MVDoppler-Pose: Multi-Modal Multi-View mmWave Sensing for Long-Distance Self-Occluded Human Walking Pose Estimation (Supplementary Material)

Jaeho Choi[1,2]   Soheil Hor[2]   Shubo Yang[2]   Amin Arbabian[2]
[1]DGIST   [2]Stanford University

jhochoi@dgist.ac.kr, {soheilh, shuboy, arbabian}@stanford.edu

https://mvdoppler-pose.github.io/

Abstract

In this supplementary material, we provide detailed descriptions for millimeter-wave (mmWave) signal pre-processing as well as implementation specifics of the proposed MVDoppler-Pose model. Furthermore, we offer additional details on data acquisition and error metrics, and present a broader set of experimental results.

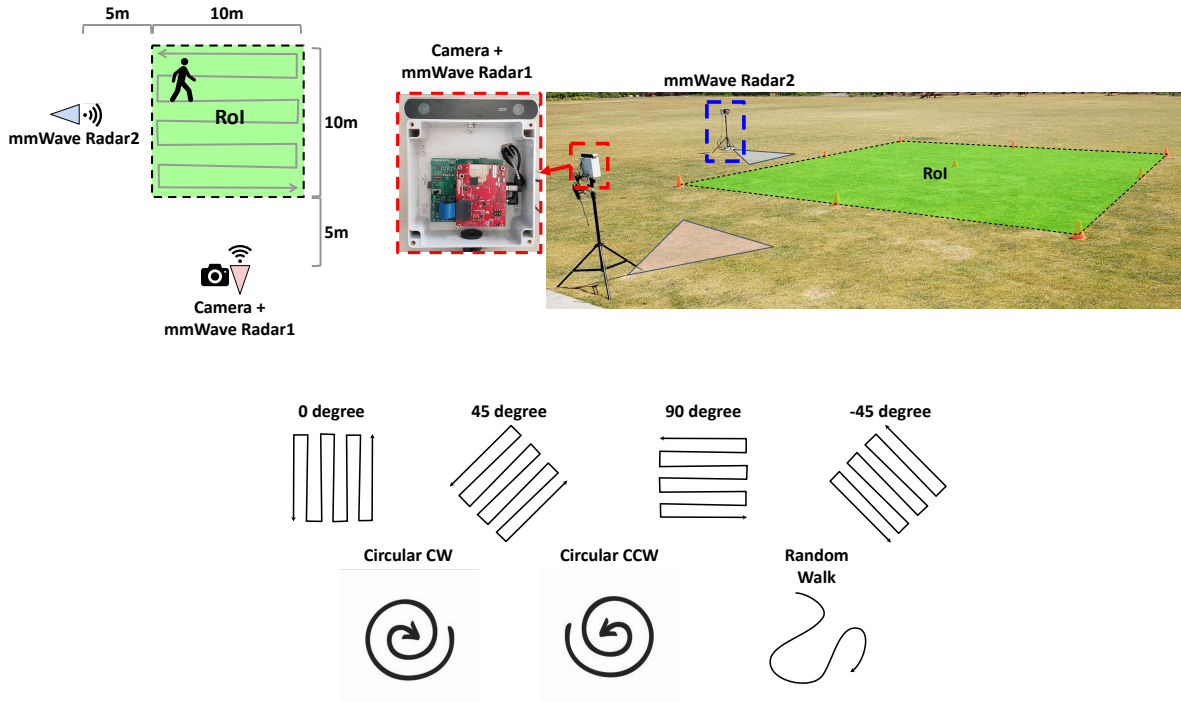## A1. Details for Experimental Setup and Dataset

Figure A1. Experimental setup. **Top**: Experimental environment including RoI and data capture setup. **Bottom**: Selected walking patterns of subjects during data capture.

In this project, we expanded the scope of the MVDoppler [5] dataset with enhanced annotations of 3D human poses, now

releasing it as MVDoppler-Pose. In this section, we detail the experimental setup as well as the methodology employed for generating precise human pose annotations corresponding to the MVDoppler data.

A1.1. Experimental Configuration

Table A1. Comparison of MVDoppler-Pose with related datasets. * denotes that the dataset is not publicly available. N/A: Not reported in the paper. Note that most of current approaches operate for subjects without global walking (*i.e.* performing activities while standing still), from limited distances between the sensor and subjects. In contrast, our MVDoppler-Pose involves freer walking trajectory patterns of subjects from much larger RoI.

Dataset	Modality		Trajectory Pattern				Distance	RoI
	mmWave (Single)	mmWave (Multi)	Standing in Place	Walking in Straight	Walking in Circle	Random		
RF-Pose3D [23] *	✓	-	✓	✓	✓	✓	N/A	N/A
mmPose [13] *	✓	-	✓	-	-	-	N/A	-
mmMesh [19] *	✓	-	✓	✓	✓	-	1.5m	N/A
MARS [1]	✓	-	✓	-	-	-	2m	-
mmBody [3]	✓	-	✓	-	-	-	3m~5m	-
mRI [2]	✓	-	✓	✓	-	-	N/A	-
mm-Fi [22]	✓	-	✓	-	-	-	3m	-
M4esh [20] *	✓	-	✓	✓	✓	-	0m~4.6m	3m×4.6m
mmGPE [21] *	✓	-	✓	✓	-	-	1.5m~4m	-
MDPose [15] *	✓	-	✓	-	-	-	4m	-
HuPR [8]	✓	✓	✓	✓	-	-	N/A	-
XRF55 [17]	✓	-	✓	✓	-	-	0.75m~3.85m	3.1m×3.1m
MVDoppler-Pose	✓	✓	✓	✓	✓	✓	**5m~15m**	**10m×10m**

