Poster: An Empirical Study in mmWave-Based 3D Human Pose Estimation

Zhenyu Wang
University of North Carolina at Chapel Hill
Chapel Hill, USA
zywang@cs.unc.edu

Shahriar Nirjon
University of North Carolina at Chapel Hill
Chapel Hill, USA
nirjon@cs.unc.edu

ABSTRACT

mmWave-based 3D human pose estimation is gaining popularity due to its non-intrusive nature and privacy-preserving capabilities. However, understanding the inner workings of these black-box models remains a challenge, especially with the unreliable and inconsistent signals from mmWave sensing. In this paper, we propose a new metric to quantify model behavior and systematically analyze the predicted pose joints. Our findings show that revealing hidden correlations between radar inputs and pose predictions can significantly enhance tasks such as human activity recognition. Incorporating this characterized information into the input of downstream model improves accuracy by 9.21%.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computer systems organization** → **Embedded and cyber-physical systems**.

KEYWORDS

mmWave sensing, joint characterization, human pose estimation

1 INTRODUCTION

As mmWave sensing matures, its applications in human-centric tasks are expanding. Ensuring the accuracy of core tasks like human pose estimation is crucial to prevent errors from affecting downstream applications such as gait analysis [1], patient monitoring [6], and posture tracking [5]. Unfortunately, state-of-the-art mmWave-based human pose estimation models [3, 4, 10] struggle with downstream activity recognition tasks, despite their impressive skeleton-wise accuracy. This is especially true when certain activity-specific joints are poorly predicted, resulting in suboptimal inputs for downstream models.

We observe that, while these models produce accurate poses, they often rely more on prior knowledge of human body structure and typical poses than on sensor signals, particularly when the reflected signal is sparse. Understanding these inherent model biases and characterizing the behavior of pose estimation models remain open challenges.

Although model bias is a known phenomenon with several mitigation techniques [3, 4, 8–10], it is somewhat inevitable in the context of mmWave sensing. mmWave radar receives highly unreliable and inconsistent reflections off the human body [2, 7]. In real-world settings, it is practically impossible to position the radar to receive adequate signals from all parts of the body, even when signal quality is augmented with software and hardware techniques [8, 11]. If a model relies solely on mmWave signals without prior knowledge, it may never fully learn the concept of pose due to the lack of dense signals from all body parts. Therefore, instead of attempting to eliminate this bias, we propose to characterize it and make it explicit in the predicted pose.

By clarifying how each body joint is estimated—whether it is sensed by the radar or generated based on statistical prior knowledge—we observe a significant improvement in the accuracy of downstream activity recognition. The recognizer benefits from the additional information on joints, enabling it to make more accurate predictions by understanding the precise distribution of bias in the pose estimator.

2 JOINT CHARACTERIZATION

Pose estimation using mmWave signals involves predicting the 3D coordinates of 19-22 body joints, such as the head, shoulders, elbows, knees, and ankles. A model's accuracy is generally assessed based on joint estimation error, which is the distance between the predicted joint coordinates and the ground truth. If this error falls below a certain threshold, the joint is classified as *positive*; otherwise, it is classified as *negative*. However, this metric does not account for whether the joint was estimated using actual radar signals or if it was inferred from the model's prior knowledge. Even if the model makes accurate predictions, its internal decision-making process remains opaque, making it challenging to fully understand how it operates.

By utilizing both signal data and ground-truth joint coordinates, we can gain insight into the model's behavior. Using the inverse square law, we can quantify how much the signals contribute to each joint's estimation. For each joint, we define *signal strength* as $\sum_{n=1}^{N} \frac{I_n}{d_{j,n}^2}$, where N is the number of points in the mmWave point cloud, I_n is the intensity of the signal, and $1/d_{j,n}^2$ is the normalized inverse squared distance between joint j and point n. If the signal strength exceeds a defined threshold, the joint is considered *sensed* by the radar. Some joints may be *indirectly sensed* if their coordinates can be derived from another directly sensed joint.

