Human activity classification with radar signal processing and machine learning

Jia, Mu; Li , Shaoxuan ; Le Kernec, Julien; Yang, Shufan; Fioranelli, Francesco; Romain, Olivier

DOI
10.1109/UCET51115.2020.9205461

Publication date
2020

Document Version
Final published version

Published in
2020 International Conference on UK-China Emerging Technologies (UCET)

Human activity classification with radar signal processing and machine learning

Mu Jia
James Watt School of Engineering
University of Glasgow
Glasgow, UK
2288945L@student.gla.ac.uk

Shaoxuan Li
James Watt School of Engineering
University of Glasgow
Glasgow, UK
2288929J@student.gla.ac.uk

Julien Le Kernec
James Watt School of Engineering
University of Glasgow
Glasgow, UK
julien.lekernec@glasgow.ac.uk

Shufan Yang
James Watt School of Engineering
University of Glasgow
Glasgow, UK
shufan.yang@glasgow.ac.uk

Francesco Fioranelli
MS3-Microwave Sensing Signals and Systems, TU Delft, Delft, The Netherlands
f.fioranelli@tudelft.nl

Olivier Romain
ETIS – Signal and Information Processing lab, University Cergy-Pontoise, Cergy, France
olivier.romain@cyu.fr

***Abstract*—As the number of older adults increases worldwide, new paradigms for indoor activity monitoring are required to keep people living at home independently longer. Radar-based human activity recognition has been identified as a sensing modality of choice because it is privacy-preserving and does not require end-users compliance or manipulation. In this paper, we explore the robustness of machine learning algorithms for human activity recognition using six different activities from the University of Glasgow dataset recorded with an FMCW radar. The raw radar data is pre-processed and represented using four different domains, namely, range-time, range-Doppler amplitude and phase diagrams, and Cadence Velocity Diagram. From those, salient features can be extracted and classified using Support Vector Machine, Stacked AutoEncoder, and Convolutional Neural Networks. The fusion of handcrafted features and features from CNN is applied to get the best scheme of classification with over 96% accuracy.**

***Keywords*—radar, signal processing, machine learning, deep learning, classification, healthcare, assisted living**

I. Introduction

Recently, the increasing number of the elderly population is a serious problem for most countries. Some of those old people have no choice but to live alone, and thus it is hard for their family and governments to take care of them, especially when some emergency happens, such as falling. To avoid those dangerous situations or further injuries, timely detection of those hazardous activities is important. If we can recognize or even predict them, medical support can be dispatched in time. Cameras are suitable sensors to realize human activity recognition, but it is affected by the lighting conditions, and some blind zones cannot be detected because of dead angles. Most importantly is the user's privacy may be violated by the camera intentionally or unintentionally. Radar is an alternative sensing modality that can perform this function while protecting individual privacy. By analyzing the radar signal reflected from the user, activity classification can be achieved, and thus dangerous actions can be detected [1].

Radar data can be transformed into different domains where useful features can be extracted for classification, namely, range-time (RT), Doppler-time (DT) amplitude and phase, and Cadence Velocity Diagram (CVD). From [2], DT is used for the activity classification from which velocity-related physical features from the target can be calculated. In [3], CVD is used for feature extraction combined with spectrograms. Phase diagrams (PD) are also proven to provide salient information regarding the change and disturbance of signals [4].

Traditional machine learning methods, such as SVM (support vector machine) with different kernel functions, are widely used in the industry. It has good performances for activity classification when the number of activities is limited to 10-20, with accuracies of in the low and mid 90% [5]. For deep learning networks in [6], the authors show that stacked autoencoder(SAE), and convolutional neural network (CNN) have excellent results of 84% and 90% separately, for dealing with a large dataset. In this paper, those 3 classifiers are all used for a comparison based on their accuracy, robustness, and combination with fusion.