The Fig. A1 shows the overall experimental setup for the MVDoppler dataset, involving an HD stereo camera (ZED from Stereolabs) and two off-the-shelf FMCW radars (AWR1843 from Texas Instruments). The experimental layout, as illustrated in the top of the figure, strategically positions the two mmWave radars in a cross-view arrangement around a designated 10m×10m region of interest (RoI), with each sensor placed 5m away from this rectangular area. To optimize sensor functionality as well as minimize cross-interference, the radars operated on different frequency bands (77-78GHz for mmWave radar1 and 79-80GHz for mmWave radar2), facilitating their simultaneous, high-duty-cycle (>95%) operation. Additionally, synchronized RGB recordings from the stereo camera, co-located with mmWave Radar1, were captured together for enhanced visual analysis of the scene. The dataset documentation [5] offers further details including IRB approval, participant statistics, and sensor specifications.

During data collection within the extensive RoI, a total of 13 participants were instructed to perform different types of hand movements, including *normal walking*, *hands in pockets*, and *texting*. These activities were distinctly recorded across seven different walking patterns (the bottom side of Fig. A1) within the entire RoI, which naturally guarantees a comprehensive representation of diverse locations, movement speeds, and walking angles of each subject.

Note that our MVDoppler-Pose is the first mmWave human pose estimation (HPE) dataset that fully involves the comprehensive coverage of both location and angle diversities for human subjects while engaging in various activities, as depicted in Fig. A2. Table A1 shows the comparison between the current radio-based HPE models with MVDoppler-Pose in the context of experimental setups. Except for RF-Pose3D [23] which utilizes a bulky, custom hardware system, most of existing mmWave HPE models operate under controlled human walking scenarios (*i.e.* without free random movements) as well as restricted RoI sizes. In contrast, our MVDoppler-Pose involves freer walking trajectory patterns from much larger RoI. By capturing a wide range of positional and angular diversity from this expanded RoI, MVDoppler-Pose sets a new benchmark for comprehensive evaluation of HPE task across a variety of possible human walking trajectory patterns, surpassing the scope and limitations of the previous mmWave-based HPE datasets.
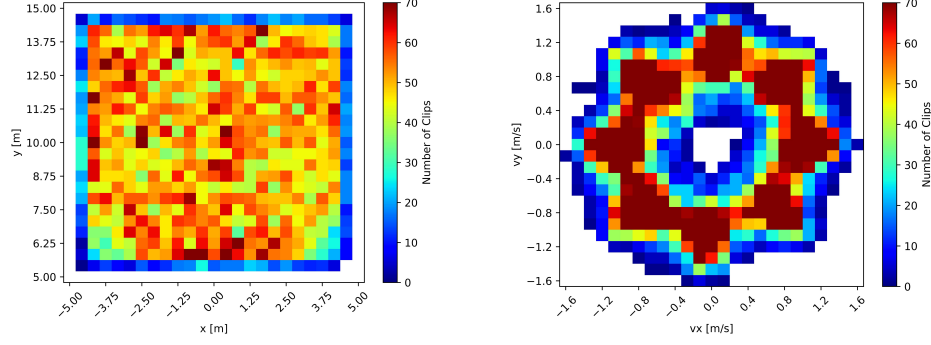
Figure A2. Dataset statistics across both location (**Left**) and angle diversities (**Right**).

A1.2. 3D Human Pose Annotation

To generate 3D human pose annotations for each dataset frame, we leveraged the overall annotation pipeline described in [2, 22]. This approach primarily involves lifting the multi-view 2D poses, derived from synchronized RGB recordings, into accurate 3D configurations through the combination of 3D triangulation techniques and optimization-based refinement processes. Here, manual 2D pose annotation of each RGB frame was used to ensure the quality and accuracy of the 2D data being converted into 3D space. The estimated 3D poses with errors are then cleaned up again through visual inspection to further improve the reliability of the annotations, yielding over 600k frames of final paired data.

A2. Details for mmWave Pre-Processing Pipeline

In this section, we introduce the basic principles of mmWave signal, outlining the pre-processing pipeline and its distinctive characteristics.

A2.1. FMCW Signal Modeling and Pre-Processing

In this subsection, we delve into the detailed pipeline for processing mmWave signals based on a frequency-modulated continuous-wave (FMCW) radar sensor. The FMCW radar operates by transmitting a periodic FMCW signal, which, after interacting with objects in its surroundings, reflects back to the sensor. The formal representation of the transmitted (Tx) and received (Rx) signals of the FMCW radar can be modeled as [4, 7, 9]

$$x_{\text{Tx}}(t_{\text{f}}) = \exp\left(j\left(2\pi f_c t_{\text{f}} + \pi \frac{BW}{T_{\text{f}}} t_{\text{f}}^2\right)\right), \tag{A1}$$

$$x_{\text{Rx}}(t_{\text{f}}, t) = \sum_i \alpha_i x_{\text{Tx}}\left(t_{\text{f}} - \frac{2R_i(t)}{c}\right), \tag{A2}$$

where x_{Tx} refers to the transmitted FMCW signal, and x_{Rx} is the received signals represented by the weighted summation of time-delayed x_{Tx}. t_{f} and t denote the fast and slow time instant, each of which reflects the spatial domain with respect to the round-trip time-of-flight delay and temporal domain with respect to the pulse repetition interval (PRI) of the radar, respectively. f_c, BW, T_{f} and c indicate the center frequency, bandwidth, pulse width, and the speed of light, respectively. α signifies the reflection coefficient of each electromagnetic (EM) scatterer and R is the radial distance between the sensor and

