Real-time Human Motion Behavior Detection via CNN using mmWave Radar

Article *in* IEEE Sensors Letters · December 2018

DOI: 10.1109/LSENS.2018.2889060

2 authors:

Renyuan Zhang
Lyft Inc

Siyang Cao
The University of Arizona

IEEE Sensors Letters VOL. 3, NO. 2, FEBRUARY 2019

Microwave/millimeter wave sensors

Real-Time Human Motion Behavior Detection via CNN Using mmWave Radar

Renyuan Zhang[1*] and Siyang Cao[1**]

Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ 85712 USA
*Student Member, IEEE
**Member, IEEE

Manuscript received October 3, 2018; accepted December 18, 2018. Date of publication December 21, 2018; date of current version February 4, 2019.

Abstract—A real-time behavior detection system using millimeter wave radar is presented in this article. Radar is used to sense the micro-Doppler information of targets. A convolution neural network (CNN) is further implemented in the detection and classification of the human motion behaviors using this information. Both the convolution layers and architecture of CNNs are presented. The analysis on loss and accuracy of training results is also shown. The experimental result indicates a precise determination of human motion behavior detection using the proposed system.

Index Terms—Microwave/millimeter wave sensors, behavior detection, convolution neural network (CNN), micro-Doppler effect, micro-Doppler signature, radar, RF.

I. INTRODUCTION

Human motion behavior monitoring has been attracting great interest among researchers, especially in surveillance, tracking, and patient monitoring. However, current methods are limited to using cameras and infrared sensors [1]. Though cameras, including thermal cameras, are accurate and reliable for surveillance usage, the disadvantages of cameras are obvious—they can leak privacy and rely on light condition [2]. A more comprehensive and accurate sensor is needed for behavior monitoring. In this article, we are using micro-Doppler signatures resulted from millimeter wave (mmWave) radar to detect the motion behavior of people.

In recent years, small and low-cost single chip consumer radar systems operating at mmWave frequencies have opened up a vast range of new applications, such as automotive radar, health monitoring radar [3], unmanned aerial vehicle (micro-UAV) radar [4], robot guidance radar, and a host of other applications [5]. Automotive radar, especially working at 77 GHz, is an attractive research area, which plays an important role in the automotive industry, autonomous cars, and driving assistance systems [6]. It can capture the micro-Doppler signature to recognize target via dynamic time warping [7], study human kinetics [8], and track objects [9], [10]. However, the studies on human motion behavior are bounded. In [11], Seifert *et al.* provided human walking style using micro-Doppler, but how to classify the behavior is not studied. In [12], Shrestha *et al.* gave a good support vector machine (SVM) classification on the human motion behavior for indoor monitoring. To further classify behavior into various classes, neural network is a promising method. Its classification result can be continuously improved as more samples are fed into the training process. However, there is only limited study using neural network for radar data [11], [12]. In this article, we use convolution neural network (CNN) to detect the behavior of people, and build a real-time monitoring system, which can provide a reliable classification of motion behaviors.

Micro-Doppler effect is induced by micromotion dynamics of a target or its structure, such as vibration, rotation, tumbling, and coning motions, which is widely existed in bulk motion of a radar target, a.k.a. Doppler effect [13]. In this article, the mmWave radar provides prediction output of the target behavior using micro-Doppler effect. A convolutional neural network is investigated to train the radar system to recognize the behavior of people. In this case, we treat micro-Doppler signature as a data image with intensities. The convolution layers train the collected data into the output matrix and optimize the parameters in the network neurons. The query micro-Doppler signatures of the incoming testing dataset then can be predicted using this CNN. A probability matrix is produced from query micro-Doppler signatures.

In this article, we are the first to build a real-time human motion behavior detection system using the commercial mmWave radar. We developed the system using robotic operating system (ROS) to control low-level device (mmWave radar) and transmit streaming data across platforms. We also investigated how CNN can be used for detecting human motion behaviors. Meanwhile, behaviors can be defined by users, and the system can be trained offline. Consequently, the proposed system can detect human motion behavior accurately in real time. The proposed system can have a wider application, such as autonomous driving, traffic monitoring, and patient cares.

The rest of this article is organized as follows. In Section II, the methodology of extracting micro-Doppler information, designing of CNN, and ROS framework are introduced. In Section III, training and experimental results are presented. In Section IV, this article is concluded.

