Continuous Human Activity Classification with Radar Point Clouds and Point Transformer Networks

Kruse, Nicolas; Fioranelli, Francesco; Yarovoy, Alexander

DOI
10.23919/EuRAD58043.2023.10289503

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the 2023 20th European Radar Conference (EuRAD)

Continuous Human Activity Classification with Radar Point Clouds and Point Transformer Networks

Nicolas C. Kruse[1], Francesco Fioranelli[2], Alexander Yarovoy[3]
MS3 Group, Delft University of Technology, the Netherlands
{[1]n.c.kruse, [2]f.fioranelli, [3]a.yarovoy}@tudelft.nl

***Abstract*— Due to numerous benefits, radar is considered as an important sensor for human activity classification. The problem of classifying continuous sequences of activities of unconstrained duration has been studied in this work. To tackle this challenge, a radar data processing method utilizing point transformer networks has been proposed. The method has been experimentally verified on a dataset of human activities, and experiments have been performed to determine its optimal implementation. Promising preliminary results on a 9-class dataset show test accuracy and macro F-1 scores in the range of 83% and 73% respectively.**

***Keywords*— Human activity recognition, machine learning, radar, point cloud processing.**

I. Introduction

Monitoring Activities of Daily Living (ADL) can provide medical professionals with important insights on a patient's well-being and can for example be utilized to perform fall detection in assisted living scenarios. Radar sensors are being considered for this task due to their functioning in darkness, their non-contact nature, and the benefit of privacy preservation they offer over e.g. cameras. Additionally, limited through-wall capabilities have been demonstrated using radar sensors [1]. Besides fall detection [2], [3], other applications include gait analysis [4], [5], vital sign monitoring [6], [7], [8], and activity monitoring [9], [10], [11]. A current challenge in activity classification for radar is the processing of sequences of continuous activities of unconstrained duration.

Transformer networks [12] are deep learning architectures that have booked substantial successes in the fields of natural language processing [13] and computer vision [14], where continuous data sequences are processed. At the core of these networks lays the so-called *attention mechanism*, which allows for correlations between input elements to influence the network weights dynamically. Transformer networks have recently received attention as a tool to improve radar-based activity classification [15], [16]. One of potential radar data representations for a transformer network is the point cloud (PC), and several adaptations of the transformer network for point cloud processing (Point Transformers, PT) have been proposed [17], [18], [19].

In this work, a method is proposed to process continuous sequences of human activities into point cloud representations suitable for classification with a PT network. Typically, radar point cloud data consists of radar returns in a 3-dimensional spatial representation. The sensors used in this work however can provide only 1-dimensional range information on the subject and thus an alternate approach is formulated. The proposed method is evaluated on an experimentally collected dataset of human activities. The main contributions of this work are as follows.

- A novel approach for continuous activity classification using PT networks;
- A method for transforming radar data with a singular range dimension to a point cloud representation suitable for a point transformer;
- A performance evaluation of the methodology on an experimental dataset of human activities.

The rest of this paper is structured as follows: In section II the experimental setup, as well as the sequences of human activities are described. Section III contains the data processing methodology as well as an introduction to the PT network. Results of the experimental validation of the method are discussed in section IV, with a conclusion in section V.

II. Experimental Setup

The dataset used for this work is publicly available [20]. The measurement setup consists of a set of five Humatics PulsON P410 pulsed UWB SISO radars arranged in a semicircle with diameter of 6.38 m, regularly spaced at 45° intervals. The measurement area is a concentric circle of diameter 4.38 m. Images of the measurement setup are available in e.g. [21]. All nodes operate at a center frequency of 4.3 GHz, a bandwidth of 2.2 GHz, and a PRF of 122 Hz, resulting in a maximum unambiguous velocity of $\pm 2.2\ \mathrm{m\,s^{-1}}$.

The data consist of labeled sequences of human activities of unconstrained duration, performed in arbitrary directions within the measurement area. The multi-node setup allows for the simultaneous capture of each activity from five different aspect angles. Nine activity classes are considered in the dataset [20], including: walking, standing still, sitting and bending, falling and standing up. 30 sequences of 2 minutes duration are captured for 14 participants. Of these, two sequences contain all nine activities, and the rest all contain various subsets of the nine activities. For a full description of the sequences and a list of the nine activities, see [2], [20].

The output of each node is a real vector which undergoes a Hilbert transform in order to obtain a complex-valued IQ vector, which is subsequently reshaped into a matrix with short-time and long-time dimensions. This matrix is used as the initial raw data for further processing in section III-A and corresponds to a range-time map if its magnitude is taken.