the corresponding scatterer. Using a frequency mixer, the received signal $s_{\text{Rx}}(t_{\text{f}})$ can be converted to a baseband signal as

$$r(t_{\text{f}}, t) = x_{\text{Tx}}(t_{\text{f}})x_{\text{Rx}}^*(t_{\text{f}})$$

$$= \sum_i \alpha_i \exp\left(j\left(2\pi f_c t_{\text{f}} + \pi \frac{BW}{T_{\text{f}}}t_{\text{f}}^2\right)\right)$$

$$\cdot \exp\left(-j\left(2\pi f_c\left(t_{\text{f}} - \frac{2R_i(t)}{c}\right) + \pi \frac{BW}{T_{\text{f}}}\left(t_{\text{f}} - \frac{2R_i(t)}{c}\right)^2\right)\right)$$

$$= \sum_i \alpha_i \exp\left(j4\pi \frac{R_i(t)}{c}\left(f_c + \frac{BW}{T_{\text{f}}}t_{\text{f}}\right) - j4\pi \frac{BW}{T_{\text{f}}}\frac{R_i^2(t)}{c^2}\right)$$

$$\approx \sum_i \alpha_i \exp\left(j4\pi \frac{R_i(t)}{\lambda}\right)\exp\left(j4\pi \frac{BW}{T_{\text{f}}}\frac{R_i(t)}{c}t_{\text{f}}\right). \tag{A3}$$

Here, $*$ is the conjugate operator and λ denotes the signal wavelength ($\lambda = c/f_c$). It should be noted that the converted baseband signal comprises sinusoidal signals with different fundamental frequencies of $2BW\,R_i(t)/(T_{\text{f}}c)$, where BW, T_{f}, and c are constant values over time. Consequently, we can transform $r(t_{\text{f}}, t)$ into a range profile signal by applying a fast Fourier transform (FFT) with respect to t_{f}:

$$s(R, t) = \mathcal{F}_{t_{\text{f}}}\{r(t_{\text{f}}, t)\}$$

$$= \sum_i \alpha_i \delta\left(f - \frac{2BW}{T_{\text{f}}}\frac{R_i(t)}{c}\right)\exp\left(j4\pi \frac{R_i(t)}{\lambda}\right)$$

$$= \sum_i \alpha_i \delta\left(R - R_i(t)\right)\exp\left(j4\pi \frac{R_i(t)}{\lambda}\right), \tag{A4}$$

where $R = (cT_{\text{f}}f)/(2BW)$ spans the radial range domain. $\mathcal{F}_{t_{\text{f}}}\{\cdot\}$ refers to the FFT operator across t_{f}-axis and δ is an impulse signal envelope. This FFT operation results in a signal where each impulse's position in the frequency domain corresponds to the radial range $R_i(t)$ of the scatterer, thereby reflecting the spatial position of objects in the scene.

Note that the range resolution of $\Delta R = c/(2BW)$ [9], remains *constant regardless of the distance of the target* R, ensuring mmWave radar's capability to capture the spatial context of subjects irrespective of their distance, provided the SNR level is sufficient.
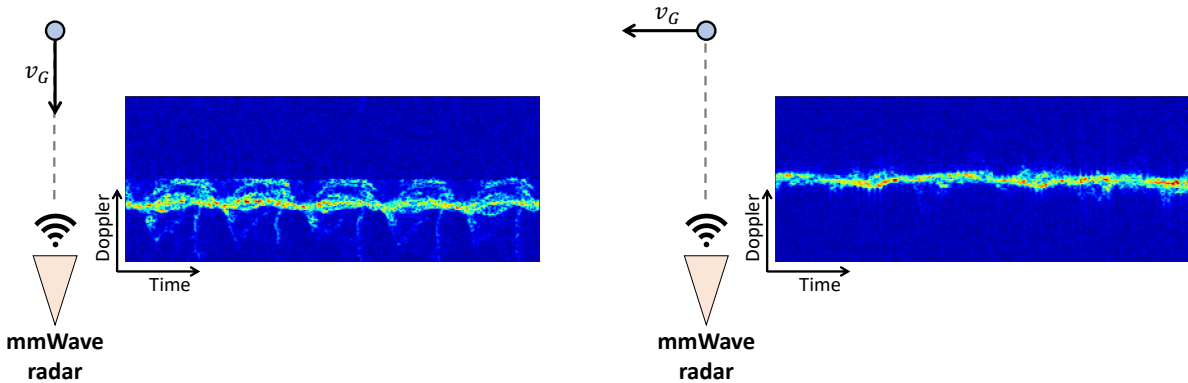
A2.2. Doppler Effect

Figure A3. Doppler measurements of a person walking towards the sensor (**Left**) or a person with tangential movements (**Right**).

Beyond the positional measurement, the Doppler property in mmWave radar enables the direct capture of *motional information* of the reflected objects. This effect is observable as phase differences over time in mmWave radar's range profiles, mathematically expressed as

$$\Delta\theta(t) = \frac{4\pi\Delta R}{\lambda} = \frac{4\pi v\Delta t}{\lambda}, \tag{A5}$$

where, v refers to the relative velocity between the mmWave sensor and each EM scatterer (*e.g.* joint of human body) in the radial direction. Given a sufficiently small time window to ensure a constant v, this constant velocity component leads to a constant periodicity in the mmWave phase that converges to a specific Doppler shift in the frequency domain. Thus, we can extract the Doppler information (*i.e.* motional information of the body scatterer) at each time window, by utilizing a short-time Fourier transform (STFT) along the slow time dimension.

