

milliMamba: Specular-Aware Human Pose Estimation via Dual mmWave Radar with Multi-Frame Mamba Fusion

Supplementary Material

Niraj Prakash Kini[†], Shiao-Rung Tsai[†], Guan-Hsun Lin[†],
Wen-Hsiao Peng[†], Ching-Wen Ma[†], Jenq-Neng Hwang[‡]

[†]National Yang Ming Chiao Tung University, Taiwan, [‡]University of Washington, USA

{nirajnycu.ee06, abc900203abc.cs12, lin710277nycu.cs14, machingwen}@nycu.edu.tw,
wpeng@cs.nctu.edu.tw, hwang@uw.edu

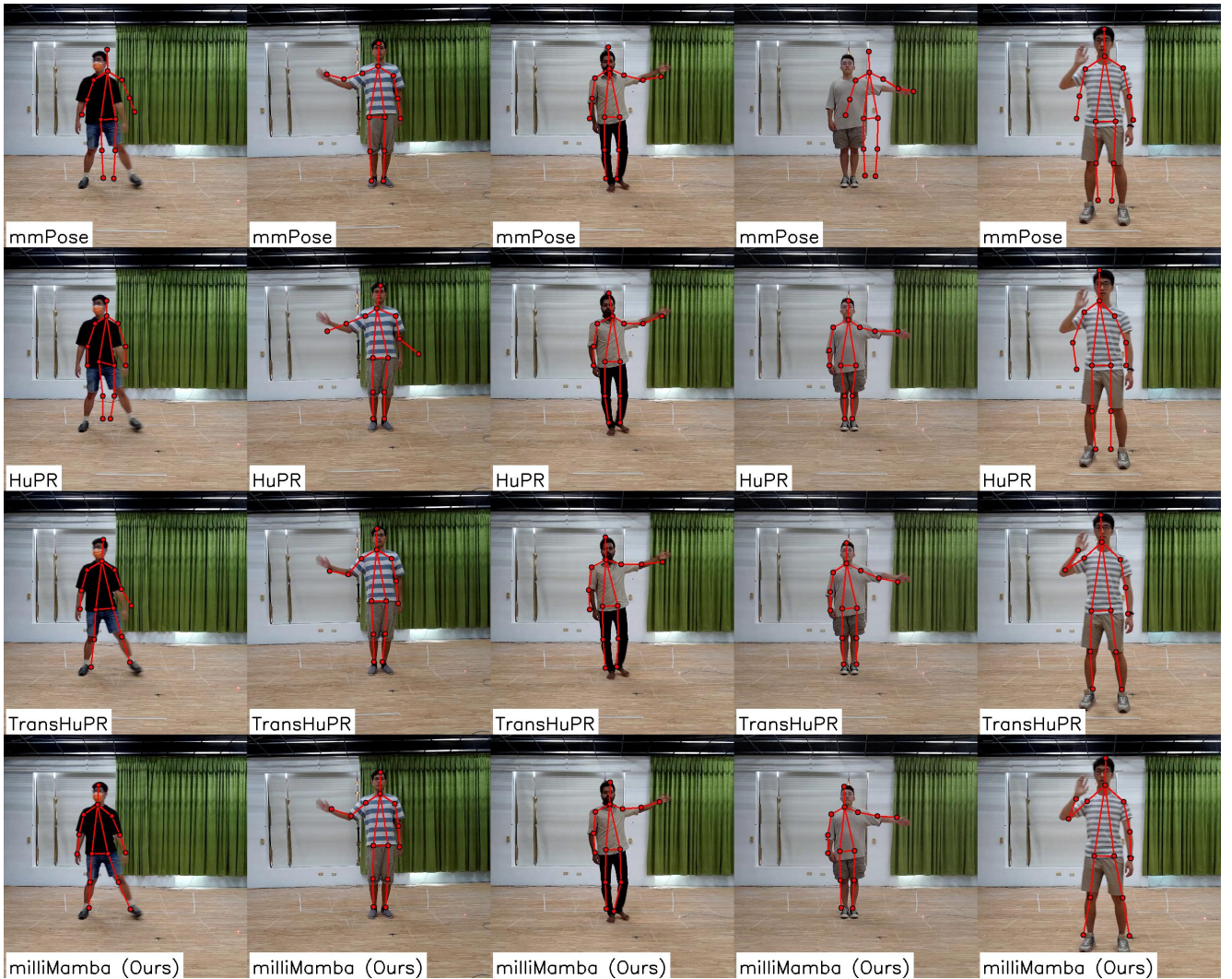


Figure 1. Qualitative results on the HuPR dataset, with each column corresponding to a different sequence.

