Human Activity Sensing from Low-rate Samples under Integrated Networking

Moh Sabbir Saadat; Sanjib Sur
Computer Science and Engineerng, University of South Carolina, Columbia, SC, USA
msaadat@email.sc.edu; sur@cse.sc.edu

***Abstract*—Indoor sensing of human activity with wireless signal has long been identified as a potential tool for privacy-aware continuous health diagnosis in all lighting conditions. The introduction of high-frequency and large bandwidth millimeter-wave (mmWave) signal has given more fuel than ever to this idea. However, there are limited efforts to truly integrate sensing on networking systems. In this work, we explore an opportunistic activity sensing scheme within a networking system.**

***Index Terms*—mmWave, Sensing, PCD, Graph Neural Networks**

I. Introduction

An effective approach to continuous health diagnosis, especially for elderly patients is indoor sensing of activities. A wide plethora of fine-grained applications come under this activity sensing umbrella: activity recognition, abnormal gait detection, posture estimation *etc.* Camera-based monitoring systems, while being highly accurate, are a bad choice for such indoor sensing applications. *Firstly,* cameras invade privacy at home since they capture clearly identifiable information. *Secondly,* cameras fail in low-light or dark conditions such as during the night. Towards this end, using wireless radio signals for human activity sensing has received significant attention, especially with short wavelength and large bandwidth mmWave signal [1]–[4]. Besides enabling privacy-aware indoor sensing in all lighting conditions, reusing mmWave signal from networking systems removes the requirement of equipping a home environment with additional hardware. This is also significant for preserving an ambient living condition.

The existing efforts in integrating sensing on networking systems either focus on novel beamforming designs [4] which would require major modifications to networking protocols, or propose reusing training beams [2]. For the latter, the authors in [2] propose modulating the beam scanning frequency based on the sensing requirements. This time-duplexing sensing and networking presents us with a tradeoff in sensing accuracy and networking throughput. To illustrate this, we simulated an IEEE 802.11ad network with periodic switch-off at every 200 ms to enable sensing for durations of 40 ms. Figure 1 (a) shows that this frequent switch to sensing costs the network an average throughput of 250 Mbps, increases the throughput standard deviation to 350 Mbps as well as introducing 40 ms latency. On the other hand, figure 1 (b) shows that even

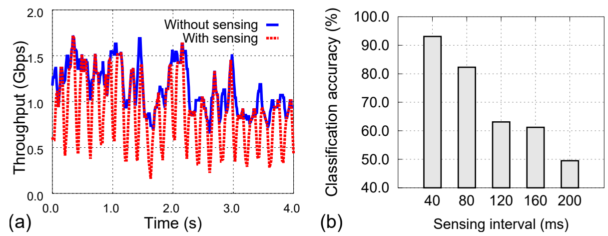

Fig. 1: *Sensing-networking tradeoff: (a) Introducing sensing reduces the networking throughput and increases the latency. Sensing interval: 200 ms (b) Increasing sensing interval degrades sensing accuracy.*

at 200 ms sensing interval, a classifier for human activity [5] (with 18 dynamic classes) achieves only 50% accuracy. At lower sensing intervals, while sensing accuracy improves, networking throughput will suffer more. To get around this tradeoff, [3] proposes reusing the networking packets instead. However, mmWave networking employs directional beams, and varies the beamwidth to tackle mobility and throughput tradeoff, whereas, activity sensing requires a broader beam to capture reflections from all parts of the subject's body. Moreover, the networking client and sensing subject may be distinct entities in which case reusing the networking beam is not a solution.

We propose *milliNetS* to break this networking-sensing tradeoff by accomplishing sensing from low-rate sensing samples. The primary intuition in *milliNetS* is that an integrated networking-sensing system can still leverage the opportunistic idle times in networking when the throughput requirement drops so that the beam can be switched and beamwidth changed to acquire sensing samples. Next, a learnable *temporal prediction* network upsamples the the resulting irregular, low-rate sample sequence in time to estimate a high-rate, regular sample sequence. Such a high-rate sample sequence improves recognition of human activity.

II. *MilliNetS* Design

Firstly, milliNetS generates a 3D representation of the key reflectors in the scene by converting the acquired sample sequence into Point Cloud Data (PCD). *Secondly, milliNetS* proposes a *temporal prediction* model based on Dynamic Graph Convolutional Neural Network (DGCNN) [6] to estimate the missing samples and reconstruct a high-rate sample

This work is partially supported by the NSF under grants CNS-1910853, CAREER-2144505, and MRI-2018966.

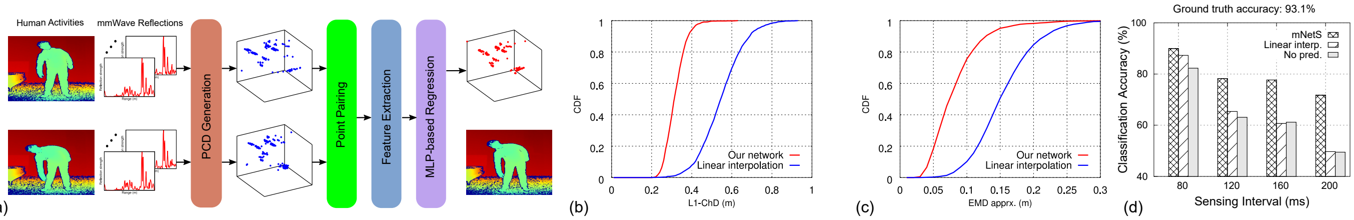

Fig. 2: (a) *milliNetS* design overview; Performance of temporal prediction in terms of: (b) L1-ChD, and (c) EMD; (d) Reconstructed high-rate sample sequence improves activity recognition

sequence. Figure 2 (a) shows an overview of the system.