The motional (*i.e.* radial velocity) resolution observed by Doppler is determined as $\Delta v = \lambda/(2T)$, where T represents the length of the time window [6]. Remarkably, this implies that the effective resolution of mmWave radar's motional information is *independent to the distance of subjects*, similar to its positional counterpart.

Meanwhile, the radial velocity v measured by Doppler can further be decomposed into $v = v_G \cos(\phi)$, where v_G is the absolute velocity of the target and ϕ denotes the angle between the target's movement direction and radar line-of-sight (LoS). That is, the motional information measured by Doppler is highly sensitive to the relative angle ϕ. For instance, a person walking directly towards ($\phi = 180°$) or away ($\phi = 0°$) from the radar maximizes the Doppler measurement, whereas one with tangential movements ($\phi = \pm90°$) minimizes it (Fig. A3).

A3. Details for Benchmark and Model Implementation

A3.1. Details on Model Implementation

In the development of MVDoppler-Pose, we harness both positional and motional mmWave measurements to achieve enhanced performance compared to previous single-modal approaches. We drive positional information from the processed range profiles $s(R, t)$ after applying clutter suppression and selective windowing for a detection range of up to 15m. To extract motional information, we first identify the subject's location using the constant false alarm rate (CFAR) algorithm [12]. Subsequently, we employ STFT within a 1.5m range window centered on the detected location, utilizing a Hamming window of 128 chirp durations, an 87.5% window overlap, and an FFT size of 128. Both the magnitude components of positional and motional data are clipped to 3 seconds, resized into 128×128 images, and normalized for consistency.

To train the MVDoppler-Pose model, we employ the Adam optimizer with an initial learning rate of 10^{-4}, a weight decay of 0.01, and a batch size of 32, across 250 epochs. The learning rate follows a cosine annealing scheduler with warm restarts, a strategy designed to finely adjust the learning rate throughout the training phases. Specifically, the parameters of the scheduler involve an initial restart interval of 25, a periodicity increase factor of 1, and a minimum learning rate threshold of 10^{-6}. The hyper-parameters regarding our cross-domain loss function were empirically selected as $\lambda_p = 1$, $\lambda_m = 0.1$, and $\mathbb{T} = \{2, 4, 6, 8\}$.

For fair comparison across all baseline benchmarks, including both vision-based HPE models and MVDoppler-Pose variations, we apply a consistent time window length for all comparative analyses. The vision-based baselines were re-implemented following their official codes to match the conditions of MVDoppler-Pose dataset, and the single-modal or single-view MVDoppler-Pose variants were developed by skipping the operations of corresponding branches. During inference, we select the central frame from the 3D pose sequences as our final output.

A3.2. Details on Error Metrics

To evaluate the quality of the 3D pose outputs, we utilize the mean per joint position error (MPJPE), a metric that quantifies the spatial distance between the predicted and actual 3D poses. In addition to MPJPE focusing primarily on the spatial quality of the estimated poses, we introduce a correlation metric (ρ), designed to assess the temporal quality and coherence of the estimated 3D joint patterns. This metric leverages Pearson's correlation coefficient to analyze the alignment between the time-series of predicted joints and the reference, calculating an average correlation across joints.

Notably, this correlation metric offers a unique advantage over traditional distance-based metrics by emphasizing the temporal fidelity of joint movements without being largely sensitive to minor offset shifts. For a more intuitive analysis, we compare the spatial distance- and correlation-based errors for three distinct estimated patterns against a given sinusoidal signal (Fig. A4). Despite the global quality of the three estimations being significantly different, the distance-based metric (*i.e.* L2 distance) indicates a uniform level of error, underscoring its inability to fully capture the global quality of the joints over time. In contrast, the correlation metric correctly reflects the temporal quality between the predicted and actual patterns, showcasing its superior capability in globally reflecting the temporal dynamics of human movement patterns.
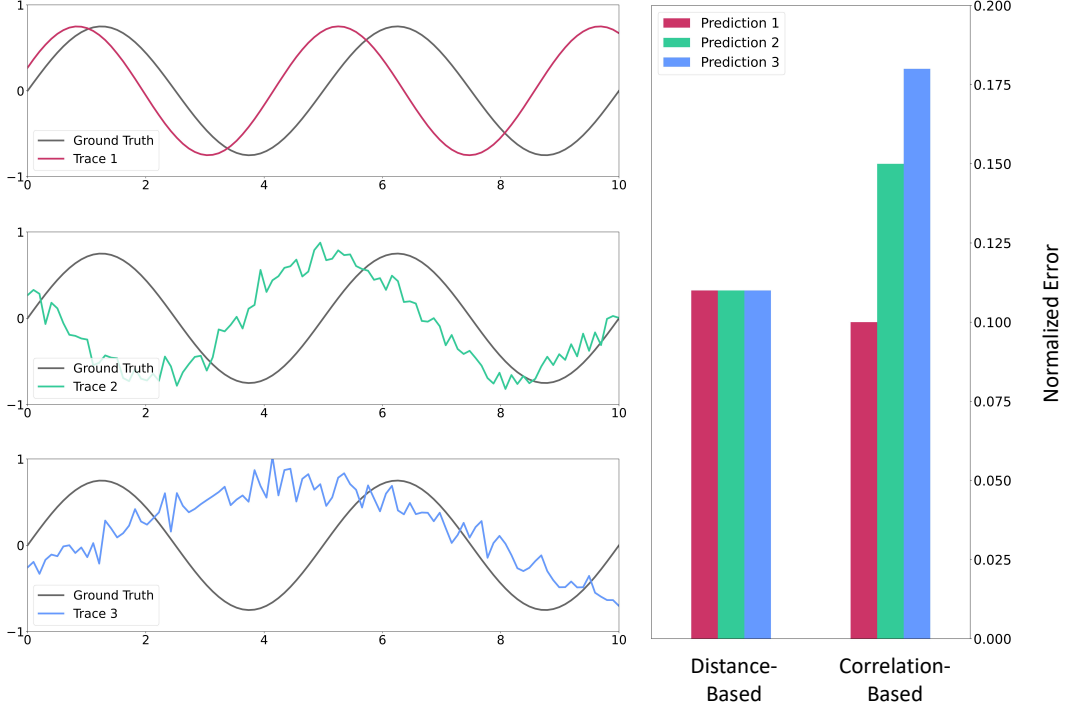
Figure A4. Comparison between distance- and correlation-based metrics. Note that the error values are normalized, and the correlation-based metrics are visualized on a negative basis (*i.e.* $1 - \rho$) for a more straightforward comparison.

