Human Motion Recognition Exploiting Radar with Stacked Recurrent Neural Network

Mingyang Wang[a], Yimin D. Zhang[b,*], Guolong Cui[a,*]

[a] *School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731 China*
[b] *Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA, 19132 USA*

Abstract

We develop a novel radar-based human motion recognition technique that exploits the temporal sequentiality of human motions. The stacked recurrent neural network (RNN) with long short-term memory (LSTM) units is employed to extract sequential features for automatic motion classification. The spectrogram of raw radar data is used as the network input to utilize the time-varying Doppler and micro-Doppler signatures for human motion characterization. Based on experimental data, we verified that a stacked RNN with two 36-cell LSTM layers successfully classifies six different types of human motions.

Keywords: Human motion recognition, radar, deep learning, recurrent neural network, long short-term memory

1. Introduction

Human motion recognition becomes increasingly attractive in many applications, such as computer gaming, smart home, elderly care, kinesiology, and secure surveillance, and is valuable in improving the quality of entertainment experiences and living quality [1, 2, 3, 4, 5]. A number of techniques have been developed with different types of sensors to sense human motions. These techniques can be classified into two major groups: wearable and unwearable. Wearable methods capture the details of human actions since sensors are directly attached to a body part[6]. However, wearable sensors are often forgotten, and easy worn out. As a result, wearable methods are inconvenient and unreliable.

Unwearable methods, on the other hand, overcome these issues and therefore become a more preferred choice[1, 2, 3, 4, 5, 7]. Unwearable sensors are typically installed in fix positions. One such technique is based on computer vision[8, 9, 10, 11, 12, 13]. By using optical and depth cameras, human motion

*Corresponding author
Email addresses: mingyangwang.uestc@gmail.com (Mingyang Wang), ydzhang@temple.edu (Yimin D. Zhang), cuiguolong@uestc.edu.cn (Guolong Cui)

features can be extracted from images and video frames for motion classification and recognition. Nevertheless, computer vision techniques are strictly limited by the light conditions and raise significant privacy concerns. Radar technique conquers these issues because it does not depend on lighting conditions and nor does it raise privacy concerns[14]. The radar senses the Doppler and micro-Doppler signatures associated with human motion[15, 16]. Specifically, Doppler refers to the frequency change due to the motion of the torso, whereas micro-Doppler refers to that due to other body parts, such as the limbs. The time-varying spectrum, analyzed using the spectrogram[17], represents all instantaneous Doppler and micro-Doppler components at each time [1, 18, 19]. Based on such spectrum, hand-crafted features, such as extreme frequency ratio[1], Boulic human walking model [20], the step repetition obtained from the cadence velocity spectrum[21, 22], Doppler frequency of trunk and arm movements[23], total Doppler bandwidth[24], can be designed to characterize the motions. However, as Doppler and micro-Doppler signatures vary with each individual, even for the same type of motions, motion classification based on hand-crafted features is not reliable. Therefore, automated and optimized feature extraction is desired for more reliable human motion characterization and classification.

Toward this end, the recently developed deep learning methods are capable to automatically learn and represent the more general and accurate characteristics from the measured data[25]. Inspired by the information processing mechanism of human nervous systems for vision and hearing, deep structures with multiple layers can be constructed and employed to extract hierarchical features of data. Such techniques have been applied in image classification, object detection, speech recognition and natural language processing with a great success[25]. Convolution neural network (CNN) is a commonly used deep learning method for image classification and object recognition without the need of designing hand-crafted features[26]. Because of the grid structure of images, CNN can automatically learn and extract structural features from a series of local small regions in an image[25, 26, 5]. For radar based human motion recognition, the Doppler signatures are usually represented in a two-dimensional joint time-frequency domain, which can be treated as an image. Hence, many of the works exploit CNN to extract motion features from such time-frequency images[27, 2, 4]. However, it is noted that every type of motion consists of a chain of coherent postures. Thus, in the underlying application, the radar echoes of human motions have high temporal sequentiality rather than the spatial grid structure in images. However, CNN cannot take the advantage of such sequentiality.

Motivated by this fact, we opt for the recurrent neural network (RNN) method which accounts for the sequential information and memorizes the relationship between the current and the historical inputs[28, 29]. In other words, RNN treats the inputs with a sequential order as state transitions. By utilizing such temporal correlation, RNN can learn the time-varying dynamic signatures and make use of the sequentiality of human motions to improve the classification and recognition performance.

In this paper, we propose a novel human motion recognition method that

employs a stacked RNN with long short-term memory (LSTM) units to automatically recognize human motions based on radar signals. In particular, we construct, train, and test a stacked RNN with two 36-cell LSTM layers to classify and recognize the following six motions: boxing, hand clapping, hand waving, piaffe (walking at a fixed spot), jogging, and walking. The advantages and the effectiveness of the proposed method are verified using experimental data.