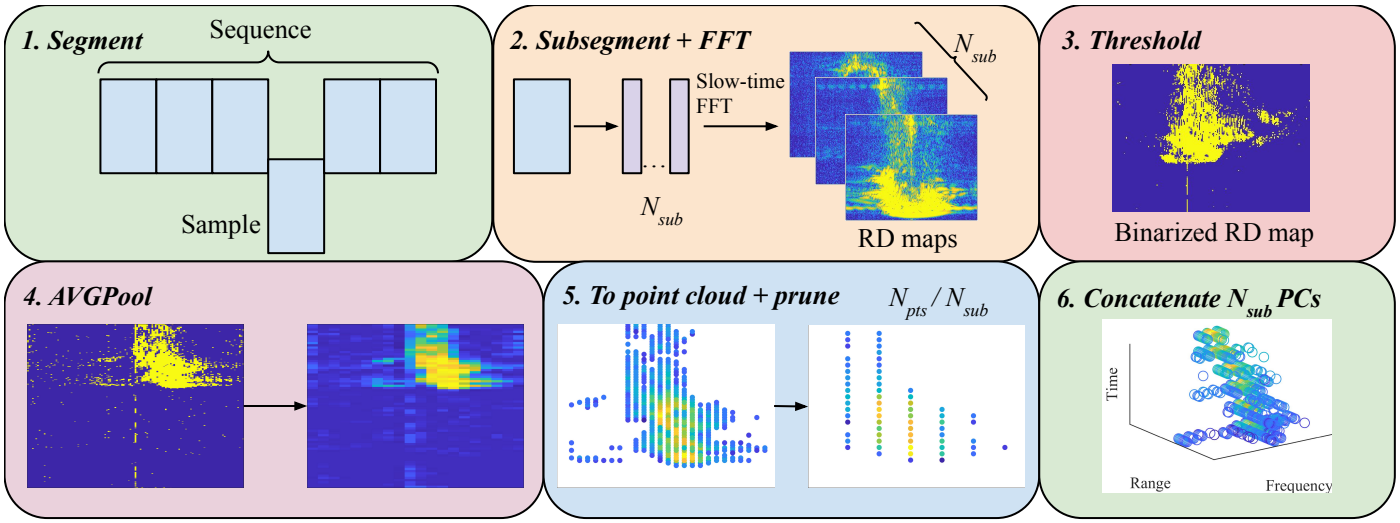

Fig. 1. Proposed data processing pipeline. A raw input sequence is segmented (1) in discrete samples. Each sample is subsegmented (2) into N_{sub} sections, on which FFTs over slow-time are performed to obtain N_{sub} Range-Doppler RD maps. A static threshold is applied (3) to each map to obtain binarized RD maps. The RD maps are downsampled through an average pooling operation (4). The downsampled map is converted to a list of points, and low-intensity points are pruned (5) until the required amount of points for the PT network is achieved. Finally, the N_{sub} subsegments are concatenated (6) in the time dimension, resulting in a full point cloud for the original sample, consisting of N_{pts} points with four features of range, Doppler, time, and intensity.

III. Proposed Method

A. Data Processing

The proposed approach to generate samples for the point transformer classifier starting from complete raw data sequences is outlined in Figure 1. First, an input sequence is split into fixed-duration segments. Each segment constitutes a *single* sample for the classifier and is thus assigned a single activity label. In the case of multiple activity labels occurring in the segment, the majority label is selected. Each segment is further subdivided into N_{sub} time-ordered subsegments in order to introduce a time evolution component within the sample. Taking the FFT along the slow time dimension of all subsegments results in N_{sub} Range-Doppler (RD) maps representing the time evolution of the sample in terms of range and Doppler. An example RD map is given in Figure 2.

Initial noise rejection is performed by means of a static threshold, set as a fraction of the maximum intensity in dB scale of the RD map as:

$$R_{i,j}^{bin} = \begin{cases} 1, & \text{if } R_{i,j} > \alpha \max\limits_{i,j} R_{i,j} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Here, α indicates the threshold level between $[0, 1]$, $R_{i,j}$ denotes an element of the RD map in dB scale, and $R_{i,j}^{bin}$ an element of the resultant binarized RD map.

