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

Jaeho Choi, Soheil Hor, Shubo Yang, and Amin Arbabian

DGIST, Stanford University

Contact
jhochoi@dgist.ac.kr

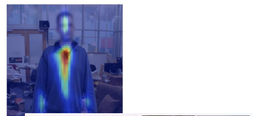
Motivation

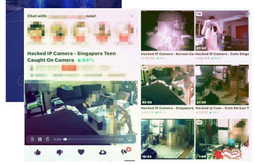
3D Pose: Intuitive Representation of Walking Person

Pedestrian Detection

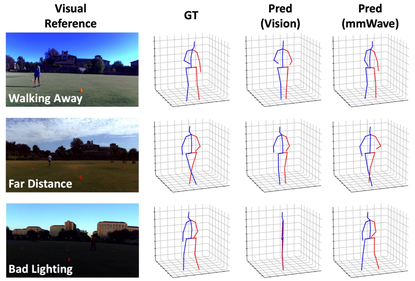
Medical Analysis

Using mmWave Modality for 3D Pose Estimation





Privacy-Preserving Monitoring



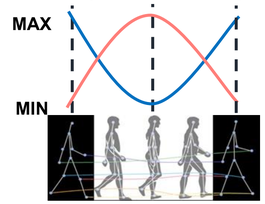
Robustness in Challenging Scenarios

mmWave radar can act as an effective alternative of vision for 3D pose estimation

Our Approach

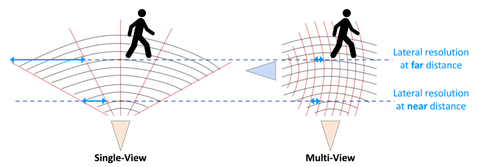
Cross-Modal Fusion (Positional + Motional)

Local **range difference** between right and left joints
Local **velocity difference** between right and left joints



- For a walking person, range and velocity are complementary each other in terms of differentiating each joint.

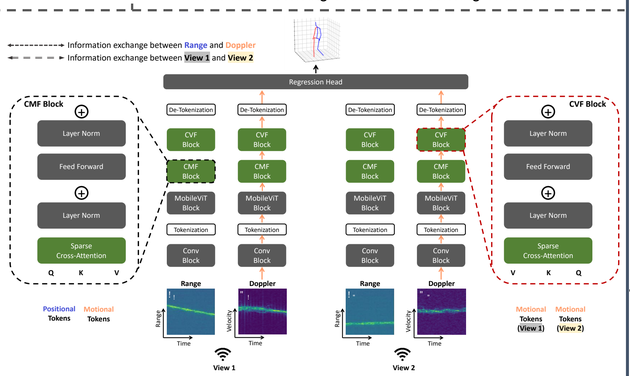
Cross-View Fusion



- Introduction of an additional sensor in the lateral direction, whose radial sensing covers cross sensing of anothe sensor.

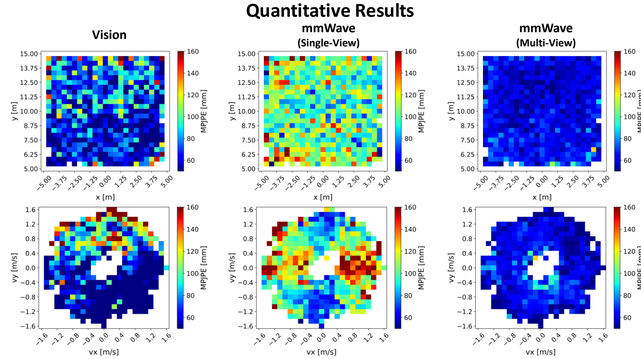
Overall Framework

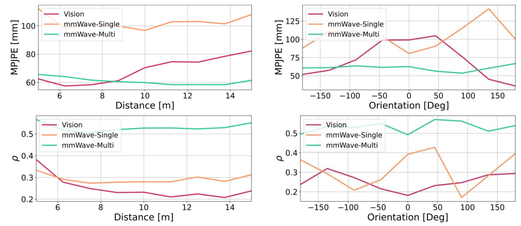
- Take time-range (positional) and time-Doppler (motional) from cross-view mmWave sensors as input.
- During iterative encoding pipeline, the 4 inputs are integrated through cross-attention-based dual-stage fusion.
- Overall model is supervised by the 3D pose generated from synchronized video.