This supplemental material provides additional qualitative and empirical results to support our submission. The submitted zip file contains this PDF and two folders: `TransHuPR_dataset_vis` and `HuPR_dataset_vis`, which include additional qualitative videos.

1. HuPR Dataset Visual Examples

Figure 1 illustrates qualitative results from selected sequences in the HuPR dataset.

2. TransHuPR_dataset_vis Folder

This folder contains qualitative videos of the same sequences shown in the static visualization of the TransHuPR dataset in the main paper. These videos highlight the temporal consistency of our predictions more clearly than static images alone.

3. HuPR_dataset_vis Folder

This folder contains qualitative videos corresponding to the sequences shown in Supplementary Figure 1, as well as additional sequences not included in the figure. These videos provide a dynamic view of various HuPR sequences, better illustrating the temporal consistency and robustness of our model’s pose predictions.

4. Why We Use the Middle Frame During Inference

Although our model predicts T poses within each sliding window during training, we retain only the center frame during inference. This is because poses near the temporal boundaries have access to only partial context: for example, the first frame sees only future inputs, while the last frame sees only past ones.

We empirically validate this by comparing prediction accuracy at different frame indices within the window. As shown in Figure 2, the accuracy peaks at the center and drops significantly at the edges, supporting our design choice.

5. Different Zigzag scanning orders

The Table 1 presents an experimental results examining how different Mamba scanning orders, i.e. the sequence in which the *frame*, *range*, and *angle* dimensions are processed, affect performance. Among all variants, the *range* \rightarrow *angle* \rightarrow *frame* pattern achieves the highest AP, AP⁵⁰, and AP⁷⁵ scores, indicating that this scanning direction best captures the underlying spatio-temporal structure. Other scanning orders yield slightly lower performance, confirming that the choice of scan pattern notably influences accuracy.

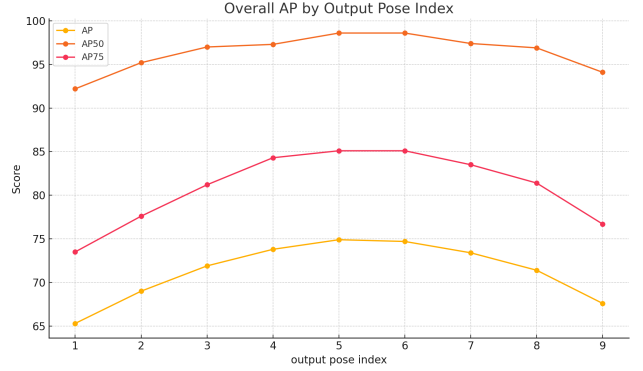


Figure 2. Prediction accuracy versus frame index within the sliding window. The middle frame consistently achieves the highest accuracy.

Table 1. Ablation study on the impact of different mamba scanning patterns.

| Scanning Pattern in Mamba | AP | AP ⁵⁰ | AP ⁷⁵ |
|--|-------------|------------------|------------------|
| Spiral Scan | 70.8 | 97.0 | 80.8 |
| <i>frame</i> \rightarrow <i>range</i> \rightarrow <i>angle</i> | 72.1 | 97.2 | 81.4 |
| <i>angle</i> \rightarrow <i>range</i> \rightarrow <i>frame</i> | 72.6 | 97.3 | 82.7 |
| <i>range</i> \rightarrow <i>angle</i> \rightarrow <i>frame</i> | 74.5 | 98.5 | 84.7 |