Given the resolution in range and Doppler for a typical binarized RD map from the chosen dataset, converting the parts of the map exceeding the threshold level directly to a point cloud results in too many points for the classifier. For this reason, an average pooling operation is first applied to the binarized RD map. The size of the filter is determined by first computing the ratio η between the amount of pixels exceeding the threshold and the final desired amount of points for the given subsegment:

$$\eta = \frac{\sum_{i,j} R_{i,j}^{bin}}{\frac{N_{pts}}{N_{sub}}}, \tag{2}$$

where N_{pts} indicates the required amount of points for the full sample, N_{sub} denotes the number of subsegments, and the sum is in effect counting the nonzero elements. The filter size is subsequently set to η, rounded down to the nearest integer.

At this stage, the remaining pixels are stored as a list of points and sorted by their respective intensity. The list is subsequently pruned until the desired amount of points is reached. Once all N_{sub} RD maps have been processed, the lists of points are joined, resulting in a point cloud for the original sample with four features, namely range, Doppler, intensity, and time.

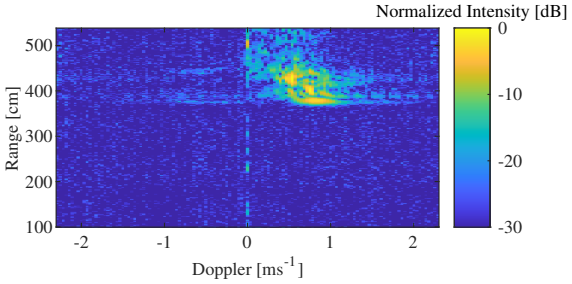

Fig. 2. Example of Range Doppler map for a walking motion, computed over a one second interval and with normalized intensity.

B. Point Transformer Network

The classification networks used in this work are based on the Point Transformer networks by Zhao et al., 2022 [17], adapted by Guo, 2022 [22]. In the former, the conventional

transformer architecture is adapted for use with point cloud data through the development of a self-attention layer for point cloud data. The architecture consists of four sequential blocks consisting of a self-attention layer and a cardinality reducing layer. The latter works through furthest point sampling, followed by a k Nearest Neighbour (kNN) operation to pool the feature vectors of the initial point cloud onto the remaining points. Following [17], it is set that $k = 16$ and the cardinality is reduced by 4 in each of the blocks.

IV. Results

In this section initial results for the proposed approach are discussed, with the parameters' values reported in Table 1. For all experiments, a sample holdout validation method is utilized, with a 80%:20% training:testing ratio over all samples in the full dataset, i.e. including data from all five radars in [20].

Table 1. Default parameters' values for the results reported in this work.

Parameter	Value	Unit
Sample Length	2.0	[s]
Static Threshold	80	[%]
Points per PC	1024	[-]

Figure 3 displays classification performance versus the amount of points per point cloud, with each point cloud representing a single, two seconds long sample. Important to note is that the Doppler resolution remains fixed throughout. It can be seen that an increase in points initially yields improved classification results, most likely due to the increased amount of information available to the classifier. After 2048 points however, a decrease in performance is seen, most likely attributable to the fact that an increasing amount of noise is included in the point clouds, visible as speckle in the example RD map in Figure 2. In addition to this, the increasing amount of points that lay in the RD map region of interest occupied by the subject has diminishing returns in terms of relevant information content for classification of the activity performed.

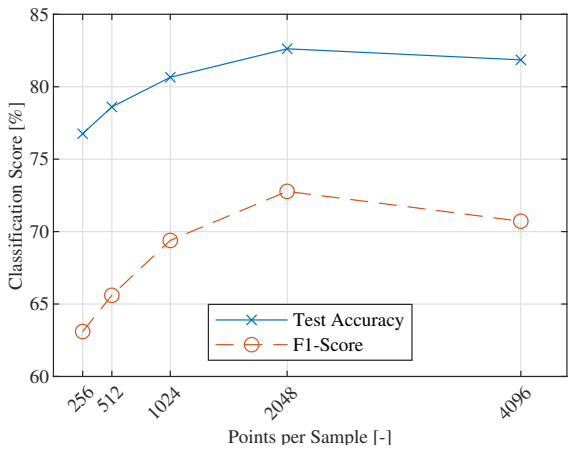

Fig. 3. Classification performance versus the amount of points in a point cloud for a single 2s-long sample. Test accuracy and macro F1-score are acquired through a 20% sample holdout over all sequences in the dataset.

