milliFlow: Scene Flow Estimation on mmWave Radar Point Cloud for Human Motion Sensing

Fangqiang Ding[1], Zhen Luo[1], Peijun Zhao[2], Chris Xiaoxuan Lu[3]

[1]University of Edinburgh, [2]Massachusetts Institute of Technology, [3]University College London

Motivation

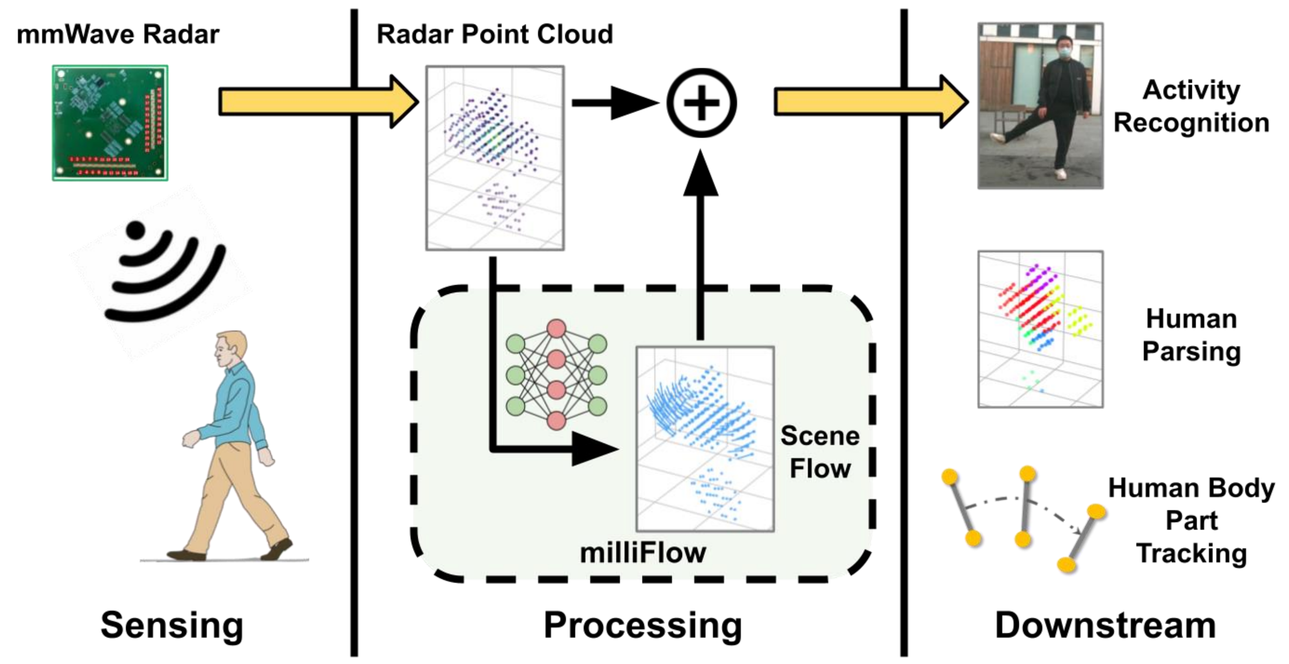
- **Advantages of mmWave radar for human sensing, over**

Camera:
- Robust to visual degradation (e.g., low lighting, smoke and fog)
- Privacy-preserving and non-intrusive for scenarios like smart house

Other RF signals (e.g., WiFi):
- More trustworthy and fine-grained data – radar point cloud

LiDAR:
- Smaller size, lower cost and power consumption
- Unaffected by airborne particles (e.g., rain, snow and smoke)

- **Scene flow as point-wise motion feature**



Research insight:
- Point-wise velocity can facilitate cross-frame movement analysis
- Estimate and use scene flow as an intermediate feature to better support radar-based human motion sensing tasks

