TU Delft

Delft University of Technology

Time-Frequency Resolution Analysis for Continuous Human Activity Recognition using Radar Networks

Guendel, Ronny G.; Fioranelli, Francesco; Yarovoy, Alexander

DOI
10.1109/iWAT57102.2024.10535806

Publication date
2024

Document Version
Final published version

Published in
2024 IEEE International Workshop on Antenna Technology, iWAT 2024

Citation (APA)
Guendel, R. G., Fioranelli, F., & Yarovoy, A. (2024). Time-Frequency Resolution Analysis for Continuous Human Activity Recognition using Radar Networks. In *2024 IEEE International Workshop on Antenna Technology, iWAT 2024* (pp. 341-344). (2024 IEEE International Workshop on Antenna Technology, iWAT 2024). IEEE. https://doi.org/10.1109/iWAT57102.2024.10535806

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

2024 IEEE International Workshop on Antenna Technology (iWAT) | 979-8-3503-1475-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/iWAT57102.2024.10535806

Time-Frequency Resolution Analysis for Continuous Human Activity Recognition using Radar Networks

Ronny G. Guendel, Francesco Fioranelli, Alexander Yarovoy
Microwave Sensing, Signals and Systems (MS3) Group, Delft University of Technology, Delft, Netherlands
{r.gundel, f.fioranelli, a.yarovoy}@tudelft.nl

***Abstract*—The effect of different time-frequency (TF) resolution values is analyzed in the context of Human Activity Recognition (HAR) using multiple radars distributed in a network. Specifically, different spectrograms computed with various Short-Time Fourier Transform (STFT) window lengths and Morse wavelet transform are compared as input representation to a Convolutional Neural Network (CNN), together with a coherent combination of multiple spectrograms. The study emphasizes the importance of selecting appropriate window sizes for TF analysis and for classification, balancing the observation time with the physical duration of the diverse activities, and also avoiding correlation between different data samples that may compromise the generalization ability of the method. The results employing this coherent sensor fusion demonstrate the efficacy of the investigated method, achieving an F1 score of 0.943 on a challenging public dataset containing 9 activities performed by 15 participants.**

***Index Terms*—Radar signal processing, distributed radar, deep learning, CNN, human activity recognition.**

I. Introduction

Human Activity Recognition (HAR) using radar technology has made notable advancements, both in terms of the realization of systems that are compact and low-power, and of the signal processing and related algorithms to elaborate their data [1]–[3]. In the context of HAR, an open research challenge is the formulation and validation of approaches to capture and recognize a wide range of human activities performed continuously, from those extended in time such as walking, to transient actions that may also be rather infrequent, such as falling. In typical radar processing pipelines for HAR [1], [4], time-frequency (TF) analysis is used to generate micro-Doppler (μD) signatures, and selecting the correct window sizes is crucial for effectively capturing the intricacies of these diverse activities [5], [6].

In this paper, various window functions for this processing step are investigated and compared, and a method is proposed to coherently sum all individually obtained μD spectrograms. Specifically, different spectrograms are computed by the Short-Time Fourier Transform (STFT) with various window lengths, using continuously recorded sequences of human activities collected with a distributed radar network. These spectrograms computed with individual windows are compared in terms of HAR performance with our proposed method to combine multiple windows, and with the Morse wavelet transform, highlighting the unique challenges and insights offered by each technique [7]. When processing continuous sequences of activities, correlation issues arising from the duration of the STFT window and hop size are discussed in the context of generating uncorrelated data for training & testing classification algorithms; for this, a sliding window approach is employed to help separate the data. The investigated approaches are validated with experimental radar data collected with a network of five distributed radar nodes, utilizing coherent sensor fusion (also known as signal fusion) to amalgamate data from all radar nodes, effectively capturing human activities from different aspect angles [8], [9]. Our comprehensive dataset and the methodology for STFT extraction from radar data are detailed in [10], and an example code[1][2] for generating a μD spectrogram is publicly available [11].

The rest of the paper is organized as follows. Section II describes the signal processing pipeline, elucidating the methodologies and techniques employed. Section III presents the results with comprehensive analysis and discussion. Finally, Section IV concludes the paper.

II. Signal model & Processing

The dataset used in this work was collected with 5 distributed monostatic radar nodes, involving 15 participants [10]. These 5 Ultra-Wide Band (UWB) radar nodes by Humatics P410 (former PulsON) are simultaneously employed with coded waveform capabilities to reduce mutual interference. The in-phase component is recovered by filter banks, with the quadrature component obtained by the Hilbert transform. The Pulse Repetition Frequency (PRF) f_{PRF} is equal to 122 Hz, with unambiguous Doppler frequency of ±61 Hz (±2.2 m/s). The radar filterbanks have a time-of-flight sampling rate of τ=61 ps. The range resolution with a bandwidth of B=2.2 GHz is 68 mm.

