Human Motion Recognition Based on Wi-Fi Imaging

Liangliang Lin, Kun Zhao(✉), Xiaoyu Ma, Wei Xi, Chen Yang, Hui He, and Jizhong Zhao

School of Computer Science and Technology, Department of Telecommunications, Xi'an Jiatong University, Xi'an 710004, People's Republic of China

Abstract. The current wireless sensing technology has some problems, such as low resolution caused by narrow signal bandwidth, poor environmental adaptability caused by multipath effect and so on. To solve the above problems, this paper provides a new perception idea, Wi-Fi imaging human actions, and then using image processing method for action recognition. In the Wi-Fi imaging part, according to the different spatial angles of different parts of the body trunk such as head, chest and legs relative to the receiving end, this paper processes the human body reflection signal received by the receiving end, obtains the signal strength corresponding to each azimuth signal in the space, and generates the human body heat map. In the stage of action recognition, firstly, the background interference is removed. According to the characteristics of imaging and action in this paper, a continuous action segmentation method is proposed. The image action features are obtained through intensive sampling. Finally, the SVM is optimized by genetic algorithm to improve the accuracy of action classification under different conditions. Through the analysis of experimental results, the method proposed in this paper can produce high-precision imaging of human body under practical application conditions. The recognition accuracy of different actions in the experiment is more than 90%.

Keywords: Motion recognition · Wi-Fi image · SVM

1 Introduction

Human motion perception and recognition refers to the technology of using devices other than the human naked eye to obtain the current dynamic information of individuals and identify them. With the proposal of virtual reality, augmented reality and other technologies, human motion perception technology is urgently needed in many fields. For example, the motion perception recognition technology is applied to the field of human-computer interaction to get the somatosensory interaction technology, which gets rid of the complex and learning cost equipment and improves the efficiency of life. In recent ten years, with the popularity of the concept of artificial intelligence technology and the rapid development of intelligent devices and the Internet of things, the application scenarios of human motion perception and recognition technology are more extensive, such as directly using simple actions to control the switch and regulation of

H. Gao and X. Wang (Eds.): CollaborateCom 2021, LNICST 406, pp. 608–627, 2021.
https://doi.org/10.1007/978-3-030-92635-9_35

home intelligent appliances in the field of smart home, so as to improve people's quality of family life; In education, conference and other scenarios, use actions directly to control PPT and documents, or intelligently adjust the conference process according to human actions [1, 2].

With the construction of intelligent devices and the Internet of things, the technology of using wireless signals to perceive human actions appears, this technology also belongs to the category of unbound sensing technology. Due to the existence of the communication module in the intelligent device, the wireless signal has the advantages of wide distribution and high signal strength in the indoor space. Because of its original design intention, the penetration of the wireless signal is very superior. When perceiving the human body, it breaks through the restriction that the optical perception must be in the line of sight of the device, and avoids the risk of personal privacy disclosure. However, as an emerging technology, wireless sensing technology also faces some challenges. For example, the main purpose of wireless signal is communication rather than human perception, which needs to be improved in physics or algorithm. Some of the current studies directly use the waveform changes of signals to perceive human actions, but due to the abstraction of waveform changes, it is difficult to distinguish different actions, which reduces the accuracy and has a great impact on the environment. The other part uses the wireless signal to image the human body and then recognize the action, but the imaging effect is poor and the recognition action is single. Therefore, this paper proposes human motion recognition based on Wi-Fi imaging, which aims to improve the imaging algorithm to realize human fine-grained imaging through Wi-Fi perception of human body information, and carry out a variety of motion recognition based on imaging, which has great application scenarios and theoretical significance, and is worthy of in-depth research.

The propagation of electromagnetic signal in space will have a certain degree of attenuation signal, which is related to the length of its propagation path and the reflection and refraction generated by penetrating the object in the propagation process. When the object moves, the generated multipath signal will also change, and the intensity of the received electromagnetic model will also change. Therefore, RSSI is highly sensitive to moving objects in the environment, and different individual movements will have different signal intensity transformation. Therefore, some research work uses RSSI information to roughly image different moving objects, and some work attempts to use RSSI for more accurate human action recognition, but the effect is not good. This is because RSSI is too rough and the granularity is too large to play a role in fine-grained perception [12–15].

