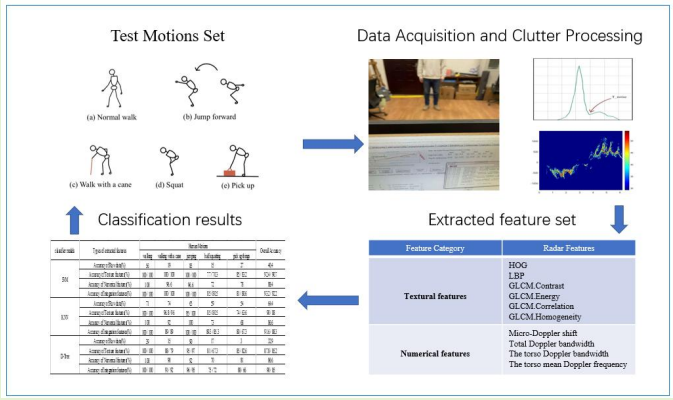
Recognition of Approximate Motions of Human Based on Micro-Doppler Features

Ziqian Wang, Aifeng Ren, Member, IEEE, Qi Zhang,
Adnan Zahid, and Qammer H. Abbasi, Sensor Member, IEEE

***Abstract*—With the aging society approaching, the daily safety of the elderly has attracted more and more people's attention, especially the elderly who live alone without anyone to take care of them. How to judge whether there is danger from the daily behavior of those who live alone and providing timely assistance is a research topic of health protection in the field of smart home care. Modern radar sensing technology and intelligent perception technology adopted in the domain of health care have started to gain significant interest, which can replace wearable sensors or cameras without feeling uncomfortable or having privacy issues. In this paper, some preliminary results were presented to develop an indoor monitoring system for human activity recognition using micro-Doppler features of radar signals. The experimental campaign involved five different with some similar activities of 10 subjects to be collected using a Doppler radar. A total of several texture features and numerical features are extracted from the spectrogram and time-frequency domain of measurements and applied three machine learning (ML) algorithms such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Decision Tree (D-Tree) for the precise classification of all five different motions. The results illustrated that the SVM classifier using the Radial Basis Function (RBF) kernel reaches the best results for all three feature sets, and the classification of mixed features can reach 93.2%, as well as 88.4% and 92.4% for numerical and texture features. The performance of SVM is 1.6% better than KNN and 3.2% better than D-Tree.**



***Index Terms*—Micro Doppler, Machine Learning, Approximate Motions, Classification.**

I. Introduction

OVER the past few years, the world's population of people living alone has been increasing, there were 37 million one-person households in the United States in 2021, the U.S. Census Bureau reported, representing 28 percent of all households across the country and 15 percent of the overall population. In 1960, just 13 percent of households were occupied by just one person [1].

Among those who live alone, the safety of the elderly receives particular attention. The population of older adults living alone has been increasing [2], and most of these older adults live alone or in nursing homes, making it difficult for their families to give them the necessary care they need. Chronic and cardiovascular diseases are common in the elderly as their body functions decline, therefore, in case of a sudden attack when they live alone, they may fall and bump at any time causing fatal injuries, which may bring more serious secondary injuries if they are not found in time to seek medical attention. Although falls cannot be completely prevented, when the detection system can accurately identify falls and issue emergency alerts, secondary injuries will be reduced. So the monitoring of human activity in daily life is very beneficial to get timely help and rescue in case of emergency, and plays an important role in daily life.

In terms of indoor motion detection, several detection schemes have been proposed in the literature. These systems typically use two different sensing sensors to record or recognize the user's motions, including wearable and non-wearable sensors [3]. Wearable device-based solutions such as wearing smartwatches, VR glasses, etc. Non-wearable device solutions consist of vision acquisition and sensing units with image and video technologies, WiFi RF equipment and those using micro-Doppler radar, etc. [4]-[8]. The focus of our work is on human motion recognition by micro-Doppler radar sensors.

Human Motions Recognition (HMR), a research topic in the field of image processing, has been a major focus of researchers. In recent years, with the widespread popularity of cameras in people's lives, visual acquisition and perception sensors using

Z. Wang is with the School of Electronic Engineering, Xidian University,Xi'an 710126, China (e-mail: Wangziqian@stu.xidian.edu.cn).

A. Ren is with the School of Electronic Engineering, Xidian University, Xi'an 710126, China (e-mail: afren@mail.xidian.edu.cn).

Q. Zhang is with the School of Electronic Engineering, Xidian University, Xi'an 710126, China (e-mail: 21021210686@stu.xidian.edu.cn).

A. Zahid, and Q. H. Abbasi are with the School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.(e-mail: a.zahid.1@research.gla.ac.uk; Qammer.Abbasi@glasgow.ac.uk).

image and video technologies are increasingly being used for human motion recognition [9]. Nevertheless, video and image technologies are highly sensitive to illumination conditions, and the motion perception effect is strictly limited by illumination conditions [10]. For instance, the motion perception ability of images and videos diminished or even vanished in dark environments. Meanwhile, with the flourishing development of Internet technology, there is a risk that the image and video information captured by the camera will be stolen and maliciously disseminated by unscrupulous elements, nowadays, the focus on image privacy protection causes people to be wary of image and video information acquisition, which to a certain extent affects the application and development of image and video technology in HMR. Compared to visual sensor technology, radar technology has minimal problems with limited illumination conditions and personal privacy violations.

In another regard, wearable sensors motion recognition methods typically require special sensing units for motion perception, thus increasing the cost and requiring people to wear specialized sensing elements at all times, which affects people's movement freedom in a certain extent [11]. In addition, the buttons on these sensors typically need to be operated by the user, but in case of an accident, people may lose their mobility and unable to complete the operation. Compared to wearable sensors technology, radar does not require additional equipment to be worn, nor additional operations to be performed.

As mentioned above, radar technology is not limited by illumination conditions and can perceive human motion normally even in dark environments, and HMR through radar technology rarely constitutes an invasion of human privacy. Furthermore, radar can be conveniently embedded in the environment and also facilitate contactless behavior recognition. Hence, radar recognition units are more suitable for domestic and indoor surveillance. These features turn radar sensing technology into an ideal sensor for human motion recognition, detection, and classification.

The radar detection process can be described as the emission of electromagnetic waves, which are reflected upon contact with the target, and through the processing of the return signal, from which the necessary information can be obtained. When contacted with a moving target, the frequency of the transmit signal will change. This frequency is only related to the carrier frequency of the transmit signal and the speed of the target. This phenomenon is called the Doppler effect. Apart from the movement of the main body, there are other parts attached to the main body performing micro-movements. These micro-movements add modulation to the Doppler frequency generated by the main body and widen the bandwidth of the Doppler frequency. This phenomenon is called the micro-Doppler effect [12]-[16].

