Radar-Based Continuous Human Activity Recognition with Multi-Label Classification

Ullmann, Ingrid; Guendel, Ronny G.; Kruse, Nicolas Christian; Fioranelli, Francesco; Yarovoy, Alexander

DOI
10.1109/SENSORS56945.2023.10324957

Publication date
2023

Document Version
Final published version

Published in
2023 IEEE SENSORS, SENSORS 2023 - Conference Proceedings

Citation (APA)
Ullmann, I., Guendel, R. G., Kruse, N. C., Fioranelli, F., & Yarovoy, A. (2023). Radar-Based Continuous Human Activity Recognition with Multi-Label Classification. In *2023 IEEE SENSORS, SENSORS 2023 - Conference Proceedings* (Proceedings of IEEE Sensors). IEEE. https://doi.org/10.1109/SENSORS56945.2023.10324957

Radar-Based Continuous Human Activity Recognition with Multi-Label Classification

Ingrid Ullmann[1], Ronny G. Guendel[2], Nicolas Christian Kruse[2], Francesco Fioranelli[2], Alexander Yarovoy[2]

[1]Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[2]Microwave Sensing, Signals and Systems Group, Delft University of Technology, 2628 CD Delft, The Netherlands

ingrid.ullmann@fau.de

Abstract—**This paper presents a novel approach to radar-based human activity recognition in continuous data streams. To date, most work in this research area has aimed at either classifying every single time step separately by means of recurrent neural networks, or using a two-step procedure of first segmenting the stream into single activities and then classifying the segment. The first approach is restricted to time-dependent data as input; the second approach depends crucially on the segmentation step. To overcome these issues we propose a new approach in which we first segment the stream into windows of fixed length and subsequently classify each segment. Since due to the fixed length, the segment is not restricted to one activity alone, we use a multi-label classification approach, which can account for multiple activities taking place in the same segment by giving multiple outputs. To obtain a higher classification accuracy we fuse several radar data representations, namely range-time, range-Doppler and spectrogram. Using a publicly available dataset, an overall classification accuracy of 95.8% and F1 score of 92.08% could be achieved with the proposed method.**

Keywords—Continuous human activity recognition; radar; multi-label classification; deep learning, ResNet; activities of daily living.

I. Introduction

With an aging society, ambient assisted living will become more and more important. To provide safety for elderly or impaired persons living independently, remote activity surveillance can help to quickly react to dangerous situations such as falls.

There are numerous sensor principles for motion capture. One possible sensor for this task is radar. By exploiting the Doppler effect, radar has a natural capability of measuring motion. Since radar data are difficult to interpret for human eyes there are less concerns about privacy than with optical sensors. In contrast to wearable devices, radar is a completely remote sensor, which is beneficial for mentally impaired people, e.g., people suffering from dementia.

Due to its benefits, radar has gained attention as a sensor for the task and radar-based activity recognition has been investigated in the research community for some time [1-3]. Numerous works have performed classification of radar data by machine learning methods. However, many of these use quite artificial setups with activities taking place one after another at predefined start and stop times. However, in reality, activities will take place continuously and with arbitrary durations. Some recent works have tried to address this problem, e.g., [4-7], see [8] for an overview. To tackle the issue of continuous data streams, two main approaches have emerged [8]. The first is to use neural networks particularly designed for time-dependent signals, e.g., recurrent neural networks (RNN) [4] such as the long-short-term-memory (LSTM) [5]. These networks take sequences as input which are often of temporal nature. Typically, radar spectrograms are used. They represent the Doppler frequency (i.e. velocity of the movement) over time and are well suited for activity classification. However, radar data offer more information, such as range-Doppler plots, which are independent of time but can provide information, which the spectrogram cannot provide. For example, spectrograms are incapable of directly discriminating between translational movements such as walking and in-place movements such as falls. For this reason, using multi-representation inputs for radar-based activity recognition has become more and more popular [7], [9]. However, range-Doppler plots are no sequence and therefore they not suited as input to RNNs.

The second approach to continuous activity recognition is to first recognize the start and stop of one activity within the time stream and then classify the activity in this segment [6-7], [10]. This second approach has the advantage that it can use the well-known classification strategies from single activity classification as well as the multi-representation input strategy. However, this approach depends critically on the separation step; and it will fail when there are several overlapping activities taking place and start and stop are not unique.

To overcome the above-named restrictions we propose a new approach for continuous activity classification: we first separate the stream into segments of fixed length. Therefore, we avoid the crucial separation step from the above-named approach. Since the segments are of fixed length, multiple activities can fall into the corresponding time. Therefore, we propose to use a multi-label classification approach. Multi-label classification is well established in computer vision. It can find several entities in an image and is therefore suited to detect several activities in a time segment. An example of multi-label classification in image processing is shown in Fig. 1. The translation to activity recognition is illustrated in Fig. 2.