The rest of this paper are arranged as follows. In Section 2, the signal model of the radar echoes reflected from a human body is established. Then, we propose the recognition method based on stacked RNN and describe the principle of the network in Section 3. In Section 4, experimental results using measured radar data are presented to validate the proposed method. Section 5 concludes this paper.

2. Signal Model

Consider a continuous-wave radar transmitting a sinusoidal waveform with frequency f_c, expressed as $s(t) = \exp(\mathtt{j}2\pi f_c t)$. For an ideal point target p at a range of R_{p0} at initial time t_0 moving with velocity $v_p(t)$ toward a direction $\varphi_p(t)$ with respect to the radar line of sight, the echo reflected from the target at time t is expressed as

$$s_p(t) = A_p(t) \exp\left[\mathtt{j}2\pi f_c \left(t - \frac{2R_p(t)}{c}\right)\right], \tag{1}$$

where $A_p(t)$ is the amplitude of the echo at time t, c is the velocity of electromagnetic wave propagating in air, and $R_p(t)$ is the instantaneous target distance, expressed as

$$R_p(t) = R_{p0} + \int_{t_0}^{t} v_p(\tau) \cos\varphi_p(\tau) \mathrm{d}\tau. \tag{2}$$

Consequently, the Doppler frequency introduced by the relative motion between the target and the radar is

$$f_D(t)_p = \frac{2v_p(t)\cos\varphi_p(t)}{c} f_c. \tag{3}$$

Since a human body can be treated as an intricate reflecting object comprised by many points, the receiving signal of the radar reflected from the human body is the summation of the echoes from all point-like targets, expressed as

$$s_{\mathrm{body}}(t) = \int_{\Sigma(t)} A_p(t) \exp\left\{\mathtt{j}2\pi f_c \left(t - t_p\right)\right\} \mathrm{d}p, \tag{4}$$

where

$$t_p = \frac{2R_p(t)}{c} = \frac{2R_{p0}}{c} + \int_{t_0}^{t} \frac{f_D(\tau)_p}{f_c} \mathrm{d}\tau, \tag{5}$$

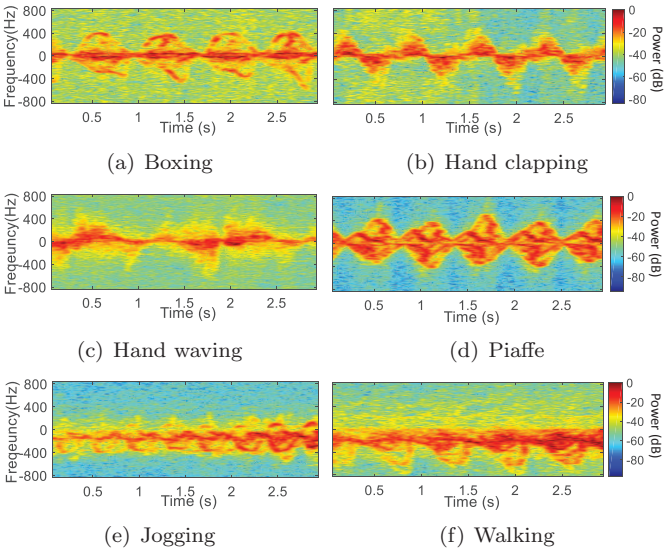
and $\Sigma(t)$ denotes the collection of scattering points in the human body at time t.

Since the velocities of distinct body parts are different and vary with time, the Doppler frequency $f_D(\tau)_p$ in Eq. (5) changes with the position of p as well as the time. As such, time-frequency spectrum is suitable to characterize the Doppler and micro-Doppler signatures. In this paper, spectrogram generated from the short-time Fourier transform is adopted, expressed as

$$S_{\text{body}}(t,f) = \left| \int_{-\infty}^{\infty} \tilde{s}_{\text{body}}(t) h(t-\tau) e^{-\text{j}2\pi f\tau} \text{d}\tau \right|^2, \tag{6}$$

where $\tilde{s}_{\text{body}}(t)$, the base-band signal corresponding to the echo $s_{\text{body}}(t)$ in Eq. (4), contains the micro-Doppler information of the motions, and $h(t)$ is a window function.

Fig. 1 shows spectrogram examples of the following six human motions: boxing, hand clapping, hand waving, piaffe (walking on a fixed spot), jogging, and walking. The details about the experimental setting is provided in Section IV-A. It is clear that each motion type has a distinct time-Doppler pattern. However, Doppler and micro-Doppler signatures vary with each individual, even for the same type of motions. Hence, automated and optimized feature extraction is desired for more reliable human motion characterization and classification.