Since the echoes of Doppler radar contain time-varying information about human motion, they can be applied to motion recognition. Most current methods obtain spectrograms using time-frequency analysis. Therefore, how to extract useful features from the micro-Doppler signal becomes a critical factor in performing HMR. Nowadays, researchers have done great work in HMR, such as walking, kneeling, standing, bending, falling, and other motions classification. Particularly for detecting falls in old people, the detection has achieved a high accuracy rate [17],[18]. Distinguishing actions with high differences are considered inter-class classification, while intra-class classification is the same as identifying different members of the same class, and intra-class classification can be understood as distinguishing actions with high similarity [18]. Researchers normally focus on classifying motions with higher variance because this results in great classification, and therefore usually neglect approximate motions.

In [19], it is classified based on extracting phase information from high-resolution Range Map images, exploiting Histogram of Oriented Gradient(HOG) features as a feature extraction method, using SVM, KNN and classifying images using Gaussian SVM classifier with phase information, with results above 92% for both gesture recognition and human behavior classification. However, no control experiment was designed to verify its superiority on the same dataset.

In [20], it conducted experiments in four different outdoor environments and classified three motions: crawling, walking, and jogging. It can be seen that the authors also designed motions with a large variance. The seven features extracted in the spectrogram are applied to human motion classification by the SVM and RBF classifier. The classification result achieved is 80%-90%. There are two main reasons for the decrease in his classification effect, the first reason is the lack of feature selection, the other reason is the lack of noise filtering, and the increase in standard deviation due to environmental noise.

In [21], the experiments were conducted in four different indoor environments, with six motions designed for walking, falling, drinking, etc. The spectrograms derived from the experiments can distinguish each movement visually. The results obtained using the above algorithms: 77.1% accuracy for SVM, 70.8% accuracy for KNN and GoogleNet. The aim of this paper is to improve the above classification results.

In this paper, we design a machine learning (ML) method that uses texture features as the main feature for classification. Although some actions have obvious visual differences, but their numerical feature may be nearly or even identical, which means that using numerical features is not the best choice for classifying some motions. The classification effect is still improved to some extent by using texture features to classify approximate motions. We collected human daily life activities such as normal walking, walking with a cane, bending down to pick up things, squatting and jumping forward, where bending down to pick up things and squatting are similar sets of motions. Finally, SVM, D-Tree and KNN classifiers are used to classify. The paper is organized as follows: Section II presents the data collection and processing work. Section III describes the feature extraction work. Section IV discusses the classification effects of various classifiers. Finally, conclusion work are highlighted in Section V.

II. Data Collection And Processing

The classification system based on micro-Doppler features is shown in Fig. 1. The data are collected by Frequency Modulated Continuous Wave (FMCW) radar as time series data, and the raw data are converted into a matrix that contains the target micro-motion information, the rows and columns of the matrix contain the target distance and velocity dimension data, respectively. The two-dimensional matrix is performed

the distance-dimensional FFT, and finally the time-dimensional data is accumulated and performed the FFT again to obtain the distance-time matrix [22].

In addition, filtering is also required to decrease the effect of static object reflections. The ideal effect is that the spectrogram only contains the Doppler information of the human body. Through setting reasonable threshold values, irrelevant information will be filtered and human body information will be more prominent. Finally, the Short-time Fourier Transfor (STFT) is applied to the samples, i. e., means that STFT is applied to the distance-time matrix. The spectrogram is presented in the form of an image and the corresponding features can be extracted, which will be described in detail in the next sections.

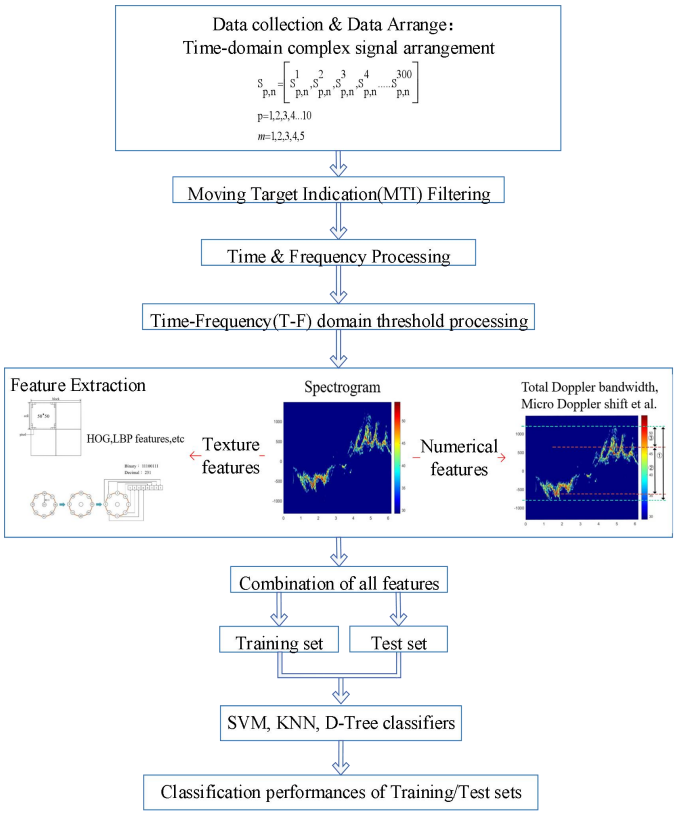


Fig. 1. Framework of the proposed system model for HMR

A. Experimental Environment and Data Collection

In this experiment, the data are collected by the Micro Doppler Radar (MDR). The radar works at 77 GHz center frequency and 5 MHz sample rate. The device is connected to a DCA1000EVM, a real-time data collector, for data collection. The experimental environment is a 7m*5m indoor student laboratory with some experimental equipment and a zone for volunteer activities.

The action of collecting is the behavior that occurs more frequently in life, as follows: normal walking, jumping, walking with a cane, half squatting and half squatting to pick up things, as shown in the Fig. 2. The collection time for each action is about six seconds, and the number of collections is 20 (collection times) × 10 (experimenter), thus the number of samples obtained is 200 (number of collections for each action) × 5 (type of action). All samples were randomly disordered to avoid the influence of sampling order or the experimenter's physical size, posture and motion habits on the action recognition result. The 70% of the training set is used to train the action recognition algorithm model, and the remaining 30% is reserved as a test data set for verification of the classification effect.

