A Survey about application based on RF signals

Jiaren Xiao
212020081200026@stu.xhu.edu.cn
Xihua University

Abstract

With the popularity of commercial wireless devices, the research based on RF signal has developed rapidly in the past decade. RF signals have great advantages in environmental perception. Wireless signals are transmitted through the transmitter. With the propagation of these signals in the medium, they are reflected from different objects and people in space to the receiver. In this process, rich environmental perception information is carried. These ubiquitous perceptual information has rich application significance and bright prospects in the fields of indoor positioning, motion perception, behavior recognition, human-computer interaction, fall detection, health monitoring, smart home, search and rescue and so on. This paper will discuss and analyze the current research results by combing some existing literature and the research status of relevant researchers in relevant fields at this stage, so as to understand the future development trend of RF signal based applications.

1 introduction

The application of wireless signal is mainly divided into two directions: WiFi Based and radar based. Compared with WiFi, radar has stronger anti-environmental interference and can carry more environmental information and object information. This paper mainly summarizes the RF applications based on radar. According to the working mode, the radar can be divided into pulse radar\cite{1,2,3,4} and continuous wave radar\cite{3,4,5,6}. Pulse radar intermittently transmits rectangular pulse periodic signals and receives reflected echo signals in the transmission gap to perceive the environment, but pulse radar has a blind area for short-range detection. Continuous wave radar can be further subdivided into single frequency continuous wave (SFCM) and frequency modulated continuous wave (FMCW)\cite{5}. It can sense the environment by transmitting continuous wave and receiving the reflected echo signal at the same time, but continuous wave radar has the problems of signal leakage and background interference. Single frequency continuous wave can only be used for velocity measurement and cannot be used for ranging. FMCW can be used not only for velocity measurement and ranging, but also for distinguishing moving targets. Appropriate radar signal technology can be flexibly selected according to specific research.

Indoor location is the basis of RF research, and the premise of many studies is the need for location information. In 2013, Fadel Adib used a 3-antenna MIMO radio with low bandwidth and low power consumption\cite{9} to determine the number of people in closed rooms and their relative positions through the wall, and showed that it has great limitations to track the human body through the RF signal generated by human motion. In 2014, an indoor 3D positioning system WiTrack\cite{10} was proposed to realize the accurate positioning of indoor human body using RF signal, but this positioning system can only locate people in motion. In 2015, based on the improvement of WiTrack, WiTrack 2.0\cite{5} was proposed, in which the accurate positioning of multi person scene and static human body was realized. In the same year, a through wall human body contour capture system RF capture\cite{11} was proposed to track the 3D position of a person’s limbs and body parts, generate a coarse-grained human body contour, and distinguish different users through the contour information.
which laid a foundation for subsequent human body posture estimation. Subsequently, with the rapid
development of perception technology and deep learning based on RF signal, a series of applications
such as sleep detection [12, 13], fall detection [14], gesture recognition [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17],
radar imaging [17, 18, 19, 20], physiological feature monitoring [12, 13, 20, 21, 22], gait recognition [23, 24],
track tracking [25, 26] have been realized. It is widely used in video games, intelligent video
surveillance, smart home, human-computer interaction, security and other fields, and has broad
application prospects and economic value.

Environment sensing technology based on wireless signal has many unique advantages. Firstly,
without wearing any sensors, behavior perception based on wireless signals is realized by detecting
the RF signal characteristics reflected by the human body. Secondly, it has good obstacle crossing
ability and realizes non line of sight perception. Third, radar signal belongs to electromagnetic wave
and is less affected by external conditions such as light, temperature and humidity. Fourth, it has
strong scalability. With the continuous development of radar technology, the bandwidth and working
frequency of radar are increasing, and the number of antennas is also increasing, which lays a good
foundation for further improving the sensing accuracy.

In this review, we mainly review the research work based on RF in recent years, and the main
contributions are as follows:

- In section 2 we introduce the ranging principle, angle measurement principle and velocity
  measurement principle based on radar signal.
- In section 3 we describe the flow and method of radar signal preprocessing.
- In section 4 we introduce applications based on RF signals. It mainly includes indoor
  positioning, gesture recognition, physiological signal monitoring, human activity recognition
  and pose estimation.

![Figure 1: Organization of this paper.](image-url)
2 Background

Radar transmits electromagnetic energy pulses through the transmitter, and RF energy propagates to and reflects from the reflected object in the medium. A small part of the reflected energy is transmitted to the receiver of the radar device, which uses this echo to determine the direction, distance and speed of the reflected object. Assuming that the FM signal is used as the transmission signal, the transmission signal can be expressed as follows:

\[ S(t) = A_t e^{j2\pi (f_c t + \frac{k}{2}t^2)} \]  

(1)

Where, \( A_t \) represents the amplitude at time \( t \), \( f_c \) represents the radar carrier frequency, and \( k \) represents the frequency modulation slope. The reflected signal propagates in space and is reflected by the target object to the receiving antenna, so received signal can be expressed as follow:

\[ R(t) = A_t e^{j2\pi (f_c (t-\tau) + \frac{k}{2}(t-\tau)^2)} \]  

(2)

Where, \( \tau = \frac{2\Omega}{C} \), represents the instantaneous time delay of echo, \( A_t \) represents the amplitude at time \( t \), \( f_c \) represents the radar carrier frequency, and \( k \) represents the frequency modulation slope.

2.1 Ranging principle

The radar transmits a radio pulse signal with very high power. The pulse signal is concentrated in one direction through the directivity of the antenna, and a speed of light propagates in a given direction. If there is a target object in this direction, part of the pulse energy will be reflected back to the radar. The radar receives this energy signal and estimates the target object distance \( R \). Specific formula as follows:

\[ R = \frac{C \cdot t}{2} \]  

(3)

Where \( C \) represents the speed of light, \( t \) represents the propagation time of the signal in the medium, and \( R \) represents the oblique departure of the target object (line of sight distance).

In addition, the distance resolution is shown in follow:

\[ R_{res} = \frac{C}{2B} \]  

(4)

Where \( C \) represents the speed of light and \( B \) represents the bandwidth. It can be seen from the equation 4 that the greater the bandwidth, the better the distance resolution.