(a) Boxing (b) Hand clapping

(c) Hand waving (d) Piaffe

(e) Jogging (f) Walking

Figure 1: Spectrograms of human motions

3. Stacked RNN with LSTM

3.1. RNN

RNN is a deep learning algorithm with a recurrent feedback structure[30]. A classical structure of RNN, which is exemplified in Fig. 2, contains three

layers: input layer, hidden layer, and output layer. Each neuron in the hidden layer includes a state feedback structure, which enables RNN to memorize historical information transformed from the input data. Therefore, RNN is more suited to deal with sequential data. Another important feature of the RNN is that, because of its recurrent structure, it can process sequences with different lengths[30].

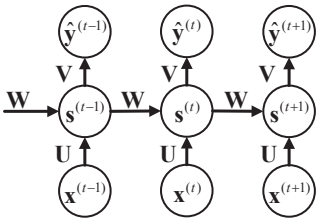


Figure 2: The RNN structure with a single hidden layer in an unfolded form

It can be observed in Fig. 1 that the radar echoes of human motions have high temporal sequentiality, and the lasting time of each motion varies. As such, the human motion spectrogram results can be treated as length-varying temporal sequences whose elements at a specific time instance form a Doppler frequency vector. Therefore, by using RNN to extract radar based human motion features with sequentiality, improved motion classification and recognition can be achieved.

In Fig. 2, the network input at time t is $\boldsymbol{x}^{(t)}$. The output of the hidden layer, $\boldsymbol{s}^{(t)}$, referred to as the state, is the result of nonlinear mapping with respect to the weighted sum of the current network input and the historical state of the previous instant, expressed as

$$\boldsymbol{s}^{(t)} = \boldsymbol{f}(\mathbf{U}\boldsymbol{x}^{(t)} + \mathbf{W}\boldsymbol{s}^{(t-1)} + \boldsymbol{b}_s^{(t)}). \tag{7}$$

Finally, the network estimation $\hat{\boldsymbol{y}}^{(t)}$ is obtained by the similar nonlinear mapping relative to the weighted sum of states, expressed as

$$\hat{\boldsymbol{y}}^{(t)} = \boldsymbol{f}(\mathbf{V}\boldsymbol{s}^{(t)} + \boldsymbol{b}_y^{(t)}). \tag{8}$$

3.2. LSTM Structure

A problem with RNN is that it cannot process long sequences because the gradient may vanish or explode during the training procedure, i.e., RNN has a short-term memory that either forgets historical information (for the gradient vanish situation) or only remembers the history at some initial instants (for the gradient explosion situation). To overcome such drawback of RNN, an LSTM structure, as shown in Fig. 3, was proposed in [31].

LSTM extends the structure of hidden neurons in RNN by introducing a novel unit called block. Within each block, one or more memory cells and multiple stream controlling gates are included. In each cell, the constant error carousel (CEC) structure shown in Fig. 3 recurrently works to update the

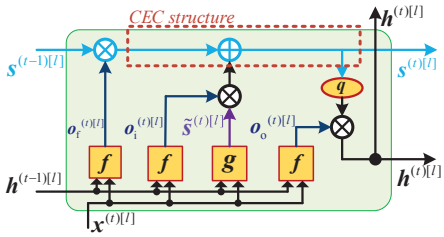


Figure 3: Structure of an LSTM block with a single cell

activation status, which is referred to as the cell state. CEC solves the vanishing gradient problem since an error stream flowing through the CEC structure maintains a constant value during the network training procedure. Multiple stream controlling gates, which can be shared by multiple cells in a single block, manage the upgrading of the information stream flowing through the CEC units. The following four gates are commonly used in an LSTM unit: input gate, output gate, forget gate, and state candidate. In this paper, only a single cell is contained in each memory block in order to simplify the complexity of network. Therefore, the workflow of an LSTM unit is expressed as (refer to Fig. 3)

$$\boldsymbol{o}_{\mathrm{f}}^{(t)[l]} = \boldsymbol{f}(\mathbf{W}_{\mathrm{f}_h}^{[l]}\boldsymbol{h}^{(t-1)[l]} + \mathbf{W}_{\mathrm{f}_x}^{[l]}\boldsymbol{x}^{(t)[l]} + \boldsymbol{b}_{\mathrm{f}}^{[l]}), \tag{9}$$

$$\boldsymbol{o}_{\mathrm{i}}^{(t)[l]} = \boldsymbol{f}(\mathbf{W}_{\mathrm{i}_h}^{[l]}\boldsymbol{h}^{(t-1)[l]} + \mathbf{W}_{\mathrm{i}_x}^{[l]}\boldsymbol{x}^{(t)[l]} + \boldsymbol{b}_{\mathrm{i}}^{[l]}), \tag{10}$$

