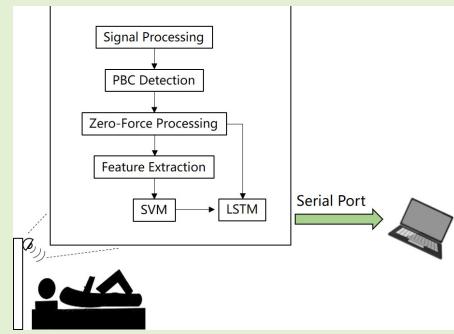


A Two-Stage Low-Complexity Human Sleep Motion Classification Method Using IR-UWB

Jialei Lai^{ID}, Zhaocheng Yang^{ID}, Member, IEEE, and Boning Guo

Abstract—With increasing concern about sleep quality, human sleep motion recognition has come into sight. Recent years have witnessed that almost all researches use deep neural networks for human motion recognition based on radar sensors. However, it is difficult to deploy these methods in edge computing in the case of limited hardware resources. This paper proposes a two-stage low-complexity method for real-time human sleep motion classification using an ultra-wideband (UWB) radar, which can be applied in edge computing. Firstly, we implement power burst curve (PBC) detection to decide whether sleep motions occur. Secondly, a method named zero-forcing processing for radar spectrum is proposed in order to reduce the number of parameters of classifiers. Later, in stage one, we propose a three-feature extraction approach based on micro-Doppler signatures and utilize support vector machine (SVM) to roughly classify micro-motion and body-movement motions. In stage two, a long short-term memory (LSTM) neural network with fewer parameters is applied to further classify body movements. Experimental results show that the proposed method has a low computational complexity, and can reach the average accuracy of the five typical sleep motions over 97% using train and test split samples, and 91.50% using the leave-two-subject-out (LTSO) samples.

Index Terms—Sleep motion detection and classification, micro-doppler signatures, low-complexity classifiers, edge computing, UWB radar.



I. INTRODUCTION

SINCE sleep quality has been more and more concerned with the increase of work and study pressure, sleep monitoring has been a hot topic. In addition to detecting respiratory and heart rate [1]–[3] when sleeping, sleep activity is also an important indicator of sleep quality. In contrast to optical sensors, radar sensors are qualified to work free from light when people sleep at night, and have a great benefit of privacy protection. Thus, more and more researchers focus on utilizing radar to classify human activities.

Almost all researchers combined the micro-Doppler signatures [4]–[6] with machine learning methods to classify human motions. Before the popularity of deep learning [7],

traditional machine learning methods [8]–[10] were usually utilized. [8] developed a support vector machine (SVM) for motion classification by extracting 6 micro-Doppler features. Besides, via some handcrafted micro-Doppler signatures, naive Bayesian [9] and K-nearest neighbours (KNN) [10] were utilized to classify unarmed/armored personnel outdoors and human motions, respectively. The main advantage of traditional methods is the low-complexity compared with deep learning, but the accuracy may reach a relatively low level.

With rapidly development of deep learning, many researches combined micro-Doppler signatures with deep neural networks to classify human motions [11]–[23]. [11] utilized deep convolutional neural networks (DCNN) for human activity classification based on micro-Doppler signatures. In addition, [12] and [13] employed DCNN for human gait recognition. In [14], a convolutional autoencoder (CAE) was proposed for human motion classification. Besides, [15] used generative adversarial networks (GAN) for radar image generation, and later DCNN was selected to classify human motions. In [17], a residual network (Res-net) was applied to classify different motions, which is a novel structure of DCNN. [18] proposed a novel end-to-end convolutional network named RadarNet for daily activity and sleep activity classification. [19] developed a real-time system of human motion detection using CNN based on a millimetre wave radar. [20] also used the millimetre wave

Manuscript received June 26, 2021; accepted July 23, 2021. Date of publication July 27, 2021; date of current version September 15, 2021. This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515011517, in part by the Science and Technology Project of Shenzhen under Grant JCYJ20190808142803565, and in part by the National Natural Science Foundation of China under Grant 61771317. The associate editor coordinating the review of this article and approving it for publication was Dr. Ying Zhang. (*Corresponding author: Zhaocheng Yang.*)

The authors are with the Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: ljl6394@163.com; yangzhaocheng@szu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2021.3100635

radar for fall detection via line kernel convolutional neural networks. Furthermore, [21] and [22] used a long short-term memory (LSTM) networks for target classification due to the characteristic of radar image.

However, it is difficult to employ deep convolutional neural networks in edge computing in the case of limited hardware resources and cost. One approach for performing DCNN is to process raw data in cloud servers, but it may cause large time delay and raw data dropout. The edge computing means that almost all algorithms are embedded in an edge device, which meets the needs of fast real-time data processing and can prevent raw data dropout. With limited hardware resources and cost, one practicable method for edge computing is to select low-complexity machine learning methods.

Considering the low-complexity of traditional method and high accuracy of deep learning, this paper implements a two-stage real-time sleep motion classification method with a low-complexity machine learning in edge computing. Firstly, sleep motion detection named power burst curve (PBC) detection is performed to decide whether sleep motions occur when sleeping. Secondly, this paper proposes a method named zero-forcing processing for radar spectrum to reduce the number of parameters of classifiers. Later, a low-complexity machine learning method which combines SVM with LSTM neural network for human sleep motion classification is proposed. In particular, in stage one, three handcrafted micro-Doppler features are proposed for SVM to roughly classify micro-motion and body movements. In stage two, a LSTM is applied to further classify body movements. Comparing with other methods, this method has the low-complexity and real-time ability, and can reach the average accuracy of the five typical sleep motions over 97% using train and test split samples, and 91.50% using the leave-two-subject-out (LTSO) samples.

The paper is presented as follows. Section II will describe the radar system and signal model. In section III, the proposed two-stage sleep motion classification is given. It consists of sleep motion detection, zero-forcing spectrum processing, SVM-based rough classification and LSTM-based body-movement classification. The real-time system is detailed in section IV. Experimental results will be presented in section V. Finally, the conclusion will be drawn in section VI.

II. RADAR SYSTEM AND SIGNAL MODEL

In this section, the radar system and signal model will be introduced. For the safety of human body, we select X4M03 radar [24] for human detection and motion classification. The radar is an impulse radio ultra-wideband (IR-UWB) radar. The power consumption of the radar is only 118 mW with low radiation. The raw echo signal is defined as radio frequency (RF) signal. By analyzing RF signal, sleep motions will be detected and classified.