This paper is organized as follows. Section II describes the details about the experimental radar and the data collection. The data pre-processing / feature extraction methodology is presented in Section III. Section IV presents the results for SVM and feature selection optimization. Section V presents the deep learning architectures and feature fusion. Finally, Section VI concludes this paper.

II. Experiment and Data Description

83 participants took part in the data collection. They were asked to perform 6 different activities and repeat them thrice. These activities are walking A01, sitting down A02, standing up A03, pick up an object A04, drinking A05 and falling A06. Those radar data were collected in safe situations at the University of Glasgow. Considering falling is a high-risk activity, the elderly volunteers were not ask to fall. Finally there are 1164 samples in our dataset [7, 8].

The radar data were collected with a frequency modulated continuous wave (FMCW) radar (Fig. 1). This radar works at 5.8GHz, with a chirp bandwidth of 400MHz and 1ms time duration. The number of samples per sweep is 128 (128 kS/s).

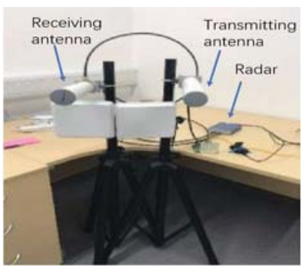

Fig. 1. View of radar used in this research, while cylinders are Yagi antenna and small blue box is radar.

III. Radar Signal Processing and Feature Extraction

A. Radar Data Processing

The aim of signal processing is to convert the raw radar data to new domains which facilitate feature extraction and classification [9]. To generate the RT domain matrix, a moving target indication filter with a cut-off frequency of ±0.0075Hz is applied to the raw data first to remove static clutter. A 128-point FFT is then used to derive the range profile. The RT domain is formed by the accumulation of range profiles over time.

After that, an STFT is applied to the RT matrix with a Hamming window of 0.2s duration, and 95% overlap for the sliding window. The accumulated DT profiles are accumulated over time with its amplitude being the DT a.k.a spectrogram (Fig. 3), where micro-Doppler (uD) features can be obtained, and its phase being the PD, which can supplement DT information. The PD is only useful if we retain the phase components with high enough amplitude in DT with thresholding and clipping frequencies beyond ±150Hz as they do not make physical sense in human activity recognition.

Finally, the CVD is obtained by applying a 981-point FFT to the DT domain along the time axis for each Doppler bin. CVD can show some information about the repetition of velocities, and thus useful periodic features can be calculated in this domain, and some key information about the shape and frequency of moving targets are also available here. Fig. 2 Summarizes the domains obtained with pre-processing.

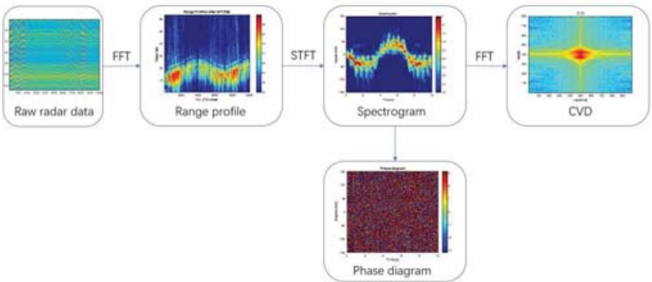

Fig. 2. The procedure of radar signal processing, from raw data matrix to range-time domain, DT/PD, and CVD.

B. Handcrafted feature extraction from RT, DT and CVD

After obtaining multi-domain information, features can be extracted. Based on [10], 36 handcrafted features are extracted, and most of them are derived from the DT. Some of them are uD features like those from DT, and some of them are periodic from CVD. Table 1 shows the list of features used in this research.