We classify each joint into one of four categories based on the accuracy of the predicted coordinates and the signal strength at that joint. Figure 1 (left) shows these four categories, defined as follows:

- *Surprise Positive:* Joints correctly estimated (close to ground truth) but not sensed (directly or indirectly).
- *Surprise Negative:* Joints incorrectly estimated (far from ground truth) despite being sensed (directly or indirectly).
- *Expected Positive:* Joints correctly estimated (close to ground truth) and also sensed (directly or indirectly).
- *Expected Negative:* Joints incorrectly estimated (far from ground truth) and not sensed (directly or indirectly).

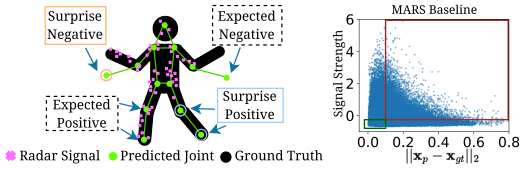

Figure 1: Joint types and their distribution for MARS model.

Model	Dataset	#Surprise Positive	#Surprise Negative	#Expected Positive	#Expected Negative
MARS (CNN-depth 6)	MARS	46,740	26,702	42,158	36,096
MARS (CNN-depth 4)	MARS	41,009	35,546	33,314	41,827
MARS (CNN-depth 6)	mmBody	115,181	35,897	82,239	32,245
mmBody (P4T-2 frames)	MARS	67,495	10,833	57,761	14,828
mmBody (P4T-5 frames)	MARS	74,230	5,299	62,213	6,838
mmBody (P4T-2 frames)	mmBody	112,117	36,748	81,388	35,177
mmBody (P4T-5 frames)	mmBody	113,580	36,834	81,302	33,318
mmMesh (3 frames)	MARS	68,327	14,786	54,074	14,509
mmMesh (10 frames)	MARS	76,909	6,450	62,410	5,927
mmMesh (3 frames)	mmBody	113,519	36,427	81,709	33,343
mmMesh (10 frames)	mmBody	116,673	37,561	80,575	30,753

Table 1: Distribution of joint types across different model variants and datasets.

Model Input	Accuracy
Joint	77.64%
Joint + Label	86.85%

(a) Labeling joints improves accuracy.

Joint Label	Surprise Positive	Surprise Negative	Expected Positive	Expected Negative
#Joint	826	122	73	1,202

(b) A closer look at test cases where joint labeling helped correct mispredictions.

Table 2: Effect of joint characteristics on activity recognizer.

3 EMPIRICAL STUDY

We examine the distribution of joint types for three leading models—MARS [3], mmBody [4], and mmMesh [10]—using publicly available datasets from MARS and mmBody, each captured with different radar technologies. To ensure fairness and consistency, the models are trained on the combined training and validation sets, and evaluated on the test sets. To align mmBody model's input format, we omit the initial frames from the input sequences. We apply thresholds of $10cm$ for joint accuracy, -0.25 for MARS signal strength, and -0.035 for mmBody signal strength, using the normalized square of signal amplitude to measure reflected intensity.

In Figure 1 (right), we present a visualization of the joints categorized by signal strength and estimation error. The results show that 30.81% of the joints fall into the surprise positive category (lower left box), while 17.60% are surprise negative (upper right box). Table 1 details the joint distributions across all three models and their variations, spanning multiple datasets. Our analysis reveals a significant presence of surprise joints, indicating underlying model behaviors that recent studies have largely overlooked.

On the MARS dataset, the average number of surprise positive joints is 62,452, and surprise negative joints is 16,603, representing 54.57% of the total positive joints and 45.35% of the negative joints. A similar distribution pattern is observed in the mmBody dataset. Moreover, as the models' precision improves, the proportion of surprise positive joints increases, while the number of surprise negative joints decreases.