2.2 Angle measurement principle

The angle of the target is determined by the directivity of the antenna. By measuring the direction of the antenna when receiving the echo, the azimuth \( \theta \) and elevation angle \( \phi \) from the radar to the target object can be determined. The simplest angle measurement principle is shown in Figure 2. The phase difference of two receiving antenna signals caused by distance difference is expressed as:

\[ \omega = \frac{2\pi}{\lambda} d \sin \theta \]  

(5)

So we can get

\[ \theta = \sin^{-1} \left( \frac{\omega \lambda}{2\pi d} \right) \]  

(6)

In this way, the angle of antenna signal transmission is obtained, where \( d \) represents the distance between receiving antennas, \( \omega \) represents the phase difference between two antennas, and \( \lambda \) represents the wavelength.

In addition, the angular resolution is approximately as follow:

\[ \theta_{res} \approx \frac{2}{N} (rad) \]  

(7)

Where \( N \) represents the virtual antenna array element. According to the equation 7, MIMO radar technology is used to form an array element of virtual receiving antenna with the product of the number of transmitting antennas and the number of receiving antennas. By adding \( N \), the angular resolution of the radar can be simply and effectively improved.
2.3 Velocity measurement principle

For a moving target, there is a frequency $\Delta f$ difference between the transmitted signal and the received signal, which is called Doppler frequency, so the target speed $V$ can compute as follow:

$$V = \frac{f_d \cdot \lambda}{2} = \frac{\lambda \omega}{4\pi T_c}$$

Where $f_d$ represents the Doppler frequency, $\lambda$ represents the wavelength, $\omega$ represents the phase difference and $T_c$ represents a chirp period. In addition, it can be deduced that the maximum measurable unambiguous speed of the radar as follow:

$$V_{max} = \frac{\lambda}{4T_c}$$

From equation 9, it can be seen that the constraint of the maximum unambiguous speed is the phase difference, which between positive and negative 180rad, and the larger measurable unambiguous speed needs more dense chirp. In addition, the velocity resolution can be compute as the follow equation:

$$V_{res} = \frac{\lambda}{2NT_c}$$

Where $\lambda$ represents the wavelength and $NT_c$ represents the time of one frame. The shorter the wavelength, the longer the time of one frame, and the higher the speed resolution.

3 Data preprocessing

3.1 Eliminating flash effect

When the equipment collects the human body reflected signal behind the wall, because the RF signal will penetrate the Wall twice, the power of the signal itself will drop by several orders of magnitude, and the reflected signal from the wall will be much larger than the reflected signal from the human body. This strong reflection from the wall will drown the analog-to-digital converter of the receiver, Preventing it from recording small changes in the reflected signal from the real human body is called the "flash effect". To solve this problem, either increase the signal bandwidth so that it can record small changes from behind the wall, isolate the reflection of the wall from the reflection of the object behind the wall, or install an additional receiving antenna behind the wall to synchronize the time of the received signal. In [9], a MIMO interference zeroing technology is proposed to eliminate the flash effect. The whole process is divided into three stages: initialization zeroing, power boost and iterative zeroing. Firstly, two known signals are transmitted on the two transmitting antennas in turn, and then the channels between the two transmitting antennas and the receiving antenna are estimated with the signals received by the receiving end, and then the two channels are used to calculate the ratio. Through this ratio, the state in the joint channel can be calculated. If no object moves, the channel is empty, which eliminates the reflected signals from all static objects.
3.2 Eliminate multi-path effect

Multi-path effect consists of two parts. One part is the static background emission in the environment, which is called static multi-path; The other part is the signal reflected by the human body to the static background, and then reflected by the static background to the receiving antenna, which is called dynamic multi-path. This multipath effect is reflected in radar imaging, which will produce multiple spots. This spot will not prevent the system from detecting humans, but will hinder human reflection and multipath reflection, thus limiting the ability to locate humans in the real environment. Since the TOF time generated by static objects in the environment does not change with time, based on this feature, the interference of static multi-path effect is eliminated by subtracting continuous RF frames. The elimination of dynamic multi-path is more complex. Because the dynamic multi-path is caused by the indirect reflection of the human body, and the TOF will change with time, the idea of dynamic multi-path elimination is that the path of the receiving antenna receiving the direct reflection from the human body at any time point is shorter than any indirect reflection path. Since distance and TOF are directly related, and TOF is related to frequency, distance and frequency are directly related. Based on this idea, the reflection through the shortest path is tracked by tracking the bottom contour (a local maximum closest to the device) in all strong reflection surfaces, so as to eliminate the dynamic multi-path.

3.3 Elimination of near-far effect

The near-far effect is a major challenge in multi-person positioning, which means that the intensity of the signal reflected by the human body close to the equipment is much higher than that far away from the equipment, which will blur the signal from the distant person, resulting in inaccurate positioning and even undetected. In order to eliminate this near far effect, SSC algorithm is proposed, which is divided into four steps: SSC detection, SSC remapping, SSC elimination and SSC iteration. Find the position of the strongest human body by superimposing the heat-maps of signals from different antennas at the same time, then remapping the position of this person into TOF, then set the TOF configuration of this point in all antenna pairs to zero, finally recalculate the heat map according to the new TOF configuration file, and then repeat the above steps until all people are found.

3.4 Filter denoising

In addition to normal reflection signals, the collected signals are often mixed with random noise signals, which are caused by the stability of equipment level and weak anti-interference. These noise signals are not of interest to us and will affect the accuracy of the whole experiment. The existence of these random noises is inevitable. The influence of noise on experimental results can only be reduced by interpolation, deletion and filtering. The idea of interpolation is that the noise is accidental, and the normal signal value before and after the abnormal value is used to replace the abnormal value. Deletion is to discard the whole abnormal signal. Since the uniform motion of the human body is continuous and smooth, a filter can be used to smooth the convex outliers, For example, Butterworth filter, discrete wavelet filter, Kalman filter etc.

4 RF application

4.1 Indoor positioning

Indoor locating is an important field of RF application. The work of early researchers mainly focused on Simulation and modeling. In the last decade, researchers began to shift their attention to mobile humans, and even through wall positioning can be realized. These indoor positioning methods can be divided into equipment positioning and no equipment positioning. For device positioning, individuals need to wear RF signal transmitting devices, such as mobile phones, smart bracelets, etc. The equipment free positioning technology does not need to wear equipment on the body. It uses the reflected RF signal of the human body to track the human body positioning. Because this positioning technology completely depends on the weak and low-power RF reflection outside the human body, which is easily blocked by the indoor complex environment, such as metal objects, mirrors, walls, etc. The work of early researchers mainly focused on Simulation and modeling. In the last decade, researchers began to shift their attention to mobile humans. In terms of positioning principle, it can be divided into RSSI, AOA and TOF. The location technology based
on RSSI mainly depends on the attenuation of the signal to determine the location. AOA determines
the position according to the angle between the human body and different antennas. TOF determines
the position by the time the signal propagates in space.

