Activity Recognition from Doppler Signatures via mmWave MIMO Radar

Alessandro Vaccari*, Carlos A. Gómez-Vega*, Gianluca Torsoli*, Maison Clouâtré†, Moe Z. Win†, and Andrea Conti*
*DE and CNIT, University of Ferrara, Italy (e-mail: a.conti@ieee.org)
†LIDS, Massachusetts Institute of Technology, USA

Abstract

Activity recognition based on sensing the reflected signals will enable novel services in next generation wireless networks. This paper proposes an approach to achieve high-accuracy activity recognition in complex wireless environments using time-frequency analysis of radar Doppler signatures. The proposed approach is validated via experimentation employing a multiple-input–multiple-output radar at millimeter waves.

I. Introduction

Sensing will be crucial in next generation wireless networks to enable novel services that rely on seamless context awareness. The 3rd Generation Partnership Project (3GPP), since Release 19, has been proposing use cases requiring sensing, and in particular, activity recognition via reflected radiofrequency (RF) signals for smart monitoring and fragile assistance [1]. However, recognizing activities performed by device-free targets based on samples of reflected RF signals is challenging, especially in complex wireless environments [2], [3].

The availability of multiple-input–multiple-output (MIMO) radars at millimeter waves (mmWaves) enables the collection of accurate sensing analytics via reflected RF signals, including the Doppler shifts generated by the dynamics and micro-movements of the target. A set of Doppler shifts collected over time contains representative information of the target activity, i.e., the activity's Doppler signature. Neural networks (NNs) promise to achieve accurate activity recognition based on Doppler signatures. NN-based methods automatically adjust the learning model parameters by minimizing a cost function that quantifies the activity recognition accuracy. This enables the processing of Doppler shifts to identify their most significant features, without requiring a mathematical description of the wireless propagation.

Doppler features are typically obtained by processing received waveform samples via the discrete Fourier transform (DFT). However, temporal correlations are crucial for activity recognition calling for time-frequency processing. The continuous wavelet transform (CWT) allows joint extraction of time and frequency information, providing an adaptable resolution according to the Heisenberg uncertainty principle.

This paper proposes a novel approach to perform activity recognition via reflected RF signals based on time-frequency analysis of Doppler signatures and a convolutional neural network (CNN). The proposed approach is validated through experimentation by employing a frequency modulated continuous wave (FMCW) MIMO radar at mmWaves.

II. Methodology

Recall that an FMCW MIMO radar measurement entails transmitting N_{c} linear chirps and then collecting their reflected echoes using multiple receiving antennas. At the receiver of an FMCW radar, the transmitted and received waveforms are multiplied using a mixer and their product is filtered using a low-pass filter (LPF). The resulting signal is referred to as intermediate frequency (IF) signal and contains information about the difference of instantaneous frequency between the transmitted and received waveforms. The IF signal corresponding to the transmission of the nth chirp is given by

$$s_{\mathrm{IF}}^{(n)}(t) = \frac{A_{\mathrm{R}}}{2} \cos\Big(2\pi(f_{\mathrm{p}}t' + f_{\mathrm{D}}T_{\mathrm{c}})\Big) \tag{1}$$

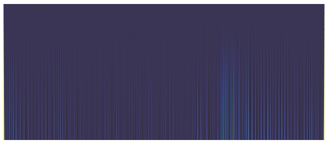
where A_{R} denotes the signal amplitude, $f_{\mathrm{p}} = 2Br/(cT_{\mathrm{c}})$ is the IF signal frequency, $t' = t - nT_{\mathrm{c}}$, $f_{\mathrm{D}} = 2f_0 v/c$ is the Doppler frequency, and T_{c} is the chirp duration. In particular, B is the chirp bandwidth, r and v denote the range and velocity of the scatterer, and c is the propagation speed. Each IF signal is sampled with period T_{s} in N_{s} samples, i.e., fast time (FT) sampling, while the entire signal sequence with period T_{c} in N_{c} samples, i.e., slow time (ST) sampling. Since a single chirp signal is received by multiple antennas, the collected IF signal samples are arranged in a 3-dimensional tensor, which is referred to as radar data cube.

The FT and angular dimension, processed by the DFT, provide both range and angle-of-arrival (AOA) measurements. The ST dimension contains the time-domain representation of the Doppler shifts collected at each considered range-angle coordinates. A location of interest can be monitored over time by performing K measurements. The ST samples collected during the measurement time frame are arranged in a vector $\boldsymbol{x} = [x_1, x_2, \dots, x_{N_{\mathrm{c}}K}]^{\mathrm{T}}$, denoting the time domain representation of the Doppler signature generated by the activity.

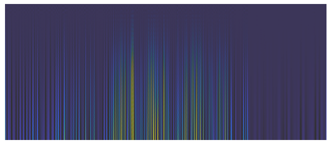
By applying the CWT to $\boldsymbol{x}$, we obtain a matrix $\boldsymbol{W}$ representing its spectrogram. Let $u \in \mathcal{U} = \{u_1, u_2, \dots, u_N\}$ and $s \in \mathcal{S} = \{s_1, s_2, \dots, s_M\}$ be the shift and scale to apply to the mother wavelet. These sets of values allow exploring different frequencies at different time instants. The element in row n and column m of $\boldsymbol{W}$ is given by

$$[\boldsymbol{W}]_{n,m} = \sum_{k=1}^{N_{\mathrm{c}}K} x_k \, \frac{1}{\sqrt{s_m}} \, \psi^*\bigg(\frac{kT_{\mathrm{c}} - u_n}{s_m}\bigg) \tag{2}$$

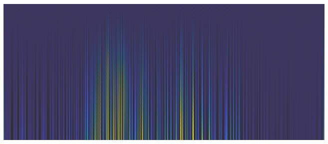
where $\psi^*(\cdot)$ is the complex conjugate of the mother wavelet $\psi(\cdot)$. In particular, each element of $\boldsymbol{W}$ represents the magni-



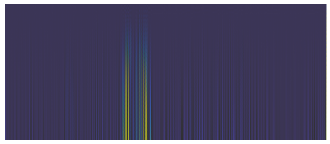
(c) TA



(a) TP



(d) AM



(b) TF

Fig. 1: Examples of time-frequency images of the considered actions obtained by processing Doppler signatures via CWT.