2023 IEEE SENSORS | 979-8-3503-0387-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/SENSORS56945.2023.10324957

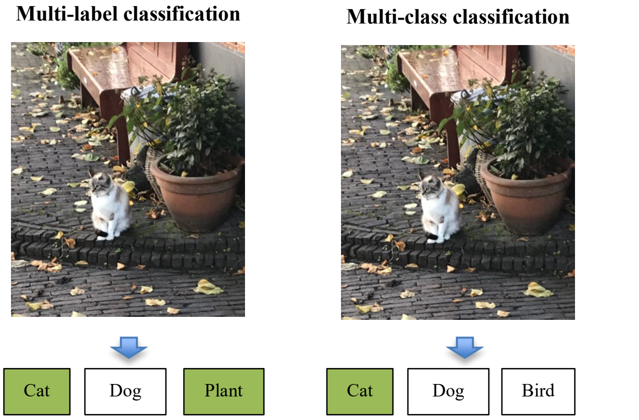

Fig. 1. Illustration of multi-label classification. It can give multiple outputs out of multiple possible labels (left side). In contrast, multi-class classification outputs one label out of multiple possible labels (right side).

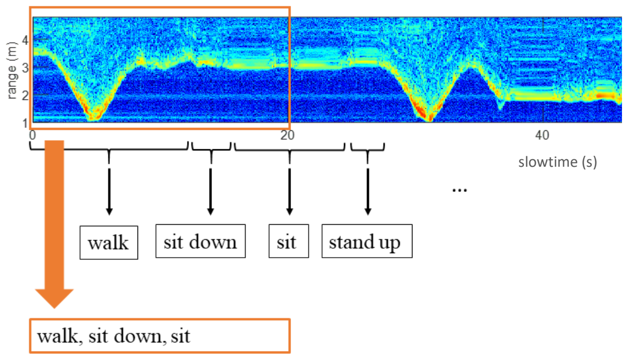

Fig. 2. Multi-label classification for activity recognition in a stream: The classification output are all activities taking place in the selected time.

II. Proposed Approach

A. Employed Dataset

We used the publicly available dataset provided by TU Delft [11]. Fig. 3 shows the measurement setup used to capture the dataset. It consists of five simultaneously operated pulsed radars at 4.3 GHz with 2.2 GHz bandwidth and 122 Hz pulse repetition frequency. The nodes are arranged in a semicircle to capture a large field of view. Details are provided in [12].

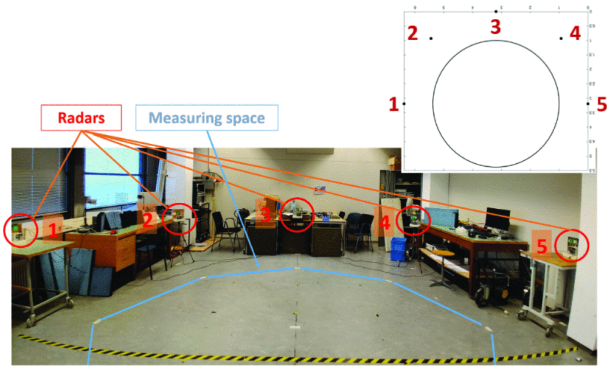

Fig. 3. Photograph and schematic of the measurement setup which was used to capture the dataset we used in this work (from [12]).

The dataset consists of data from 15 subjects performing 9 different activities: Walking, being stationary (no activity), sitting down, standing up from sitting, bending from sitting, bending from standing, falling from walking, standing up, and falling from standing. Additionally there is a null-class with undefined activity.

B. Data Preprocessing

Each data stream from the dataset was split into segments of fixed length along the time axis. Overlapping windows were used to safely capture all activities and avoid misclassification when activities take place at the very beginning or end of the fixed time frame. In our implementation, we used overlapping windows starting every 10 seconds and lasting 30 seconds each. The segmentation process is illustrated in Fig. 4.