Channel state information is the channel attribute in electromagnetic signal transmission, which is used to reflect the linear superposition of the influence of multipath effect caused by reflection, scattering and other reasons on the signal during the propagation of each subcarrier signal. CSI contains the changes of signal propagation to a certain medium (such as air or human body). According to different factors, it can be divided into the influence of attenuation caused by distance, the influence of attenuation of surrounding environment and the influence of object surface scattering. Therefore,

CSI has higher sensitivity to individual movement than RSSI. CSI can therefore play a role in a subtler field of perception. Wihear uses directional antenna to acquire CSI changes caused by slight movement of mouth to perceive different words emitted by individuals; Falldefi uses OFDM as a wireless signal to obtain the changes of CSI and RSSI caused by different human fall actions, and finally identify them [16, 19].

The typical technology of imaging using RF signal is the military radar system and the prohibited article inspection device in subway and train. It uses the characteristics of high reflectivity of metal to electromagnetic signal to search the information of metal objects such as aircraft position and whether to carry knives. The imaging system for human body information and its movement is basically inseparable from professional multi antenna equipment, and the modulation mode is customized to make the transmitted signal ultra-high frequency and ultra-large bandwidth. These systems are essentially different from Wi-Fi based imaging systems because their working scene range is very narrow. The characteristics of high frequency and high penetration make the signal penetrate or be absorbed by the human body rather than reflect the signal. Therefore, it will do varying degrees of harm to human health. These technologies have certain advantages in accuracy, such as CT examination devices in hospitals. In the imaging research using Wi-Fi like signals, researchers have also made some pioneering achievements. RF capture uses the customized waveform of Wi-Fi like signal frequency to achieve a fairly accurate imaging of the human body, which is a big step in the feasibility of Wi-Fi imaging technology. In CARM, the real Wi-Fi signal is used to realize human imaging, but the ideal effect cannot be achieved [20, 30, 32].

The use of wireless RF signals, especially Wi-Fi signals for human motion perception, almost achieves full signal coverage in specific application scenarios, and makes use of existing routing equipment without reinstallation and deployment. Therefore, it is a popular trend in current research. However, in the previous work, some of them directly use Wi-Fi information for motion perception. They need to process the collected continuous waveform data and distinguish the actions through different waveforms, which leads to single recognition action and great difficulty in distinguishing. For example, detecting falling action may be confused with bending, squatting and other actions. The other part uses imaging, the human body is imaged first, and then the action recognition is carried out. However, due to the rough imaging, the detection of actions with large amplitude has a good effect, but the limb movements cannot be detected. This paper improves the previous imaging method, images the human body more finely, obtains more diverse human actions, and proposes a continuous action recognition method to extract different action features and classify actions through SVM to achieve the purpose of action recognition.

Our contribution can be summarized as follows:

1) Wi-Fi imaging using multiple antenna arrays. Visible electromagnetic signals are transmitted to the surface of the object and reflected. The reflected light enters the retina or the optical perceptron such as photosensitive material in the camera, and the inverted image of the object is generated on the retina or photosensitive material. This is the principle of optical sensing imaging. Similarly, the invisible Wi-Fi signal and the

visible light belong to the electromagnetic signal and have the same propagation characteristics. The Wi-Fi signal will also be reflected when it propagates to the human body surface. The reflected signal is received by the receiver, and the image of the object is obtained by processing the signal received by the receiver. In detail, the human trunk and limbs do not belong to the same plane as the signal receiving end, so the reflected signal reflected by the human trunk will form an incident angle with the signal receiving end, and the reflected signals of different parts of the human trunk, such as head, chest and legs, are also different from the incident angle of the signal receiving end, and different individuals are different because of their height and body shape, The intensity and incident angle of the reflected signals of their different body parts are also quite different. All the reflected signals in the environment are obtained by using multiple antennas at the receiving end. Because the filtered reflected signals of the surrounding static objects, the received signals with greater intensity are the reflected signals of the human body. The incident angle of these signals is calculated by the algorithm, the positions of these different body parts can be determined in two-dimensional space, so as to obtain the heat map of the human body. Therefore, this paper uses angle measurement to simplify the complex problem of final imaging using accurate ranging. This paper extends the multi signal classification algorithm applied in one-dimensional space, that is, MUSIC algorithm, to make it applicable in two-dimensional space, and then improves it to improve the human imaging effect.