Table 1: List of the extracted features

Feature List of Radar	#
Energy curve	3
Skewness	1
Centroid	2
Bandwidth	2
SVD (Singular Value Decomposition)	12
SVD from RT domain	12
Entropy	1
SRF(Step repetition frequency) of CVD	1
Peak value of CVD	2
Total	36

Most features are extracted from the DT domain. According to previous work, the centroid of Doppler in DT is of the highest robustness, which can show the movement of the torso and limbs. It is given by (1).

$$f_C(j) = \frac{\Sigma_i f(i) \int S(i,j)}{\Sigma_i S(i,j)} \quad (1)$$

$S(i,j)$ is the element of row i and column j of the DT matrix and $f(i)$ is the Doppler frequency the i^{th} Doppler bin. The bandwidth is the difference of the extreme value of DT domain representing the modulation around the mean Doppler component. Other features, including skewness and entropy, are utilized to give a metric of grey levels of the DTs. For other information, singular value decomposition (SVD) has been used to exploit more useful features containing a spectral projection of the time and frequency domains, respectively. SVD is also applied to the RT domain, calculating the mean value and standard deviation of its components as features. The frequency of velocity is extracted from CVD after an FFT along the time axis of DT, which represents the periodicity of step for the walking A01 action, for example.

IV. Classification and Feature Selection

The aforementioned handcrafted features from RT, DT, and CVD are classified using SVM for activity recognition. Different kernel functions are compared in this part to get the best classifier. Finally, a sequential backward selection (SBS) is used to increase accuracy and reduce the complexity of the algorithm and time efficiency. The framework of this part is shown in Fig. 3.

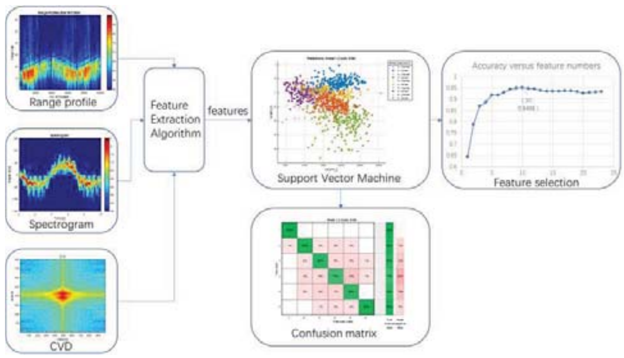

Fig. 3. Feature extraction, classification and feature selection

A. Support Vector Machine

SVM with handcrafted features is a traditional machine learning method to realize classification. The training time of SVM is much lower than deep learning methods, while it has good performances when the number of data types is small. In this paper, 36 features extracted from 3 different domains are used as input. SVM with 5 different kernels are compared based on 10-fold cross-validation. From Fig. 4, the cubic SVM performs best for this dataset with 91.6% accuracy.

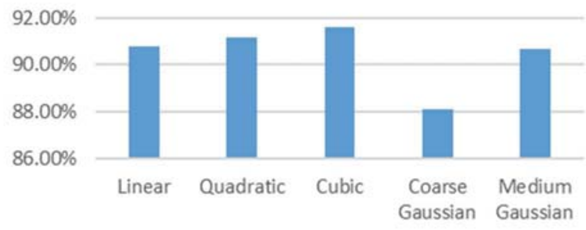

Fig. 4. Results of SVM with different kernel functions

B. *Feature Selection*

According to [11], not all the features have a positive contribution to the current accuracy. To choose the most salient features and save computational power, feature selection is introduced. SBS is applied to the former classifier based on [12]. The SBS algorithm tests all combinations of features sequentially starting from the whole feature set and removing one feature and keeping the subset of features providing the best accuracy and repeating the process until there is only one to optimize the classification by identifying the subset of features yielding the best accuracy as shown in Fig.5. The improvement in accuracy is 3.58% when there are 14 features selected yielding 95.24%, which saves about 60% in computation power.