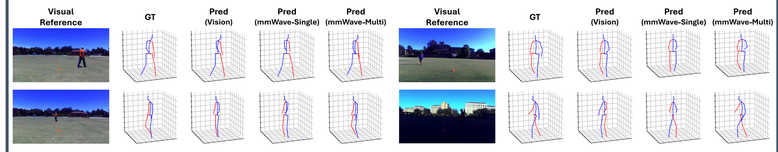
Preliminary Results and Analysis

Quantitative Results





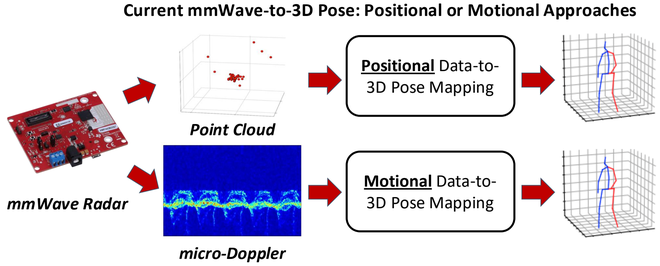
Qualitative Results



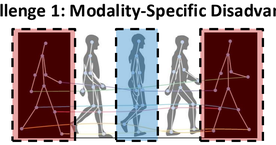
No dependency with respect to human range and orientation!

Current Approaches & Challenges

Current mmWave-to-3D Pose: Positional or Motional Approaches



Challenge 1: Modality-Specific Disadvantages

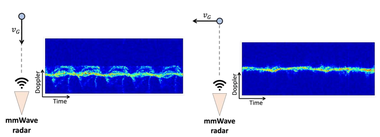


: Challenging scenario for Positional modality

: Challenging scenario for Positional modality

Hard to capture full dynamics of human motion

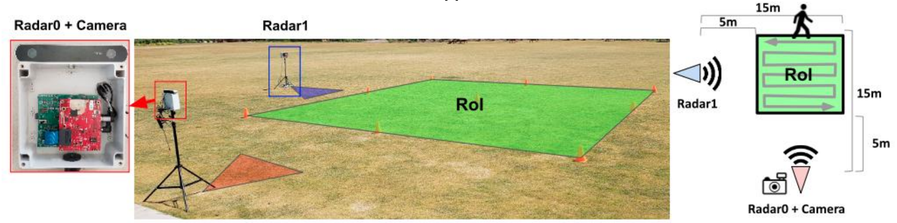
Challenge 2: Directionality of mmWave data



Both modalities work well in radial motions, but struggle with tangential motions

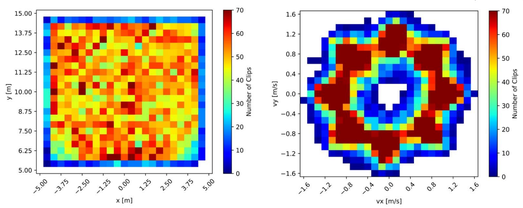
Experiment Design

Dataset: MVDoppler-Pose Dataset



- Dataset capture with 2 FMCW radar (TI AWR1843) and 1 HD stereo camera (Stereolabs ZED)
- Subjects were instructed to walk at natural speed in entire RoI while doing 3 activity classes (Normal/Pockets/Texting)
- Generate ground truth of 3D keypoint using video-based estimation (PoseFormer) and post-processing

Location and Trajectory Diversity



- Covers -5m < X < 5m, 5m < Y < 10m, and walking speeds of 0.75 - 1.25 m/s
- Virtually uniform coverage of all walking directions
- Generate ~20K episodes from 13 people

Summary

- Propose mmWave-Pose model based on cross-modal, cross-view fusion
- Estimate 3D pose of walking person with virtually no dependency on range and angle

Future Directions

- Camera and Radar fusion for advanced estimation
- Human mesh recovery with multi-view mmWave radars
- Combine with efficient transformer for 3D perception at the edge

References

- S. Hor, S. Yang, J. Choi, and A. Arbabian, "MVDoppler: Unleashing the Power of Multi-View Doppler for MicroMotion-Based Gait Classification," in Neural Information Processing Systems (NeurIPS), New Orleans, LA, USA, Dec. 2023.
- Zheng, Ce, et al. "3D Human Pose Estimation with Spatial and Temporal Transformers." in IEEE/CVF International Conference on Computer Vision (ICCV), Virtual, Oct. 2021.