The X4M03 radar operates with carrier frequency of 7.29GHz and bandwidth of 1.4GHz for data collection. With a sampling frequency of 23.328GHz, the number of maximum sample points reaches 1536 per frame. The maximum detection range is 9.9m, and the range between each sample point is 0.00644m. With the bandwidth of 1.4GHz, the range resolution is 0.107m. **Table I** depicts the parameters configured. The raw

TABLE I
RADAR PARAMETERS CONFIGURED

Radar parameters	Value	Unit
Carrier frequency	7.29	GHz
Bandwidth	1.4	GHz
Range resolution	0.107	meter
FPS for PBC detection	17	frame
FPS for spectrum	102	frame
Power consumption	118	mW
Main-lobe width	-60~60(-10dB)	degree
Detection range	2.51	meter

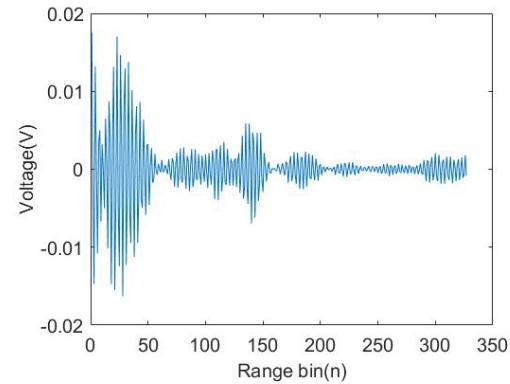


Fig. 1. The RF signal of one frame.

echo signal of X4M03 is a Gaussian pulse signal defined by:

$$g(t - t_r) = V \exp\left(-\frac{(t - t_r)^2}{2\tau^2}\right) \cos(2\pi f_c(t - t_r)) \quad (1)$$

where V denotes pulse amplitude, t means fast time, t_r is time delay of the transmit signal, and f_c denotes carrier frequency. Besides, τ determines the -10 dB bandwidth:

$$\tau = (2\pi f_B (\log_{10}(e))^{\frac{1}{2}})^{-1} \quad (2)$$

Here, f_B is bandwidth, and e is Euler's number.

After hardware sampling in X4M03, the raw RF signal is defined as $r(m, n)$, where m denotes the frame index and n denotes the range bin. **Fig. 1** shows the signal containing all range bins at one frame.

In order to remove static target and suppress the background clutter, a method named running average [25] is applied. The method has a low computation complexity, and is much suitable for this real-time application. The running average algorithm is described as

$$c(m, n) = \alpha c(m - 1, n) + (1 - \alpha)r(m, n), \quad (3)$$

$$y(m, n) = r(m, n) - c(m, n), \quad (4)$$

where $c(m, n)$ is the background clutter signal, $y(m, n)$ denotes the clutter-subtracted signal and α is an update coefficient. We set α as 0.95, meaning that micro-Doppler features will arise rapidly while the clutter will remain longer.

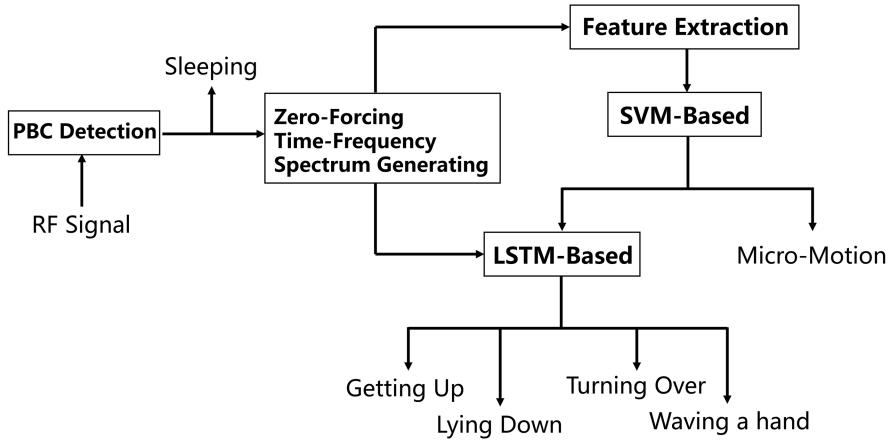


Fig. 2. The workflow of proposed two-stage sleep motion classification method.

III. PROPOSED TWO-STAGE HUMAN SLEEP MOTION CLASSIFICATION

After clutter suppression, the RF signal is used to perform the proposed two-stage sleep motion classification. This section will detail the proposed method, of which the workflow is shown in Fig. 2.

A. Power Burst Curve Detection

After clutter suppression, the RF signal of frames per second (FPS) 17 is converted to range-frequency map using fast Fourier transform at each range bin for human motion detection. The range-frequency map is described as

$$X(f, n) = \sum_m y(m, n) e^{-j2\pi f m}, \quad (5)$$

where f denotes frequency.

When sleeping, the frequency of human body is mainly produced by respiration and heartbeat whose frequency ranges from $\pm 0.2\text{Hz}$ to $\pm 2\text{Hz}$. Beyond these frequencies, the power of other frequencies is mainly produced by body movements, which is named as power burst curve (PBC). The PBC is used to detect whether human motion occurs when sleeping and is defined by

$$Z(n) = \ln(\sum_{f=\Psi} |X(f, n)|) \quad \Psi = \{|f| > \gamma\}, \quad (6)$$

where γ is defined as high frequency threshold and is set as 2Hz.

[26] proposed an adaptive double threshold constant-false-alarm-rate (ADT-CFAR) detector based on PBC. The detector firstly sums the PBC values of all range bins, and then counts these values. Later a probability distribution is fitted when body movements do not occur in a certain environment. When body movements occur, the detector can catch the change that the PBC values do not obey the probability distribution.

However, the statistical data of PBC may be different under various environments, and the probability distribution may show a relatively deviation. Therefore, this paper adopts a simple and effective method for PBC detection. When human body movement occurs, the frequency generated by body

movement is higher than that produced by respiration and heartbeat. What's more, the power of these frequencies is usually much higher than that produced by noise. Therefore, it is feasible to extract the power of the frequency generated by body movements.