2) Adding subcarriers to improve the effect of MUSIC algorithm. The angle measurement accuracy of MUSIC algorithm is positively correlated with the number of antennas, but due to the small number of antennas in ordinary home routers in real conditions, the imaging effect will not reach the expectation. There is no doubt that the multipath signals generated by different parts of the human torso will cause measurable phase shift on different physical antennas of the receiver array, which is the physical basis of MUSIC algorithm angle measurement. Due to the characteristics of Wi-Fi transmission, the carrier and subcarrier of the signal will also generate phase shift due to different frequencies, even if the causes of phase shift are different, However, it can be comprehensively calculated mathematically and does not affect the properties of the equation. Therefore, the influence caused by subcarriers can be added to the MUSIC algorithm, because subcarriers play the same role as physical antennas in the algorithm, which essentially increases the number of antennas and improves the measurement accuracy. Here, the space is saved and the algorithm effect is improved without increasing the number of physical antennas. However, the effect of MUSIC algorithm on coherent signals reflected by human body is very poor. Combined with the advantages of different decoherence, this paper proposes a global decoherence method for Wi-Fi human body imaging to further improve the imaging effect, and then verifies each improved effect through simulation.

3) Experimental results and analysis. Firstly, this paper realizes the prototype of the algorithm by using the notebook computer loaded with 9300 network card and multi heel antenna and PicoScenes platform. Then, different volunteers are used for downlink imaging and action recognition in different scenes for the imaging algorithm proposed in this paper, and the influence of other influencing factors such as distance and signal frequency on the model is analyzed to evaluate the effect and robustness of the algorithm.

2 Wi-Fi Imaging Algorithm Based on 3D Virtual Array

The imaging method used in this paper is different from the traditional imaging method, which requires accurate ranging. Its principle is that the Wi-Fi signal is transmitted to different parts of the object and reflected. For the receiving end, the incoming wave directions of the reflected signals of different parts of the object in physical space must be different. If these incoming wave directions can be estimated, the positions of different parts of the object in space can be determined reversely, Combined to get a complete image of the object. As shown in Fig. 1, just as optical perception will not image substances without entities, the surrounding static reflection will be filtered out in this paper, so the image is only a moving human body. The three main different parts of the moving human body (head, trunk and limbs) will reflect the signal, and the receiving end not in the same plane with the human body will receive the reflected signals of these different parts. These reflected signals will have different entry angles with the antenna array plane. Analyze the received signal by some method to obtain the azimuth of the reflected signal from the human body the pitch angle reverses the relative positions of different parts of the human body, and then obtains the heat map of the whole human body.

2.1 Imaging Algorithm Based on Virtual 3D Array

In this paper, the human body needs to be imaged in two-dimensional space. Due to the limitation of antenna, the previous imaging algorithm has poor estimation effect on the similar incident angle in two-dimensional plane, so the performance of the algorithm in human body imaging is still poor, and cannot achieve the goal of human motion imaging.

According to the principle of imaging algorithm, the number of antennas must be greater than the number of sources, and the more antennas, the more accurate the estimation of sources. Therefore, directly increasing the number of physical antennas is the simplest solution, but it is limited to the implementation conditions and application scenarios. If the imaging granularity can be improved without adding physical antennas, it will be of great benefit to the subsequent recognition accuracy and future applications. Therefore, in this section, by analyzing the characteristics of subcarriers, an imaging algorithm based on virtual three-dimensional array is proposed to improve the imaging accuracy. Then, according to the characteristics of 3D virtual array imaging algorithm, a new decoherence algorithm is designed, which improves the effect of imaging algorithm to a greater extent.

It can be seen that the core of the imaging algorithm in this paper is to calculate the guidance vector matrix, which is composed of the phase shift caused by the distance between physical antennas. Wi-Fi technology uses multiple different subcarriers to transmit signals, and the signals are modulated on different subcarriers to achieve the purpose of parallel transmission. Different from the principle that the signal generates phase shift due to time delay on the physical antenna, when the reflected signal of L signal sources is received by the receiving array composed of M antennas, different subcarriers in the same signal reach an array element of the array at the incident angle,

and phase shift will also occur due to different frequencies. Take a subcarrier as the origin, and its phase shift is as follows:

$$\Omega = \frac{2\pi(\mathrm{L}-1)\mathrm{d}(f_1 - f_2)\cos(\varphi)}{c} \tag{1}$$

where f_1 and f_2 represent frequencies of these two different subcarriers, c indicating the speed of light. Because the Wi-Fi signal belongs to a narrow-band signal, the subcarrier frequencies are relatively similar and the denominator order of magnitude is much larger than the numerator. Therefore, even if all 30 subcarriers are calculated into the formula, the maximum phase obtained is less than 0.06 radian, which can hardly participate in the calculation, which has little effect on the improvement of angle measurement algorithm.