Table A2. Quantitative comparison of MVDoppler-Pose-variants against the vision- and mmWave-based models for all walking scenarios (between -180° and 180°) according to different levels of target distance.

Method	Modality	$y <$**8.5m**		**8.5m**$\leq y \leq$**11.5m**		**11.5m**$< y$	
		MPJPE↓	ρ↑	MPJPE↓	ρ↑	MPJPE↓	ρ↑
Sun *et al.* [14]	RGB	73.37	0.29	79.95	0.23	90.56	0.21
Wei *et al.* [18]	RGB	71.10	0.31	77.29	0.23	87.44	0.22
Li *et al.* [10]	RGB	80.86	0.37	86.43	0.32	103.68	0.32
Zhao *et al.* [24]	RGB	78.75	0.37	85.84	0.33	101.56	0.32
Wang *et al.* [16]	mmW-Position	122.30	0.24	121.46	0.24	125.51	0.23
Zhou *et al.* [25]	mmW-Motion	124.41	0.27	125.93	0.27	127.09	0.26
Tang *et al.* [15]	mmW-Motion	121.80	0.28	120.32	0.27	125.60	0.27
Ours (Single-View)	mmW-Position	117.99	0.25	114.03	0.25	114.77	0.26
Ours (Single-View)	mmW-Motion	120.53	0.29	118.38	0.27	124.94	0.27
Ours (Single-View)	mmW-Multi	104.89	0.30	98.59	0.28	104.02	0.29
Ours (Multi-View)	mmW-Position	69.67	0.47	66.47	0.48	64.75	0.48
Ours (Multi-View)	mmW-Motion	71.48	0.52	67.78	0.51	66.13	0.52
Ours (Multi-View)	mmW-Multi	**64.20**	**0.53**	**59.99**	**0.52**	**58.80**	**0.53**

A4. Additional Results

A4.1. Results for All Walking Scenarios

Note that the quantitative analysis presented in the Table 1 of the main paper primarily addressed comparisons within self-occluded scenarios (walking angles between -45° and 45°). We extend our analysis to encompass all walking scenarios (walking angles between -180° and 180°), which are summarized in Table A2. This broader analysis (*i.e.* not only involving challenging self-occlusion scenarios but also easier samples such as normal walking) shows general performance enhance-

ments in camera-based models. Nevertheless, despite these improvements, they still continue to experience performance degradation with increasing distance, unlike the mmWave baselines which maintain more consistent results.
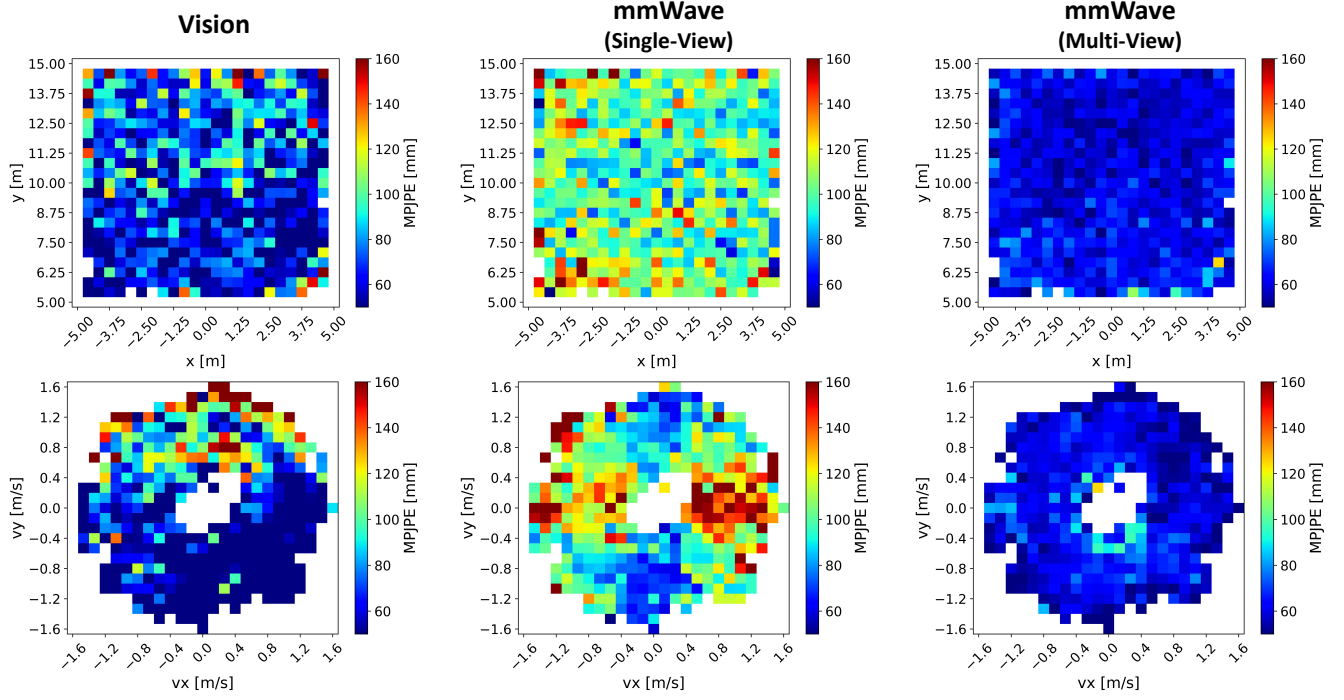
A4.2. Effect of Location and Walking Angle