II. METHODOLOGY OF RADAR MICRO-DOPPLER SIGNATURE OBSERVATION OF REAL TIME

A. Real Time Frequency-Modulated Continuous-Wave (FMCW) Micro-Doppler Signature Processing

For an mmWave radar not considering to resolve angles, the received signal processing can be classified into two different scenarios: without constant false alarm rate (CFAR) detection and with CFAR detection.

The first scenario (without CFAR case) is to use the entire range-Doppler data. The received data cube (i.e., a series of range-Doppler data) is passed through Doppler-time extraction. Accumulated real-time data are then went through a CNN prediction node. An example can be seen in Fig. 1, where the mmWave radar obtains the micro-Doppler signature of a human walking by raw sampling on board and

This work is supported by The University of Arizona. Corresponding author: Siyang Cao (e-mail: caos@email.arizona.edu).
Associate Editor: S. Tadigadapa.
Digital Object Identifier 10.1109/LSENS.2018.2889060

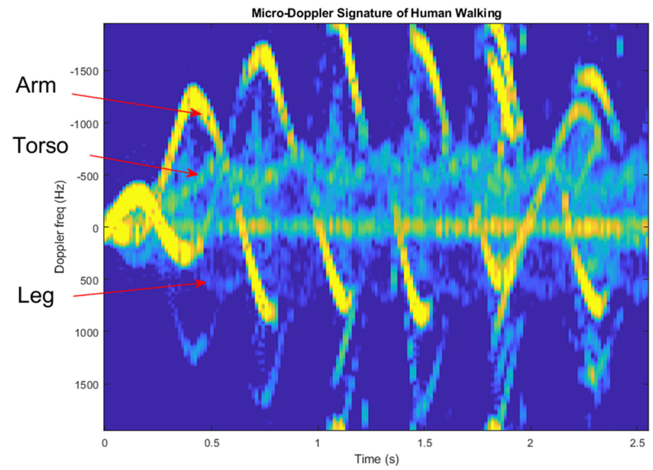


Fig. 1. Micro-Doppler signature of a human walking.

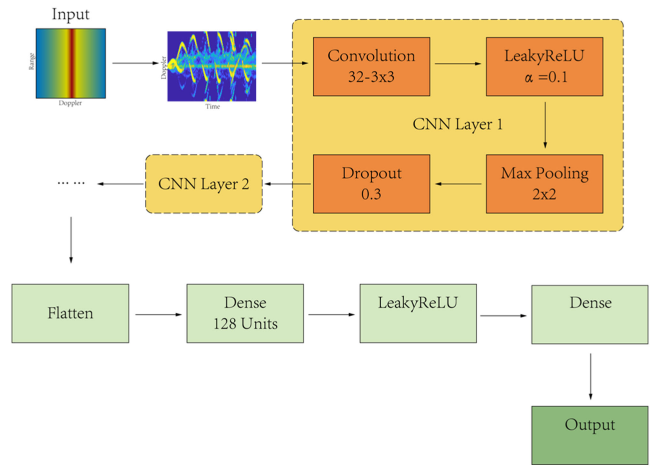


Fig. 2. CNN framework.

processing through a host computer. Arm, leg, and torso can be clearly recognized in the entire Doppler data, though the clutter is not totally removed. In the proposed system, we predefined the number of chirps, and pulse repetition interval to guarantee the Doppler resolution is sufficient in detecting human motion behaviors.

The second scenario (with CFAR case) is to use the point cloud data, i.e., the raw range-Doppler data are processed by the integrated CFAR algorithm on the radar board, which only picks the radar detection points and produces point cloud data with Doppler information. The point cloud data have much less of a data volume for data transmission and can be used for the real-time application.

In the rest of this article, we use the point could data and transmit the data to the host computer. The host computer can process the point cloud data using grouping and clustering algorithms and form the micro-Doppler signature data. With a fixed time frame, the micro-Doppler data are passed through a trained network. In this way, the human motion behavior can be detected in real time.

A series of sequential range-Doppler frames are used to analyze micro-Doppler signatures. The frames are buffered and stored on a host computer as input to the CNN network. The buffered data are updated by incoming data and processed through the network. It makes the real-time motion behavior detection and classification feasible.

B. Convolution Neural Network

CNN is a class of deep and feed-forward artificial neural networks with multilayer perceptrons. Conventionally, a CNN is used as a common classifier for images or visually imagery because of its convolution process emulates the response of an individual neuron to visual stimuli [14], [15]. In this article, the CNN is used on unvisualized micro-Doppler signatures. Since the micro-Doppler signatures are time variant, the input layer neurons for CNN are depending on time. The Doppler effects from multitargets are clustered using density-based spatial clustering of applications with noise (DBSCAN) or other clustering algorithms before the micro-Doppler signature classification.