4.1.1 Base on RSSI

When the distance is close, the RSSI attenuation is faster, and when the distance is far, the RSSI
attenuation is slower. The positioning technology based on RSSI estimates the signal strength value
between the transmitting point and the receiving point according to this characteristic, and then
converts the propagation loss into the distance between the receiving point and the transmitting point,
and then calculates the location of the unknown node according to the relevant algorithm. In [28],
a positioning framework based on hybrid dual frequency RSSI is proposed. By setting multiple
reference positioning points in advance, and then using a smart phone to collect and preprocess the
dual RF RSSI, a dual RF RSSI fingerprint model is established to improve the positioning accuracy.
This positioning technology based on RSSI has low hardware requirements, low power consumption
and easy implementation. It is suitable for simple environment. Once the indoor environment is
complex, the positioning accuracy is greatly reduced. Therefore, this positioning technology is
generally suitable for small-scale and simple environment.

4.1.2 Base on AOA

In [9], a Wi-Vi system based on AOA is proposed. This method can even go through the wall to
determine the number of people and their relative position in an enclosed space, and recognize the
simple actions made by others behind the wall. Through tracking calculation with conventional
antenna array $\Theta$ In contrast to locating stationary users, Wi-VI uses a receiving antenna and uses
inverse synthetic aperture radar technology (show in Figure 3) to compare the position of mobile users
at different times as an antenna array for calculation $\Theta$. Locate and track dynamic users.

![Figure 3: ISAR. The moving object itself emulates an antenna array as an inverse synthetic aperture. We leverages this principle to locate the moving target object.](image)

4.1.3 Base on TOF

**Acquisition of TOF:** The measurement of TOF [10, 5] is difficult. Because the RF signal propagates
in space at the speed of light, the difference of tens of centimeters in space may be the reflection time
difference of tens of microseconds in time dimension, and the measurement of tens of microseconds is
difficult. The most direct way is to design a sub nanosecond high-speed digital to analog converter to
sample the transmitted pulse signal and obtain the TOF time. However, this method is expensive and
cost-effective. Therefore, a TOF estimation method based on FMCW is designed. As shown in the
Figure 4, the change in time is transformed into the change in frequency by using the characteristics
of FMCW. The carrier frequency of a series of chirp signals transmitted by FMCW changes linearly
in the cycle time, showing a repeated frequency scanning operation in the whole time dimension.
FMCW converts the time difference into the carrier frequency offset, which is easy to observe in the
spectrum of the received signal.
Figure 4: **FMCW operation.** For the carrier frequency sweep signal of the transmission period of the transmitting antenna, the received signal $f_y(t)$ has a time-shifted relative to the transmission signal $f_x(t)$, which is reflected as a frequency-shifted in the carrier frequency. TOF can be obtained by this frequency-shift $\Delta f$.

**Single person positioning:** The WiTrack\[10] system calculates the TOF through the frequency difference between the transmitted signal and the reflected signal, and then performs single person positioning in combination with the geometric position of the horizontal antenna. Through the estimated TOF calculation, a set of user positions is obtained. Mapping the set of positions into 2D space can obtain an ellipse. The ellipses obtained by two groups of different horizontal antennas will produce two groups of intersections in space, and then a group of intersection positions can be excluded according to the orientation of the antenna, so as to realize 2D positioning (as shown in the Figure 5). 3D positioning is to add a set of vertical antenna positions on the basis of two-dimensional.

Figure 5: **2D Localization.** The TOF estimate from a receiving antenna defines an ellipse whose foci are the transmitting antenna and the receiving antenna. It can uniquely localize a person using the intersection of two ellipses.

**Multi person positioning and static user positioning:** WiTrack2.0\[5] has been improved on the basis of WiTrack, which can not only locate multiple people, but also locate the static human body through the slight movement generated by human breathing. Previous positioning systems either needed dense and measured sensor grids to cover the whole space, or could not locate multiple users, which made them unsuitable for home environment. In addition, they require users to move constantly to detect their presence, requiring extensive prior calibration or training. WiTrack 2.0 solves these problems. It is a device free positioning system with multiple transmit and receive antennas. In the multi person positioning, the biggest challenge is how to eliminate the signal interference in the space medium and correctly match the TOF with the antenna. In multi person positioning, different people have different distances from the same antenna, and the same antenna has different distances from different people. As a result, multiple peaks will appear in the TOF profile of the received signal in the receiving antenna, and multiple TOF measurements will be obtained. If we can’t correctly match different TOF measurements with the correct antenna pair, a set of wrong ellipses will be generated,
which will locate people in the wrong position. In order to solve this challenge, it can be set a transmission interval $\tau$ to eliminate the signal interference in the space medium and $\tau > TOF_{\text{limit}}$ ($TOF_{\text{limit}}$ is the largest TOF in the environmental reflected signal), antennas transmit signals in different transmission $\tau$ successively at intervals (as show in Figure 6), so that different TOFs can be correctly matched with antenna pairs to accurately locate multiple users in the environment, get the correct spatial location of the user. Static user positioning is different from non-static user positioning. It uses long window subtraction for positioning and frames with long interval for static user positioning. The static user is positioned according to the small motion generated by inhalation and exhalation. Because the human respiratory motion can be regarded as static in a very short time, the use of short window subtraction will produce very fuzzy results.

![Figure 6: Multi-shift FMCW.](image)

Transmits FMCW signals from different transmitting antennas after inserting virtual delays $\tau$ between them. Each delay must be larger than the highest time-of-flight($TOF_{\text{limit}}$).