For the sake of implementing real-time system, a sliding window of 1 second overlap is applied to calculate fast Fourier transform with fixed length of window (1.5 second), which is defined as short-time Fourier transform (STFT). Within one sliding window, we set a threshold η to decide whether body movement occurs. If the ratio of the power of high frequency to total power is more than the threshold, it is determined that the body movement occurs:

$$\frac{\sum_n Z(n)}{\ln(\sum_n \sum_f |X(f, n)|)} \begin{cases} \geq \eta, & H_1 \\ < \eta, & H_2 \end{cases} \quad (7)$$

where H_1 means that body movement occurs while H_2 means oppositely.

However, it is inevitable that the detector may lead to false alarm and missed alarm. In view of the range between each range bin is 0.00644m, the radar is much sensitive to tiny movements. Thus, the threshold η can be rather small, resulting that there are almost no missed alarms due to the high sensitivity of radar. Meanwhile, the detector is capable of catching some tiny movements because of the false alarm. To solve the problem of false alarms, a category named micro-motion is introduced. The micro-motion means that sleeper may be shaking body, talking in his dream, or performing other tiny movements.

B. Zero-Forcing Time-Frequency Spectrum Generating

If human body movement occurs, digital frequency mixing and low-pass filtering are performed to convert the RF signal to time-range (TR) matrix defined as $z(m, n)$. Specially, the FPS of RF signal for classification is set as 102 to avoid blind speed. The STFT is used to generate radar time-frequency

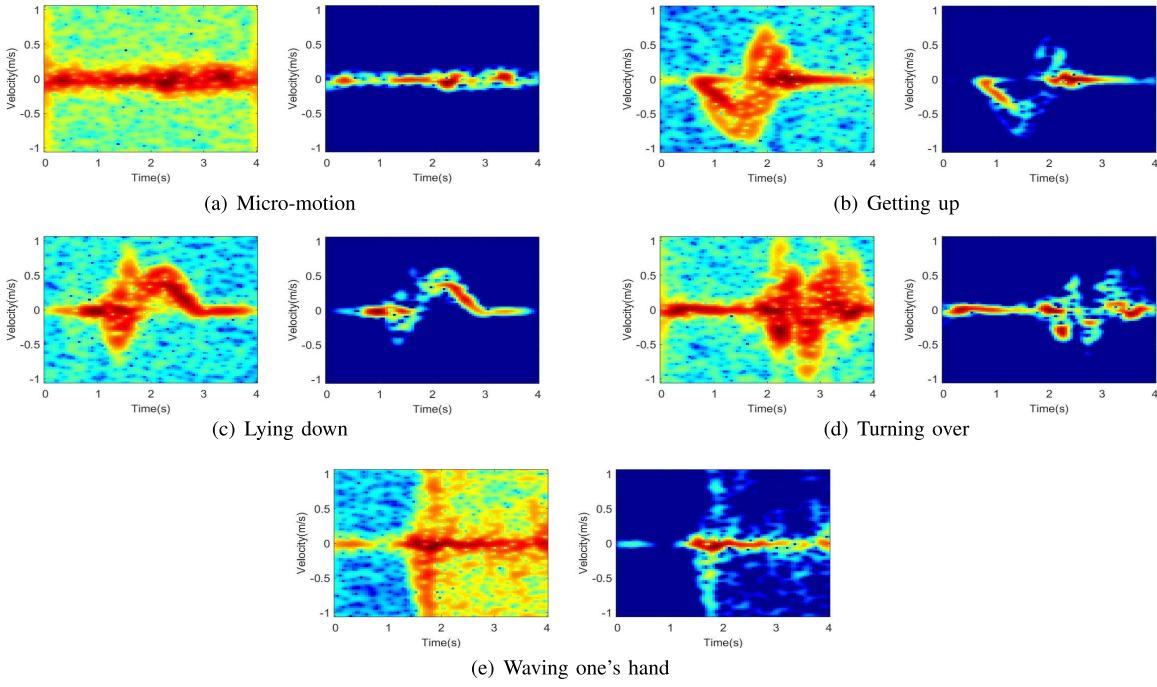


Fig. 3. The comparisons of spectrums containing all motions.

spectrum:

$$S(m, f_d) = \left| \sum_{l=-\infty}^{+\infty} \sum_{n=1}^R z(m+l, n) w(l) e^{-j2\pi f_d l} \right| \quad (8)$$

where $w(l)$ is a Gaussian window function, R is the number of range bins, and f_d denotes the Doppler frequency. The length of Gaussian window is 80, and 75-point overlap is applied.

Almost all existing methods for classification directly utilize radar spectrum without processing. Since deep neural networks have the powerful ability of self-learning and feature extraction, the classifiers are capable of recognition with high precision without regard to the signal-to-noise-ratio (SNR) of radar. Because of low radiation of X4M03 radar, the SNR will be also low, causing that the spectrum is full of noise. Therefore, to reduce the number of parameters of classifiers and deploy algorithms in edge computing, this paper proposes a method named zero-forcing processing for radar spectrum. The zero-forcing processing tries to force the values of noises on zero and keeps effective values of micro-Doppler signatures. The detailed method for zero-forcing processing is described in Algorithm 1. Fig. 3 (a)-(e) show micro-motion and other four body movements, including getting up, lying down, turning over and waving one's hand. In the figure, the left is raw spectrum which exists much noise and the right shows the spectrum processed that is cleaner.

After zero-forcing processing, the spectrum is $S'(m, f_d)$, and will be relatively clean and the difference of signatures between various classes will be enlarged. To take full advantages of signatures among these classes and reduce the computation, the low-complexity machine learning method is proposed. The low-complexity machine learning classifiers

consist of SVM and LSTM. The detailed classifiers will be drawn as follows.

C. Two-Stage Low-Complexity Classification Approach

If human motions occur, the SVM will firstly predict the presence of micro-motion or body movements. Only if the body movements occur, the LSTM will further classify all body movements. Thus, the proposed classifiers are capable of reducing the computation because the probability of performing micro-motion is much higher than that of performing body movements when sleeping. Five categories will be recognized, including micro-motion and four body movements.

1) Stage 1: SVM-Based Rough Classification: Firstly, the rough classification is carried out by SVM. According to micro-Doppler signatures generated by different motions, three features are proposed to classify micro-motion and body movements.