Fig. 2 shows the structure of the CNN network from the radar raw range-Doppler response. The summarization of the CNN using micro-Doppler from Fig. 2 is listed in Table 1.

In Fig. 2 and Table 1, we show a typical frame of convolution layer, which contains the following.

1) *Convolution layer:* It performs convolution calculation based on the micro-Doppler data. The convolution coefficients are trained to calculate features.
2) *Activation layer* (leaky ReLU function)*:* Leaky ReLU is the abbreviation of leaky version of Rectified Linear Unit. It is widely used in neural network as the activation function. It overcomes the "dying ReLU" issue and is a nonsaturating activation function, as given by

$$f(x) = \begin{cases} \alpha x, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{1}$$

where α is a very small number (typically equals 0.01) to avoid negative values not mapping appropriately.
3) *Max pooling layer:* The spatial pooling reduces the dimensionality of each feature map but retains the most important information. The objective of subsampling is to get an input representation by reducing its dimensions, which helps in reducing overfitting.
4) *Dropout layer:* The dropout layer is added to avoid overfitting.
5) *Fully connected layer* (with flatten and dense layers)*:* Several convolution layers are then merged to a fully connected layer. The goal of the fully connected layer is to flatten the high-level features that are learned by convolutional layers. It combines all the features from the original input.

As the CNN network is setup, as shown in Fig. 2,we collected the training data via performing different types of human motion behaviors, such as walking, swinging hands, standing/sitting, and shifting. The training data are stored, and the training processing is performed in a high-performance computing (HPC) unit with on-chip graphics processing units (GPUs). As coefficients are generated through the training data, the neural network model is stored as an h5py file and uploaded to the computer for implementing the detection and classification in real time.

C. ROS Framework

The implementation of the real-time micro-Doppler behavior monitoring is done on ROS framework. ROS provides hardware and low-level device control, message passing, and source management. For the proposed system, we implemented low-level device control of the mmWave sensor (Texas Instruments' AWR1642) on ROS and transmitted data via ROS messages. In this way, the real-time monitoring can be achieved. The workflow can be seen in Fig. 3. The radar data are transmitted into the host computer for behavior detection. It passes through data transmission via ROS node and generates micro-Doppler frames in micro-Doppler signature node as input for the CNN

Table 1. CNN Summarize.[a]

Layer (type)	Output Shape	Parameter #
conv2d_1 (Conv2D)	(None, 128, 25, 32)	320
leaky_re_lu_1 (LeakyReLU)	(None, 128, 25, 32)	0
max_pooling2d_1 (MaxPooling)	(None, 64, 13, 32)	0
dropout_1 (Dropout)	(None, 64, 13, 32)	0
conv2d_2 (Conv2D)	(None, 64, 13, 64)	18496
leaky_re_lu_2 (LeakyReLU)	(None, 64, 13, 64)	0
max_pooling2d_2 (MaxPooling)	(None, 32, 7, 64)	0
dropout_2 (Dropout)	(None, 32, 7, 64)	0
conv2d_3 (Conv2D)	(None, 32, 7, 128)	73856
leaky_re_lu_3 (LeakyReLU)	(None, 32, 7, 128)	0
max_pooling2d_3 (MaxPooling)	(None, 16, 4, 128)	0
dropout_3 (Dropout)	(None, 16, 4, 128)	0
flatten_1 (Flatten)	(None, 8192)	0
dense_1 (Dense)	(None, 128)	1048704
leaky_re_lu_4 (LeakyReLU)	(None, 128)	0
dropout_4 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 3)	387

[a]Total parameters: 1 141 763
Trainable parameters: 1 141 763
Nontrainable parameters: 0

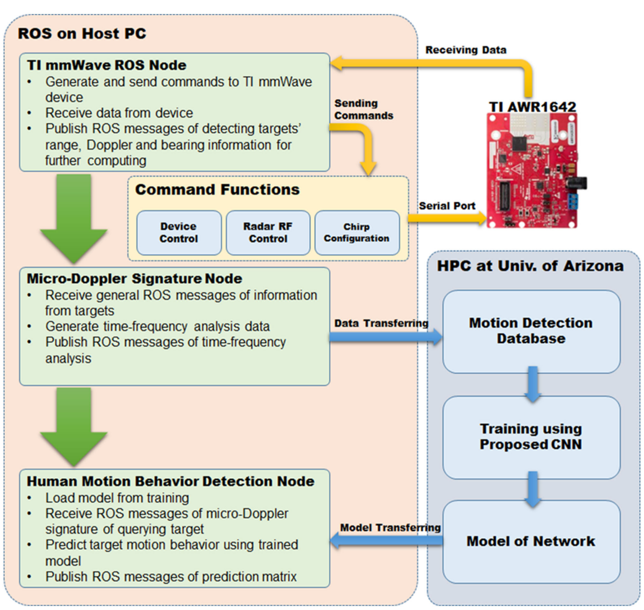


Fig. 3. Whole ROS implementation on mmWave radar sensors.