$$\tilde{\boldsymbol{s}}^{(t)[l]} = \boldsymbol{g}(\mathbf{W}_{\tilde{\mathrm{s}}_h}^{[l]}\boldsymbol{h}^{(t-1)[l]} + \mathbf{W}_{\tilde{\mathrm{s}}_x}^{[l]}\boldsymbol{x}^{(t)[l]} + \boldsymbol{b}_{\tilde{\mathrm{s}}}^{[l]}), \tag{11}$$

$$\boldsymbol{s}^{(t)[l]} = \boldsymbol{o}_{\mathrm{f}}^{(t)[l]} \odot \boldsymbol{s}^{(t-1)[l]} + \boldsymbol{o}_{\mathrm{i}}^{(t)[l]} \odot \tilde{\boldsymbol{s}}^{(t)[l]}, \tag{12}$$

$$\boldsymbol{o}_{\mathrm{o}}^{(t)[l]} = \boldsymbol{f}(\mathbf{W}_{\mathrm{o}_h}^{[l]}\boldsymbol{h}^{(t-1)[l]} + \mathbf{W}_{\mathrm{o}_x}^{[l]}\boldsymbol{x}^{(t)[l]} + \boldsymbol{b}_{\mathrm{o}}^{[l]}), \tag{13}$$

$$\boldsymbol{h}^{(t)[l]} = \boldsymbol{o}_{\mathrm{o}}^{(t)[l]} \odot \boldsymbol{q}(\boldsymbol{s}^{(t)[l]}), \tag{14}$$

where superscript $[l]$ denotes the l-th hidden layer; $\boldsymbol{f}, \boldsymbol{g}$ and $\boldsymbol{q}$ stand for different activation functions; $\boldsymbol{s}$ denotes the cell states; $\mathrm{f}, \mathrm{i}, \tilde{\mathrm{s}}$ and o appearing in subscripts respectively represent the forget gate, input gate, state candidates, and output gate; and $\odot$ denotes the Hadamard product operation.

Eq. (12) shows the upgrading procedure of the cell states controlled by the forget gate and the input gate. When $\boldsymbol{o}_{\mathrm{f}}^{(t)[l]}$ approaches to zero, the forget gate closes and most historical memories are blocked to flow into the current cell. Thus, old memories are "forgotten" and their contributions less affect the current network output. On the contrary, when $\boldsymbol{o}_{\mathrm{f}}^{(t)[l]}$ approaches to one, the forget gate opens and the historical information is accessed to pass through the current cell. As such, old memories play a vital role in the network processing.

Similarly, the value of $\boldsymbol{o}_{\mathrm{i}}^{(t)[l]}$, which determines the opening or the closing of the input gate, controls appending of new information generated by state candidate $\tilde{\boldsymbol{s}}^{(t)[l]}$ to the cell state. Consequently, the LSTM structure can well learn the dynamic signatures from long time-dependent sequential data.

3.3. Proposed Structure

Since abstracted features can be extracted at each time instance, the output of an LSTM unit is also a sequence. Thus another LSTM layer can be stacked on the top of the current LSTM layer to extract more generalized sequential features. Based on this idea, we propose a novel method applying a deep neural network via stacking multiple RNN hidden layers with LSTM units to learn dynamic motion features and improve human motion classification and recognition. The proposed stacked RNN structure with two LSTM layers is shown in Fig. 4.

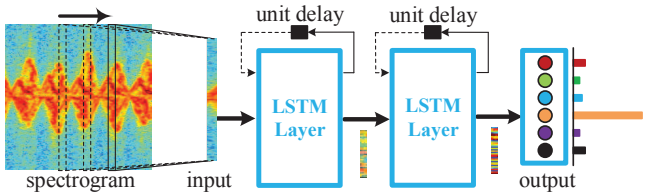


Figure 4: The structure of stacked RNN with LSTM layers

The raw radar data is first preprocessed to generate the spectrograms. Because of the high dynamic range of spectrograms, we perform the logarithm operation and normalization on spectrograms, and the results are fed into the network. Then, the stacked RNN with multiple LSTM layers extracts dynamic motion signatures. Finally, the output layer provides the probability of each motion and predicts a specific motion class at each time instant.

The stacked RNN with LSTM layers can be trained by utilizing the back propagation through time algorithm [32], and the network parameters are optimized through Adadelta algorithm[33] with an adaptive learning rate.