The first feature is the beyond-vital-sign Doppler power (BVSDP):

$$F_1 = \sum_{m=1}^M \sum_{f_d=4}^{F-4} S'(m, f_d), \quad \text{if } \sum_{f_d=1}^F S'(m, f_d) > 0, \quad (9)$$

where $S'(m, f_d)$ means the spectrum after zero-forcing processing, M is time sequences and F is the number of FFT points which is set to 80. The feature F_1 is extracted to catch the power of the Doppler frequencies that are beyond vital sign.

The second feature is the beyond-vital-sign Doppler information (BVSDI):

$$F_2 = - \sum_{f_d=4}^{F-4} \frac{S'(m, f_d)}{\sum_{f_d \in \varphi} S'(m, f_d)} \ln \left(\frac{S'(m, f_d)}{\sum_{f_d \in \varphi} S'(m, f_d)} \right), \quad (10)$$

Algorithm 1 Zero-Forcing Processing

Input: a matrix of raw spectrum described as S .
Output: a matrix of spectrum defined as S' after zero-forcing processing.

```

//row is the number of matrix rows and is equal to 80.
//col is the number of matrix columns or time sequences.
//Th1 is the noise threshold.
//Th2 is power threshold.
S' ← zeros matrix
max_power ← max(S)
S ← log10(S)
for i = 0; i < row; i ++
    for j = 0; j < col; j ++
        if S(i, j) > max_power - Th1
            S'(i, j) = (S(i, j) - max_power + Th1)/Th1
    //normalization
    end if
end for
end for

energy ← Th2*sum(S',1)/row
//energy is a row vector with length of time sequences.
for i = 0; i < row; i ++
    for j = 0; j < col; j ++
        if S'(i, j) < energy(i)
            S'(i, j) = 0
        else
            S'(i, j) = (S'(i, j)-energy(i))/(1-energy(i))
    //normalization
    end if
end for
end for

```

where φ denotes all frequency components. The feature F_2 is extracted to catch the information entropy of the Doppler frequencies that are beyond vital sign. The information entropy is a quantitative index of information content of a system.

The third feature is the ratio of BVSDB to total Doppler power (RBVSDP):

$$F_3 = \frac{F_1}{\sum_{m=1}^M \sum_{f_d \in \varphi} S'(m, f_d)}. \quad (11)$$

The feature F_3 means the ratio of BVSDB to total micro-Doppler power. As shown in Fig. 4, it is obvious that there exists a hyperplane that distinguishes two types of samples. Thus, a SVM is much qualified for classification. In particular, the Gaussian kernel function with Gamma 0.5 is used, and the penalty parameter is 0.1 when training. We use sklearn framework for training.

2) Stage 2: LSTM-Based Body-Movement Classification:

The neural networks are employed to further classify body movements if SVM predicts that the body movements occur. Considering that the spectrum is a set of signal vectors with time varying, LSTM neural networks can learn these signal vectors effectively. The LSTM is a type of recurrent neural

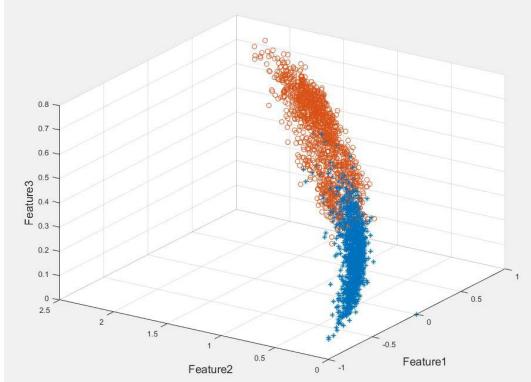


Fig. 4. Three-feature distributions of micro-motion and body movements.

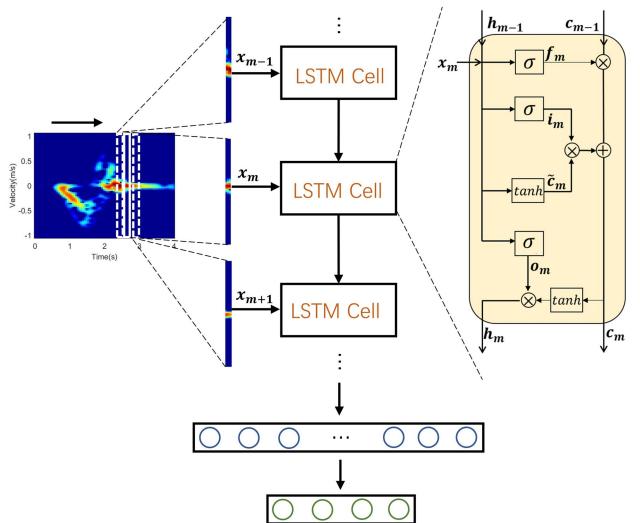


Fig. 5. The structure and the input vector of neural networks.

network, which can extract complex long-term dependencies over time.

The structure and the input vector of neural networks is shown in Fig. 5. The input x_m is a vector of time series and is equal to $S'(m, 1 : F)$. The $S'(m, 1 : F)$ means a series of vectors at a certain time step m . There are three gates in a LSTM cell [22], including forget gate f_m , input gate i_m and output gate o_m . Besides, \tilde{c}_m updates the hidden state, c_m denotes the update to current cell state, and h_m means the output of current cell state. The optimal number of neurons of the LSTM is 240, which will be detailed in section V.

By iterating weights and biases, the LSTM learns the distribution of micro-Doppler signatures. After one fully-connected layer, the last layer of networks is a softmax layer which predicts the type of sleep motions. In the training procedure, adam optimizer [27] is utilized for cross entropy minimization and all parameters are initialized with zero. We select Tensorflow framework to train networks with the use of python.

IV. PROPOSED REAL-TIME EDGE RADAR SYSTEM

As shown in Fig. 6, we design two experimental scenes in order to be close to real living scene, including a folding bed

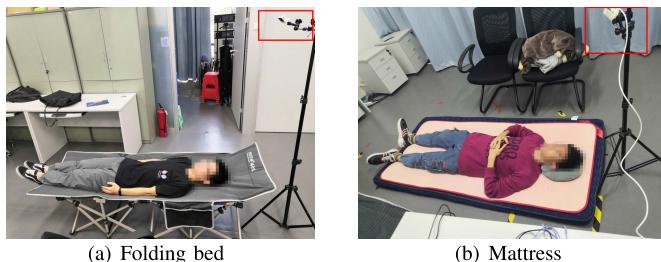


Fig. 6. Experimental scenes.