In terms of signal model, the integration of complex Range-Time (RT) matrices $\mathbf{r}(t, r)$ from each individual radar unit can be represented mathematically as follows:

$$\mathcal{R}(t, r) = \frac{1}{N_r} \sum_{n_r=1}^{N_r} \mathbf{r}(t, r)^{(n_r)} \tag{1}$$

[1]https://github.com/rgundel/micro-Doppler-STFT-PulsON-Radar
[2]https://doi.org/10.5281/zenodo.10402956

Authorized licensed use limited to: TU Delft Library. Downloaded on February 11,2025 at 11:32:50 UTC from IEEE Xplore. Restrictions apply.

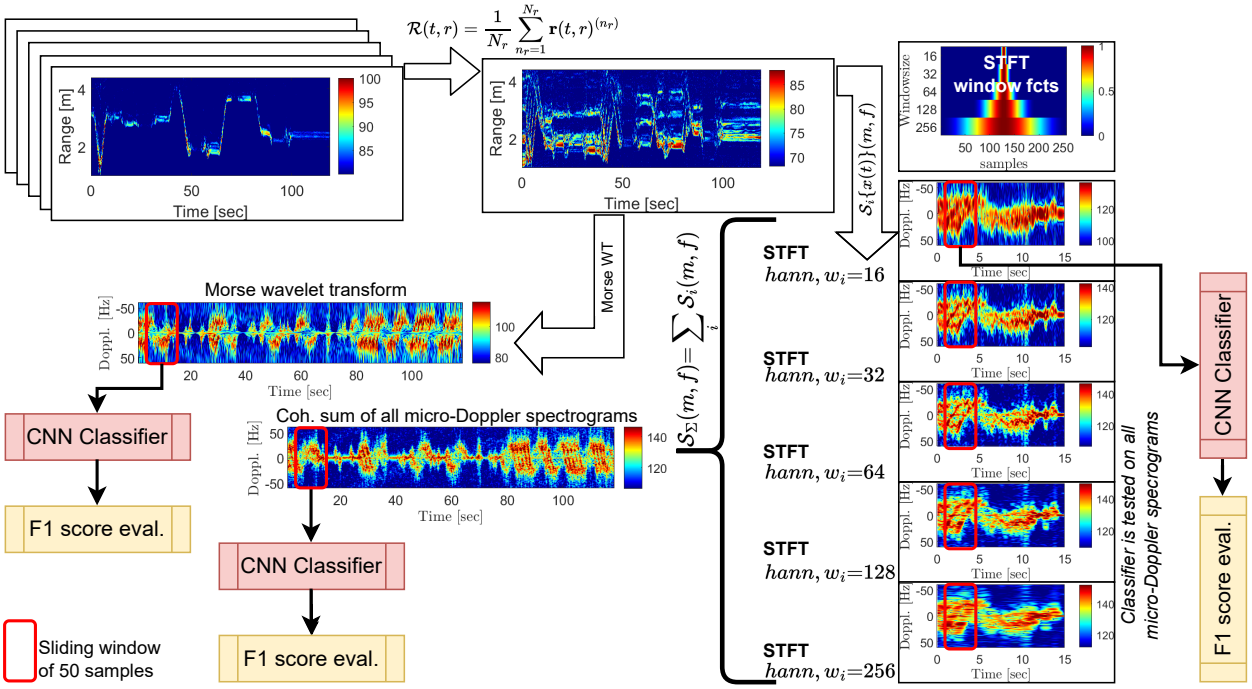


Fig. 1: Processing block diagram for acquiring radar data: This involves transitioning from a range-time (RT) map of single radar units to extracting time-frequency (TF) representations using STFT windows ranging from 16 to 256 samples (equivalent in time to 131ms to 2,099ms). The process also includes the coherent summation of all individually computed µD spectrograms and the application of the Morse wavelet transform before feeding the data into a Convolutional Neural Network (CNN).

Here, $\mathcal{R}(t,r)$ represents the composite RT matrix, synthesizing data across all five radar nodes, where n_r represents each radar node, ranging from 1 to N_r (typically assumed to be 5), and t and r represents the slow-time and fast-time index, respectively. The resulting matrix $\mathcal{R}(t,r)$ is subsequently utilized to compute a µD (μD) spectrogram, as illustrated in the block diagram of Fig. 1.