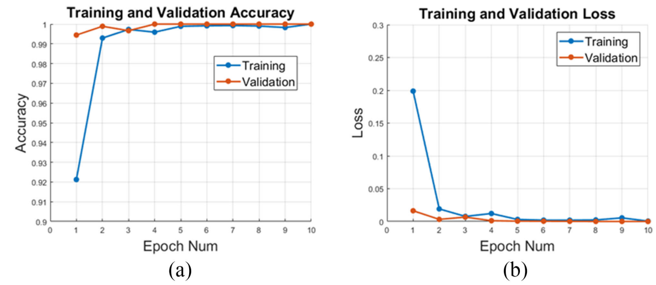


Fig. 4. Training and validation. (a) Accuracy. (b) Loss.

network. The CNN network is implemented in human motion behavior detection node. The neural network model is stored as an h5py file and transmitted from HPC to host computer before launching the human motion behavior detection node.

The texas instruments (TI) mmWave radar is operating at the frequency of 77 GHz with 25 Hz frame rate. For indoor application, we configure the radar parameters to a maximum range of 17.788 m. The Doppler resolution is 0.097 m/s, and the Micro-Doppler data are at a frame rate of 25 Hz.

III. RESULTS

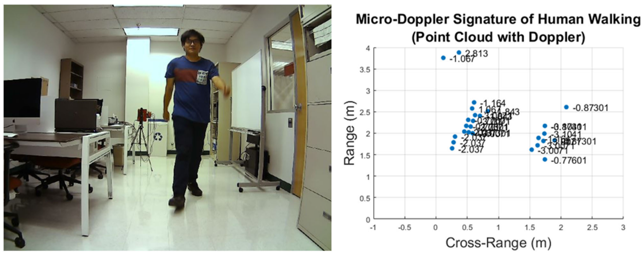
A. Training Results

The training data obtained from testing researchers, which consist hundreds of human motion behavior samples, such as walking, waving hands, clapping, sitting/standing, bending, falling, *etc.* In this study, behaviors such as walking, swinging hands for help, and falling are our interest. Different behaviors are recorded and stored in our database. The database is then trained with the proposed CNN on HPC. The training time on HPC costs around 37 min. For the first ten epochs of CNN, the training and validation accuracies, and losses are shown in Fig. 4.

From Fig. 4, validation loss and validation accuracy both are in sync with training loss and training accuracy. The model is not overfitting from the reason that validation loss is decreasing and the gap between training and validation accuracies are relatively small. Unlike the accuracy, loss is not a percentage. The training loss is a summation of the errors made for each example in training or validation sets. Loss value implies how well or poorly a certain model behaves after each iteration of optimization. Ideally, one would expect the reduction of loss after each, or several, iteration(s). The accuracy of a model is usually determined after the model parameters are learned and fixed and no more learning is taking place, i.e., loss cannot further decrease. In this study, the outcome produces an accuracy above 99.5% and loss less than 0.02. Therefore, after the model is converged, the proposed system can produce a good human motion behaviors detection result.

B. Testing Results

The testing is conducted in the laboratory of the Department of Electrical and Computer Engineering of the University of Arizona. The human motion behavior testing is conducted by a researcher who is trying to move into the detection area and perform different behaviors, such as walking and waving hands. The mmWave radar is set at the height of 80 cm and pointing to the room area. The system running ROS human motion detection node is in real time, and the experiment shows the real-time motion behavior detection of people. The actual



(a) (b)

Fig. 5. Real-time micro-Doppler prediction scene. (a) Camera vision. (b) Radar point cloud output with Doppler velocities in meter per second.