Fig. 7. The data flow of X4M03 in real-time system.

and a mattress. The radar is set around 1~1.5m above the subject.

After training, the parameters of SVM and LSTM will be saved, and be employed in real-time system. The real-time system is implemented in XTMCU02 that is a part of the module of X4M03. The X4M03 is composed of X4A02 antenna module, X4SIP02 acquisition module and XTMCU02 main control module. The X4A02 receives RF signal and X4SIP02 collects the sample data. The core of XTMCU02 is ARM Cortex M7, and the XTMCU02 works at 300MHz, which includes 2048kb flash memory and 384Kb SRAM. The XTMCU02 will regularly read the sample data from the X4SIP02. Users can perform specific data processing algorithms in XTMCU02 through the application programming of real-time operating system (RTOS). The data flow is shown in Fig. 7.

All the algorithms are embedded in XTMCU02 operated by C language without cloud servers. The final output of the real-time system is a recognition result transmitted by a serial port to computer, which effectively proves that the system can be applied in edge computing. Fig. 8 shows the real-time system for sleep motion classification at the scene of mattress, where the computer presents the final output. In addition, Fig. 9 shows the scene of folding bed. In Fig. 8 and Fig. 9, from number one to number nine, nine motions are recognized. In the left part, the predicted label, transmitted by the serial port, is shown by green line, while the real label is shown by red circle. The right part of each figure match the recognition results of the left.

V. EXPERIMENTAL RESULTS

There are ten subjects who participate in the experiment, including seven males and three females. **Table II** shows the physical information of ten subjects. The samples of each class are given in **Table III**. The total training samples is 6553, while the leave-two-subject-out (LTSO) samples is 459. Training samples come from people A-H, and LTSO samples are from people I-J. All the experiments are carried out by training 80 percent of samples and testing 20 percent of samples.

TABLE II
THE PHYSICAL INFORMATION OF SUBJECTS

People	A	B	C	D	E	F	G	H	I	J
Gender(F/M)	F	F	M	M	M	M	M	M	F	M
Height(cm)	150	158	165	169	169	174	180	180	160	174
Weight(kg)	44	53	60	56.5	60	67.5	65	70	57.5	60

TABLE III
THE NUMBER OF SAMPLES OF EACH MOTION TYPE

Motion type	The number of samples	The number of LTSO samples	Label
Micro-motion	1205	94	0
Lying down	1305	88	1
Getting up	1302	96	2
Turning over	1434	90	3
Waving	1307	91	4
Total	6553	459	-

Five-fold cross validation is performed to analyze the experimental results and later the average accuracy can be computed.

A. Hyper-Parameter Configuration

The window length and overlapping in STFT are important vectors influencing the performance. On the one hand, with the same sliding window points, the long window length can improve the frequency resolution, but will reduce the time resolution. On the other hand, with the same window length, the more overlapping points will improve the time resolution while can have a high computational complexity. With the same sliding window points of 5, Fig. 10 shows the spectrums of getting up with different window length of 40, 60, 80 and 100. It is evident that the window length of 80 may be the most suitable intuitively. With the same window length of 80, Fig. 11 shows the spectrums of getting up with different sliding window points of 3, 5, 7 and 9. It is evident that although the sliding window points of 3 shows the best performance, the sliding window points of 5 may be the most suitable with the slightly lower complexity.

The number of hidden layer neurons deeply influences the performance. The large size of networks can improve the accuracy, but when the accuracy reaches a certain level, more parameters may be ineffective and increase computing and storage resources. We apply different number of parameters to figure out the optimal model. The LSTM model is used to classify four body-movement motions. As shown in [Table IV](#), the accuracy will be presented with the different number of neurons. [Fig. 12](#) shows the accuracy changes with training epochs using fold1 samples. It is evident that the optimal number of neurons is 240 and we use 240 neurons for related experiments.

The time duration of spectrum is an important factor which has an impact on performance. Short time duration may cause poor performance while long time duration will lead to large time delay and high complexity. With the increasing of time

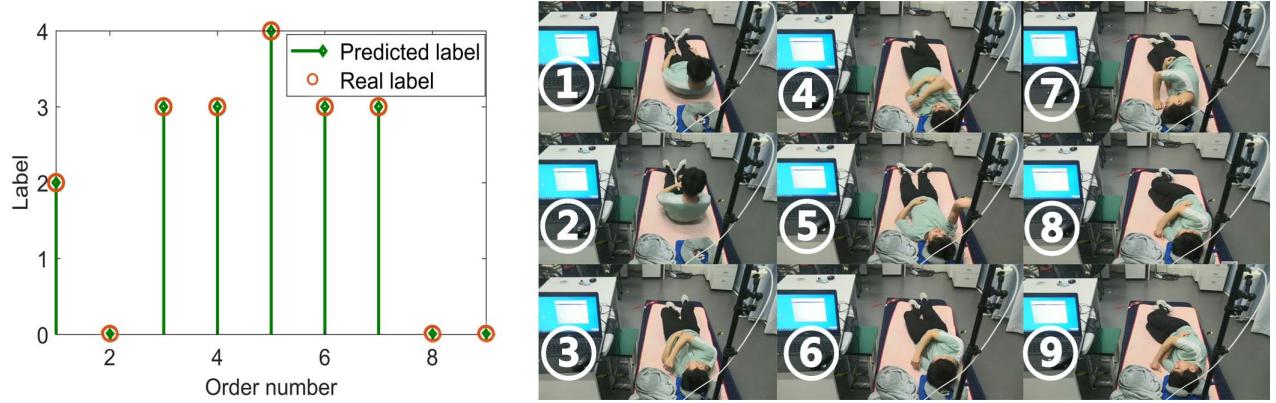


Fig. 8. Real-time results transmitted by a serial port at the scene of mattress.