The echo signal has a total of 300 frames. Each frame has 128 frequency modulation (FM) periods. Each FM period has

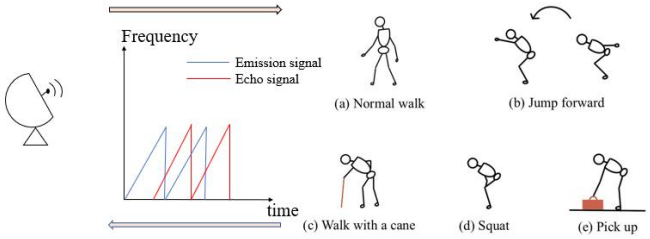


Fig. 2. Description of the five different human activities

256 sampling points, and the radar FM period has 160 microseconds. Therefore, the time to capture 300 frames of data is 6.1s. Arrange the collected raw data into a data matrix of $M \times N$ as follows (1).

$$S_{p,n} = \begin{bmatrix} S_{p,n}^{1} \\ S_{p,n}^{2} \\ S_{p,n}^{3} \\ \cdots \\ S_{p,n}^{300} \end{bmatrix} \quad (1)$$

Where p = 1, 2, 3..... 10 indicates the number of the experimental subjects. n = 1, 2, 3, 4, 5 indicates the motion number. M is equal to 128×300, indicating the total number of FM period, and N is equal to 256, which indicates the number of sample points in one cycle. $S_{p,n}$ means that experimenter p performs the specified action number n. The label in the upper right corner indicates the number of frames.

B. Moving Target Indication (MTI) Filtering

Before performing the STFT, it is necessary to minimize the influence of the echoes of stationary objects on the echoes of human motion. The spectrum of these stationary clutter waves can cover up the spectrum of human motion. Since the spectral peak of the center frequency of a stationary object is constant, in contrast, the spectral peak of the center frequency of a moving target is influenced by the Doppler frequency modulation of the target that the spectral peak varies in the form of a cosine envelope. Therefore, Moving Target Indication Filtering (MTI) can be considered to suppress stationary clutter [23].

MTI is a clutter suppression technique that uses adjacent FM

period frequency pair cancellation to remove static clutter interference while retaining the moving target spectrum. In [24], this paper cites an MTI filter for FMCW radar: the echo signal from the previous arrival was subtracted with the current echo signal to achieve a simple single-delay MTI filter.

$$\Delta S_{B,K}(f) = S_{B,K}(f)[\exp(j2\pi f_d T_r) - 1] \tag{2}$$

As shown in (2), it can be seen that for a static target with zero Doppler frequency, i.e., f_d is equal to zero, the spectral difference of adjacent FM cycles is zero. In contrast, for moving targets, the spectral difference of adjacent FM cycles is not zero. When MTI processes the echo signal, it causes a decrease in signal intensity of static targets, while maintaining the signal intensity of moving targets. Thus, MTI processing can effectively remove clutter from static targets while retaining Doppler information from non-static targets like people.

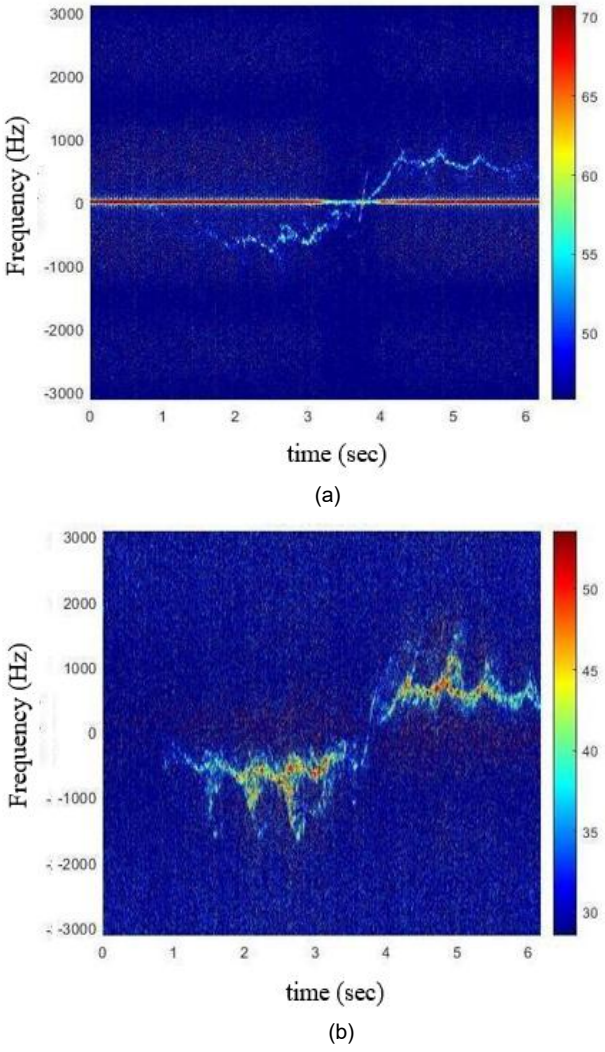


Fig. 3. (a) Spectrogram without MTI processing. (b) MTI processed spectrogram

In Fig. 3(a), as you can see that the human body Doppler information has been covered by clutter and no valid information can be extracted. In Fig. 3(b), human Doppler information dominates after filtering. However, we can also see that there are still some clutter after processing, so further processing is needed.

C. STFT and Threshold Processing

The classical Fourier transform actually compares the signal with a complex sinusoidal function, and the complex sinusoidal function is more limited in the time domain and cannot describe the time-domain characteristics of time-varying spectral signals. While STFT, as a commonly used linear time-frequency analysis method, can overcome the limitations of the classical Fourier transform [25].

The STFT analyzes the unfolded signal with a set of elementary functions $\{\gamma_{t,w}(\tau) = \gamma(\tau - t)e^{j\omega\tau}\}$ that are well aggregated in the time-frequency domain [26].

The spectrogram for displaying radar echoes can be obtained after STFT processing. These spectrograms are useful for analyzing frequency changes with time, thus capturing the micro-Doppler features of various human activities. Then the STFT of the $s(t)$ is shown in (3).

$$STFT(t,\omega) = \int s(\tau)\gamma^*(\tau - t)e^{-j\omega\tau}d\tau \tag{3}$$

The STFT expression reflects the similarity between the $s(t)$ and $\gamma(\tau - t)e^{j\omega\tau}$, where $\gamma(t)$ is called the window function and lasts for a short time. The choice of window size is crucial as it describes the resolution of time and frequency. The STFT first divides the time domain signal $x(n)$ into M segments by a window function of length L and overlap H, and then performs an N-point fast Fourier transform (FFT) on each segment. A series of FFT results are obtained by the window function, as well as the results are arranged into a two-dimensional matrix, i.e., the spectrogram. In this paper, the data matrix composed of (1) is sampled at N points for each frequency modulation period, i.e., the STFT is done for each row to obtain the distance amplitude spectrum, then the distance information is summed, and STFT is done again after traversing the sequence of M frequency modulation periods to obtain the spectrum containing Doppler modulation information.

With experimental comparison, if the classification work is performed with the STFT transformed spectrogram, the traditional numerical features cannot ideally distinguish the two actions due to the highly similar motion amplitude, squatting degree, etc. between half-squatting and bending down to pick up something. For this reason, it is a better choice to perform classification according to texture features, yet the quality of the spectrogram texture has a certain requirement, which we need to further minimize the environmental clutter. The weak clutter processing can effectively accomplish the above, and the critical to weak clutter processing is to determine a noise threshold [27].