Table 2. CNN Prediction Accuracies.[a]

Layer (type)	Accuracy
Human walking and vanish from radar	96.32%
Human waving hands when standing or sitting	99.59%
Human sitting to standing and walking transition	64%
Human walking back and forth	91.18%
No micro-Doppler detections	97.84%
Complex detections including all behaviors	95.19%

[a]Collected multiple test samples and averaging results among samples.

test is shown in Fig. 5. In Fig. 5(a), the camera vision is provided for reference. For this project, the CNN is only applied on micro-Doppler signatures, and at the testing stage, vision analysis can provide a good reference. In Fig. 5(b), the detected human with Doppler information is provided. Bulk motion of body is detected, as well as the micro-motions from arms and legs. Different parts of the body are clearly detected with appropriate configuration on chirps and CFAR.

The testing results of CNN prediction on human motion behavior monitoring using mmWave radar can be seen in Table 2. Each accuracy is calculated from the prediction from node and the actual behavior labelled from host.

The testing results in Table 2 show a very precise prediction over different continuous behaviors. The low accuracy may be due to the insufficient data in training for the transition behaviors, such as standing and walking. These behaviors happen in a short amount of time, and we are training using only 1896 samples, instead of 9973 samples for other behaviors. For other behaviors, the accuracy is relatively high. As micro-Doppler signatures are far more than human motion behaviors, such as walking and waving hands, the proposed system can be extended to include other behaviors.

IV. CONCLUSION

In this article, a real-time human motion behavior monitoring system using mmWave radar is introduced. It can perform as a ward in an interested area to collect human motion behavior information. Future work can apply this application to automotive radars, which allows automotive radar to classify car, pedestrian, and bicyclist. In addition, more human motion behavior can be added to the network for better recognition of the environment and the detection.

REFERENCES

[1] E. Benli, Y. Motai, and J. Rogers, “Human behavior-based target tracking with an omni-directional thermal camera,” in *Proc. IEEE Trans. Cogn. Developmental Syst.*, 2017, doi: 10.1109/TCDS.2017.2726356.

[2] M. S. Hossain, “Patient status monitoring for smart home healthcare,” in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, 2016, pp. 1–6.

[3] D. Schreurs, M. Mercuri, P. J. Soh, and G. Vandenbosch, “Radar-based health monitoring,” in *Proc. IEEE MTT-S Int. Microw. Workshop Ser. RF Wireless Technol. Biomed. Healthcare Appl.*, Singapore, 2013, pp. 1–3.

[4] M. Caris, S. Stanko, S. Palm, R. Sommer, and N. Pohl, “Synthetic aperture radar at millimeter wavelength for UAV surveillance applications,” in *Proc. IEEE Int. Forum Res. Technol. Soc. Ind. Leveraging Better Tomorrow*, Turin, Italy, 2015, pp. 349–352.

[5] R. J. Evans *et al.*, “Consumer radar: Opportunities and challenges,”in *Proc. 11th Eur. Radar Conf.*, Rome, Italy, 2014, pp. 5–8.

[6] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, “Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band,” *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 845–860, Mar. 2012.

[7] G. E. Smith, K. Woodbridge, and C. J. Baker, “Radar micro-Doppler signature classification using dynamic time warping,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 3, pp. 1078–1096, Jul. 2010.

[8] A. K. Singh and Y. H. Kim, “Analysis of human kinetics using millimeter-wave micro-Doppler radar,” *Procedia Comput. Sci.*, vol. 84, pp. 36–40, 2016.

[9] T. Wagner, R. Feger, and A. Stelzer, “Radar signal processing for jointly estimating tracks and micro-Doppler signatures,” *IEEE Access*, vol. 5, pp. 1220–1238, 2017.

[10] D. Belgiovane and C. Chen, “Micro-Doppler characteristics of pedestrians and bicycles for automotive radar sensors at 77 GHz,” in *Proc. IEEE Eur. Conf. Antennas Propag.*, Paris, France, 2017, pp. 2912–2916.

[11] A. Seifert, A. M. Zoubir, and M. G. Amin, “Radar classification of human gait abnormality based on sum-of-harmonics analysis,” in *Proc. IEEE Radar Conf.*, Oklahoma City, OK, USA, 2018, pp. 940–945.

[12] A. Shrestha, J. Le Kernec, F. Fioranelli, E. Cippitellii, E. Gambi, and S. Spinsante, “Feature diversity for fall detection and human indoor activities classification using radar systems,” in *Proc. Int. Conf. Radar Syst.*, 2017, pp. 1–6.

[13] V. C. Chen, F. Li, S. Ho, and H. Wechsler, “Micro-Doppler effect in radar: Phenomenon, model, and simulation study,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, Jan. 2006.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst.*, vol. 1, pp. 1097–1105, 2012.

[15] D. Maturana and S. Scherer, “VoxNet: A 3D convolutional neural network for real-time object recognition,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 922–928.