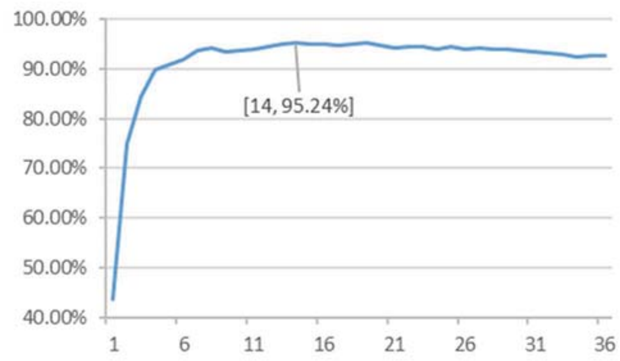

Fig. 5. Feature extraction, classification and feature selection

V. Deep Learning Architecture

SVM seems to have good performances, but if the complexity of the dataset increases, their performance may sharply decrease. Deep learning has better robustness in those conditions [13]. Two methods, Stacked AutoEncoder (SAE) and Convolutional Neural Network (CNN) are used to classify those activities.

A. *Stacked Autoencoder*

An autoencoder consists of an encoder and a decoder. The aim of the SAE is compressing the input into a middle layer and then decompressing it back, and make the output as similar to the input as possible. SAE is several autoencoders stacked together. The input to SAE is the figure of DT, and it is converted to a 1-dimensional array for the input layer of SAE. The structure of this 3-layers SAE is shown in Fig. 6. After training, the middle layer is drawn out containing the compressed representation of the input is connected to a SoftMax layer to realize classification using a Bayes optimizer. After optimization, an SAE yielded 92% accuracy.

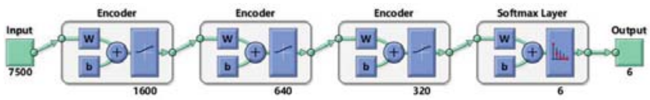

Fig. 6. Classification network from 3-layers SAE

B. *Convolutional Neural Network*

CNN is a widely used architecture for the feature extraction of images. Compared with SAE, it is more focused on global features. The optimization for activity classification is obtained with gradient descent and backpropagation of error [14]. The CNN is tested with DT and PD as inputs, and those matrices are converted to compressed figures of size 32x32x3. After Bayes optimization, the CNN network yielded 92.21% accuracy for the DT input, while the CNN with PD as input yielded only 81.9% accuracy. The structure of DT input is shown in Fig. 7. The depth of this net is 2 with a 3×3 filter size. Larger filter sizes with fewer network depth may have the same theoretical result, but this structure performs better and more efficiently.

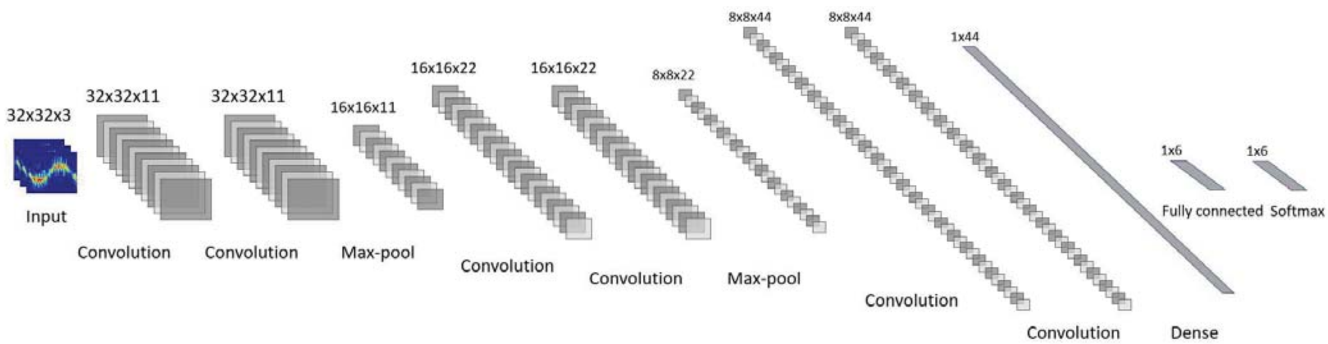

Fig. 7. CNN structure with DT input

C. *Robustnesss comparison and retraining*

Although those classifiers have good results for this dataset, some of them may have a loss in performance when there are changes in the data which would indicate that the classifier did not generalize the model enough and therefore overfitted to the training set. For instance, if the number of activities increases, the classifier based on handcrafted features may not perform as well have bad results because they cannot generalize well.