The spectrogram generated after STFT are not the same for different human motions, whereas the noise spectrogram of the same experimental environment are certain, therefore, the spectrogram of different motions can be compared with the same environmental spectrogram to obtain the noise threshold. Fig. 4 is the signal intensity graph of the experimental environment without human motion and when a human

performs motion,where the motion is selected as representative of normal walking. In Fig. 4, the signal intensity profile is presented as a Gaussian-like distribution, thus we can use the lowest signal energy intensity value in the background environment intensity profile that deviates from the Gaussian distribution as the noise threshold. The signal strength of the background environment is substantially weakened around 10dB, while normal walking is only weakened around 40dB, with a stronger distribution between 10dB and 40dB. Since the signal intensity values of different motions are distributed differently, the noise threshold will be slightly different and not necessarily the same, so it is necessary to determine different noise thresholds according to the different motions to compare with the environmental signal intensity, which can effectively remove weak clutter and extract useful human motion signals.

As shown by the Fig. 5, the spectrogram visually shows a large gap due to the different movements. It is observed that the strongest part of the echo of the spectrogram comes from the torso, while the periodic micro-Doppler modulation around the torso comes from the motion of the limbs. Among all activities, jumping causes the largest echo frequency, which can reach about 2KHz, and the difference between the maximum micro-Doppler frequency caused by limb swing and the maximum Doppler frequency caused by torso movement is large during jumping, indicating that the limb swing amplitude of the action is large. In particular, the negative value is large due to the radar-oriented jump, the arm swings from down to above the head during the jump, and the swing amplitude is the largest in the course of the action. In contrast, the regular oscillation of normal walking limbs causes a more symmetrical spectrogram and more uniform micro-Doppler shift. In the case of walking with a cane, the Doppler frequency shift is basically caused by the trunk because of the small swing of the limbs in this movement. as the trunk is also no significant movement, the maximum Doppler frequency compared to the jump is also much smaller.

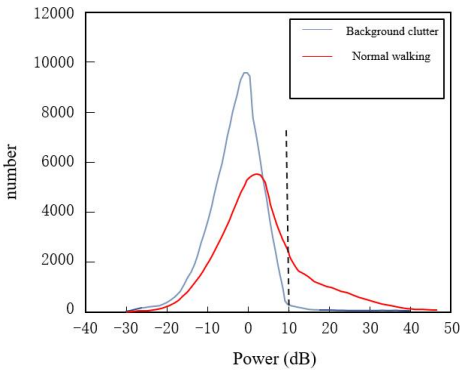


Fig. 4. Human motion and experimental environment signal strength curves

The Fig. 5 are the spectrograms after thresholding, and now we compare the effect before and after processing on the normal walking spectrograms. Fig. 3(b) is unprocessed, while Fig. 5(a) is processed. The characteristic information can be further evident.

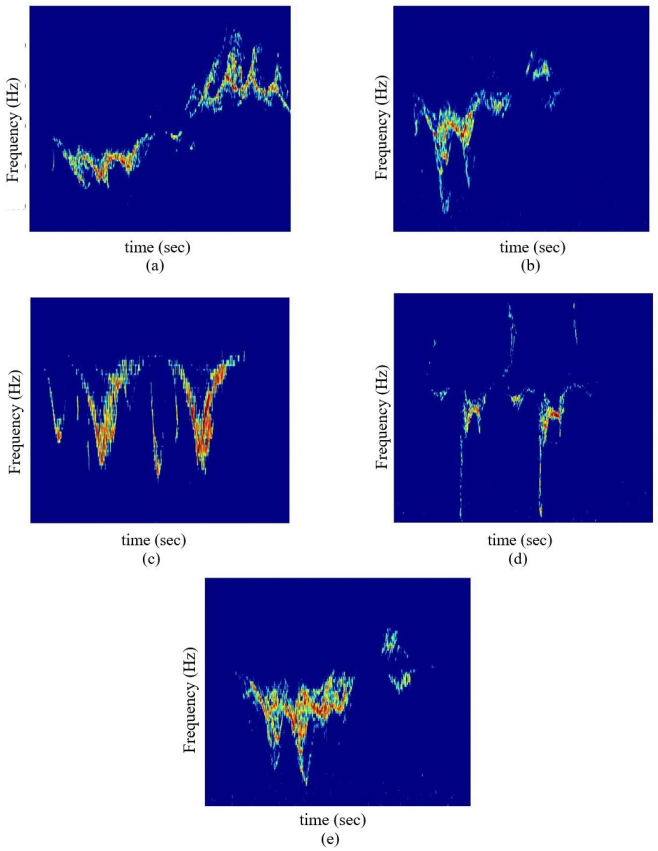


Fig. 5. Spectrograms of various human motions (a) normal walking. (b) half squatting. (c) walking with a cane. (d) jumping. (e) half squatting to pick up things

III. Feature Extraction

The spectrogram contains various state information of human motion over a period of time, which can characterize the whole process of human motion, not limited to a specific moment. The spectrogram can be directly used as input to the classifier without extracting features to identify different human motions. However, such high-dimensional data can make the internal structure of the classifier very complex, and the training process can become lengthy. Therefore, selecting appropriate features can reduce the dimensionality of the data and make the classifier work more efficiently. In the previous literature, workers have proposed many features for extracting human micro-Doppler information[28]-[31], mainly numerical features that cannot be visualized, and these features perform good in inter-class motions classification work and capture the state of human motion during the whole process of motions. In the form of images, it is visually more intuitive to distinguish inter-class motions, the experimental results also show that the texture feature performs better. In general, texture features are easily recognized by the human visual system, they can make full use of the spatial information of images. There are various methods for texture feature extraction, the statistical methods are standard methods for texture analysis and have been proven to be effective [32]. Commonly used measures include entropy, histogram moments, directionality, contrast, correlation, inertia, and energy [33]. This paper will select four numerical features and seven texture features to characterize the micro-Doppler features, as shown in Table I.

TABLE I
UNITS FOR MAGNETIC PROPERTIES

Feature Category	Radar Features
Textural features	Histogram of Oriented Gradient(HOG) Local Binary Pattern(LBP) GLCM.Contrast GLCM.Energy GLCM.Correlation GLCM.Homogeneity Entropy
Numerical features	Micro-Doppler shift Total Doppler bandwidth The torso Doppler bandwidth The torso mean Doppler frequency

A. HOG Feature

We can extract the histogram of the gradient distribution of the image as a feature, which can reflect the edge Information contained in the image to some extent [34]. The HOG describes the image by a set of local histograms, these histograms characterize information about the direction and magnitude of the image gradients, especially at the image edges and corner points where the gradient values are large. Since the image is thresholded, the Doppler information is concentrated in a certain range, and the effect of noise is reduced. Therefore HOG will not detect the edge of the noise spectrum.