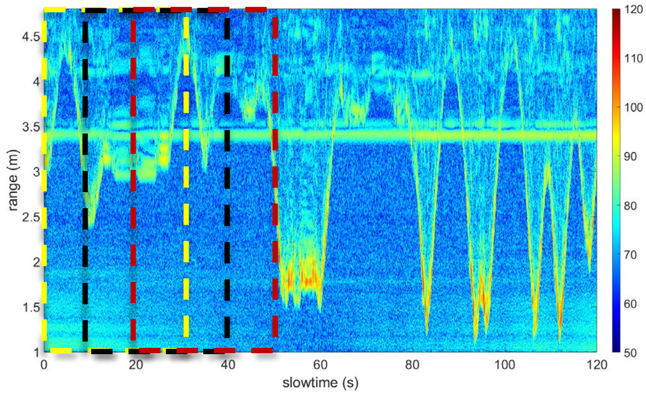

Fig. 4. Illustration of the proposed segmentation process: The data stream is split into windows of 30 s, starting every 10 s. In this example, the first frame is marked by the yellow dashed line, lasting from time $t = 0$ to $t = 30$ s. The second frame (black line) lasts from $t = 10$ s to $t = 40$ s; the third frame (red line) lasts from $t = 20$ s to $t = 50$ s and so on.

We computed three representations of the radar data, namely range over time, range-Doppler, and spectrogram (Doppler over time). The range-Doppler data are computed from the range-time data by means of a Fourier transform along the slowtime dimension. The spectrogram data are obtained from a short-time Fourier transform. In our implementation, we used windowing with a Hann window of 80 bins and 75 % overlap.

The dataset contains sequences of mixed activities as well as sequences with only one activity. Both types were used equally for training and testing. The data of all five radar nodes were treated independently, without any fusion. That way, the network generalizes to various aspect angles of the movements w.r.t. the radar.

C. Deep Learning Network

We used a Residual Neural Network (ResNet) with 50 layers for classification. ResNet is a deep learning network often used in image classification. The network was implemented in MATLAB R2022a.

For each of the three radar data representations an individual network was used for training and classification. Fusing the data was performed in a post processing step which will be explained in detail in Sec. III-A.

The input for each network were the data stream segments of 30 seconds. A total of 13275 training samples was fed to each network. The classifier output of each network comprises the above-named 9 activity classes plus the null-class. Particular of the multi-label classification approach, any one as well as any combination of the 10 classes is possible for each time segment. Therefore, the ground-truth label of each segment is a 1x10 binary vector indicating wheteher the respective activity is present in the selected time frame or not.

Whereas single-label classification typically uses a softmax function, for the multi-label classification task, a sigmoid layer is required, followed by a binary cross-entropy-loss output layer to generate the multi-label output. The multi-label classification output is a 1x10 vector with a probability for each activity class. In order to compare these to the ground truth, a threshold value must be selected. If the probability is equal to or above this threshold, the final output value is set to 1, which means that the activity was found within the segment. If the probability is below the threshold, the binary output is set to 0, which means that the activity is not within the segment. In our implementation we set the threshold to 0.5 to account for the binary decision. We investigated other values as well and it turned out that the range of 0.4 - 0.6 yielded best results.

III. Results

A. Multi-Representation Postprocessing

After training the three networks with 10 persons' data, their performances were evaluated by leave-person-out testing [12], using the remaining 5 persons' data as test data. For each representation, multi-label classification of the test data was performed. The output for each representation is a binary incidence vector as described above. These results can be compared to the ground truth for each representation (range-time, range-Doppler, spectrogram) alone. In this work we additionally fused these representations by averaging the probabilities obtained from the three representations for each class. Comparing the average value to the threshold yields the binary output of the fusion process. This process is illustrated in Fig. 5.

B. Classification Performance

Table I lists the obtained accuracies and micro-averaging F1 scores [13] for the three data representations range vs time, range-Doppler plot and Doppler vs time alone and for the fusion of all three as described above. The fusion outperforms the individual representations. An overall classification accuracy of 95.8 % and an F1 score of 92.08 % were achieved with the fusion approach, which indicates the feasibility of the proposed concept.

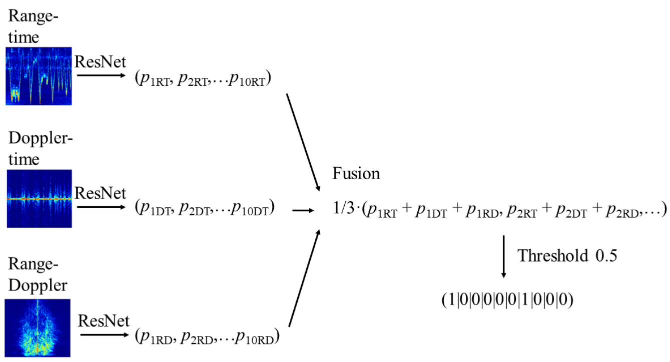

Fig. 5. Illustration of multi-representation fusion in this work: All three representations are classified individually, the output of the classifier being a probability vector for the activities. Fusion is performed by averaging the representations' probabilities for each activity. A binary output for the presence / absence of every activity is obtained from thresholding. Here, as an example, the final output is activities 1 and 7.