## 4.2 Gesture recognition and physiological sign monitoring

### 4.2.1 Gesture recognition

Gestures can be divided into micro gestures and macro gestures. Micro gestures generally refer to the small actions generated by a part of the hand. Macro gestures refer to the large actions of the whole hand moving from one position to another. People’s research on gestures mainly focuses on the field of single hand micro movement, and the gesture recognition of simultaneous movement of both hands remains to be explored. With the penetration of computer technology in all aspects of life, human-computer interaction has become an inevitable trend. Because radar has the ability of detecting micro motion with high precision, the use of radar and other RF sensors to develop human-computer interaction interface (HCI) based on gesture recognition (HGR) has attracted the attention of researchers. In the past decade, radar based HGR has witnessed a great climax and rapid development. From the type of radar, gesture recognition can be divided into ultra bandwidth pulse radar[1, 2, 3, 4] and continuous wave radar[6, 7, 15, 16, 8]. For pulse radar, it mainly uses the original data-driven deep learning method to recognize gestures, and the related feature extraction work is less. There are various feature extraction technologies in continuous radar. Appropriate classifiers are designed to recognize gestures through manually extracted classification features. These classifiers are based on supervised and unsupervised machine learning.

The main challenge of gesture recognition is the extraction of gesture features. Good gesture features can greatly improve the accuracy of recognition. Select appropriate radar technology for gesture motion acquisition, and then filter and format the received signal data to convert the data into appropriate format. The data format for gesture feature extraction[29] is mainly divided into the following six types:

- **Time-Amplitude**: The Time-Amplitude mainly uses the time-varying amplitude of the received signal to extract the gesture motion contour. This type of gesture signal is one-dimensional and can be used to develop a simple classifier based on signal processing[30].
- **Range-Amplitude**: Range-Amplitude is to extract features by using reflection intensity at different distances[31].
• **Time-Range**: The Time-Range is to obtain one-dimensional and two-dimensional signals to classify gestures by using the Time-Varying Distance of the received signal. That is, the amplitude of the distance of the recorded hand over time.

• **Time-Doppler**: Time-Doppler uses Time-Varying Doppler frequency shift to extract gesture features, and takes the change of Doppler frequency with time as the input of CNN for gesture recognition.

• **Range-Doppler frequency / Time-Doppler velocity**: Range-Doppler frequency / Time-Doppler velocity extracts gesture features by the change rate of Doppler frequency shift relative to distance.

• **Time-Frequency**: Time-Frequency is to detect gestures by observing the change of frequency with time.

Finally, after these data formats are converted into the required shapes, they can be input into an appropriate classifier for gesture classification.

### 4.2.2 Physiological sign monitoring

With the development of big data technology and the concept of healthy life, more and more people begin to care about their health. In the past few years, researchers have become more and more interested in ubiquitous health detection. By monitoring people’s physiological signals, such as sleep posture, sleep state, respiration and heart rate, using these information to estimate a person’s physical health and emotional state, they have made great contributions to the prevention of many major diseases. It has had a great impact on the medical system. At present, many monitoring methods are either based on health monitoring equipment and need to be installed on the human body, such as finger pulse oximeter, respiratory belt, smart watch, which brings inconvenience to people’s daily life. Or based on vision, this monitoring scheme not only has strong invasiveness and destroys people’s privacy, but also has strong requirements for the environment, which is basically ineffective in weak light and dark environment. RF based system is welcomed by researchers because of its non-contact, low invasiveness, strong resistance to environmental factors and sensitivity to micro movement. It is widely used in the monitoring field of daily physiological signs, such as respiratory and heartbeat rate monitoring, sleep quality and state, etc.

The monitoring of respiration and heartbeat rate is based on the RF signal changes caused by small movements generated by inhalation, exhalation and heartbeat. The changes caused by such movements are regular and periodic, but this change is far less than the changes caused by static background and dynamic background in the environment. How to separate the target signal Noise elimination is a major challenge in this application field. In [20], the multi antenna FMCW radar is used to monitor the breathing and heartbeat rate of a single person at the same time. The unnecessary motion and interference noise in the background are reduced through the superposition of multi antenna signals and the breathing and heart rate are detected through frequency filtering and the elimination of aperiodic signals. In [19], the interference caused by multiple RF signals is modeled, the original respiratory signals are separated by independent component analysis (ICA), and then the respiratory signals are optimally matched with the corresponding people to realize the monitoring of multi person respiration and heart rate. In [22], the system first locates each user in the environment, then amplifies the reflected signal of each user and analyzes its changes to extract respiration and heart rate. [21] cross ambiguity function (CAF) and depth transmission network (DTN) When the RF signal acts on the human body, it will be reflected from the human body according to the direction of the human body to form a multi-path signal indicating the posture of the human body. By separating the signals of different paths, these signals are combined with the received signals Associate with the subjects and construct an inverse mapping to predict the sleeping position of the human body [12]. [13] combine convolution network and recursive neural network to extract the subjective change characteristics of sleep and obtain the sleep process time to predict the sleep stage.

### 4.3 Human activity recognition

Human activity recognition is not only an important branch in the field of computer vision, but also a hot spot in the research of computer vision. It is widely used in the fields of game entertainment, smart home, human-computer interaction, security and so on. Simple activity recognition is action classification. Given a section of data signal, it is correctly classified into several known action
models. The activity recognition of complex points is behavior recognition. The data signal contains not only one action category, but multiple action categories. First identify the start time and end time of the action, and then classify the data action. The ultimate goal of activity recognition is to analyze who is at what time, where and what is doing in the signal. Human activity recognition can be simply divided into vision based and RF based from the form of input data.

Vision based Human activity recognition is a very extensive branch of activity recognition research at this stage. There have been many excellent methods and frameworks. These methods and frameworks can be simply divided into template based methods[37, 38], spatiotemporal attention based methods[39, 40], trajectory based methods[41, 42], double convolution flow based methods[43, 44] and three-dimensional convolution based methods[45, 46]. The template based method[37, 38] uses a group of templates to represent the actions to be identified, and classifies the actions to be tested by calculating the similarity between the actions to be tested and the template actions. The limitation of this method is the definition of standard template actions, which is only suitable for classifying simple actions. The method based on spatiotemporal attention[39, 40] uses the methods of detection[47, 48, 49, 50, 51] and segmentation[52, 53, 54, 55, 56] to extract the features of the region of interest in the video image, and send the extracted features to the classifier for classification. This method requires high accuracy of detection and segmentation, which is suitable for the situation with complex background. The idea of trajectory based method[41, 42] is that a trajectory will be generated with human movement, which can be expressed by the changes of the coordinates of the key points of the human skeleton, and the actions are classified by extracting the eigenvalues of this trajectory. The method based on dual stream network convolution[43, 44] extracts the features of the image through two independent streams - spatial convolution stream and temporal convolution stream. Finally, the features are fused and the actions are classified. This method is not suitable for the extraction of complex optical flow information and has a large amount of calculation. Three dimensional convolution[45, 46] expands one-dimensional acquisition time information on the basis of two-dimensional convolution. This method is simpler than that based on two stream convolution network. It can capture short-time time characteristics more directly, but it is not suitable for capturing long-time time information.