To compare those classifiers, a new dataset with 155 samples is introduced with the same activities. This dataset is collected in a different environment with 7 different participants while the activities are the same as the former one. Classifiers are tested by this new data and compared with each other, as shown in Fig. 8. The accuracy of SVM declines from 95.24% to 52.73%. SAE and CNN show better robustness while their accuracy also decreases slightly from 91.23% and 92.21% to 69.39% and 77.17%, respectively. To improve this situation, half of the new data are put into the training set to retrain those classifiers, and the results of all the classifiers increased to 70.89% for SVM, 80.31% for SAE, and 90.29% for CNN. In this case, CNN has shown its robustness compared to SAE and SVM.

D. *Feature Fusion*

Finally, different feature-fusion schemes are compared combining handcrafted features (H), features from CNN (DT and PD) are combined together as input to an SVM. There are 4 different combinations with 3 possible pairings (H+DT, H+PD, DT+PD), and all the features combined. The cross-validation result is shown in Fig. 9. The fusion of handcrafted features and CNN-features from the DT achieves the highest accuracy at 96.65%, which is better than the optimized SVM by 1.41% and SAE and CNN with DT as input by 5.42% and 4.44%, respectively.

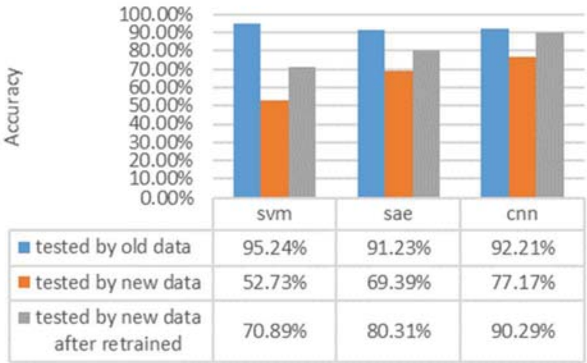

Fig. 8. Comparison between SVM, SAE and CNN

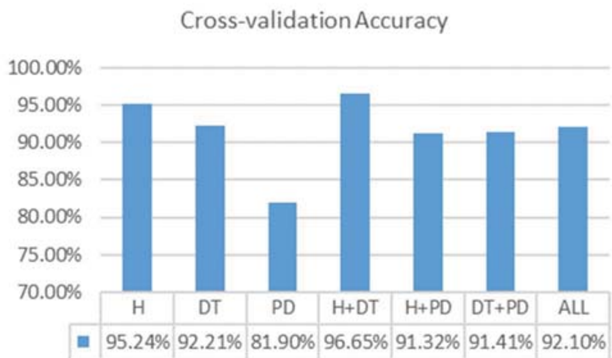

Fig. 9. Comparison between different feature fusion schemes. H means handcrafted features, DT means features from the CNN with DT as input, and PD means features from the CNN with PD as input.

VI. Conclusion

In this paper, the radar data are collected with an FMCW radar worked at 5.8GHz with 400 MHz bandwidth and 1 ms duration, including 6 different kinds of activities. In the data processing part, it is converted into 4 different domains: range-time, Doppler-time, PD and CVD. 36 features are extracted from those domains and then input to SVM. After the comparison of 5 kernel functions, cubic SVM has a cross-validation accuracy of 91.26%. SBS algorithm is used to select features, and 14 features are left to increase the accuracy to 95.24% while saving 60% computation power. For deep learning architecture, SAE and CNN are applied to this dataset. SAE yields 91.23% accuracy and CNN 92.21%. By introducing a new test dataset, the robustness of classifiers was tested showing that the CNN has the best robustness while SVM is the least robust. Finally, a feature fusion of handcrafted features and CNN features is used, increasing the classification accuracy to 96.65% obtained with handcrafted features combined with the CNN features obtained from DT input.