4. Experiments and Analysis

4.1. Experiment Setup

We exploit Ancortek Software Defined Radio (SDR) 2500B kit[34] as the radar to transmit continuous-wave sinusoidal signal with a carrier frequency of 25 GHz for sensing human motions. A sampling rate of 64 kHz is selected in order to ensure that detailed motions of human body parts are captured while keeping a minimum volume of acquired data for the subsequent process.

We establish a dataset for radar based human motion recognition at the Advanced Signal Processing Lab, Temple University. The experimental scene is shown in Fig. 5. During the data acquisition, each human subject performed

motions along the radar line of sight. The dataset contains the following six motion classes for two subjects: (a) boxing, (b) hand clapping, (c) hand waving, (d) piaffe, (e) jogging, and (f) walking. Subjects performed each motion for 3 seconds so as to contain enough motion cycles with necessary features. Each activity was repeated 100 times for each subject and the corresponding spectrograms were generated with an 8192-point Hamming window and a 90% overlap, which results in 1200 examples and each example contains 225 temporal frames. In addition, since the maximum velocity for all the six types of motions is no more than 5 meter/second, we discard the spectrogram data whose absolute velocity exceeds 5 meter/second. Therefore, the size of the preprocessed spectrogram is 225×214.

We employ 4-fold cross validation to evaluate the learning ability of the stacked RNN with LSTM units for human motion classification. The training and cross validation datasets respectively contain 900 and 300 examples. The number of hidden LSTM layers and the LSTM cell size are hyperparameters chosen by cross validation. The activation functions $\boldsymbol{g}$ and $\boldsymbol{q}$ are chosen as *tanh*, and $\boldsymbol{f}$ is chosen as *sigmoid*.

Because of the large volume of data and enormous number of network parameters, we use Keras [35] with Tensorflow [36] backend, and utilize NVIDIA GPU with CUDA library called cuDNN [37] to accelerate the training procedure. The learning rate of Adadelta [33] is set to 0.5. Early stopping and dropout [38] with a probability of 50% added to the last LSTM layer are applied to overcome overfitting problem. The core configures of computer used for training are NVIDIA GeForce GTX 1080 Ti GPU (with 11 GB memory) and 2.4 GHz Intel Xeon CPU E5-2640 v4.

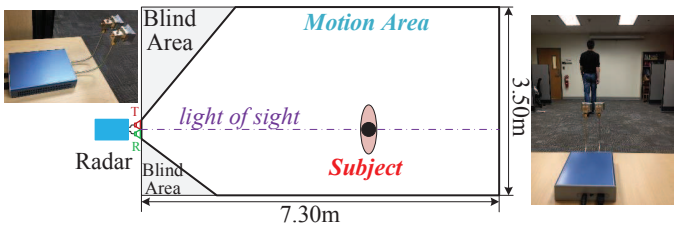


Figure 5: The setup of the experiments

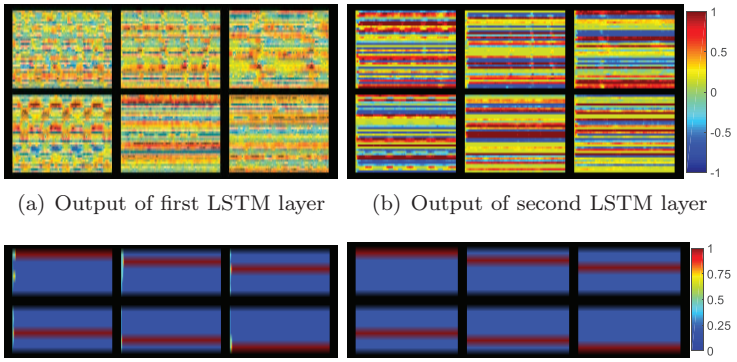
4.2. Hyperparameters and Extracted Features

After a heuristic search on the hyperparameters, the best model includes two hidden LSTM layers, each contains 36 cells. The average accuracy on the 4-fold cross validation set over all temporal frames is 98.97%. This accuracy verified that the stacked RNN with LSTM units is capable to classify human motions based on Doppler and micro-Doppler without manual feature extraction.

Fig. 6 shows the output of LSTM layers as well as the classified results of the output layer in the proposed network. The output of the first LSTM layer shown in Fig. 6(a) illustrates that the extracted features are of sequentiality,

and different motions have distinct feature sequences. From the output of the second LSTM layer depicted in Fig. 6(b), the features become more abstract and more sparse. Also, these features get less time-variant and yield more clear structures. As such, the function of the LSTM layer resembles that of an "encoder" to translate different motion signals into a set of distinct features or codes, each representing a motion type. The output layer acts as a "decoder" trained by those codes to interpret features into motion classes. When classifying an unknown motion signal, the whole network works like a "codec" to perform effective motion classification.