In this work, various window functions, denoted as w_i, are utilized in conjunction with the Short-Time Fourier Transform (STFT) to derive spectrograms $\mathcal{S}$ from the input signal $x(t)$. This is obtained from the composite RT matrix and is defined as:

$$\mathcal{S}_i\{x(t)\}(m,f) = \int_{-\infty}^{\infty} x(t)\cdot w_i(t-m)\cdot e^{-j2\pi ft}\,dt \quad (2)$$

where f represents the frequency, and m is the slow-time shift. In addition, the individual spectrograms obtained using various window lengths are coherently summed up as follows:

$$\mathcal{S}_\Sigma(m,f) = \sum_i \mathcal{S}_i(m,f) \quad (3)$$

This generates a new possible input for subsequent classification algorithms, alongside the output of the Morse wavelet transform [12] which is also considered in this study.

III. Experimental Results & Discussion

Specifically for the STFT, Hanning window sizes of 16, 32, 64, 128, and 256 samples were used, with each sample corresponding in time to 8.2 ms, with a consistent hop size of 8.2 ms. These parameters are listed in Table I. Additionally, the coherent sum of µD spectrograms, as described in Eq. (3), was computed from the STFT outputs computed individually with each window size from 16 up to 256 samples.

TABLE I: Signal processing parameters for computing time-frequency (TF) spectrograms and sliding window parameters for the CNN input.

Parameter:	TF	Sliding window
Win. size	16,...,256 (131ms,...,2,099ms)	50 (410ms)
Win. type	Hanning	rectangle
Hops	1 (8.2ms)	4 (32.8ms)
Resize dim.	–	64x64

The resulting spectrograms as well as the output of the Morse wavelet transform (WT) are 2D matrices that represent continuous sequences of activities. In order to leverage the effectiveness of Convolutional Neural Networks (CNN) for classification, these matrices are segmented into separated image samples with an additional sliding window. Specifically, we employed a sliding window of 50 samples, equivalent in time to 410 ms, with a hop size of 4 samples (32.8 ms) as depicted in Fig. 1 and summarized in Table I. Using these two windows, one for TF analysis and one to segment the spectrograms, can introduce correlation within

Authorized licensed use limited to: TU Delft Library. Downloaded on February 11,2025 at 11:32:50 UTC from IEEE Xplore. Restrictions apply.

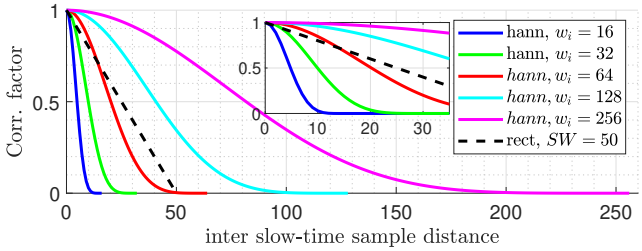


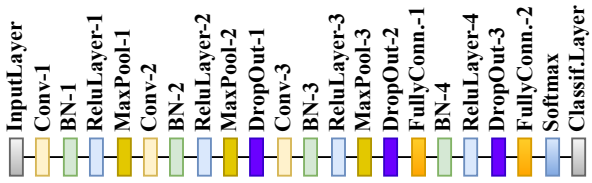
Fig. 2: Correlation factors for varying slow-time sample distances across different Hanning window sizes used in the STFT, denoted as w_i, and a rectangular sliding window of 50 samples for segmentation, denoted as SW.

the data used to train, validate, and test the CNN. The potential risk of introducing redundant and correlated features into the CNN requires careful consideration. The correlations for the windows used in the STFT and the additional sliding window are graphically illustrated in Fig. 2.

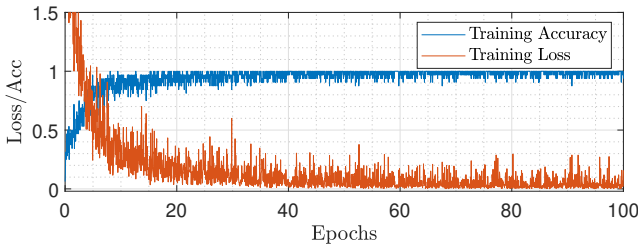
The chosen CNN architecture, detailed in Fig. 3a, was designed to optimize classification performance for this study. The network begins with an image input layer, adjusted for the selected 64x64 image-like data. Subsequently, the network includes multiple convolutional layers with batch normalization and ReLU activation, each followed by a max pooling layer to extract and down-sample features. To mitigate overfitting, dropout layers are strategically placed within the network, with increasing dropout rates towards the deeper layers. The architecture terminates in fully connected layers leading to a softmax output, tailored to the number of classes in our dataset. Class weights are integrated into the final classification layer to address class imbalance. The network was trained using Stochastic Gradient Descent with Momentum (SGDM) optimizer, incorporating a learning rate of 0.01 and L2 regularization to prevent overfitting. Training parameters such as epochs, batch size, and validation frequency were carefully chosen to ensure robust learning, while monitoring the training progress through validation data and periodic checks. The associated learning curves for training loss and training accuracy are shown in Fig. 3b, illustrating the CNN performance over 100 epochs.