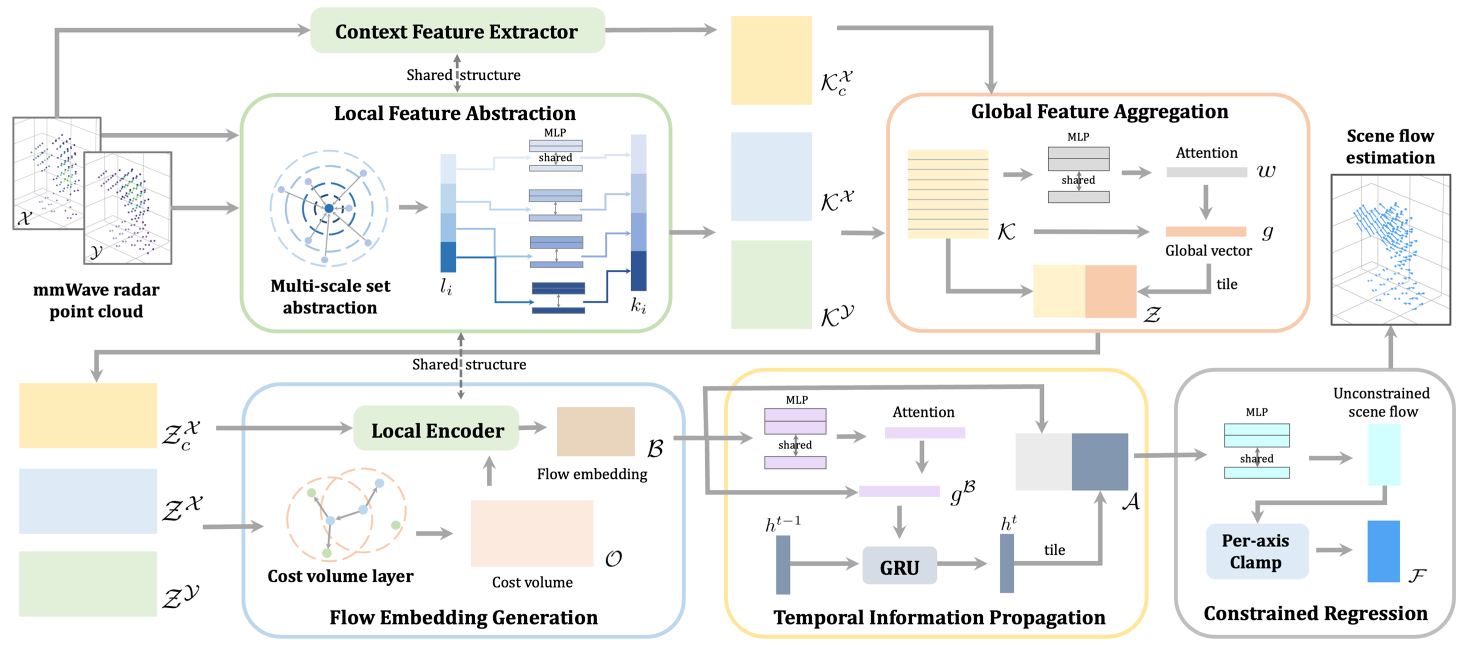
Problem Formulation

- **Goal:** estimate a 3D motion field that describes the scene dynamics
- **Input**: two consecutive 3D point clouds collected by mmWave radar
- **Output:** a set of 3D translation vectors that represent the inter-frame displacement for each point in the first frame

Proposed Method

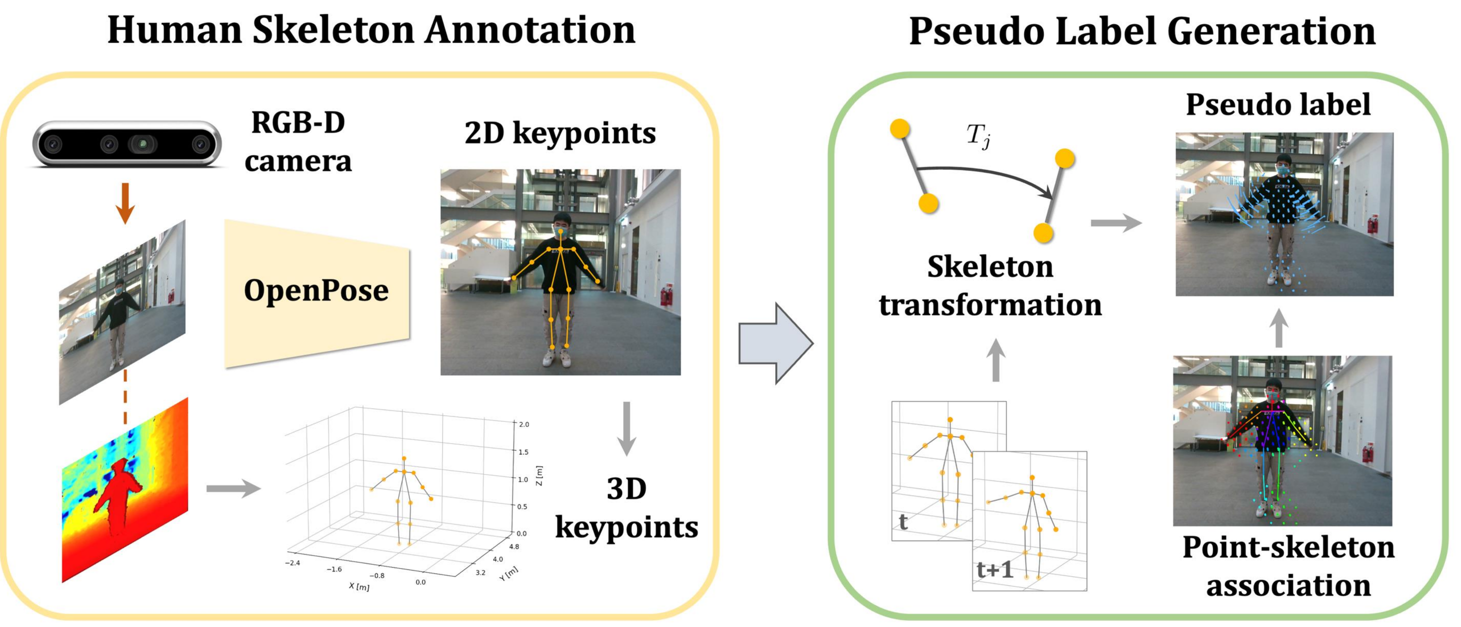
- **Technical challenges**
  - **Sparsity and noise**: ∼100 points per frame, often missing data for specific body parts; multi-path effect, presence of ghost points
  - **Lack of temporal cues**: low-resolution or no Doppler velocity; absence of consistent radar point data across frames
  - **Scene flow annotation**: labor-intensive and expensive; lack real-world correspondence; non-rigid nature of human movement