Comparing the classification results of the network depicted in Fig. 6(c) with the ground truth of the motions illustrated in Fig. 6(d), the proposed network can accurately classify the motion types at all instants except at some beginning instants. The misjudgments in these instants are due to the absence of any prior motion information because the memory of the network at the initial time is blank. As we continue to process the following motions, perfect classification is achieved during all the subsequent time instants.



(a) Output of first LSTM layer (b) Output of second LSTM layer

(c) Classified results of network (d) Ground truth

Figure 6: Extracted features and classified results of the stacked RNN with LSTM units: The six motions (boxing, hand clapping, hand waving, piaffe, jogging, and walking) are in the order of left to right and top to bottom.

4.3. Evaluation and Comparison on Testing Set

To illustrate the salient generalization performance of the stacked RNN with LSTM units, we evaluate the proposed model and compare the results with the output of a deep convolution neural network (DCNN), which is similar to the model in [2], on a group of testing data by using the average accuracy.

In order to make the distribution of the testing data consistent with that of the training data, the same six motions of training dataset are collected from the same object. We randomly concatenate the spectrogram results corresponding to the six motions and achieve a 135-second data stream. A 3-second sliding window is adopted and the data samples within the window is fed into the input layer in both the proposed model and the DCNN.

For fair comparison, the training and cross validation procedures for the DCNN are the same as those of our method. The examples fed into DCNN are identical to the ones fed into the proposed network. The fine-tuned DCNN contains three convolution layers with four 5×5 filters, three maxpooling layers with a 2×2 window and 2×2 strides, and two dense layers with 100 neurons and 6 neurons respectively. The average accuracy of the 4-fold cross validation is 98.98%. Therefore, the DCNN achieved the same level of training performance with essentially identical cross validation accuracy as compared with the proposed method.

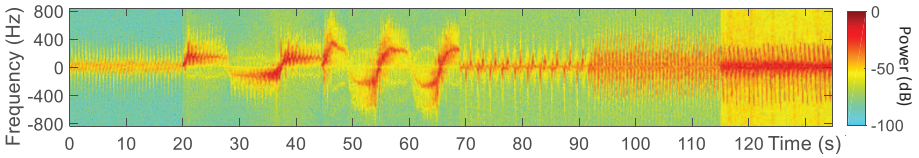
The classification results obtained using the two networks are shown in Fig. 7. From Fig. 7, the average accuracy of the testing classification is 92.65% for the stacked RNN with LSTM units and 82.33% for the DCNN. From the confusion matrices shown in Fig. 7(e) and (f), we can obtain that the micro-F1 score of our model, which is 0.4632, is higher than that of the DCNN, which is 0.4117. Moreover, the response time of the motion type switching for the DCNN is about 2.071 seconds, which is almost 4.34 times longer than that for the proposed model, which is about 0.4769 seconds. Since DCNN does not have the capability to use the sequentiality with memory, the generalization performance of the DCNN for radar based human motion recognition is lower than the proposed method. Therefore, the stacked RNN with LSTM units is more suitable for real-time operations than the DCNN. Moreover, the number of parameters in the DCNN, which is 314,822, is much higher than that in the proposed network, which is 47,878. Therefore, compared with the DCNN, the stacked RNN with LSTM units requires much less memory resources.

5. Conclusion

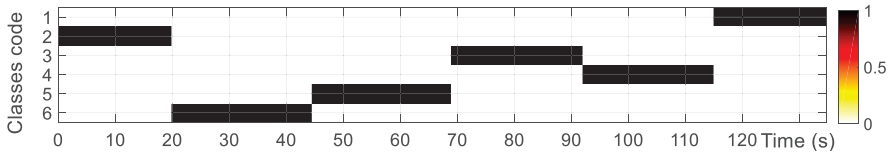
This paper addressed the problem of human motion recognition exploiting a Doppler radar. To utilize the temporal sequentiality of human motions, we proposed a novel human motion classification method which employs the stacked RNN with LSTM units to automatically extract sequential features and classify motion types. Since the time-varying Doppler and micro-Dopper signaures can represent human motions well, the spectrogram of raw radar data is used as the inputs of network. We considered six different motions, and trained a stacked RNN with two 36-cell LSTM layers. Experimental results verified attractive human motion recognition performance with an overall classification accuracy of 92.65%. In the future, we will apply and evaluate our proposed method for more motion types including some categories of fine, small and precise actions.