TABLE I. Classification Accuracies and F1 Scores

	Range-time	Doppler-time	Range-Doppler	**Fusion**
Accuracy	95.25 %	86.49 %	93.79 %	**95.80 %**
F1 score	91.38 %	74.31 %	88.26 %	**92.08 %**

IV. Conclusion

In this work a novel approach to radar-based human activity recognition within continuous time streams was introduced. We proposed a two-step procedure of first segmenting the stream into windows of fixed length and then classifying the segments by multi-label classification. Experimental results with a publicly available dataset demonstrated the feasibility of the approach with accuracy of 95.8 % obtained.

To investigate the proposed concept further, parameter studies w.r.t. various durations and overlaps of the segments will be investigated in the future, as well as different fusion approaches and thresholds.

Furthermore, our approach has the potential to unlock novel opportunities: By utilizing the suggested segmentation and multi-label classification, it becomes feasible to classify concurrent activities involving multiple individuals. Advancing the understanding of multi-person scenarios will be a next step towards practical application of radar-based activity sensing.

Acknowledgment

This work was partly funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1483 – Project-ID 442419336, EmpkinS.

REFERENCES

[1] S. Z. Gurbuz and M. G. Amin, "Radar-Based Human-Motion Recognition With Deep Learning: Promising Applications for Indoor Monitoring," in *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16-28, July 2019, doi: 10.1109/MSP.2018.2890128.

[2] J. . -C. Chiao *et al*., "Applications of Microwaves in Medicine," in *IEEE Journal of Microwaves*, vol. 3, no. 1, pp. 134-169, Jan. 2023, doi: 10.1109/JMW.2022.3223301.

[3] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve and K. Chetty, "SimHumalator: An Open-Source End-to-End Radar Simulator for Human Activity Recognition," in *IEEE Aerospace and Electronic Systems Magazine*, vol. 37, no. 3, pp. 6-22, 1 March 2022, doi: 10.1109/MAES.2021.3138948.

[4] S. Zhu, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Continuous human activity recognition with distributed radar sensor networks and CNN–RNN architectures," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–15, 2022, Art. no. 5115215.

[5] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, "Bi-LSTM network for multimodal continuous human activity recognition and fall detection," IEEE Sensors J., vol. 20, no. 3, pp. 1191–1201, Feb. 2020.

[6] S.-W. Kang, M.-H. Jang, and S. Lee, "Identification of human motion using radar sensor in an indoor environment," Sensors, vol. 21, no. 7, 2021, Art. no. 2305.

[7] E. Kurtoglu, A. C. Gurbuz, E. A. Malaia, D. Griffin, C. Crawford, and S. Z. Gurbuz, "ASL trigger recognition in mixed activity/ signing sequences for RF sensor-based user interfaces," IEEE Trans. Human-Mach. Syst., vol. 52, no. 4, pp. 699–712, Aug. 2022

[8] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli and A. Yarovoy, "A Survey on Radar-Based Continuous Human Activity Recognition," in *IEEE Journal of Microwaves*, doi: 10.1109/JMW.2023.3264494.

[9] L. Cao, S. Liang, Z. Zhao, D. Wang, C. Fu, K. Du, "Human Activity Recognition Method Based on FMCW Radar Sensor with Multi-Domain Feature Attention Fusion Network". *Sensors* **2023**, *23*, 5100. https://doi.org/10.3390/s23115100

[10] N. Kruse, R. Guendel, F. Fioranelli and A. Yarovoy, "Segmentation of Micro-Doppler Signatures of Human Sequential Activities using Rényi Entropy," *International Conference on Radar Systems (RADAR 2022)*, Hybrid Conference, Edinburgh, UK, 2022, pp. 435-440, doi: 10.1049/icp.2022.2357.

[11] R. G. Guendel, M. Unterhorst, F. Fioranelli, A. Yarovoy, "*Dataset of continuous human activities performed in arbitrary directions collected with a distributed radar network of five nodes*". (2021 [Online]. doi: https://doi.org/10.4121/16691500.v2. Available from: https://data.4tu.nl/articles/dataset/Dataset_of_continuous_human_activities_performed_in_arbitrary_directions_collected_with_a_distributed_radar_network_of_five_nodes/16691500/2

[12] R. G. Guendel, M. Unterhorst, E. Gambi, F. Fioranelli and A. Yarovoy, "Continuous human activity recognition for arbitrary directions with distributed radars," *2021 IEEE Radar Conference (RadarConf21)*, Atlanta, GA, USA, 2021, pp. 1-6, doi: 10.1109/RadarConf2147009.2021.9454972.

[13] M. Sokolova and G. Lapalme. "A Systematic Analysis of Performance Measures for Classification Tasks." *Information Processing & Management* 45, no. 4 (2009): 427–437