Furthermore, to combat the risk of exploding gradients in our CNN caused by large weights, we implemented random under-sampling of the majority classes. This effectively balanced the training data and mitigated weight-related instability. Specifically, the main data imbalances arose from the large prevalence of class samples such as 'walking' & 'stationary', on which we applied this under-sampling strategy. This step harmonized the activity representation in the mixed activity sequence dataset, reducing skewness. For validation, we employed a random 20% holdout for testing and divided the remaining 80% into training (75%) & validation (25%) sets. This partitioning ensured that the CNN was trained and validated on representative samples.

The results, as depicted in Fig. 4, show a testing and validation accuracy and F1 score higher than 90% achieved



(a) Network architecture.



(b) Training accuracy and loss.

Fig. 3: Fig. 3a is the designed network architecture. Fig. 3b shows the training accuracy & loss, with the final validation accuracy (95.4217) and loss (0.1612) achieved with spectrograms obtained with a STFT window size of 256 samples.

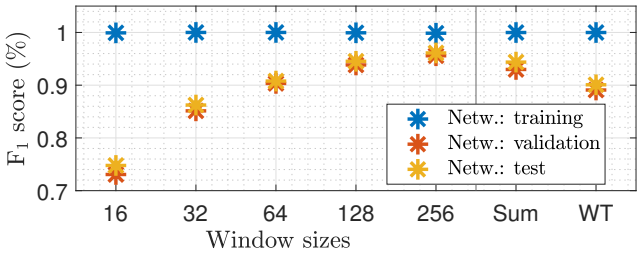
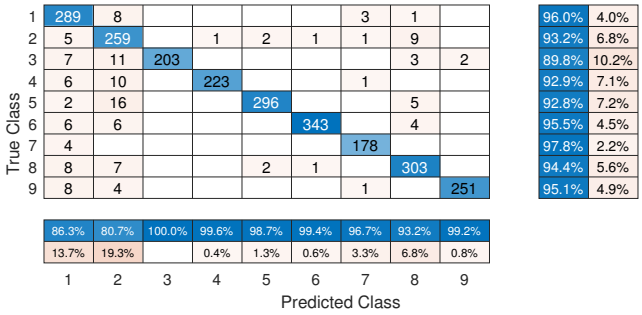


Fig. 4: F1 score obtained by using different STFT window sizes in addition to the coherent summation of all STFT windows ('Sum') and the wavelet transform.

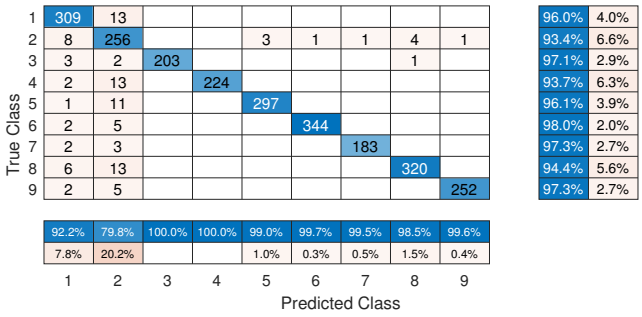
with a window size of 64 samples. A slightly better performance was observed with larger windows up to 256 samples, although this led to the critical correlation issue between different data samples previously described. Additionally, the proposed coherent integrated μD spectrogram, denoted by 'Sum' in Fig. 4, also demonstrates good performance. On the other hand, the Morse wavelet transform (WT) did not perform as well as expected. Comparing the WT with the STFT in terms of TF analysis is challenging due to their inherent differences. The STFT employs a singular basis function —$[w_i(t-m)\cdot e^{-j2\pi ft}]$— with a fixed resolution to offer uniform analysis, but this may overlook details in non-stationary signals. Conversely, the WT, using a family of wavelets, adapts its resolution based on the signal content, making it more suitable for analyzing signals with dynamic frequency content. These differences in basis functions and adaptability lead to divergent interpretations of results, particularly for signals with complex TF characteristics, thus complicating direct comparisons between the two methods.

When analyzing the classification matrices in Fig. 5, which

Authorized licensed use limited to: TU Delft Library. Downloaded on February 11,2025 at 11:32:50 UTC from IEEE Xplore. Restrictions apply.