Acknowledgment

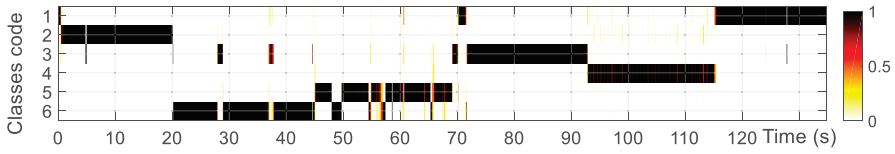
This work was supported by the China Scholarship Council for M. Wang's stay at Temple University, in part by Chang Jiang Scholar Program, in part by the National Natural Science Foundation of China under Grants 61771109, 61701088 and 61501083, and in part by the 111 project No.B17008. We thank



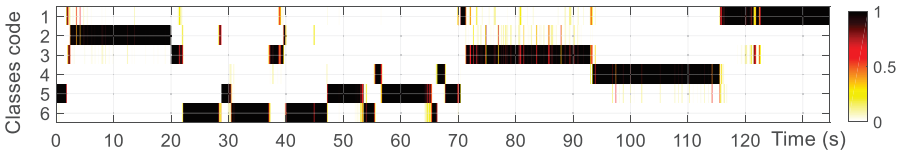
(a) The 135-second testing data stream



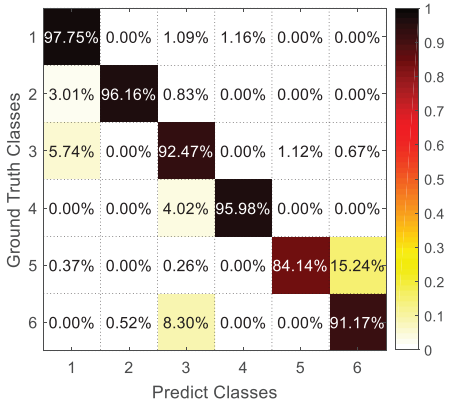
(b) Ground truth



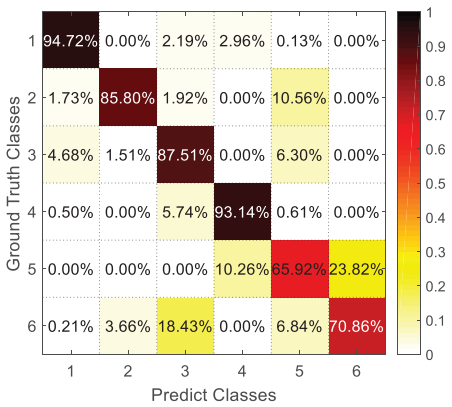
(c) Classified results of the stacked RNN with LSTM units



(d) Classified results of the DCNN



(e) Confusion matrix of the stacked LSTM



(f) Confusion matrix of the DCNN

Figure 7: Classification performance comparison between the stacked RNN with LSTM units and the DCNN. Classes codes represent 6 motions: 1 for boxing, 2 for hand clapping, 3 for hand waving, 4 for piaffe, 5 for jogging and 6 for walking.

the colleagues in the Advanced Signal Processing Lab at Temple University for their assistance in radar based human motion data collection.

References

[1] Q. Wu, Y. D. Zhang, W. Tao, M. G. Amin, Radar-based fall detection based on Doppler time–frequency signatures for assisted living, IET Radar, Sonar Navi. 9 (2) (2015) 164–172.

[2] Y. Kim, T. Moon, Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks, IEEE Geosci. Remote Sens. Lett. 13 (1) (2016) 8–12.

[3] M. G. Amin, Y. D. Zhang, F. Ahmad, K. D. Ho, Radar signal processing for elderly fall detection: The future for in-home monitoring, IEEE Signal Process. Mag. 33 (2) (2016) 71–80.

[4] B. Jokanovic, M. Amin, B. Erol, Multiple joint-variable domains recognition of human motion, in: Proc. IEEE Radar Conf., Seattle, WA, USA, 2017, pp. 948–952.

[5] E. Mason, B. Yonel, B. Yazici, Deep learning for radar, in: Proc. IEEE Radar Conf., Seattle, WA, USA, 2017, pp. 1703–1708.

[6] G. Orengo, A. Lagati, G. Saggio, Modeling wearable bend sensor behavior for human motion capture, IEEE Sensors J. 14 (7) (2014) 2307–2316.

[7] R. Lun, W. Zhao, A survey of applications and human motion recognition with microsoft kinect, Int. J. of Pattern Recognit. Artif. Intell. 29 (5) (2015) 1555008.

[8] U. Güdükbay, İbrahim Demir, Y. Dedeoğlu, Motion capture and human pose reconstruction from a single-view video sequence, Digital Signal Process. 23 (5) (2013) 1441–1450.

[9] T. Hachaj, M. R. Ogiela, Full body movements recognition unsupervised learning approach with heuristic r-gdl method, Digital Signal Process. 46 (2015) 239–252.

