.

Currently, our uncertainty modeling is limited to rota-

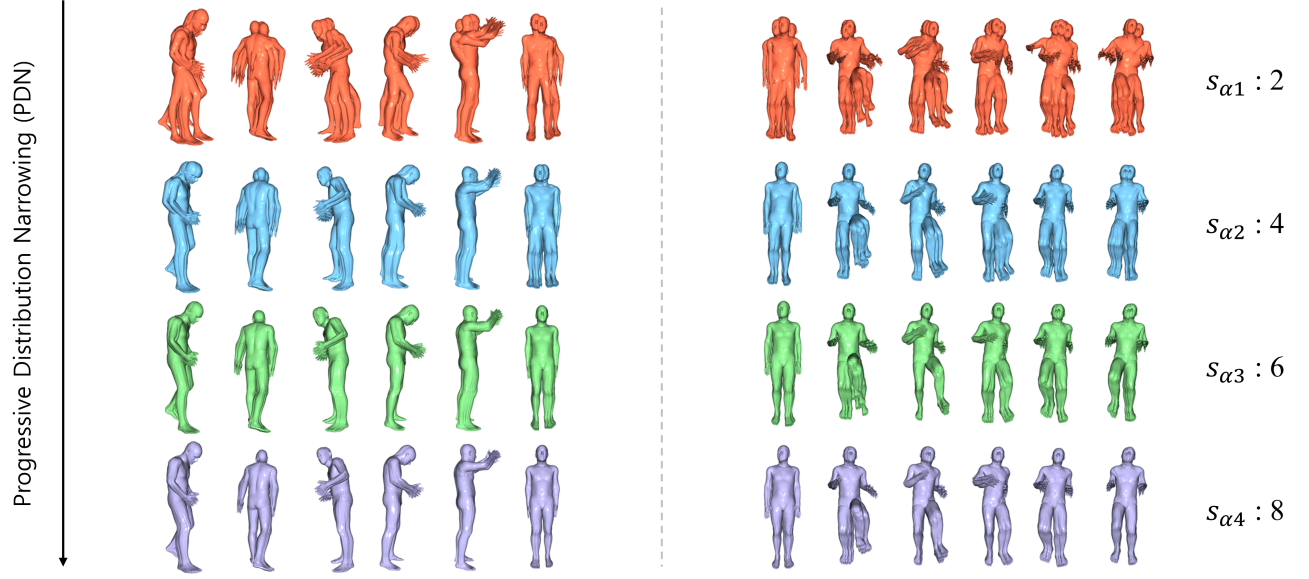


Figure 11. Qualitative results for distribution sample poses in different layers during Progressive Distribution Narrowing (PDN)

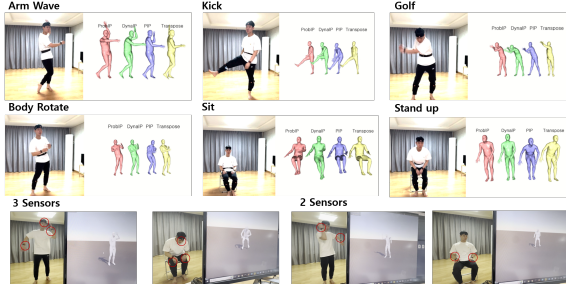


Figure 12. Real-time performance comparison with state-of-the-art (SOTA) models and representative reconstructed motion frames using extremely sparse IMU configurations (2 and 3 sensors).

tional components and does not directly influence translation prediction, leading to occasional discrepancies between predicted motion and velocity. A future direction could involve coupling probabilistic samples with their corresponding velocity estimates to achieve better coherence between motion dynamics and velocity prediction.

In summary, our model represents a significant step toward detailed motion modeling in wearable-based systems, particularly with sparse sensor data. Future enhancements, including integrating physical constraints, improved translation uncertainty modeling, and computational optimizations, would offer directions to refine and expand its applicability across diverse motion types and conditions.