The first step of HOG feature extraction is to calculate the gradient value of each pixel point to obtain the gradient histogram. The image gradient is defined as the directional change in pixel intensity on the x and y axes. The gradient vector of a pixel point can be denoted by (4).

$$\nabla f(x,y)=\begin{bmatrix} g_x \\ g_y \end{bmatrix}=\begin{bmatrix} f(x+1,y)-f(x-1,y) \\ f(x,y+1)-f(x,y-1) \end{bmatrix} \tag{4}$$

$f(x,y)$ represents the pixel value of the pixel point, while g_x and g_y are the gradients of the pixel point in the x-direction and y-direction. The magnitude $M(x,y)$ and direction $D(x,y)$ of the combined gradient in the x and y directions can be calculated using the following formula.

$$M(x,y)=\sqrt{g_x^2+g_y^2} \tag{5}$$

$$D(x,y)=\arctan\frac{g_y}{g_x} \tag{6}$$

The next step is to divide the image into some blocks, each of which is made up of some cells, which are made up of smaller pixels. In this study, a cell consists of 50×50 pixels. As the input image size is 400×350, using a 50×50 block of pixels produces eight cells in the horizontal direction and seven cells in the vertical direction. In order to reduce the effect of illumination, a final normalization process is required. The L2-norm normalization factor normalizes the HOG feature vector within the block. The L2-norm can be represented by the (7).

$$f=\frac{v}{\sqrt{\|v\|^2+e^2}} \tag{7}$$

where v is a concatenation of all HOG feature vectors within a block, f is the normalization factor, and e is a tiny constant to avoid a denominator of 0. After normalization, a huge vector contains the HOG feature vectors of all blocks of the image.

B. LBP Feature

The LBP effectively reflects the local structure of an image by comparing each pixel on the image with its neighboring pixels. The original LBP operator is defined by defining a 3×3 window, and the pixel value of the center point inside the window is used as a criterion to compare with the size of the other eight pixel values inside the window. The value more larger than this criterion is one, and less than is zero. Eight points form a binary number (usually converted to decimal representation), which is the LBP feature value of the center point[35][36].

However, the computation area of the original LBP operator is 8 points around the pixel point, which will have a large deviation when the image size changes and cannot correctly reflect the texture information around the pixel point. Therefore, an improved LBP operator, circular LBP operator, is used in this paper[37]. By replacing the previous square neighborhood with a circular neighborhood, when the points on the circular boundary are not integers or do not fall within a certain pixel grid, the gray value of the point can be calculated using the bilinear difference[37][38].The (8) can define this LBP operator.

$$LBP_{P,R}(x_c,y_c)=\sum_{p=1}^{P}s\left(I(p)-I(c)\right)\times 2^P \tag{8}$$

where P represents the number of pixels, R is the radius of the circle, (x_c,y_c) denotes the center point coordinates, and *I(p)* and *I(c)* denote the grayscale value of the middle path sample point and the center point, respectively. In this study, 8 pixels (P=8) are surrounding the central pixel with a radius of 1 (R=1). As can be seen in Fig. 6, comparing the size of the central pixel value with the other eight pixel values, greater than the central value is 1 and vice versa is 0. According to the rotation invariance of circular LBP, the value of LBP feature is obtained as 63.

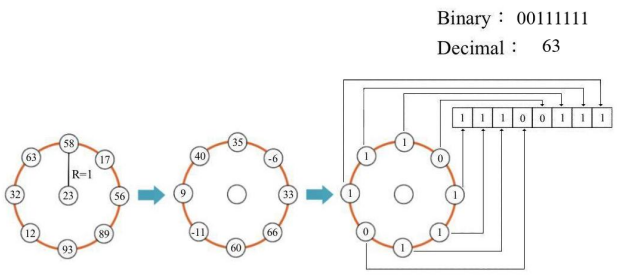


Fig. 6. Illustration of circular LBP operator

C. GLCM Features

The Gray Level Co-occurrence Matrix (GLCM) is a matrix that describes the gray level relationship between a pixel in a local area or a whole area of an image and the neighboring pixels or pixels within a certain distance. In short, the number of times a pixel pair with the same gray value at a specified distance and direction appears in the entire image [39]. In this study, the features used in GLCM are contrast, energy, correlation and homogeneity.

1) Contrast: It reflects the total local grayscale variation of the image. If the local pixel pair has a greater difference in grayscale, the greater the contrast of the image and the clearer the visual effect of the image[40]. (9) is the formula for calculating the contrast ratio.

$$Contrast = \sum_{i=0}\sum_{j=0}\left|i-j\right|^2 p(i,j) \tag{9}$$

2) Energy: It reflects the degree of uniformity of image gray distribution and texture coarseness. A large energy value indicates that the current texture is a more stable texture with regular changes. (10) is the formula for calculating energy.

$$Energy = \sum_{i=0}\sum_{j=0} p^2(i,j) \tag{10}$$

3) Correlation: It describes the degree of similarity between the row or column elements in GLCM, so the magnitude of the value reflects the local grayscale correlation, and the larger the value, the larger the correlation. (11) is the formula for calculating correlation.

$$Correlation = \sum_{i=0}\sum_{j=0}\frac{(i-\mu i)\times(j-\mu j)\times p(i,j)}{\sigma i \sigma j} \tag{11}$$

4) Homogeneity: It is a measure of the grayscale homogeneity of the image, and the value is larger if the grayscale of the image is locally homogeneous. (12) is the formula for calculating homogeneity.

$$Homogeneity = \sum_{i=0}\sum_{j=0}\frac{p(i,j)}{1+\left|i-j\right|} \tag{12}$$

D. Other Features

In addition to extracting the six texture features mentioned above, the entropy of the image is also extracted as a texture feature. It reflects the amount of average information in the image. The entropy of the image can be used as a feature since the amount of information varies depending on the action performed by the person. We also extracted four common numerical features based on [27][41]. This includes torso mean Doppler frequency, the Micro-Doppler shift, Total Doppler bandwidth, and the torso Doppler bandwidth. The torso mean

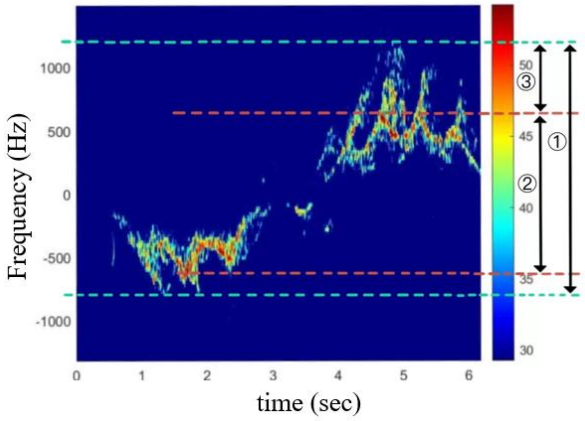


Fig. 7. Numerical features of human walking spectrograms

Doppler frequency corresponds to the speed of the volunteer. Human torso velocity is a basic but important piece of information because torso velocity varies greatly from one activity to another. Total Doppler bandwidth ① is related to the speed of limb movement, where a rapid swing of the arm or leg produces a large bandwidth. Micro-Doppler shift ③ is a measure of the asymmetry between forwarding and backward limb movement. The torso Doppler bandwidth ② represents the Doppler bandwidth of the trunk only. It characterizes the oscillatory motion of the torso as the body performs the activity.