[10] F. Harjanto, Z. Wang, S. Lu, A. C. Tsoi, D. D. Feng, Investigating the impact of frame rate towards robust human action recognition, Signal Process. 124 (2016) 220–232.

[11] Y. Ji, Y. Yang, X. Xu, H. T. Shen, One-shot learning based pattern transition map for action early recognition, Signal Process. 143 (2018) 364–370.

[12] X. Ji, J. Cheng, W. Feng, D. Tao, Skeleton embedded motion body partition for human action recognition using depth sequences, Signal Process. 143 (2018) 56–68.

[13] Y. Zheng, H. Yao, X. Sun, S. Zhao, F. Porikli, Distinctive action sketch for human action recognition, Signal Process. 144 (2018) 323–332.

[14] Z. Chen, G. Li, F. Fioranelli, H. Griffiths, Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks, IEEE Geosci. Remote Sens. Lett. 15 (5) (2018) 669–673.

[15] B. Jokanovic, M. G. Amin, Y. D. Zhang, F. Ahmad, Multi-window time-frequency signature reconstruction from undersampled continuous-wave radar measurements for fall detection, IET Radar Sonar Navig. 9 (2) (2015) 173–183.

[16] S. Z. Gürbüz, C. Clemente, A. Balleri, J. J. Soraghan, Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems, IET Radar Sonar Navig. 11 (1) (2017) 107–115.

[17] B. Boashash, Time-frequency signal analysis and processing: a comprehensive reference, Academic Press, 2016.

[18] V. C. Chen, F. Li, S.-S. Ho, H. Wechsler, Micro-Doppler effect in radar: Phenomenon, model, and simulation study, IEEE Trans. Aerosp. Electron. Syst. 42 (1) (2006) 2–21.

[19] F. H. C. Tivive, A. Bouzerdoum, M. G. Amin, A human gait classification method based on radar doppler spectrograms, EURASIP J. Adv. Signal Process. 2010 (1) (2010) 389716.

[20] R. Boulic, N. M. Thalmann, D. Thalmann, A global human walking model with real-time kinematic personification, The Visual Comput. 6 (6) (1990) 344–358.

[21] R. Ricci, A. Balleri, Recognition of humans based on radar micro-Doppler shape spectrum features, IET Radar Sonar Navig. 9 (9) (2015) 1216–1223.

[22] B. Erol, S. Z. Gürbüz, M. G. Amin, Automatic data-driven frequency-warped cepstral feature design for micro-doppler classification, IEEE Trans. Aerosp. Electron. Syst. 54 (4) (2018) 1724–1738.

[23] F. Li, C. Yang, Y. Xia, X. Ma, T. Zhang, Z. Zhou, An adaptive S-method to analyze micro-doppler signals for human activity classification, Sensors 17 (29) (2017) 2769–1–18.

[24] M. Zenaldin, R. M. Narayanan, Radar micro-doppler based human activity classification for indoor and outdoor environments, Proc. SPIE 98291B (2016) 1–10.

[25] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[26] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

[27] M. S. Seyfioğlu, S. Z. Gürbüz, Deep neural network initialization methods for micro-Doppler classification with low training sample support, IEEE Geosci. Remote Sens. Lett. 14 (12) (2017) 2462–2466.

[28] M. Zhang, Y. Yang, Y. Ji, N. Xie, F. Shen, Recurrent attention network using spatial-temporal relations for action recognition, Signal Process. 145 (2018) 137–145.

[29] M. Wang, G. Cui, X. Yang, L. Kong, Human body and limb motion recognition via stacked gated recurrent units network, IET Radar, Sonar Navi. 12 (9) (2018) 1046–1051.

[30] Z. C. Lipton, A critical review of recurrent neural networks for sequence learning (2015).
URL http://arxiv.org/abs/1506.00019

[31] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (8) (1997) 1735–1780.

[32] P. J. Werbos, Backpropagation through time: What it does and how to do it, Proc. IEEE 78 (10) (1990) 1550–1560.

[33] M. D. Zeiler, ADADELTA: An adaptive learning rate method (2012).
URL https://arxiv.org/abs/1212.5701

[34] Ancortek Inc. [link].
URL http://ancortek.com/sdr-kit-2500b

[35] F. Chollet, et al., Keras (2015).
URL https://github.com/fchollet/keras

[36] M. Abadi, et al., TensorFlow: Large-scale machine learning on heterogeneous systems (2015).
URL https://www.tensorflow.org

[37] S. Chetlur, et al., cuDNN: Efficient primitives for deep learning (2014).
URL https://arxiv.org/abs/1410.0759

[38] N. Srivastava, et al., Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learning Research 15 (1) (2014) 1929–1958.