An interesting direction to pursue for future research on the robustness of classification would be the comparison of FMCW radar with an OFDM radar when multiple radars operate in the same vicinity [15].

References

[1] Radar for indoor monitoring : detection, classification, and assessment. Boca Raton: CRC Press, Taylor & Francis Group, 2018, pp. xiii, 390 pages.

[2] H. Li, A. Shrestha, F. Fioranelli, J. Le Kernec, and H. Heidari, "FMCW radar and inertial sensing synergy for assisted living," The Journal of Engineering, vol. 2019, no. 20, pp. 6784-6789, 2019, doi: 10.1049/joe.2019.0558.

[3] R. Ricci and A. Balleri, "Recognition of humans based on radar micro-Doppler shape spectrum features," IET Radar, Sonar & Navigation, vol. 9, no. 9, pp. 1216-1223, 2015.

[4] Munteanu M., Rusu C., Vladareanu L., Petreus D., Rusu V., and D. M., Eds. EKG Analysis Using STFT Phase (International Conference on Advancements of Medicine and Health Care through Technology. IFMBE Proceedings. Berlin, Heidelberg: Springer, 2009.

[5] T. D. Bufler and R. M. Narayanan, "Radar classification of indoor targets using support vector machines," IET Radar, Sonar & Navigation, vol. 10, no. 8, pp. 1468-1476, 2016.

[6] S. Z. Gürbüz, B. Erol, B. Tekeli, and B. Çağlıyan, "Operational assessment and adaptive selection of micro-Doppler features," IET Radar, Sonar & Navigation, vol. 9, no. 9, pp. 1196-1204, 2015, doi: 10.1049/iet-rsn.2015.0144.

[7] F. Fioranelli, S. A. Shah, H. Li, A. Shrestha, S. Yang, and J. Le Kernec. Radar signatures of human activities, doi: 10.5525/gla.researchdata.848.

[8] D. F. Fioranelli, D. S. A. Shah, H. Li, A. Shrestha, D. S. Yang, and D. J. L. Kernec, "Radar sensing for healthcare," Electronics Letters, vol. 55, no. 19, pp. 1022-1024, 2019, doi: 10.1049/el.2019.2378.

[9] V. C. Chen, D. Tahmoush, and W. J. Miceli, "Radar Micro-Doppler Signatures: Processing and Applications," in Micro-Doppler Signatures - Review, Challenges, and Perspectives: The Institution of Engineering and Technology, 2014, ch. Radar Micro-Doppler Signatures: Processing and Applications.

[10] 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 Journal, vol. 20, no. 3, pp. 1191-1201, 2020, doi: 10.1109/jsen.2019.2946095.

[11] C. Ding et al., "Inattentive Driving Behavior Detection Based on Portable FMCW Radar," IEEE Transactions on Microwave Theory and Techniques, vol. 67, no. 10, pp. 4031-4041, 2019, doi: 10.1109/tmtt.2019.2934413.

[12] H. I. Huang, Y. S. Wu, Y. K. Chan, and C. H. Lin, "Study on image feature selection: A genetic algorithm approach," in IET International Conference on Frontier Computing. Theory, Technologies and Applications, 4-6 Aug. 2010 2010, pp. 169-174.

[13] M. S. Seyfioglu, A. M. Ozbayoglu, and S. Z. Gurbuz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," IEEE Transactions

on Aerospace and Electronic Systems, vol. 54, no. 4, pp. 1709-1723, 2018, doi: 10.1109/taes.2018.2799758.
[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," in Intelligent Signal Processing, S. Haykin and B. Kosko Eds.: IEEE Press, 2001, pp. 306-351.
[15] J. Le Kernec and O. Romain, "Performances of multitones for ultra-wideband software-defined radar," IEEE Access, vol. 5, pp. 1-1, 2017, doi: 10.1109/access.2017.2693300.