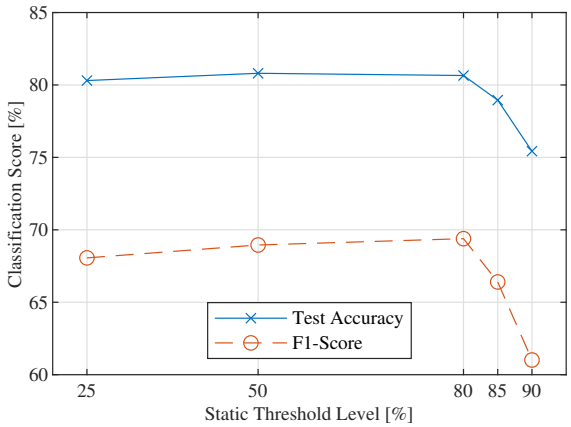

Fig. 4. Classification performance versus the threshold level used to obtain the binarized RD maps, as illustrated in Figure 1. Test accuracy and macro F1-score are acquired through a 20% sample holdout over all sequences in the dataset.

In Figure 4 the results are shown for an analysis of classification performance as a function of the static threshold level α in Eq. (1) that is used to binarize the input RD maps. For threshold levels below 80%, classification performance remains largely unaffected, as point selection is ultimately performed based on point intensity as described in section III-A. Although a lower threshold means that more points originating from noise in the RD maps are eligible for inclusion in the final point cloud, Figure 3 demonstrates that this only becomes problematic for point clouds with a very large amounts of points. Increasing the threshold to values above 80% results in a sharp decrease in classification performance. Visual inspection of binarized RD maps at these threshold values reveals that regions in the RD map originating from the subject are rejected, resulting in sub-optimal retention of information for the activity to be classified.

Figure 5 shows classification performance versus the quantity of subsegments per sample. All samples are two seconds long, as noted in Table 1. A singular subsegment entails a complete loss of temporal information within the sample, leading to degraded classification performance. For higher values, i.e. 8 and 10 subsegments per sample, a decrease in performance is also noted. The cause for this may be the decrease in Doppler resolution corresponding to the shorter observation windows for each subsegment. For the case of two second samples, 8 and 10 subsegments result in Doppler resolutions of 15 $\mathrm{cm\,s^{-1}}$ and 19 $\mathrm{cm\,s^{-1}}$ respectively. Due to the highly dynamic and non-stationary nature of human activities, ideal observation window durations are situational. However, the analysis in Figure 5 reveals that observation windows in the order of 0.33 s maximize classification performance for this dataset.

V. Conclusion

In this work, radar-based classification of continuous sequences of human activities is explored with a point transformer network, using a novel data processing method

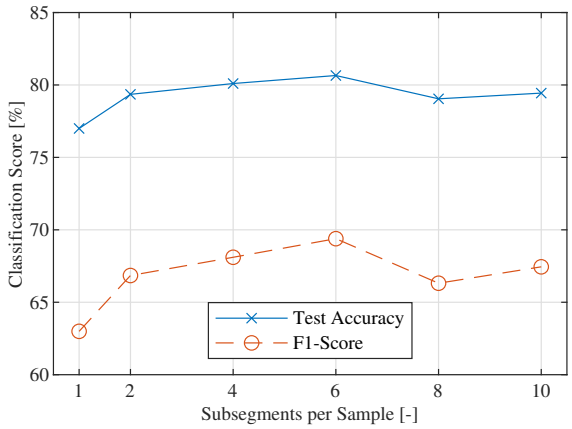

Fig. 5. Classification performance versus the amount of subsegments per each 2s-long sample. Test accuracy and macro F1-score are acquired through a 20% sample holdout over all sequences in the dataset.

for radar point cloud representations. It is shown that, with the proposed pipeline, a test accuracy and macro F-1 score of 83% and 73% respectively can be achieved on a challenging publicly available experimental dataset [20]. Experiments are performed to gauge the classification performance of the method and PT classifier, and to explore the limits and stability of the results for various parameters of the pipeline.

In future work, sensor fusion will be explored to more effectively utilize the multi-aspect measurements that the dataset offers. Additionally, various methods of point selection can be investigated, such as e.g. prioritizing higher Doppler points rather than using intensity. Improving the fixed-window segmentation method currently employed may decrease sample ambiguity and is also expected to improve performance.

Acknowledgment

This research is funded by the Dutch Research Council (NWO) through the project *RAD-ART* (Radar-aware Activity Recognition with Innovative Temporal Networks).

References

[1] X. Wang, Y. Wang, S. Guo, L. Kong, and G. Cui, "Capsule Network With Multiscale Feature Fusion for Hidden Human Activity Classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.