However, using the flight time of the incoming wave signal can make the discrimination of subcarrier phase shift more obvious. Similarly, after introducing the flight time, the original phase shift formula is transformed into the following formula

$$\Omega = 2\pi(f_1 - f_2)\tau \tag{2}$$

τ to indicates the flight time. By introducing new parameters, the discrimination of phase difference is greatly enhanced, with a maximum of about 3.5 radians, which meets the measurement standard. As the number of unknowns increases, the difficulty of solving the equations increases. However, under the experimental conditions in this paper, the number of multipath reflected signals generated by the reflection of signals on the human body is generally no more than 7. Because the equation group added with subcarriers is 30 groups, which is much larger than the increased number of unknowns and does not interfere with the solution, the three-dimensional virtual array imaging algorithm is feasible.

We have discussed the phase difference caused by the physical distance of different antennas, but the propagation distance of the same signal to the same antenna will not change. Set the phase of a certain subcarrier as the phase of the original subcarrier, and the phase difference will be caused by the different flight time and frequency of different subcarriers:

$$\Omega(\tau_i) = 2\pi(n-1)f_\delta\tau_1 \tag{3}$$

where f_δ describes the frequency difference of different subcarriers. The phase difference of all N subcarriers is composed of a vector, expressed as:

$$\mathrm{h}(\tau_i) = \left[e^{j\Omega_1}, e^{j\Omega_2}, \ldots, e^{j\Omega_N}\right]^T \tag{4}$$

It can also be added into the formula $X(t) = \mathrm{HK}(t) + N(t)$, and we get is:

$$\begin{bmatrix} x_{11}(t) \\ \vdots \\ x_{1N}(t) \\ \vdots \\ x_{M1}(t) \\ \vdots \\ x_{MN}(t) \end{bmatrix} = \begin{bmatrix} e^{j\Phi_{11}} e^{j\Omega_{11}} & \cdots & e^{j\Phi_{1L}} e^{j\Omega_{1L}} \\ \vdots & \ddots & \vdots \\ e^{j\Phi_{\mathrm{L1}} e^{j\Omega_{N1}}} & \cdots & e^{j\Phi_{\mathrm{ML}}} e^{j\Omega_{NL}} \end{bmatrix} * \begin{bmatrix} s_1(t) \\ \vdots \\ s_L(t) \end{bmatrix} + \begin{bmatrix} n_{11}(t) \\ \vdots \\ n_{1N}(t) \\ \vdots \\ n_{\mathrm{LN}}(t) \\ \vdots \\ n_{MN}(t) \end{bmatrix} \tag{5}$$

where $x_{mn}(t)$ represents the received signal value of the m-th antenna and the n-th subcarrier with a subcarrier of an antenna as the origin received by the antenna array. The physical meaning of Φ_{mi} is the phase difference of the i-th reflected signal on the m-th antenna relative to the origin subcarrier of the origin antenna, and Ω_{ni} represents the phase difference of the i-th reflected signal on the n-th subcarrier relative to the origin subcarrier of the origin antenna. Mathematically, the phase shift caused by subcarriers is equivalent to that caused by physical antennas. In fact, it increases the number of equations and makes it easier to solve. In the physical sense, it increases the number of antennas and improves the accuracy of angle measurement. The next step is still to obtain the guidance vector matrix of the received signal, and then perform eigenvalue decomposition on the matrix, which is consistent with the steps after MUSIC algorithm to obtain the spatial spectral function:

$$P(\varphi, \theta, \tau) = \frac{h^H(\varphi, \theta, \tau) h(\varphi, \theta, \tau)}{h^H(\varphi, \theta, \tau) U U^H h(\varphi, \theta, \tau)} \tag{6}$$

A new parameter time of flight τ is introduced, and the physical meaning of the spectral function is that each direction in two-dimensional space combines the signal density of time of flight. As long as the possible azimuth and pitch angles of the source with the strongest signal are found in all flight times, the imaging heat map in two-dimensional space can be obtained.