IV. Result And Discussions

In this study, three maximum likelihood classifiers, namely SVM, KNN, and D-tree are considered to analyze the and D-tree are considered to analyze the performance of various feature datasets. The Table II describes the classification performance of the data set with various features using the classifier. The 70% of the data are used for training and 30% for testing. Upon detailed analysis, the classification performance of the classifier on the original dataset is quite low, due to the existence of many redundant information and high dimensionality of the original data, which results in long classification time and low accuracy. Texture features and numerical features are approximately effective in classifying inter-class motions, while texture features are more effective in classifying intra-class motions, in the D-Tree classifier for the motion of half-squat, the classification accuracy is increased by 11% using texture features compared to numerical features. The classification effect is increased to some extent after the hybrid features are performed, mainly due to two reasons, one is due to the increased feature types, and then the texture features perform better than the numerical features.

In order to verify the positive effect of threshold processing, the experiments compare the results of thresholded and untreated, and the green font in the table represents the accuracy rate without threshold processing. Since numerical features are more sensitive to micro-Doppler information, the threshold processing is not appropriate for the optimization of numerical features, and the feature values may be excessively wide apart before and after processing, resulting in incorrect feature information. The threshold processing is more effective in classifying texture features, and the overall accuracy after

TABLE II

CLASSIFICATION RESULTS OF THREE CLASSIFIERS ON THREE FEATURE DATASETS

(the green font in the table represents the accuracy rate without threshold processing,and the black represents the result with thresholding)

classifier models	Types of extracted features	Human Motions					Overall Accuracy
		walking	walking with a cane	jumping	half squatting	pick up things	
SVM	Accuracy of Raw data(%)	56	19	85	15	27	40.4
	Accuracy of Texture features(%)	100 / 100	100 / 100	100 / 100	77 / 70.3	85 / 83.2	92.4 / 90.7
	Accuracy of Numerical features(%)	100	96.6	96.6	72	78	88.4
	Accuracy of Integration features(%)	100 / 100	100 / 100	100 / 100	85 / 80.5	81 / 80.6	93.2 / 92.2
KNN	Accuracy of Raw data(%)	71	74	65	59	54	64.4
	Accuracy of Texture features(%)	100 / 100	96.8 / 96	95 / 100	85 / 80.5	74 / 63.6	90 / 88
	Accuracy of Numerical features(%)	100	92	100	73	68	86.6
	Accuracy of Integration features(%)	100 / 100	89 / 89	100 / 100	89.3 / 85.3	80 / 67.3	91.6 / 88.3
D-Tree	Accuracy of Raw data(%)	39	15	90	17	3	32.9
	Accuracy of Texture features(%)	100 / 100	88 / 79	95 / 97	81 / 67.3	85 / 82.6	87.8 / 85.2
	Accuracy of Numerical features(%)	100	90	92	70	81	86.6
	Accuracy of Integration features(%)	100 / 100	91 / 92	96 / 95	75 / 72	88 / 66	90 / 85

processing is more than that of the unprocessed ones, by up to 15%. Nevertheless, it does not mean that the texture features are not interested in micro-Doppler information, for the jump, the accuracy of KNN and D-Tree exhibits less accuracy after thresholding than the unprocessed results, which may be caused by the fact that thresholding filters out some micro-Doppler information, yet this decrease is acceptable for the improvement of the overall accuracy rate.

Depending on the results of the three classifiers, the SVM classifier using the RBF kernel function reaches the best results for all three feature sets, and the classification of mixed features can reach 93.2%, as well as 88.4% and 92.4% for numerical and texture features. The performance of SVM is 1.6% better than KNN and 3.2% better than D-Tree. Further comparing the results without thresholding, SVM improves 3.9% over KNN and 7.2% over D-Tree, once again illustrating that the performance improvement after thresholding is substantial.

To verify the effectiveness of the proposed method, the model of this paper is compared with four existing methods, including Alexnet and Googlenet based feature extraction methods [21], which send the features to different classifiers and networks for classification. The data augmentation and principal component analysis (DA-PCA) are used for dimensionality reduction in [42], which processes the spectrograms and then directly the inputs of classifiers such as Random Forest, long short-term memory (LSTM), bi-directional Long short-term memory (BiLSTM), convolutional neural network (CNN) and other classifiers. Finally, the generalized PCA (GPCA) is used as features [43], GPCA features are derived from range map and spectrograms respectively, and at last the corresponding two sets of features are connected into a single feature vector as the input of KNN classifier. The data collected in this paper were applied to the above model, and the classification comparison results are shown in Table III.

This dataset was collected in an environment similar to room B and room D in [21], so the results are closer. When using the model in [42], the optimal results were also obtained using CNN, but the overall results for each classifier were lower than the results in the reference paper, probably due to the different data set acquisition environments and acquisition actions. In the comparison of all models, the integration features (mixed texture features and numerical features) proposed in this paper still achieved the best results, while the classification results using texture features also outperformed the other models, confirming the effectiveness of texture features and the superiority of the present model through comparative experiments.

TABLE III

CLASSIFICATION RESULTS OF THE SAME DATA SET IN FOUR MODELS

Feature Extraction	Classificaton	Accuracy(%)
Alexnet	SVM	85.2
	KNN	83.6
Googlenet	Googlenet	74.7
DA-PCA	Random Forest	87.6
	KNN	85.3
	SVM	69.8
	LSTM	80.3
	BiLSTM	85.5
	CNN	92.6
GPCA	KNN	86.3
Numerical features	SVM	88.4
	KNN	86.6
	D-Tree	86.6
Texture features	SVM	92.4
	KNN	90
	D-Tree	87.8
Integration features	SVM	**93.2**
	KNN	91.6
	D-Tree	90

The F1 score matrix is used to measure the accuracy of the classification model, which is the quantitative result of combining the precision and recall metrics, with a maximum of

TABLE IV

F1 SCORE RESULTS

(the green font in the table represents the accuracy rate without threshold processing, and the black represents the result with thresholding)