(a) Confusion matrix of µD spectrogram with coherent summation {F1 score: 0.943}.



(b) Confusion matrix of µD spectrogram with window size 256 samples {F1 score: 0.961}.

Fig. 5: Confusion matrices for the test results for the proposed coherent summation of all µD spectrograms in Fig. 5a, and the µD spectrogram obtained with a STFT window size of 256 samples in Fig. 5b.

specifically shows Fig. 5a for the coherent summation of all µD spectrograms and Fig. 5b for the highest classification results using a STFT window of 256 samples, noticeable outliers were identified in the classes of (1) Walking and (2) Stationary. This may result from the forced under-sampling of these classes to balance the dataset, thus reducing their inter-class correlation as initially discussed. The other in-place classes are (3) Sitting Down, (4) Standing Up from Sitting, (5) Bending from Sitting, (6) Bending from Standing, (7) Falling from Walking, (8) Standing Up from the Ground, and (9) Falling from Standing, respectively. The achieved test macro F1 scores were 0.943 and 0.961 for Fig. 5a and Fig. 5b, respectively.

IV. Conclusion

In this study, the effectiveness of a coherent signal fusion approach in integrating Range-Time (RT) data from multiple radar nodes in a network was successfully demonstrated. This technique is crucial for fusing radar data captured from different aspect angles into unified representations in the time-frequency (TF) spectrogram domain. Specifically, in this work varying window sizes for the TF analysis via STFT were considered and compared, together with an alternative coherent summation of all individually computed spectrograms and Morse Wavelet Transform. The resulting data were subsequently used as input to a CNN for classification purposes. The validation of the investigated approach was performed leveraging on a publicly available dataset containing continuous sequences of human activities recorded by five radars and 15 participants [10].

The results show that the proposed summation of all spectrograms yielded a satisfactory F1 score of 0.943. The use of a long window size of 256 samples showed even higher results, but these findings should be interpreted with caution. The equivalent time duration of approximately 2 seconds, coupled with a high inter-sample correlation, although beneficial in certain scenarios, may not be ideally suited for accurately capturing the nuances of in-place, fast and infrequent activities. This paves the way to implementations of adaptive processing schemes where the window length and other parameters related to the TF analysis of the radar data can be dynamically adjusted, depending on the scenario under test and the specific kinematics of the observed people.

References

[1] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

[2] C. Li, Z. Peng, T.-Y. Huang, T. Fan, F.-K. Wang, T.-S. Horng, J.-M. Muñoz-Ferreras, R. Gómez-García, L. Ran, and J. Lin, "A review on recent progress of portable short-range noncontact microwave radar systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 5, pp. 1692–1706, 2017.

[3] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, T. Macks, and F. Fioranelli, "Cross-frequency training with adversarial learning for radar micro-doppler signature classification (rising researcher)," in *Radar Sensor Technology XXIV*, K. I. Ranney and A. M. Raynal, Eds., vol. 11408, International Society for Optics and Photonics. SPIE, 2020, p. 114080A.

[4] Y. Yang, C. Hou, Y. Lang, T. Sakamoto, Y. He, and W. Xiang, "Omnidirectional motion classification with monostatic radar system using micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3574–3587, 2020.

[5] B. Jokanovic, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Multi-window time–frequency signature reconstruction from undersampled continuous-wave radar measurements for fall detection," *IET Radar, Sonar & Navigation*, vol. 9, no. 2, pp. 173–183, 2015.

[6] P. Klaer, A. Huang, P. Sévigny, S. Rajan, S. Pant, P. Patnaik, and B. Balaji, "An investigation of rotary drone HERM line spectrum under manoeuvering conditions," *Sensors (Basel)*, vol. 20, no. 20, p. 5940, 2020.

[7] S. Scholl, "Fourier, Gabor, Morlet or Wigner: Comparison of time-frequency transforms," 2021.

[8] R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Distributed radar fusion and recurrent networks for classification of continuous human activities," *IET Radar, Sonar and Navigation*, vol. 16, no. 7, pp. 1144–1161, 2022.

[9] 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)*, 2021, pp. 1–6.

[10] 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/datasets/eda41444-9ae8-4655-aac5-4d2541b2bcd8

[11] R. G. Guendel, "micro-Doppler-STFT-PulsON-Radar," Dec. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.10402956

[12] J. M. Lilly and S. C. Olhede, "Generalized Morse wavelets as a superfamily of analytic wavelets," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 6036–6041, 2012.

Authorized licensed use limited to: TU Delft Library. Downloaded on February 11,2025 at 11:32:50 UTC from IEEE Xplore. Restrictions apply.