PCD Generation: To sense the strong reflections from the human body and resolve them in the 3D spatial domain, a large bandwidth signal is transmitted from multiple channels, and the received signal is combined with a copy of the transmitted signal. This combined signal is then processed through a series of radar signal processing to generate a *Range-Doppler-Azimuth-Elevation* heatmap. A naive thresholding on this heatmap does not suffice to acquire the strong reflections from human activity. This is due to two reasons: *Firstly,* the heatmap must be filtered to remove the strong reflections from the static objects in the surroundings. This is done by zeroing out the heatmap elements corresponding to Doppler velocities near zero. *Secondly,* the heatmap is still very noisy, and naively picking out local peaks will generate too few points to represent the human activity. Thus, *milliNetS* employs the Constant False Alarm Rate (CFAR), a well-established adaptive algorithm used in radar detection. Furthermore, besides the three spatial dimensions, the dimensionality of the PCDs are extended by adding the signal intensity and the Doppler channels.

Temporal prediction: *milliNetS* reconstructs high-rate sample sequence by estimating missing samples from nearest ground truth samples. Let's assume a sample with PCD representation P_i was acquired at time t_i and the next sample, P_j was acquired at time t_j. *milliNetS* proposes that, to estimate all missing samples at all times t_k where $t_i < t_k < t_j$, high dimensional feature maps, F_i^m and F_j^m (m is the dimensionality of the feature maps) of the two PCDs are combined by weighted addition based on their respective temporal nearness to P_k:

$$\hat{F}_{k,m} = w_i \cdot F_{i,m} + w_j \cdot F_{j,m} \tag{1}$$

$$where, w_i = \frac{t_j - t_k}{t_j - t_i}; \quad w_j = \frac{t_k - t_i}{t_j - t_i} \tag{2}$$

To ensure that this weighted combination combines feature maps of nearest point pairs between P_i and P_j, a point pairing operation is carried out before feature mapping. The estimated high-dimensional feature map, $\hat{F}_k^m$ of P_k is then passed through a stack of MLP-based decoder to obtain $\hat{P}_k$ as an estimation of P_k. This is in contrast to the temporal interpolation of PCD sequences in computer vision where an estimation of pointwise trajectory mapping is done. Such pointwise trajectory mapping from one frame to the next is not suitable for mmWave PCD sequences due to the highly noisy nature of mmWave reflections due to specular and variable reflectivity, resulting in unstable point sets in successive PCD samples. Moreover, the feature maps of the PCDs are obtained through a stack of customized DGCNN layers [6]. This handles the structural irregularity and order invariance of PCDs which traditional CNN layers cannot.

Loss function: The loss function which trains the *temporal prediction* network must be able to quantifiably distinguish two PCD samples based on pointwise displacement, measured by L1 Chamfer distance (L1-ChD) as well as overall structural similarity, measured by Earth Mover's Distance (EMD). This is because the ground truth PCD samples from mmWave reflections are highly noisy, and thus, the network must learn to map the global geometric structure as well.

$$L = \lambda_C \cdot L_{ChD} + \lambda_E \cdot L_{EMD} \tag{3}$$

III. Activity Recognition and Future Work

milliNetS shows an improvement of 24 cm and 7 cm in terms of L1-ChD and EMD over linear trajectory interpolation (2 [b-c]). To validate the reconstructed high-rate sample sequences' (sample at every 40 ms) ability to improve activity recognition, we use the resulting high-rate samples in an activity classification network, designed to recognize seven classes of activity. The network applies an LSTM layer on the feature maps of the PCD sequences. Figure 2 (d) shows that PCD sequences estimated from *milliNetS* improve activity recognition over trajectory-based linear interpolation from 49.5% to 72% at ground truth sampling every 200 ms.

The next step in human activity sensing is acquiring the sequence of 3D posture. The posture sequence of a person is a strong marker of physical health, and with such information, an AI agent can perform fine-grained health diagnosis such as detecting early symptoms of age-related diseases.

References

[1] Arindam Sengupta, et. al., "mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns," *IEEE Sensors Journal*, 2020.

[2] Jacopo Pegoraro, et al., "RAPID: Retrofitting IEEE 802.11ay Access Points for Indoor Human Detection and Sensing," 2022. [Online]. Available: https://arxiv.org/abs/2109.04819

[3] ——, "SPARCS: A Sparse Recovery Approach for Integrated Communication and Human Sensing in mmWave Systems," in *IEEE IPSN*, 2022.

[4] Fan Liu, et al., "Toward Dual-functional Radar-Communication Systems: Optimal Waveform Design," *IEEE Transactions on Signal Processing*, 2018.

[5] Edward M Sitar, et al., "A Millimeter-Wave Wireless Sensing Approach for at-Home Exercise Recognition," in *ACM MobiSys*, 2022.

[6] Yue Wang, et al., "Dynamic Graph CNN for Learning on Point Clouds," *ACM Transactions on Graphics (TOG)*, 2019.