(a) AM

(b) TF

Fig. 2: Examples of recognized activities performed by a human target in the experimentation environment.

tude of the correlation between the signal $\boldsymbol{x}$ and the scaled and translated wavelet function $\psi(\cdot)$.

The spectrogram $\boldsymbol{W}$ is the time-frequency representation of an activity performed during the measurement time frame and is used as input of a CNN to perform activity recognition.

III. Experimentation

The proposed approach is validated at the University of Ferrara Department of Engineering in an indoor environment characterized by several scatterers. For the experimentation, we employ an off-the-shelf FMCW MIMO radar equipped with 2 transmitting and 16 receiving antennas organized in a linear array. A measurement is repeated every $100\,\mathrm{ms}$ and consists of 128 chirp signals in the frequency range $76-77\,\mathrm{GHz}$ with a chirp duration $T_{\mathrm{c}} = 120\,\mu\mathrm{s}$. The employed mother wavelet is the Mexican Hat and the scale set is $\mathcal{S} = \{2^{j/v}, \forall\, v = 1, 2, \dots, 256\}$ with $j \in \{0.1, 1, 2, 3, 4\}$. The set $\mathcal{U}$ is defined according to Heisenberg uncertainty principle.

The activities to recognize are *target absence* (TA), *target presence* (TP), *arm movement* (AM), and *target falling* (TF) Such activities are performed in a fixed location at a range $r = 1.5\,\mathrm{m}$ perpendicular to the antenna array. Fig. 1 shows an example of spectrograms obtained via CWT. Fig. 2 shows a human target moving an arm and falling. For every activity, we performed 100 experiments with a duration of $3\,\mathrm{s}$ each.

The implemented CNN architecture is composed of 7 layers: (i) a 2D convolutional input layer with 32 filters and kernel size of 32 elements; (ii) a max pooling layer with pool size of 8; (iii) a 2D convolutional layer with 16 filters and kernel

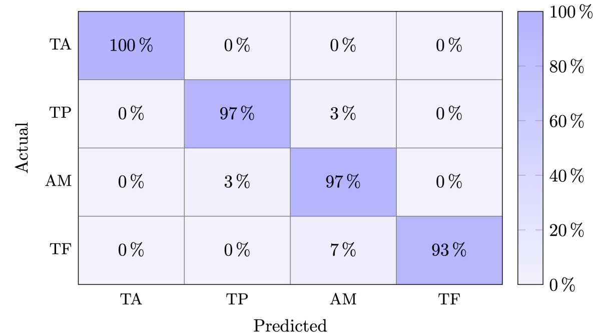


Fig. 3: Confusion matrix for the activity recognition accuracy.

size of 32; (iv) a max pooling layer with pool size of 2; (v) a dropout layer with 20% of dropping rate; (vi) a fully connected layer with 16 units; and (vii) an output layer. The inner layers of the CNN employ ReLU activation functions, while the output layer employs the softmax function. The training phase is performed employing the categorical cross-entropy as the empirical loss function and the Adam optimizer with a learning rate of 10^{-3}. The experimental data is divided in 70% for training and 30% for testing.

Fig. 3 shows the confusion matrix obtained performing activity recognition with the proposed approach. The average activity recognition accuracy achieved via experimentation is 97%. In particular, TA is identified with 100% accuracy. TP and AM are identified with 97% accuracy, while TF is recognized with 93% accuracy. In 7% of the occasions, TF was confused with AM, since an arm movement is also observed as a reaction during the fall.

IV. Conclusion

This paper presents a novel approach to perform activity recognition via reflected RF signals. The proposed approach employs a CNN to classify actions based on a Doppler signature spectrogram obtained using the CWT. The approach is validated via experimentation employing an off-the-shelf MIMO radar at mmWaves, showing an accuracy of 97%. The proposed approach for activity recognition enables the design and deployment of innovative services based on sensing for next generation wireless networks.

Acknowledgment

The fundamental research described in this paper was supported, in part, by the Office of Naval Research under Grant N62909-22-1-2009, by the National Science Foundation under Grant CNS-2148251, and by federal agency and industry partners in the RINGS program.

References

[1] *Technical Specification Group Services and System Aspects; Feasibility Study on Integrated Sensing and Communication*, 3rd Generation Partnership Project 3GPP™ TR 22.837 V19.0.0, Jun. 2023, Release 19.

[2] S. Bartoletti, Z. Liu, M. Z. Win, and A. Conti, “Device-free localization of multiple targets in cluttered environments,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 5, pp. 3906–3923, Oct. 2022.

[3] M. Chiani, A. Giorgetti, and E. Paolini, “Sensor radar for object tracking,” *Proc. IEEE*, vol. 106, no. 6, pp. 1022–1041, Jun. 2018.