Human Motions	SVM			KNN			D-Tree		
	NF	TF	IF	NF	TF	IF	NF	TF	IF
walk	0.97	0.95 / 0.94	1 / 1	1	0.95 / 0.95	0.96 / 0.96	0.87	0.86 / 0.82	0.93 / 0.92
walking with a cane	0.98	1 / 1	1 / 1	0.95	0.97 / 1	0.91 / 0.86	0.94	0.93 / 0.88	0.97 / 0.96
jumping	0.98	1 / 1	1 / 1	0.95	0.97 / 1	1 / 1	0.95	0.97 / 0.98	0.97 / 0.97
half squatting	0.73	0.83 / 0.76	0.83 / 0.8	0.7	0.8 / 0.71	0.83 / 0.76	0.73	0.79 / 0.69	0.8 / 0.7
pick up things	0.75	0.83 / 0.79	0.82 / 0.8	0.72	0.78 / 0.7	0.83 / 0.73	0.75	0.83 / 0.77	0.83 / 0.68

1 and a minimum of 0. The specific results are shown in Table IV, where each action is classified as a binary model, where NF, TF and IF are abbreviations for numerical features, texture features and integration features, respectively.

V. CONCLUSION

This paper presents a non-contact and non-privacy-invasive indoor human motions detection system, which classifies daily actions by combining ML with micro-Doppler radar. For this purpose, the MDR obtains the motion target echo data, performs a time-frequency transformation, and extracts numerical features in the frequency domain as well as texture features of the image from the spectrogram, and classifies them using SVM, KNN and D-Tree classifiers to detect motion target specific actions. Furthermore, the SVM classifier based on RBF kernel outperforms KNN and D-Tree for all three feature sets, and the results indicate that the classification accuracy of texture features outperforms numerical features for approximate actions, while there is little variation between the two performances for inter-class actions. Compared with [19][20][21], the inclusion of weak clutter suppression in addition to MTI clutter suppression intends to maximize the clutter suppression, which makes the classification results more desirable and stable compared to the next. Moreover, the classification performance of micro-Doppler features and texture features are compared in the experimental part, which confirms the contribution of texture features to classification.

The technique proposed in this paper demonstrates the great potential of texture features in target motion detection, and for future research, more texture features will be considered to be extracted , and features that are more suitable for motion classification will be selected to further improve the classification effect.

REFERENCES

[1]*More Americans living alone as milestones slip later in life*; Accessed on: December. 1, 2021. [online]. Available: https://thehill.com/homenews/state-watch/583723-more-americans-living-alone-as-milestones-slip-later-in-life/

[2]Bloom D E, Boersch-Supan A, McGee P, et al, "Population aging: facts, challenges, and responses," *Benefits and compensation International*, vol. 41,no. 1,pp. 22, 2011.

[3]L.Wang, D.Q. Huynh and P.Koniusz, "A Comparative Review of Recent Kinect-Based Action Recognition Algorithms," *IEEE Transactions on Image Processing*, vol. 29, pp. 15-28, 2020.

[4]S. Kocaoˇglu and Y. G¨uven , "Elderly Fall Detection and Fall Direction Detection via Various Machine Learning Algorithms Using Wearable Sensors," *Hittite Journal of Science and Engineering*, vol. 8, no. 3, pp. 197-205, Sep. 2021.

[5]S. Z. Gurbuz and M. G. Amin, "Radar-Based Human-Motion Recognition With Deep Learning: Promising Applications for Indoor Monitoring," *IEEE Signal Processing Magazine,* vol. 36, no. 4, pp. 16-28, July 2019.

[6]W. Taylor, K. Dashtipour, S. A. Shah, A. Hussain, Q. H. Abbasi, and M. A. Imran, "Radar Sensing for Activity Classification in Elderly People Exploiting Micro-Doppler Signatures Using Machine Learning," *Sensors*, vol. 21, no. 11, p. 3881, Jun. 2021.

[7]K. Saho, K. Shioiri, M. Fujimoto and Y. Kobayashi, "Micro-Doppler Radar Gait Measurement to Detect Age- and Fall Risk-Related Differences in Gait: A Simulation Study on Comparison of Deep Learning and Gait Parameter-Based Approaches," *IEEE Access*, vol. 9, pp. 18518-18526, 2021.

[8]A. K. Seifert, M. G. Amin and A. M. Zoubir, "New analysis of radar micro-Doppler gait signatures for rehabilitation and assisted living," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 4004-4008.

[9]K. Sehairi, F. Chouireb and J. Meunier, "Elderly fall detection system based on multiple shape features and motion analysis," 2018 International Conference on Intelligent Systems and Computer Vision (ISCV), 2018, pp. 1-8.

[10]Kong, Chen, Wang, Chen, Meng, and Tomiyama, "Robust Self-Adaptation Fall-Detection System Based on Camera Height," *Sensors*, vol. 19, no. 17, p. 3768, Aug. 2019.

[11]Y. Zhang, Z. Zhang, Y. Zhang, J. Bao, Y. Zhang and H. Deng, "Human Activity Recognition Based on Motion Sensor Using U-Net," *IEEE Access*, vol. 7, pp. 75213-75226, 2019.

[12]V. C. Chen, W. J. Miceli, and D. Tahmoush, Radar micro-Doppler signatures: processing and applications. *The Institution of Engineering and Technology*, 2014.

[13]C. Clemente, L. Pallotta, A. De Maio, et al, "A novel algorithm for radar classification based on doppler characteristics exploiting orthogonal Pseudo-Zernike polynomials," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 417-430, January 2015.

[14]Chen V C, Li F, Ho S S, et al, "Analysis of micro-Doppler signatures," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 150, no. 4, pp. 271-276, 2003.

[15]Pallotta L, Cauli M, Clemente C, et al, "Classification of micro-Doppler radar hand-gesture signatures by means of Chebyshev moments," 2021 IEEE 8th International Workshop on Metrology for AeroSpace (MetroAeroSpace), 2021, pp. 182-187.

[16] V. C. Chen, The micro-Doppler effect in radar, *Artech house*, 2019.

[17]B. Erol, M. G. Amin, and S. Z. Gurbuz, "Automatic data-driven frequency-warped cepstral feature design for micro-doppler classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1724–1738, 2018.

[18]Amin M G, Ravisankar A, Guendel R G, "RF sensing for continuous monitoring of human activities for home consumer applications," *Big Data: Learning, Analytics, and Applications*, vol. 10989. SPIE, 2019.

[19]Guendel R G, Fioranelli F, Yarovoy A, "Phase-based classification for arm gesture and gross-motor activities using histogram of oriented gradients," *IEEE Sensors Journal*, vol. 21,no. 6, pp. 7918-7927, 2020.