Figure A5. MPJPE of each baseline model across the RoI (**Top**) and walking velocity (**Bottom**). Vision performance is degraded both from long-distance and self-occluded (*i.e.* heading backward) cases, while single-view mmWave is sensitive to the movements with perpendicular direction. Multi-view mmWave approach maintains great independence for both the subject's location and moving direction.

For a more granular analysis with respect to the effect of locations and walking angles of the subjects, we dissect the pose estimation result of each baseline model according to 2D location (x-y) and 2D velocity (vx-vy), as showcased in Fig. A5. The spatial distribution analysis (x-y plot) uncovers a significant decline in the performance of the vision-based model with increasing distance, attributed to the diminishing information content in each pixel at extended ranges. Furthermore, an exploration of the vx-vy plot, which directly reflects the impact of a subject's orientation, highlights the struggles of vision approaches in occlusion scenarios, such as movements directed away from the camera. Single-view mmWave model, on the other hand, maintains robustness in distance variations, but experiences significant degradation when individuals move in a tangential direction to the radar. The multi-view mmWave model consistently shows significant resilience across nearly all scenarios both in x-y and vx-vy plots, achieving trajectory-agnostic HPE within the entire coverage.

Fig. A6 provides a detailed examination of MVDoppler-Pose in multi-view setup utilizing additional complementary metrics. This extended evaluation encompasses Procrutes Analysis MPJPE (PA-MPJPE), temporal correlation coefficient (ρ), percentage of correct keypoints (PCK) with a threshold of 150mm, and global orientation error. Across all examined metrics, MVDoppler-Pose demonstrates considerable stability across *x-y* and *vx-vy* variations.

A4.3. Effect of Time Window

Table A3 represents the performance of MVDoppler-Pose with respect to different temporal window lengths. The table demonstrates that increasing the window sizes enriches the temporal context, improving overall model performance.

A4.4. Effect of Backbone Model

Table A4 shows the performance of MVDoppler-Pose when trained and evaluated with different MobileViT architectures [11]. MVDoppler-Pose with larger backbones tends to achieve more accurate predictions but at higher computation costs.
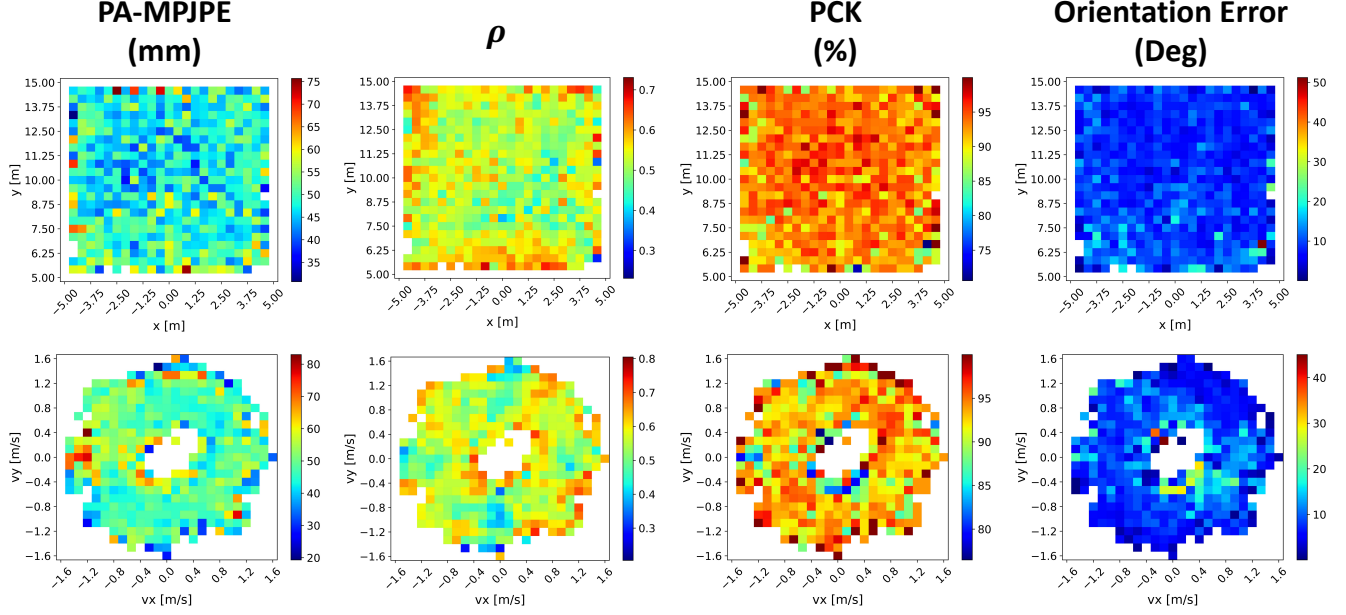
Figure A6. MVDoppler-Pose analysis with additional evaluation metrics across the RoI (**Top**) and walking velocity (**Bottom**). MVDoppler-Pose shows considerable resilience for both the subject's location and direction variations.