- **Scene flow network**



- Overcome the sparsity and noise by local-global feature integration
- Address the lack of temporal cues by propagate temporal information

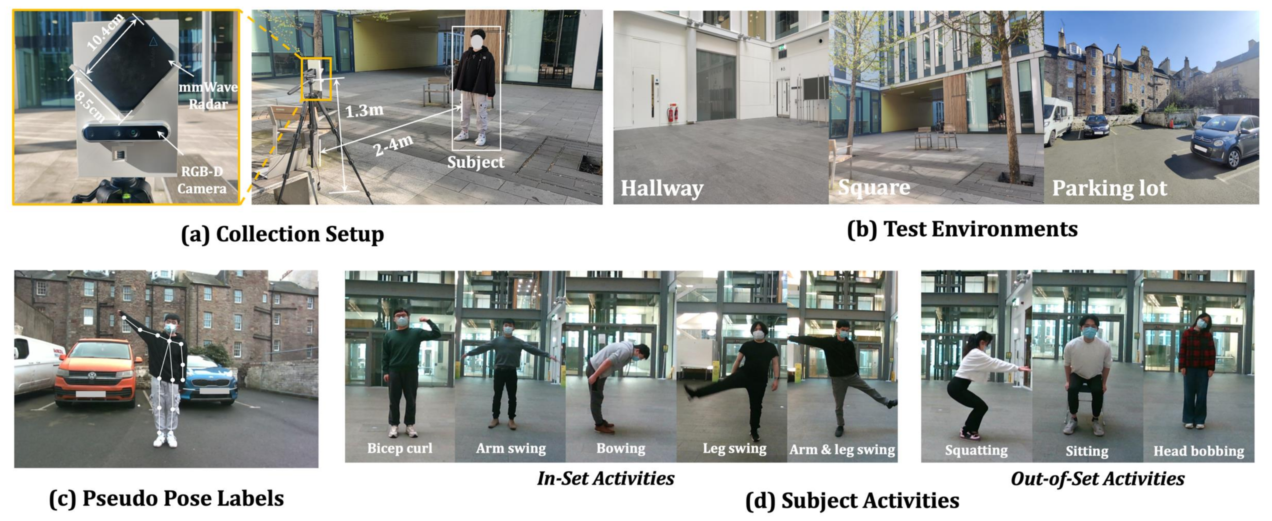
- **Automatic scene flow labelling**



- **Assumption**: The non-rigid human body can be segmented into multiple rigid-motion skeletons, which induce the scene flow in their vicinity.
- Generate pseudo labels from 3D skeletons without any anno. efforts.