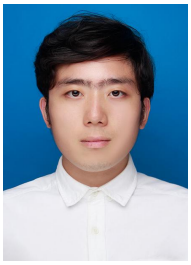
[20]Zenaldin, M. and Narayanan, R.M, "Radar micro-Doppler based human activity classification for indoor and outdoor environments," *Radar Sensor Technology XX,* 2016, pp. 364-373.
[21]S. A. Shah and F. Fioranelli, "Human Activity Recognition : Preliminary Results for Dataset Portability using FMCW Radar," *2019 International Radar Conference (RADAR)*, Toulon, France, 2019, pp. 1-4.
[22]R. P´erez, F. Schubert, R. Rasshofer and E. Biebl, "Deep Learning Radar Object Detection and Classification for Urban Automotive Scenarios," 2019 Kleinheubach Conference, 2019, pp. 1-4.
[23]C. Will, P. Vaishnav, A. Chakraborty and A. Santra, "Human Target Detection, Tracking, and Classification Using 24-GHz FMCW Radar," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7283-7299, September. 2019.
[24]M. Ash, M. Ritchie and K. Chetty, "On the Application of Digital Moving Target Indication Techniques to Short-Range FMCW Radar Data," *IEEE Sensors Journal*, vol. 18, no. 10, pp. 4167-4175, May. 2018.
[25]S. Neemat, O. Krasnov and A. Yarovoy, "An Interference Mitigation Technique for FMCW Radar Using Beat-Frequencies Interpolation in the STFT Domain," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 3, pp. 1207-1220, March. 2019.
[26]Mateo, C., Talavera, J.A. "Bridging the gap between the short-time Fourier transform (STFT), wavelets, the constant-Q transform and multiresolution STFT," *Signal, Image and Video Processing*, vol. 14, no. 8, pp. 1535-1543, 2020.
[27]Y. Kim and H. Ling, "Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328-1337, May 2009.
[28]B. Tekeli, S. Z. Gurbuz and M. Yuksel, "Information-Theoretic Feature Selection for Human Micro-Doppler Signature Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2749-2762, May 2016.
[29]F. Fioranelli, M. Ritchie, S. Z. G¨urb¨uz and H. Griffiths, "Feature Diversity for Optimized Human Micro-Doppler Classification Using Multistatic Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 2, pp. 640-654, April 2017.
[30]H. Li, A. Mehul, J. Le Kernec, S. Z. Gurbuz and F. Fioranelli "Sequential Human Gait Classification With Distributed Radar Sensor Fusion," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7590-7603, March. 2021.
[31]W. Li, B. Xiong and G. Kuang, "Target classification and recognition based on micro-Doppler radar signatures," 2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL), 2017, pp. 1679-1684.
[32]Shi, X, Zhou, F, Liu, L, and Zhao, B, "Textural feature extraction based on time –frequency spectrograms of humans and vehicles," *IET Radar,Sonar & Navigation*, vol. 9, no. 9, pp. 1251-1259, December. 2015.
[33]F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 722-733, July 1996.
[34]F. Suard, A. Rakotomamonjy, A. Bensrhair and A. Broggi, "Pedestrian Detection using Infrared images and Histograms of Oriented Gradients," *2006 IEEE Intelligent Vehicles Symposium*, 2006, pp. 206-212.
[35]T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July. 2002.
[36]Hassaballah, M, Alshazly, H.A, and Ali, A.A, "Ear recognition using local binary patterns: A comparative experimental study," *Expert Systems with Applications*, vol. 15, no. 118, pp. 182-200, Mar. 2019.
[37]Guo Z, Zhang L, Zhang D, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern recognition*, vol. 43, no. 3, pp. 706-719, 2010.
[38]M. N. Khan, A. Das, M. M. Ahmed, "Multilevel weather detection based on images: A machine learning approach with a histogram of oriented gradient and local binary pattern-based features," *Journal of Intelligent Transportation Systems*, vol. 25, no. 5, pp. 513-532, 2010.
[39]Kumar, Dharmender. "Feature extraction and selection of kidney ultrasound images using GLCM and PCA," *Procedia Computer Science*, vol. 167, pp. 1722-1731, 2020.
[40]M. Bhave and M. Wanjari, "IRetrieval: Image Retrieval Based on Color Feature and Texture Feature," *International Journal of Advanced Research in Computer Science*, vol. 5, no. 6, pp. 165-167, 2014.
[41]H. Li, A. Mehul, J. Le Kernel, S. Z. Gurbuz and F. Fioranelli, "Sequential Human Gait Classification With Distributed Radar Sensor Fusion," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7590-7603, 15 March15, 2021.
[42]W. Taylor, K. Dashtipour, S. A. Shah, A. Hussain, Q. H. Abbasi, and M. A. Imran, "Radar Sensing for Activity Classification in Elderly People Exploiting Micro-Doppler Signatures Using Machine Learning," *Sensors*, vol. 21, no. 11, p. 3881, Jun. 2021.
[43]B. Erol and M. Amin, "Generalized PCA Fusion for Improved Radar Human Motion Recognition," *2019 IEEE Radar Conference (RadarConf)*, Boston, MA, USA, 2019, pp. 1-5.

Ziqian Wang is working toward the master's degree in electronic and communication engineering, School of Electronic Engineering, Xidian University, Xi'an, China.

His research interests in micro-Doppler human motion recognition.

Aifeng Ren, received the Ph.D. degree in information and telecommunication engineering from Xi'an Jiaotong University, Xi'an, China, in 2007.

He is currently an Professor with School of Electronic and Engineering, Xidian University. He has authored/co-authored over 40 journal and conference publications. His research interests include embedded wireless sensor networks, terahertz sensing technology, complex brain network, and signal and image processing.



Qi Zhang, a student, is pursuing the Master's degree in Electronic Science and Technology from Xidian University, Xi'an, China.

He is interested in micro-Doppler human motion recognition.

Adnan Zahid, received his B.Sc. (Hons) in Electronics and Communications Engineering from Glasgow Caledonian University, and MSc degree in Electronics and Electrical Engineering from University of Strathclyde, in 2016. He is currently pursuing his Ph. D Research Degree in University of Glasgow. His research interests include Machine learning enabled Terahertz sensing for precision agriculture technology at cellular level.

Qammer H. Abbasi (SM'16), received his BSc and M. Sc. Degrees electronics and telecommunication engineering from the University of Engineering and Technology (UET), Lahore, Pakistan, and Ph.D. degree in electronic and electrical engineering from the Queen Mary University of London (QMUL), U.K. , in 2012. He is currently a Lecturer (Assistant Professor) with the School of Engineering, University of Glasgow, U.K. He has contributed to a over 250 leading international technical journal and peer reviewed conference papers, and 8 books.

He is an Associate editor for IEEE Journal of Electromagnetics, RF, and Microwaves in Medicine and Biology, the IEEE Sensors Journal, IEEE open access Antenna and Propagation, IEEE Access jouranl and acted as a guest editor for numerous special issues in top notch journals.