This vision based human activity recognition is vulnerable to the influence of the environment, will completely lose its effect in the weak light or dark environment, can not work, and is vulnerable to the influence of human appearance, clothing and other characteristics. Therefore, the human activity behavior based on RF has attracted the attention of researchers. The high-frequency RF signal has a strong ability to perceive the environment and human activities, and has a strong ability to resist environmental interference. The collected RF signals are preprocessed, and then the corresponding features are extracted for action classification.

4.3.1 Identity recognition

Identity recognition[23, 24, 25, 26, 57, 58, 59, 60, 61] is an important application field in RF applications. Identity recognition is a prerequisite for multi person activity recognition. Target identity is recognized by extracting the physical characteristics of the target, which is generally used to identify people and people or human and non-human targets. The continuous wave radar is used to collect the target signal data. After eliminating the irrelevant interference and filtering and denoising, the signal in the time domain is transformed into spectrum through FFT[62], and the features related to the physical features of the target, such as the frequency, amplitude, step, signal bandwidth and signal intensity distribution of the spectrum, are extracted from the spectrum to predict the target identity. [58] considered MD signature to identify individuals. [59] reported the results of identifying 8 individuals using a K-NN classifier and two manual features (subject’s stride and trunk line). In [60], Gaussian mixture models (GMMs) were used to identify individuals and distinguish male and female subjects according to manual characteristics. A total of 20 recordings of 30 subjects were used to train and test the developed model. Similarly, in [61], eight individuals were identified using GMM. SVM is trained by extracting the physical features of the target, which is used to classify the target, but it can not recognize the people who move aperiodically or fast. [24] MD signature is used to represent the two-dimensional spatial structure. As the input of DCNN, gait features are extracted to identify people moving freely in space, which makes the model more robust to deal with environmental changes. [57] Taking the RF track as the input, a hierarchical attention module is used to generate the feature map, and the multi-task learning method of recognizable features is used as supervision to extract a long-term recognition feature to recognize the target identity.
4.3.2 Activity recognition

**Action recognition:** In the traditional behavior recognition model, the features are usually extracted manually [63, 64, 65, 66, 14, 67, 68], and then classified by classifier. In [66], a human motion recognition framework based on environmental radar is proposed. A high-frequency radar is used to sample the reflected signal to capture the fine state of human activities. It has great limitations to classify some simple actions according to the manually extracted slope and relative position features. In [63], a method of using RF data projection splicing to improve the accuracy of behavior recognition is proposed and tested in two environments. In [64] solved the problem of continuity of human movement and used the nature of human activities to eliminate some impossible movements. In this study, the action is the conversion process of two states. The behavior can be described by the change from one state to another. The micro Doppler features and target distance extracted from the radar signal spectrum are input into two-dimensional PCA, and then classified by K-NN classifier. [65] used the Super bandwidth biological radar to classify human activities based on Doppler signals, and proposed a location information index classifier (PIIC), which improved the performance of the classifier. In [69], a DTL based method is proposed to separate motion activities from in-situ activities, then 2D PCA is used to extract features, and decision time classifier is used to classify actions. [14] used STFT to extract spectral features and SVM to detect falls.

**Behavior recognition:** With the increasing amount of data and the complexity of experimental environment, the limitations of traditional behavior recognition patterns are becoming more and more prominent. In the past few years, the progress of deep learning [26, 70, 71, 72, 73, 74] has promoted the progress of behavior recognition at an amazing speed. Therefore, researchers use deep learning to extract more advanced features to identify human behavior, and many excellent behavior recognition models based on deep learning have emerged, such as RF action [72], duet [70]. Marlo [26, 75] describes a data acquisition method, which generates a series of three-dimensional images according to the signal intensity reflected by the human body, inputs this continuous signal image into CNN in [73], extracts the spatial feature sequence, and extracts the temporal feature into RNN for human behavior recognition. [67] input radar data in the form of continuous time series of Doppler or range time information into Bi-LSTM or LSTM, and analyze continuous human activity sequences by using the time dependence between samples. [71] A behavior situation classification model based on RNN is proposed. In [72], a neural network model for detecting human behavior through the wall is proposed, and the action recognition model based on RF skeleton is used to understand people’s behavior and interaction. The RF signal is used as the input of the network to generate 3D human bones, which are input into the spatiotemporal attention module to extract the features of people of interest, and then these features are transmitted and input into the multi proposal to identify multi person actions and interactive actions. [25] Based on the progress of passive wireless location in the past, by continuously tracking people’s trajectory information to study family social behavior, an algorithm combining time and space is designed to generate human trajectory. The interaction behavior in three situations is analyzed in the form of trajectory + RF signal. [26] A multi-modal system is designed to infer the location and behavior of users through four components - identity matching subsystem, HMM, first-order logic framework and radio localization subsystem. Firstly, the trajectory is generated according to the RF signal, the trajectory is used as the input of HMM, and the identity of the trajectory data is matched according to the intermittent interaction between the user and the mobile phone to output the probability. The first-order logic component obtains the output of identity matching and HMM, and generates the best user behavior explanation. [74] proposed a framework for generating caption description behavior using RF signal, which is composed of four modules. The first module encodes and embeds RF skeleton and floormap together as the feature of the first module. The second module uses the video encoder to encode the paired and unpaired video after pooling as the feature of the second module. The third module aligns the video features and RF features to form a unified feature spectrum. The fourth module is the language generation module, which inputs the unified feature spectrum into the language generation network to generate words to describe human behavior.

4.4 Human pose estimation

**Voxel power:** [11] introduces a method of capturing human body contour to generate skeleton, RF capture, which scans 3D space by using 2D antenna array and FMCW chirps, first performs coarse resolution scanning to obtain the area with high reflection power, and then performs fine resolution scanning to amplify the area with reflection power, so as to obtain RF frames. The coordinates of
voxels in each 3D space are represented by spherical coordinates \((r, \theta, \phi)\). The power of each voxel in the frame can be calculated by the following formula:

\[
P(r, \theta, \phi) = \left| \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{t=1}^{T} S_{m,n,t} e^{j \frac{2\pi}{\lambda} r t} e^{j \frac{2\pi}{\lambda} \sin(\theta) (nd \cos \phi + md \sin \phi)} \right| \tag{11}
\]

In this formula, \(M, N\) represents the number of transmitting antennas and the number of receiving antennas respectively, \(S_{m,n,t}\) represents the signal from the transmitting antenna \(m\) to the receiving antenna \(n\) at time \(t\).