[2] S. Zhu, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Continuous Human Activity Recognition With Distributed Radar Sensor Networks and CNN–RNN Architectures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[3] Y. Yao, C. Liu, H. Zhang, B. Yan, P. Jian, P. Wang, L. Du, X. Chen, B. Han, and Z. Fang, "Fall Detection System Using Millimeter Wave Radar Based on Neural Network and Information Fusion," *IEEE Internet of Things Journal*, pp. 1–1, 2022.

[4] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, and D. Martelli, "Continuous Human Activity Recognition and Step-Time Variability Analysis with FMCW Radar," in *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, sep 2022, pp. 01–04.

[5] S. Hor, N. Poole, and A. Arbabian, "Single-Snapshot Pedestrian Gait Recognition at the Edge : A Deep Learning Approach to High-Resolution mmWave Sensing," in *2022 IEEE Radar Conference (RadarConf22)*. New York, NY, USA: IEEE, 2022, pp. 1–6.

[6] K.-C. Peng, M.-C. Sung, F.-K. Wang, and T.-S. Horng, "Noncontact Vital Sign Sensing Under Nonperiodic Body Movement Using a Novel Frequency-Locked-Loop Radar," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 11, pp. 4762–4773, nov 2021.

[7] G. Su, N. Petrov, and A. Yarovoy, "Dynamic estimation of vital signs with mm-wave fmcw radar," in *2020 17th European Radar Conference (EuRAD)*, 2021, pp. 206–209.

[8] Y. Han, A. Yarovoy, and F. Fioranelli, "An Approach for Sleep Apnea Detection based on Radar Spectrogram Envelopes," in *2021 18th European Radar Conference*, no. April, London, 2022, pp. 17–20.

[9] G. Chutia, S. Biswas, D. A. Palanivel, and S. Gopalakrishnan, "LW-μDCNN: A Lightweight CNN Model for Human Activity Classification using Radar micro-Doppler Signatures," in *2022 IEEE International Symposium on Smart Electronic Systems (iSES)*. IEEE, dec 2022, pp. 73–77.

[10] C. Ding, L. Zhang, H. Chen, H. Hong, X. Zhu, and C. Li, "Human Motion Recognition With Spatial-Temporal-ConvLSTM Network Using Dynamic Range-Doppler Frames Based on Portable FMCW Radar," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 5029–5038, nov 2022.

[11] Y. Zhao, A. Yarovoy, and F. Fioranelli, "Angle-insensitive Human Motion and Posture Recognition Based on 4D imaging Radar and Deep Learning Classifiers," *IEEE Sensors Journal*, pp. 1–1, 2022.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[13] S. Karita, N. Chen, T. Hayashi, T. Hori, H. Inaguma, Z. Jiang, M. Someki, N. E. Y. Soplin, R. Yamamoto, X. Wang, S. Watanabe, T. Yoshimura, and W. Zhang, "A comparative study on transformer vs rnn in speech applications," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2019, pp. 449–456.

[14] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87–110, 2023.

[15] Y. Zhao, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Distributed Radar-based Human Activity Recognition using Vision Transformer and CNNs," in *2021 18th European Radar Conference*, no. April, London, 2022, pp. 301–304.

[16] K. Thipprachak, P. Tangamchit, and S. Lerspalungsanti, "Privacy-Aware Human Activity Classification using a Transformer-based Model," in *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, dec 2022, pp. 528–534.

[17] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point Transformer," Dec 2020.

[18] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, "PCT: Point cloud transformer," *Computational Visual Media*, vol. 7, no. 2, pp. 187–199, apr 2021.

[19] N. Engel, V. Belagiannis, and K. Dietmayer, "Point transformer," *IEEE Access*, vol. 9, pp. 134 826–134 840, 2021.

[20] R. G. Guendel, M. Unterhorst, F. Fioranelli, and A. Yarovoy, "Dataset of continuous human activities performed in arbitrary directions collected with a distributed radar network of five nodes," Nov 2021. [Online]. Available: https://data.4tu.nl/articles/dataset/Dataset_of_continuous_human_activities_performed_in_arbitrary_directions_collected_with_a_distributed_radar_network_of_five_nodes/16691500

[21] R. G. Guendel, M. Unterhorst, E. Gambi, F. Fioranelli, and A. Yarovoy, "Continuous human activity recognition for arbitrary directions with distributed radars," in *2021 IEEE Radar Conference (RadarConf21)*. IEEE, May 2021, pp. 1–6.

[22] Z. Guo, "Point transformer-based human activity recognition using high-dimensional radar point clouds," Master's thesis, Delft University of Technology, 2022. [Online]. Available: https://repository.tudelft.nl/islandora/object/uuid%3Ab3074b25-f0de-49e5-888c-5d8d03606501