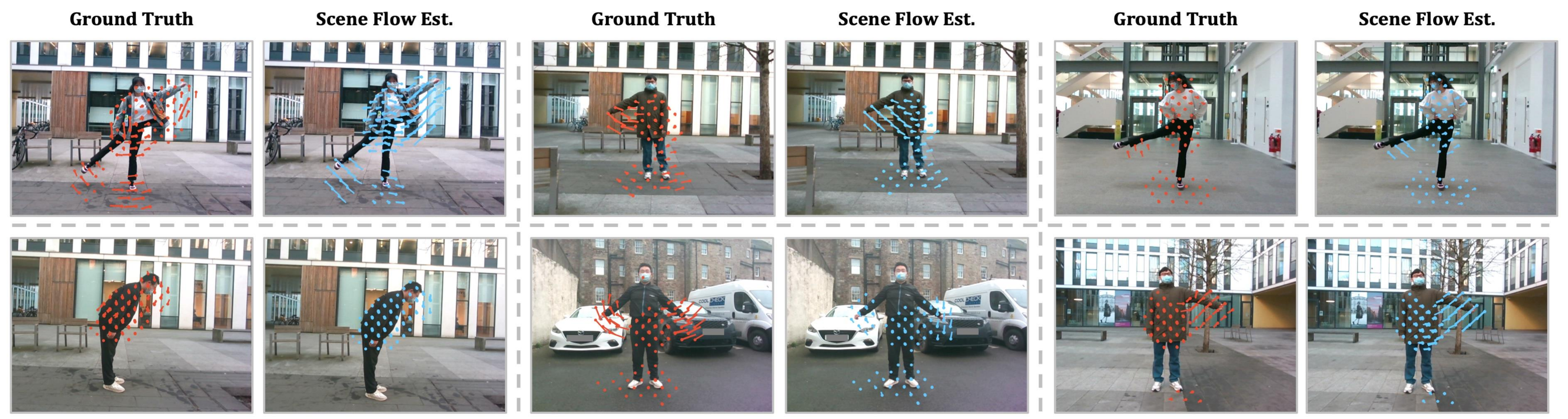
Evaluation

- **Dataset collection**
  - **Vayyar** vTrigB imaging radar (bespoke designed for fine-grained sensing)
  - **Diversity** in subjects, activities, environments.



- **Scene flow results**
  - **cm-level** accuracy though trained with pseudo labels automatically generated
  - **Generalizable** to 'out-of-set' activities

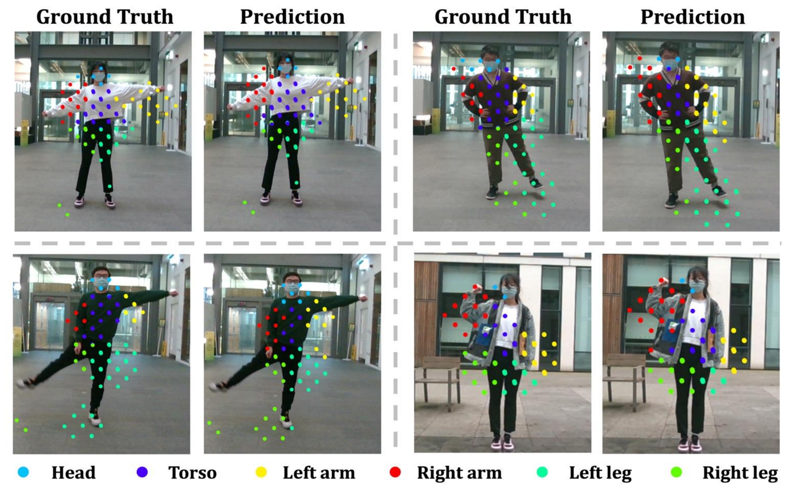
Method	EPE [m] ↓	AccR ↑
RaFlow (sup. version)	0.107	0.427
NSFP (optim. based)	0.197	0.143
Bi-FPNet (ECCV'22)	0.159	0.264
milliFlow (Ours)	**0.046**	**0.703**



- **Downstream task results**

- Human parsing:
  S1 – add point-level scene flow
  S2 – learn and use latent features

Method	Raw	w. S1	w. S2
mIoU (%)	49.09	**52.72**	51.04
oA (%)	65.75	**69.27**	68.21



- Human activity recognition

Method	Raw	w. S1	w. S2
Ours	47.32	**57.88**	57.78
MMPointGNN	52.46	**60.16**	59.94
RadHAR	44.65	49.98	**50.53**
Average	48.14	56.01	**56.08**

- Human body part tracking

Activity	Tracking length – mJE (m) 1	2	3	4
Arm swing	0.028	0.076	0.097	0.124
Leg swing	0.016	0.071	0.105	0.130
Arm & leg	0.030	0.108	0.146	0.178