**Pose estimation:** RF based pose estimation faces a huge challenge - there is no large amount of annotation data. In order to train a reliable network to estimate human pose, we need to use the idea of cross-modal supervision. As an intermediate representation feature, skeleton provides the basis for cross-modal supervision. The human body is an irregular body. In the process of human movement, some RF signals will not be directly reflected to the receiving antenna. The human body information in a single RF frame is incomplete and only contains some limb information. This is the first challenge for RF skeleton generation. \[11\] Through the cross frame joint method, the complete human body information can be obtained, and then a coarse-grained 2D RF-skeleton can be obtained through depth compensation, swing compensation, human body segmentation and bone suture. \[13\] is a device free attitude estimation system RF-pose, which does not need to wear any sensors to extract accurate two-dimensional human posture by analyzing RF signals. The biggest problem of using RF to estimate human posture is that there is no labeled data, so it is impossible to generate a two-dimensional skeleton with high reliability. In this study, a cross-modal supervisory network model Teacher-Student is proposed to solve the problem of no labeled data. Synchronize RF and video streams by connecting a camera to a wireless sensor. The video stream is used as the input of the Teacher network and the RF stream is used as the input of the Student network. The information of human joint points is extracted from the video stream as the supervised training Student network to generate fine-grained 2D RF skeleton. After training, attitude estimation can be performed using only RF signals as input. In \[76\], human posture is estimated by generating 3D RF skeleton. The RF signal input by the system is a 4D tensor, which will greatly increase the amount of calculation and can not dynamically display the skeleton action in real time. Therefore, RF-pose3D projects this 4D tensor to different planes and decomposes it into two 3D tensors, using horizontal RF frames and vertical RF frames as the input of RF network. The computation of convolution network is greatly reduced. A multi-camera coordination system is designed to generate a video based 3D skeleton to train the 3D RF skeleton for estimating human posture. Based on the real-time 3D human posture estimation system with equipment \[77, 78\], the human posture is constructed by mapping the position of the additional RFID tag to 3D coordinates. Since the original RFID phase data will be seriously distorted due to channel jump and phase winding, first calibrate the phase, and use high-precision low rank tensor completion (HaLRTC) to interpolate the missing RFID data, and then conduct down sampling operation to synchronize with the 3D pose time series obtained by Kinect 2.0, so as to ensure the high accuracy and real-time performance of pose estimation. Finally, the calibrated phase data is used as input to train the depth neural network of human skeleton reconstruction, and the labeled visual data collected by Kinect 2.0 device is used as supervision.

## 5 Conclusion

With the popularity of security awareness, people pay more and more attention to their privacy. The environment sensing technology based on wireless signal has attracted extensive attention and attention of scholars because of its unique advantages - no need to wear any sensor, non line of sight sensing, less affected by external conditions, strong scalability, strong privacy protection and so on. It can be seen from the above articles that researchers try to start from two directions and hope to design and build an environment independent, location independent and life-oriented RF based environment perception technology in the future. One is the development direction of hardware, which designs an ultra-high frequency and ultra-bandwidth device to collect signals with richer characteristics. The other is the direction of software development. By considering more influencing factors, we design an ultra-high-performance algorithm to eliminate interference and apply it in all aspects of life.
References

[1] A. Arbabian, S. Callender, S. Kang, M. Rangwala, and A. M. Niknejad, “A 94 ghz mm-wave-to-baseband pulsed-radar transceiver with applications in imaging and gesture recognition,” IEEE Journal of Solid-State Circuits, vol. 48, no. 4, pp. 1055–1071, 2013.

[2] S. Heunisch, L. O. Flager, and L.-E. Wernersson, “Millimeter-wave pulse radar scattering measurements on the human hand,” IEEE Antennas and Wireless Propagation Letters, vol. 18, no. 7, pp. 1377–1380, 2019.

[3] J. Park and S. H. Cho, “Ir-uwb radar sensor for human gesture recognition by using machine learning,” in 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2016, pp. 1246–1249.

[4] S. Y. Kim, H. G. Han, J. W. Kim, S. Lee, and T. W. Kim, “A hand gesture recognition sensor using reflected impulses,” IEEE Sensors Journal, vol. 17, no. 10, pp. 2975–2976, 2017.

[5] F. Adib, Z. Kabelac, and D. Katabi, “Multi-person localization via {RF} body reflections,” in 12th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 15), 2015, pp. 279–292.

[6] Y. Sun, T. Fei, X. Li, A. Warnecke, E. Warsitz, and N. Pohl, “Real-time radar-based gesture detection and recognition built in an edge-computing platform,” IEEE Sensors Journal, vol. 20, no. 18, pp. 10 706–10 716, 2020.

[7] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, “Soli: Ubiquitous gesture sensing with millimeter wave radar,” ACM Transactions on Graphics (TOG), vol. 35, no. 4, pp. 1–19, 2016.

[8] A. A. Pramudita et al., “Time and frequency domain feature extraction method of doppler radar for hand gesture based human to machine interface,” Progress In Electromagnetics Research C, vol. 98, pp. 83–96, 2020.

[9] F. Adib and D. Katabi, “See through walls with wifi!” in Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM, 2013, pp. 75–86.

[10] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, “3d tracking via body radio reflections,” in 11th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 14), 2014, pp. 317–329.

[11] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, “Capturing the human figure through a wall,” ACM Transactions on Graphics (TOG), vol. 34, no. 6, pp. 1–13, 2015.

[12] S. Yue, Y. Yang, H. Wang, H. Rahul, and D. Katabi, “Bodycompass: Monitoring sleep posture with wireless signals,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 2, pp. 1–25, 2020.

[13] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, “Learning sleep stages from radio signals: A conditional adversarial architecture,” in International Conference on Machine Learning. PMLR, 2017, pp. 4100–4109.

[14] A. Alnaeb, R. S. A. R. Abdullah, A. A. Salah, A. Sali, N. E. A. Rashid, and I. Pasya, “Detection and classification real-time of fall events from the daily activities of human using forward scattering radar,” in 2019 20th International Radar Symposium (IRS). IEEE, 2019, pp. 1–10.