Table A3. Performance comparison for different window sizes

Window Size	MPJPE↓	PCK↑	ρ↑
1-s	65.80	89.42	0.51
3-s	60.96	93.24	**0.53**
5-s	**58.89**	**93.61**	0.53

Table A4. Performance comparison for different baseline models

Backbone Network	MPJPE↓	PCK↑	ρ↑
MobileViT-XXS [11]	61.21	93.08	0.53
MobileViT-XS [11]	60.96	93.24	0.53
MobileViT-S [11]	**58.42**	**94.17**	**0.54**

A4.5. Additional Visualization

We provide additional qualitative results of the proposed MVDoppler-Pose in Fig. A7. The figure displays the predicted outputs of MVDoppler-Pose (in multi-view setup) against their corresponding RGB references over time.

References

[1] Sizhe An and Umit Y. Ogras. MARS: mmWave-based assistive rehabilitation system for smart healthcare. *ACM Trans. Embed. Comput. Syst.*, 20, 2021. 2

[2] Sizhe An, Yin Li, and Umit Ogras. mRI: Multi-modal 3D human pose estimation dataset using mmWave, RGB-D, and inertial sensors. In *NeurIPS*, pages 1–13, 2022. 2, 3

[3] Anjun Chen, Xiangyu Wang, Shaohao Zhu, Yanxu Li, Jiming Chen, and Qi Ye. mmBody benchmark: 3D body reconstruction dataset and analysis for millimeter wave radar. In *ACM MM*, page 3501–3510, 2022. 2

[4] Jae-Ho Choi, Ki-Bong Kang, and Kyung-Tae Kim. Remote respiration monitoring of moving person using radio signals. In *ECCV*, pages 253–270, 2022. 3
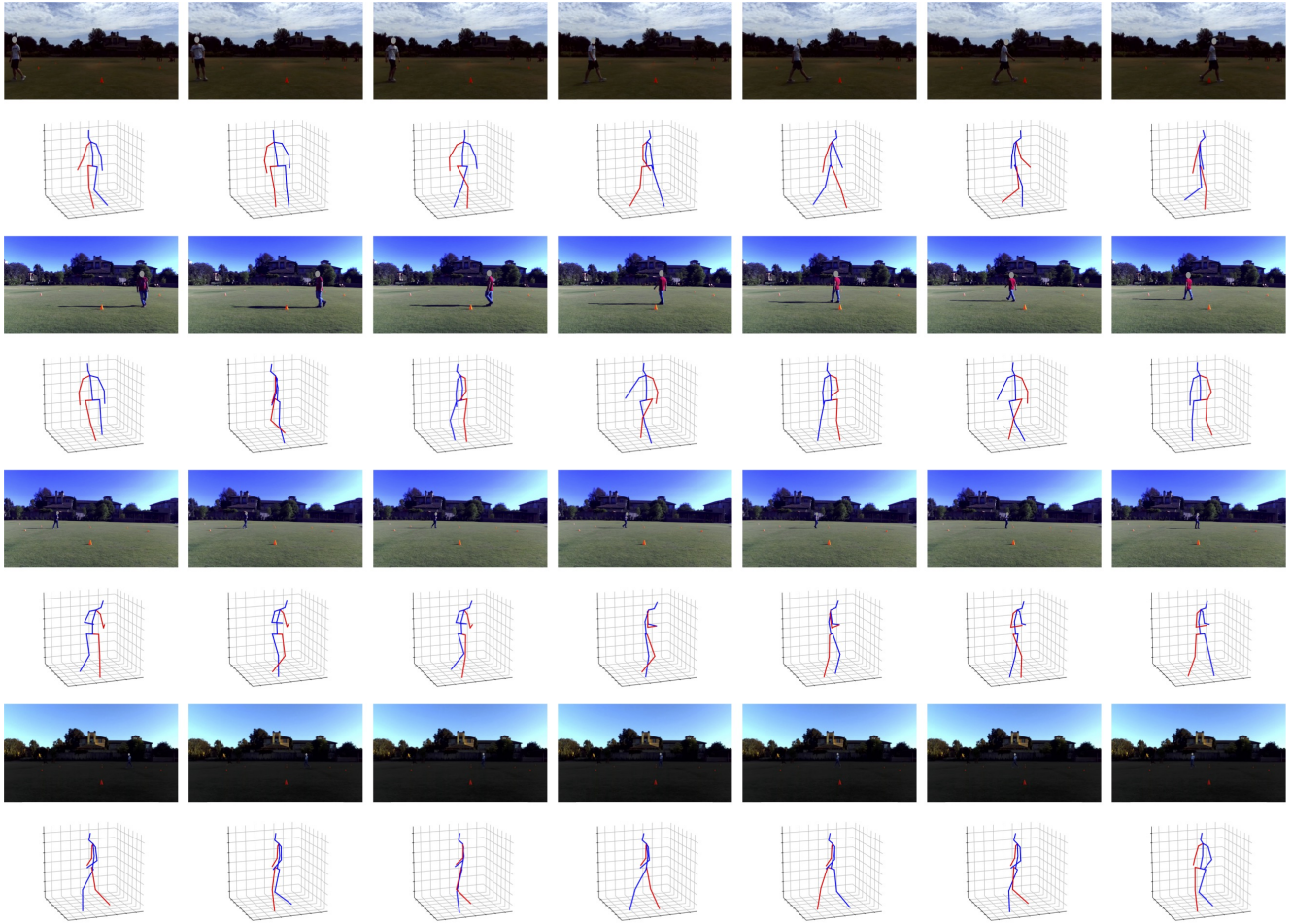
Figure A7. Additional visualization of MVDoppler-Pose predictions over time.