We design a 10-class activity recognizer that uses human body poses, represented by the 3D coordinates of 19 body joints, as input. The poses are estimated using the pre-trained MARS model, and the activities are sourced from the MARS dataset, covering both upper and lower body movements. The baseline recognizer achieves an accuracy of 77.64%. To improve performance, we augment the model by adding joint types as extra input features, while only adjusting the input layer to ensure fairness. After training and evaluating on the same dataset, this modification leads to a 9.21% improvement in accuracy, as shown in Table 2 (a).

To explore the source of this improvement, we analyze the distribution of joint types in instances where the baseline model misclassifies but the augmented model correctly predicts. As shown in Table 2 (b), most of the joint labels are either surprise positive or expected negative. This suggests that the augmented classifier effectively leverages joint type information to improve activity recognition, particularly when the input signals are weak. In contrast, the baseline model, which relies more on prior statistical knowledge when signals are sparse, generates poses containing many expected negative and surprise positive joints. These joint labels reveal patterns in the model's correct and incorrect predictions, providing critical information to improve the accuracy of downstream applications.

4 CONCLUSION AND FUTURE WORK

We introduce a new method to characterize the latent behavior of current mmWave-based human pose estimation systems by analyzing the correlation between predicted pose joints and received signal strength. Additionally, we demonstrate that incorporating this joint-wise label information with existing inputs can enhance downstream task. For future work, developing a systematic approach to generate this information without relying on ground truth is essential.

5 ACKNOWLEDGEMENTS

This work was supported, in part, by grants NSF CAREER-2047461 and NIH 1R01LM013329-01.

REFERENCES

[1] M. A. Alanazi, A. K. Alhazmi, O. Alsattam, K. Gnau, M. Brown, S. Thiel, K. Jackson, and V. P. Chodavarapu. Towards a low-cost solution for gait analysis using millimeter wave sensor and machine learning. *Sensors*, 22(15):5470, 2022.

[2] S. An, Y. Li, and U. Ogras. mri: Multi-modal 3d human pose estimation dataset using mmwave, rgb-d, and inertial sensors. *Advances in Neural Information Processing Systems*, 35:27414–27426, 2022.

[3] S. An and U. Y. Ogras. Mars: mmwave-based assistive rehabilitation system for smart healthcare. *ACM Transactions on Embedded Computing Systems (TECS)*, 20(5s):1–22, 2021.

[4] A. Chen, X. Wang, S. Zhu, Y. Li, J. Chen, and Q. Ye. mmbody benchmark: 3d body reconstruction dataset and analysis for millimeter wave radar. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3501–3510, 2022.

[5] H. Kong, X. Xu, J. Yu, Q. Chen, C. Ma, Y. Chen, Y.-C. Chen, and L. Kong. m3track: mmwave-based multi-user 3d posture tracking. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, pages 491–503, 2022.

[6] A. Sengupta, F. Jin, R. Zhang, and S. Cao. mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns. *IEEE Sensors Journal*, 20(17):10032–10044, 2020.

[7] S. Wang, D. Cao, R. Liu, W. Jiang, T. Yao, and C. X. Lu. Human parsing with joint learning for dynamic mmwave radar point cloud. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(1):1–22, 2023.

[8] H. Xue, Q. Cao, Y. Ju, H. Hu, H. Wang, A. Zhang, and L. Su. M4esh: mmwave-based 3d human mesh construction for multiple subjects. In *SenSys*, pages 391–406, 2022.

[9] H. Xue, Q. Cao, C. Miao, Y. Ju, H. Hu, A. Zhang, and L. Su. Towards generalized mmwave-based human pose estimation through signal augmentation. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023.

[10] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, and L. Su. mmmesh: Towards 3d real-time dynamic human mesh construction using millimeter-wave. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pages 269–282, 2021.

[11] K. Zheng, W. Zhao, T. Woodford, R. Zhao, X. Zhang, and Y. Hua. Enhancing mmwave radar sensing using a phased-mimo architecture. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services*, pages 56–69, 2024.