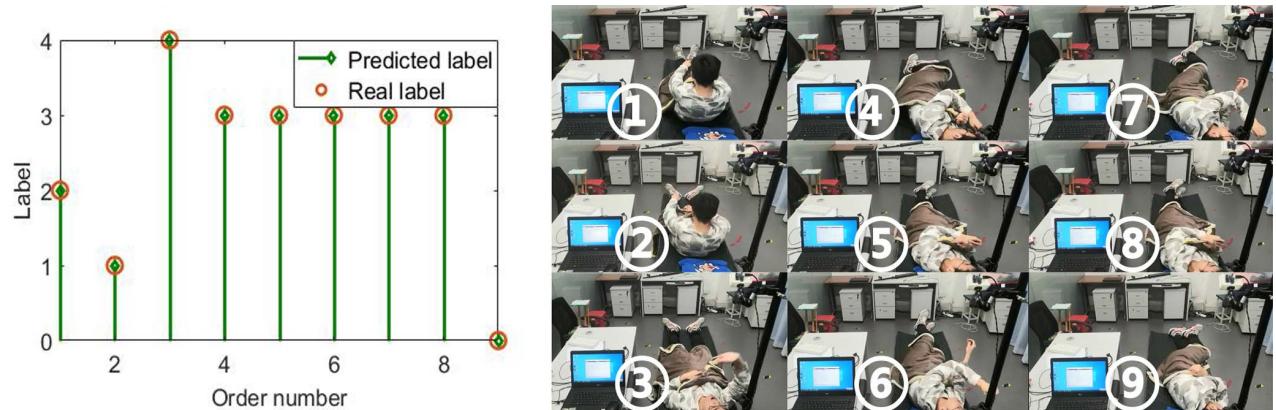
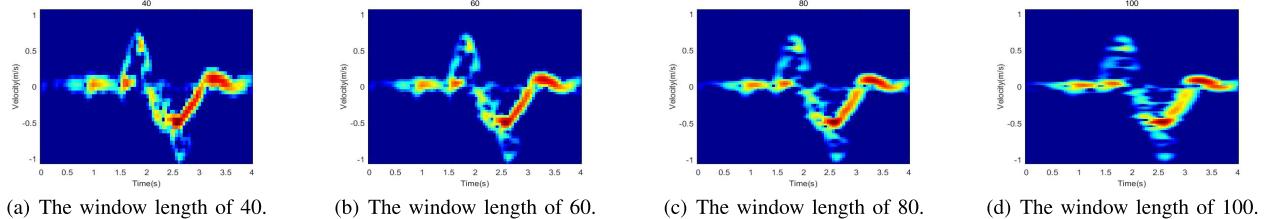


Fig. 9. Real-time results transmitted by a serial port at the scene of folding bed.



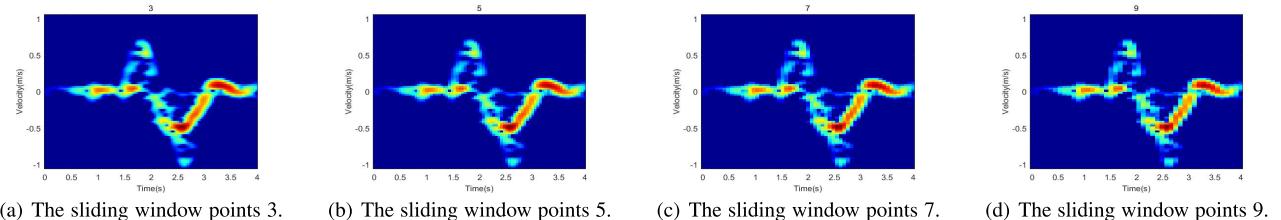
(a) The window length of 40.

(b) The window length of 60.

(c) The window length of 80.

(d) The window length of 100.

Fig. 10. The comparison of different window length in STFT.



(a) The sliding window points 3.

(b) The sliding window points 5.

(c) The sliding window points 7.

(d) The sliding window points 9.

Fig. 11. The comparison of different overlapping in STFT.

duration, some information such as noise may be useless and also degrade performance. Besides, because the time duration of most motions is around 3 seconds, while a few motions, e.g. waving, may take 4 seconds, there may be no significance to exceed time duration of 4 seconds. It is proved that the optimal time duration is 4 seconds, which is shown in Fig. 13.

B. Comparison With Other State-of-the-Art Methods

The SVM-only model is inspired by [8], which employed SVM to classify human motions using handcrafted micro-Doppler signatures. Similarly, we use three features described in section III for classification. Table IX shows the result that the accuracy is the lowest. The reason may be that these

TABLE IV
ACCURACY OF PROPOSED METHOD WITH DIFFERENT NEURONS

Neurons	Fold1	Fold2	Fold3	Fold4	Fold5	Average
40	69.22	53.98	80.64	93.26	74.93	74.41
80	94.29	89.15	85.03	94.67	95.23	91.67
120	95.88	97.47	96.73	96.63	96.45	96.63
160	96.26	98.41	95.88	97.57	96.91	97.01
200	97.29	97.38	97.94	97.94	97.57	97.62
240	98.13	97.94	98.13	97.29	97.85	97.87
280	98.13	96.82	98.32	98.13	97.85	97.85
320	96.91	97.85	97.75	97.38	98.22	97.62
360	96.63	97.66	98.22	98.22	97.75	97.70
400	98.04	96.63	98.13	97.85	97.75	97.68

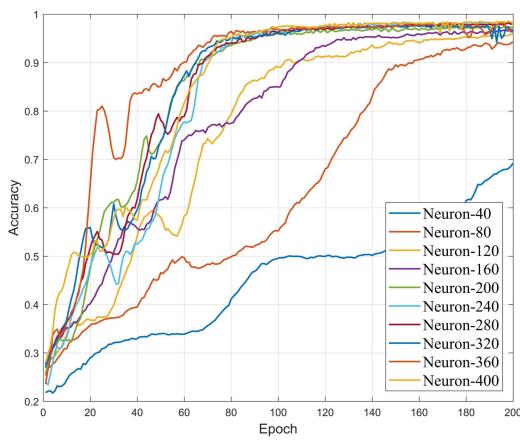


Fig. 12. The accuracy changes with training epochs.

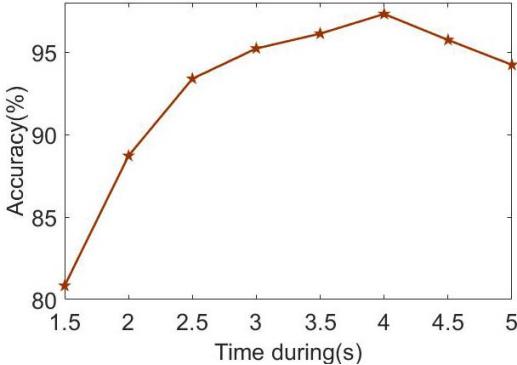


Fig. 13. The comparison of different time duration.

features are optimal for classifying micro-motion and body movement motions rather than all motions.