[5] Soheil Hor, Shubo Yang, Jae-Ho Choi, and Amin Arbabian. MVDoppler: Unleashing the power of multi-view doppler for micromotion-based gait classification. In *NeurIPS*, pages 1–11, 2023. 1, 2

[6] Cesar Iovescu and Sandeep Rao. The fundamentals of millimeter wave sensors. *Texas Instrum.*, pages 1–8, 2017. 5

[7] Chengkun Jiang, Junchen Guo, Yuan He, Meng Jin, Shuai Li, and Yunhao Liu. mmVib: Micrometer-level vibration measurement with mmWave radar. In *ACM Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, pages 1–13, 2020. 3

[8] Shih-Po Lee, Niraj Prakash Kini, Wen-Hsiao Peng, Ching-Wen Ma, and Jenq-Neng Hwang. HuPR: A benchmark for human pose estimation using millimeter wave radar. In *WACV*, pages 5715–5724, 2023. 2

[9] Jian Li and Petre Stoica. *MIMO radar signal processing*. John Wiley & Sons, Hoboken, New Jersey, USA, 2008. 3, 4

[10] Wenhao Li, Hong Liu, Hao Tang, Pichao Wang, and Luc Van Gool. Mhformer: Multi-hypothesis transformer for 3d human pose estimation. In *CVPR*, pages 13147–13156, 2022. 6

[11] Sachin Mehta and Mohammad Rastegari. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. In *ICLR*, pages 1–26, 2022. 7, 8

[12] Hermann Rohling. Radar CFAR thresholding in clutter and multiple target situations. *IEEE Trans. Aerosp. Electron. Syst.*, AES-19 (4):608–621, 1983. 5

[13] Arindam Sengupta, Feng Jin, Renyuan Zhang, and Siyang Cao. mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs. *IEEE Sens. J.*, 20(17):10032–10044, 2020. 2

[14] Xiao Sun, Bin Xiao, Fangyin Wei, Shuang Liang, and Yichen Wei. Integral human pose regression. In *ECCV*, pages 536–553, 2018. 6

[15] Chong Tang, Wenda Li, Shelly Vishwakarma, Fangzhan Shi, Simon Julier, and Kevin Chetty. MDPose: Human skeletal motion reconstruction using WiFi micro-Doppler signatures. *IEEE Trans. Aerosp. Electron. Syst.*, pages 1–12, 2023. 2, 6

[16] Fei Wang, Sanping Zhou, Stanislav Panev, Jinsong Han, and Dong Huang. Person-in-WiFi: Fine-grained person perception using WiFi. In *ICCV*, 2019. 6

[17] Fei Wang, Yizhe Lv, Mengdie Zhu, Han Ding, and Jinsong Han. XRF55: A radio frequency dataset for human indoor action analysis. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (IMWUT)*, 8(1), 2024. 2

[18] Wen-Li Wei, Jen-Chun Lin, Tyng-Luh Liu, and Hong-Yuan Mark Liao. Capturing humans in motion: Temporal-attentive 3D human pose and shape estimation from monocular video. In *CVPR*, 2022. 6

[19] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. mmMesh: towards 3D real-time dynamic human mesh construction using millimeter-wave. In *MobiSys*, page 269–282, 2021. 2

[20] Hongfei Xue, Qiming Cao, Yan Ju, Haochen Hu, Haoyu Wang, Aidong Zhang, and Lu Su. M4esh: mmWave-based 3D human mesh construction for multiple subjects. In *ACM Conf. Embedded Netw. Sens. Syst. (SenSys)*, page 391–406, 2023. 2

[21] Hongfei Xue, Qiming Cao, Chenglin Miao, Yan Ju, Haochen Hu, Aidong Zhang, and Lu Su. Towards generalized mmwave-based human pose estimation through signal augmentation. In *ACM Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2023. 2

[22] Jianfei Yang, He Huang, Yunjiao Zhou, Xinyan Chen, Yuecong Xu, Shenghai Yuan, Han Zou, Chris Xiaoxuan Lu, and Lihua Xie. MM-Fi: Multi-modal non-intrusive 4D human dataset for versatile wireless sensing. In *NeurIPS*, pages 1–13, 2023. 2, 3

[23] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. RF-based 3D skeletons. In *Conf. ACM Special Interest Group Data Commun. (SIGCOMM)*, page 267–281, 2018. 2

[24] Qitao Zhao, Ce Zheng, Mengyuan Liu, Pichao Wang, and Chen Chen. PoseFormerV2: Exploring frequency domain for efficient and robust 3D human pose estimation. In *CVPR*, pages 8877–8886, 2023. 6

[25] Xiaolong Zhou, Tian Jin, Yongpeng Dai, Yongkun Song, and Zhifeng Qiu. MD-pose: Human pose estimation for single-channel UWB radar. *IEEE J. Biom. Behavior Id. Sci.*, 5(4):449–463, 2023. 6