[15] S. Skaria, A. Al-Hourani, M. Lech, and R. J. Evans, “Hand-gesture recognition using two-antenna doppler radar with deep convolutional neural networks,” IEEE Sensors Journal, vol. 19, no. 8, pp. 3041–3048, 2019.

[16] E. Miller, Z. Li, H. Mentis, A. Park, T. Zhu, and N. Banerjee, “Radsense: Enabling one hand and no hands interaction for sterile manipulation of medical images using doppler radar,” Smart Health, vol. 15, p. 100089, 2020.
[17] M. Bocca, O. Kaltiokallio, N. Patwari, and S. Venkatasubramanian, “Multiple target tracking with rf sensor networks,” *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1787–1800, 2013.

[18] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, “Through-wall human pose estimation using radio signals,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7356–7365.

[19] S. Yue, H. He, H. Wang, H. Rahul, and D. Katabi, “Extracting multi-person respiration from entangled rf signals,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, pp. 1–22, 2018.

[20] Q. Liu, H. Guo, J. Xu, H. Wang, A. Kageza, S. AlQarni, and S. Wu, “Non-contact non-invasive heart and respiration rates monitoring with mimo radar sensing,” in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[21] Q. Chen, Y. Liu, B. Tan, K. Woodbridge, and K. Chetty, “Respiration and activity detection based on passive radio sensing in home environments,” *IEEE Access*, vol. 8, pp. 12426–12437, 2020.

[22] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, “Smart homes that monitor breathing and heart rate,” in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 837–846.

[23] Y. Kim, S. Ha, and J. Kwon, “Human detection using doppler radar based on physical characteristics of targets,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 289–293, 2014.

[24] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, “Indoor person identification using a low-power fmcw radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.

[25] C.-Y. Hsu, Y. Liu, Z. Kabelac, R. Hristov, D. Katabi, and C. Liu, “Extracting gait velocity and stride length from surrounding radio signals,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 2116–2126.

[26] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, “Enabling identification and behavioral sensing in homes using radio reflections,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[27] D. Sasakawa, N. Honma, K. Nishimori, T. Nakayama, and I. Shoichi, “Evaluation of fast human localization and tracking using mimo radar in multi-path environment,” in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2016, pp. 1–6.

[28] H. Sun, X. Zhu, Y. Liu, and W. Liu, “Construction of hybrid dual radio frequency rssi (hdrf-rssi) fingerprint database and indoor location method,” *Sensors*, vol. 20, no. 10, p. 2981, 2020.

[29] S. Ahmed, K. D. Kallu, S. Ahmed, and S. H. Cho, “Hand gestures recognition using radar sensors for human-computer-interaction: A review,” *Remote Sensing*, vol. 13, no. 3, p. 527, 2021.

[30] X. Gao, J. Xu, A. Rahman, E. Yavari, A. Lee, V. Lubecke, and O. Boric-Lubecke, “Barcode based hand gesture classification using ac coupled quadrature doppler radar,” in *2016 IEEE MTT-S International Microwave Symposium (IMS)*. IEEE, 2016, pp. 1–4.

[31] H.-S. Yeo, G. Flamich, P. Schrempf, D. Harris-Birtill, and A. Quigley, “Radarcat: Radar categorization for input & interaction,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, 2016, pp. 833–841.

[32] C. Zheng, T. Hu, S. Qiao, Y. Sun, J. Huangfu, and L. Ran, “Doppler bio-signal detection based time-domain hand gesture recognition,” in *2013 IEEE Mtt-S International Microwave Workshop Series on RF/And Wireless Technologies for Biomedical And Healthcare Applications (Imws-Bio)*. IEEE, 2013, pp. 3–3.
[33] S. Ahmed and S. H. Cho, “Hand gesture recognition using an ir-uwb radar with an inception module-based classifier,” *Sensors*, vol. 20, no. 2, p. 564, 2020.

[34] Y. Kim and B. Toomajian, “Hand gesture recognition using micro-doppler signatures with convolutional neural network,” *IEEE Access*, vol. 4, pp. 7125–7130, 2016.

[35] Y. Sang, L. Shi, and Y. Liu, “Micro hand gesture recognition system using ultrasonic active sensing,” *IEEE Access*, vol. 6, pp. 49 339–49 347, 2018.

[36] M. Zhao, F. Adib, and D. Katabi, “Emotion recognition using wireless signals,” in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 95–108.

[37] B. Luo, H. Li, F. Meng, Q. Wu, and K. N. Ngan, “An unsupervised method to extract video object via complexity awareness and object local parts,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 7, pp. 1580–1594, 2017.

[38] A. Klaser, M. Marszałek, and C. Schmid, “A spatio-temporal descriptor based on 3d-gradients,” in *BMVC 2008-19th British Machine Vision Conference*. British Machine Vision Association, 2008, pp. 275–1.

[39] S. Nazir, M. H. Yousaf, and S. A. Velastin, “Evaluating a bag-of-visual features approach using spatio-temporal features for action recognition,” *Computers & Electrical Engineering*, vol. 72, pp. 660–669, 2018.

[40] X. Peng, L. Wang, X. Wang, and Y. Qiao, “Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice,” *Computer Vision and Image Understanding*, vol. 150, pp. 109–125, 2016.

[41] H. Wang and C. Schmid, “Action recognition with improved trajectories,” in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 3551–3558.

[42] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, “Dense trajectories and motion boundary descriptors for action recognition,” *International journal of computer vision*, vol. 103, no. 1, pp. 60–79, 2013.

[43] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” *arXiv preprint arXiv:1406.2199*, 2014.

[44] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition,” in *Proceedings of the Neural Information Processing Systems (NIPS)*, 2015.

[45] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3d convolutional networks,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.

[46] H. Zhang, Y. Li, P. Wang, Y. Liu, and C. Shen, “Rgb-d based action recognition with light-weight 3d convolutional networks,” *arXiv preprint arXiv:1811.09908*, 2018.

[47] R. Zhang, F. Meng, H. Li, Q. Wu, and K. N. Ngan, “Category boundary re-decision by component labels to improve generation of class activation map,” *Neurocomputing*, 2021.

[48] F. Meng, K. Huang, H. Li, S. Chen, Q. Wu, and K. N. Ngan, “Hierarchical class grouping with orthogonal constraint for class activation map generation,” *Neural Computing and Applications*, vol. 33, no. 13, pp. 7371–7380, 2021.