[22] uses LSTM-only for human motions and we utilize this model for five-motion classification. The input of LSTM is raw spectrum. Table V shows the results. It is presented that the optimal number of neurons is 720 and proved that the proposed model can effectively reduce parameters.

The confusion matrices using LTSO samples for the proposed method, LSTM-only and SVM-only are shown in Table VI, Table VII and Table VIII, respectively. In the

TABLE V
ACCURACY OF LSTM-ONLY WITH DIFFERENT NEURONS

Neurons	Fold1	Fold2	Fold3	Fold4	Fold5	Average
160	71.00	79.05	73.81	89.62	64.27	75.55
240	78.30	89.62	93.55	90.27	80.92	86.53
320	91.67	82.79	79.89	88.68	86.62	85.93
400	94.48	95.04	93.17	96.26	94.01	94.59
480	93.66	94.20	95.34	93.05	93.59	93.97
640	96.34	91.60	96.03	97.10	95.04	95.22
720	96.87	96.34	95.80	95.88	96.41	96.26
800	96.87	95.11	97.86	94.43	95.27	95.91

TABLE VI
THE CONFUSION MATRICES USING LTSO SAMPLES
FOR THE PROPOSED METHOD

Accuracy(%)	Micro-motion	Lying down	Getting up	Turning over	Waving
Micro-motion	90.43	0	0	9.57	0
Lying down	0	86.36	0	9.09	4.55
Getting up	0	0	96.88	3.13	0
Turning over	0	0	1.11	94.44	4.44
Waving	10.99	0	0	0	89.01

TABLE VII
THE CONFUSION MATRICES USING LTSO
SAMPLES FOR THE LSTM-ONLY

Accuracy(%)	Micro-motion	Lying down	Getting up	Turning over	Waving
Micro-motion	72.34	0	0	26.60	1.06
Lying down	0	95.45	0	3.41	1.14
Getting up	2.08	3.13	90.63	3.13	1.04
Turning over	0	1.11	6.67	92.22	0
Waving	0	0	6.59	1.10	92.31

confusion matrix of SVM-only, the best performance of five motions is micro-motion, proving that these three features are much suitable for classification of micro-motion and others. Besides, in the confusion matrix of LSTM-only, the worst performance is also micro-motion. It is suspected that the networks may overfit the features of micro-motion, because micro-motions have greater randomness. Instead, the hand-crafted features can alleviate the problem of overfitting. Thus, the results of the proposed method achieve the best performance.

We compare our method with other state-of-the-art methods. [11] used raw spectrum as input and trained convolutional neural networks which consist of three convolutional layers and two fully-connected layers. In [18], a novel convolutional neural network was proposed to match radar raw data and has a better raw data representation rather than using STFT. All the experimental results including LTSO are shown in Table IX. The accuracy of proposed method using LTSO samples reaches 91.50%, and the average accuracy with train

TABLE VIII
THE CONFUSION MATRICES USING LTSO
SAMPLES FOR THE SVM-ONLY

Accuracy(%)	Micro-motion	Lying down	Getting up	Turning over	Waving
Micro-motion	88.89	0.46	0	5.09	5.56
Lying down	0	49.40	45.38	4.02	1.21
Getting up	0	36.98	56.27	6.04	0.75
Turning over	4.40	9.12	6.92	73.58	5.97
Waving	4.58	10.69	4.20	3.82	76.72

TABLE IX
THE ACCURACY OF DIFFERENT MODELS

Model	Fold1	Fold2	Fold3	Fold4	Fold5	Average	LTSO
Proposed	97.79	97.78	95.11	96.87	97.55	97.02	91.50
SVM-only	65.80	69.62	68.63	68.63	67.94	68.12	68.63
LSTM-only	96.87	96.34	95.80	95.88	96.41	96.26	88.45
DCNN[11]	95.04	97.94	97.48	93.89	97.63	96.40	84.97
RadarNet[18]	91.90	92.14	91.83	90.84	90.54	91.45	78.65

TABLE X
THE COMPLEXITY OF DIFFERENT MODELS

Model	LTSO(%)	Multiplications	Additions	Parameters
Proposed	91.50	2,845,357	2,801,091	39,101
SVM-only	68.63	35,238	19,576	15,661
LSTM-only	88.45	15,602,220	15,469,380	188,825
DCNN[11]	84.97	25,154,900	24,982,500	1,123,565
RadarNet[18]	78.65	4,982,432	4,882,400	85,888

and test split reaches 97%. The reasons why the proposed method reaches high performance using LTSO samples and have a relatively low complexity may be that the zero-forcing processing is much significant for low SNR radar such as X4M03. Furthermore, the features extracted by this paper are suitable for classification of micro-motion and others, and compared with DCNNs, the LSTM network is more capable of classifying the spectrum because of its characteristic of time sequence.

Besides, we explain the feasibility of proposed method in an aspect of complexity. To employ models in edge computing, the complexity is required to be slightly low while the accuracy reaches a high level. We will quantify the time complexity and spatial complexity to describe the complexity. The time complexity is mainly described with the number of multiplications and additions, while the spatial complexity is described with the number of parameters. Specially, the parameters of SVM are composed of dual coefficients, support vectors and intercept. The parameters of networks consist of weights and biases that convolution layers, fully-connected layers and recurrent neurons have. As is shown in Table X, we use the same one sample to compare the complexity of different models. Compared with other neural networks, it is shown that the proposed method takes the least number of both multiplications and additions, while the number of parameters is also the least. Besides, although the complexity of the

SVM-only model is much lower than that of the proposed method, the accuracy of the proposed method is much higher.

VI. CONCLUSION

This paper proposes a two-stage human sleep motion classification method with a low-complexity using UWB radar. Firstly, we perform PBC detection for human motion detection when sleeping. Later, we propose a method named zero-forcing processing for radar spectrum to reduce the parameters of classifiers. Combining handcrafted feature extraction with neural networks, a low-complexity machine learning method is employed to classify sleep motions. In stage one, the SVM is used to roughly classify micro-motion and body movements. In stage two, the LSTM is utilized to further classify body movements. Experimental results show that the proposed method can reach the average accuracy of the five typical sleep motions over 97% using train and test split samples, and 91.50% using the leave-two-subject-out (LTSO) samples. Eventually, all the algorithms can be implemented through the application programming of real-time operating system, proving that our method can be applied in edge computing without cloud servers.

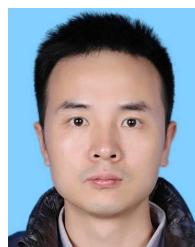
REFERENCES

- [1] K.-K. Shyu, L.-J. Chiu, P.-L. Lee, T.-H. Tung, and S.-H. Yang, "Detection of breathing and heart rates in UWB radar sensor data using FVPIEF-based two-layer EEMD," *IEEE Sensors J.*, vol. 19, no. 2, pp. 774–784, Jan. 2019.
- [2] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomed. Opt. Exp.*, vol. 6, no. 5, pp. 1565–1588, 2015.
- [3] H. Maki, H. Ogawa, S. Tsukamoto, Y. Yonezawa, and W. M. Caldwell, "A system for monitoring cardiac vibration, respiration, and body movement in bed using an infrared," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Buenos Aires, Argentina, Aug. 2010, pp. 5197–5200.
- [4] V. C. Chen, "Analysis of radar micro-Doppler with time-frequency transform," in *Proc. 10th IEEE Workshop Stat. Signal Array Process.*, Pocono Manor, PA, USA, Aug. 2000, pp. 463–466.
- [5] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Analysis of micro-Doppler signatures," *IEE Proc.-Radar; Sonar Navigat.*, vol. 150, no. 4, pp. 271–276, Aug. 2003.
- [6] V. C. Chen, F. Li, S.-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.
- [7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
- [8] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.
- [9] F. Fioranelli, M. Ritchie, and H. Griffiths, "Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-Doppler," *IET Radar; Sonar Navigat.*, vol. 10, no. 9, pp. 1702–1710, Apr. 2016.
- [10] Z. Zhang, R. Zhang, W. Sheng, Y. Han, and X. Ma, "Feature extraction and classification of human motions with LFMCW radar," in *Proc. IEEE Int. Workshop Electromagnetics: Appl. Student Innov. Competition (iWEM)*, Nanjing, China, May 2016, pp. 1–3.
- [11] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016.
- [12] Y. Shao, Y. Dai, L. Yuan, and W. Chen, "Deep learning methods for personnel recognition based on micro-Doppler features," in *Proc. 9th Int. Conf. Signal Process. Syst. (ICSPS)*, 2017, pp. 94–98.
- [13] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 669–673, May 2018.

- [14] M. S. Seyfioğlu and S. Z. Gürbüz, "Deep neural network initialization methods for micro-Doppler classification with low training sample support," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2462–2466, Dec. 2017.
- [15] I. Alnujaim, D. Oh, and Y. Kim, "Generative adversarial networks for classification of micro-Doppler signatures of human activity," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 396–400, Mar. 2020.
- [16] Y. Kim, J. Park, and T. Moon, "Classification of micro-Doppler signatures of human aquatic activity through simulation and measurement using transferred learning," *Proc. SPIE*, vol. 10188, May 2017, Art. no. 101880V.
- [17] M. S. Seyfioğlu, B. Erol, S. Z. Gurbuz, and M. G. Amin, "DNN transfer learning from diversified micro-Doppler for motion classification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 5, pp. 2164–2180, Oct. 2019.
- [18] W. Ye, H. Chen, and B. Li, "Using an end-to-end convolutional network on radar signal for human activity classification," *IEEE Sensors J.*, vol. 19, no. 24, pp. 12244–12252, Dec. 2019.
- [19] R. Zhang and S. Cao, "Real-time human motion behavior detection via CNN using mmWave radar," *IEEE Sensors Lett.*, vol. 3, no. 2, pp. 1–4, Feb. 2019.
- [20] B. Wang, L. Guo, H. Zhang, and Y.-X. Guo, "A millimetre-wave radar-based fall detection method using line kernel convolutional neural network," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13364–13370, Nov. 2020.
- [21] G. Klarenbeek, R. I. A. Harmann, and L. Cifola, "Multi-target human gait classification using LSTM recurrent neural networks applied to micro-Doppler," in *Proc. Eur. Radar Conf. (EURAD)*, Nuremberg, Germany, Oct. 2017, pp. 167–170.
- [22] J. Craley, T. S. Murray, D. R. Mendat, and A. G. Andreou, "Action recognition using micro-Doppler signatures and a recurrent neural network," in *Proc. 51st Annu. Conf. Inf. Sci. Syst. (CISS)*, Baltimore, MD, USA, Mar. 2017, pp. 1–5.
- [23] J. Lai and Z. Yang, "Human motion recognition based-on micro-Doppler simulations using transfer learning," in *Proc. Inst. Eng. Technol. Int. Radar Conf.*, Chongqing, China, 2020, pp. 1–6.
- [24] N. Andersen *et al.*, "A 118-mW pulse-based radar SoC in 55-nm CMOS for non-contact human vital signs detection," *IEEE J. Solid-State Circuits*, vol. 52, no. 12, pp. 3421–3433, Dec. 2017.
- [25] M. Piccardi, "Background subtraction techniques: A review," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, The Hague, Netherlands, Oct. 2004, pp. 3099–3104.
- [26] Y. Cheng, Z. Yang, D. Hong, and J. Chen, "Sleep anomaly events detection based on adaptive double threshold CFAR detector using UWB radar," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Dalian, China, Sep. 2019, pp. 1–5.
- [27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.



Jialei Lai received the B.E. degree from the School of Electronic Information Engineering, Shenzhen University, Shenzhen, China, in June 2019, where he is currently pursuing the M.E. degree. His research focuses are radar signal processing and pattern recognition, especially about the radar-based human target detection and motion recognition.



Zhaocheng Yang (Member, IEEE) received the B.E. degree in information engineering from the Beijing Institute of Technology, Beijing, China, in 2007, and the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2013.

From November 2010 to November 2011, he was a Visiting Scholar with the University of York, York, U.K. From June 2013 to August 2015, he was a Lecturer with the School of Electronics Science and Engineering, National University of Defense Technology. He is currently an Associate Professor with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China. His research interests include the area of signal processing, including array signal processing, adaptive signal processing, compressive sensing, machine learning, and its applications to radar systems.



Boning Guo received the B.E. degree from the School of Instrument and Electronics, North University of China, in June 2019. He is currently pursuing the M.E. degree with the School of Electronic Information Engineering, Shenzhen University. His research focuses are radar signal processing, especially the determination of human state and estimation of vital signs based on radar.