[49] H. Qiu, H. Li, Q. Wu, F. Meng, K. N. Ngan, and H. Shi, “A2rmnet: Adaptively aspect ratio multi-scale network for object detection in remote sensing images,” *Remote Sensing*, vol. 11, no. 13, p. 1594, 2019.

[50] X. Chen, H. Li, Q. Wu, F. Meng, and H. Qiu, “Bal-r2cnn: High quality recurrent object detection with balance optimization,” *IEEE Transactions on Multimedia*, 2021.
[51] W. Li, H. Li, Q. Wu, F. Meng, L. Xu, and K. N. Ngan, “Headnet: An end-to-end adaptive relational network for head detection,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 482–494, 2019.

[52] X. Xu, F. Meng, H. Li, Q. Wu, K. N. Ngan, and S. Chen, “A new bounding box based pseudo annotation generation method for semantic segmentation,” in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020, pp. 100–103.

[53] H. Shi, H. Li, Q. Wu, and K. N. Ngan, “Query reconstruction network for referring expression image segmentation,” *IEEE Transactions on Multimedia*, vol. 23, pp. 995–1007, 2020.

[54] C. Shang, Q. Wu, F. Meng, and L. Xu, “Instance segmentation by learning deep feature in embedding space,” in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2444–2448.

[55] L. Yang, H. Li, Q. Wu, F. Meng, and K. N. Ngan, “Mono is enough: Instance segmentation from single annotated sample,” in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2020, pp. 120–123.

[56] L. Yang, F. Meng, H. Li, Q. Wu, and Q. Cheng, “Learning with noisy class labels for instance segmentation,” in *European Conference on Computer Vision*. Springer, 2020, pp. 38–53.

[57] L. Fan, T. Li, R. Fang, R. Hristov, Y. Yuan, and D. Katabi, “Learning longterm representations for person re-identification using radio signals,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 699–10 709.

[58] G. Garreau, C. M. Andreou, A. G. Andreou, J. Georgiou, S. Dura-Bernal, T. Wennekers, and S. Denham, “Gait-based person and gender recognition using micro-doppler signatures,” in *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2011, pp. 444–447.

[59] D. Tahmoush and J. Silvious, “Radar micro-doppler for long range front-view gait recognition,” in *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*. IEEE, 2009, pp. 1–6.

[60] K. Kalgaonkar and B. Raj, “Acoustic doppler sonar for gait recognition,” in *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2007, pp. 27–32.

[61] Z. Zhang and A. G. Andreou, “Human identification experiments using acoustic micro-doppler signatures,” in *2008 Argentine School of Micro-Nanoelectronics, Technology and Applications*. IEEE, 2008, pp. 81–86.

[62] A. Carr, L. Cuthbert, and A. Olver, “Digital signal processing for target detection fmcw radar,” in *IEE Proceedings F-Communications, Radar and Signal Processing*, vol. 128, no. 5. IET, 1981, pp. 331–336.

[63] D. Avrahami, M. Patel, Y. Yamaura, and S. Kratz, “Below the surface: Unobtrusive activity recognition for work surfaces using rf-radar sensing,” in *23rd International Conference on Intelligent User Interfaces*, 2018, pp. 439–451.

[64] M. G. Amin and R. G. Guendel, “Radar human motion recognition using motion states and two-way classifications,” in *2020 IEEE International Radar Conference (RADAR)*. IEEE, 2020, pp. 1046–1051.

[65] F. Qi, F. Liang, M. Liu, H. Lv, P. Wang, H. Xue, and J. Wang, “Position-information-indexed classifier for improved through-wall detection and classification of human activities using uwb bio-radar,” *IEEE antennas and wireless propagation letters*, vol. 18, no. 3, pp. 437–441, 2019.

[66] S. Zhu, J. Xu, H. Guo, Q. Liu, S. Wu, and H. Wang, “Indoor human activity recognition based on ambient radar with signal processing and machine learning,” in *2018 IEEE international conference on communications (ICC)*. IEEE, 2018, pp. 1–6.

[67] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, “Continuous human activity classification from fmcw radar with bi-lstm networks,” *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 607–13 619, 2020.
[68] K. Chetty, Q. Chen, M. Ritchie, and K. Woodbridge, “A low-cost through-the-wall fmcw radar for stand-off operation and activity detection,” in Radar Sensor Technology XXI, vol. 10188. International Society for Optics and Photonics, 2017, p. 1018808.

[69] R. G. Guendel, F. Fioranelli, and A. Yarovoy, “Derivative target line (dtl) for continuous human activity detection and recognition,” in 2020 IEEE Radar Conference (RadarConf20). IEEE, 2020, pp. 1–6.

[70] D. Vasisht, A. Jain, C.-Y. Hsu, Z. Kabelac, and D. Katabi, “Duet: Estimating user position and identity in smart homes using intermittent and incomplete rf-data,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 2, no. 2, pp. 1–21, 2018.

[71] M. Z. Uddin, F. M. Noori, and J. Torresen, “In-home emergency detection using an ambient ultra-wideband radar sensor and deep learning,” in 2020 IEEE Ukrainian Microwave Week (UkrMW). IEEE, 2020, pp. 1089–1093.

[72] T. Li, L. Fan, M. Zhao, Y. Liu, and D. Katabi, “Making the invisible visible: Action recognition through walls and occlusions,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 872–881.

[73] H. Guo, N. Zhang, S. Wu, and Q. Yang, “Deep learning driven wireless real-time human activity recognition,” in ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020, pp. 1–6.

[74] L. Fan, T. Li, Y. Yuan, and D. Katabi, “In-home daily-life captioning using radio signals,” in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer, 2020, pp. 105–123.

[75] H. Guo, N. Zhang, W. Shi, S. AlQarni, and S. Wu, “Real time 3d indoor human image capturing based on fmcw radar,” arXiv preprint arXiv:1812.07099, 2018.

[76] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, “Rf-based 3d skeletons,” in Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, 2018, pp. 267–281.

[77] C. Yang, X. Wang, and S. Mao, “Subject-adaptive skeleton tracking with rfid,” in 2020 16th International Conference on Mobility, Sensing and Networking (MSN). IEEE, 2020, pp. 599–606.

[78] C. Yang, X. Wang, and S. Mao, “Rfid-pose: Vision-aided three-dimensional human pose estimation with radio-frequency identification,” IEEE Transactions on Reliability, 2020.