Although the imaging accuracy is improved according to the measurable phase between subcarriers, the introduction of subcarriers brings new problems. When the source signal is a coherent signal, it is impossible to obtain the guidance vector and then the angle of arrival of the reflected signal according to the principle of the imaging algorithm. There is almost no difference in frequency between subcarriers, so it has coherence and interferes with imaging, so it is necessary to decoherence the signal.

2.2 Description of Improved 3D Decoherence Algorithm

The previous imaging algorithms in this paper mainly estimate the incoming direction of incoherent signals, but through the analysis in Sect. 3.1, the reflected signal of Wi-Fi signal reflected by human body is coherent. According to Eq. (3–9), for coherent signals, complex constant vectors can be extracted, so that the properties of reflected

signal matrix will change, Finally, the whole equation cannot use the traditional mathematical method to calculate the guidance vector. Finally, the effect of the imaging algorithm on the coherent source is very poor, so the decoherence algorithm needs to be introduced to decoherence. Among the mainstream decoherence algorithms, the spatial leveling algorithm is simple to implement, but it has estimation defects for a large number of signals. The non-dimensionality reduction processing algorithm is also easy to implement, but the deviation is large [34, 35].

In order to avoid the above defects of smoothing algorithm, a three-dimensional decoherence algorithm based on Toeplitz matrix reconstruction is proposed in this paper. It is still l sources. From the derivation, we can get the rank loss of the signal covariance matrix in the case of signal coherence. Here, the subcarrier is also regarded as a virtual array element, and the Toeplitz matrix is constructed by taking the correlation function of the data received by each array element to decoherence the signal. Formula (5) has described that for the received signal values of rows and columns in the virtual antenna array, its correlation function with the received data of the origin array element is:

$$\begin{aligned} r_{mn} &= E\left[x_{11}x_{mn}^{\mathrm{H}}\right] \\ &= h_{11}E\left[\mathbf{S}\mathbf{S}^{\mathrm{H}}\right]h_{mn}^{\mathrm{H}} + \sigma^2\boldsymbol{I} \\ &= h_{11}\boldsymbol{R}_s h_{mn}^{\mathrm{H}} + \sigma^2\boldsymbol{I} \end{aligned} \tag{7}$$

where $\boldsymbol{R}_s$ is the source autocovariance matrix, then the correlation function vector satisfies:

$$[r_{11}, r_{12}, \cdots, r_{MN}] = h_{11}R_s[h_{11}^H, h_{11}^H, h_{12}^H, ..., h_{MN}^H] \tag{8}$$

It can be seen that the correlation function vector contains all source information. The Hermitian Toeplitz matrix is constructed from these correlation functions through conjugate operation to restore the rank of the covariance matrix:

$$\begin{bmatrix} r_{11}(1) & r_{12}^*(2) & r_{13}^* & \cdots & r_{MN}^* \\ r_{12}(2) & r_{11}(1) & r_{12}^* & \cdots & r_{MN-1}^* \\ r_{13}(3) & r_{12}(2) & r_{11} & \cdots & r_{MN-2}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{MN} & r_{MN-1} & r_{MN-2} & \cdots & r_{11}(1) \end{bmatrix} \tag{9}$$

Due to the nature of Toeplitz matrix, the conjugate operation does not affect the eigenvalue decomposition. Therefore, the eigenvalue decomposition of the new matrix is carried out to obtain the spectral function and estimate the signal angle of arrival. The reconstruction of covariance matrix avoids the damage of smoothing algorithm to array aperture, makes full use of all received information, realizes complete decoherence and improves the imaging effect.

3 Environment Adaptive Human Continuous Motion Recognition

All actions in human daily life are inseparable from the changes of limbs and trunk, and most complex actions are composed of simple actions such as arm waving, leg raising and squatting. They are the basic action unit for human interaction with the outside world. Through different combinations of basic actions, human beings have realized the transformation of the material world. Therefore, this paper sets walking, arm waving, leg raising squat and bend down, and collect the data of individual continuous different movements.

Fig. 1. Schematic diagram: walking, waving, leg lifting, squatting and bending.

3.1 Continuous Action Segmentation

Generally, the human body action will have a process of action start, action reset and then the next action. In this process, the area of the two-dimensional human body heat map will change regularly. Therefore, the area can be used to segment the beginning and end of human action, and a rectangular box can be used to measure the area change in the process of human action, so as to obtain the curve about the area. The maximum point and minimum point in the area transformation are the segmentation points of action. Then the law of area change is more obvious through smoothing.

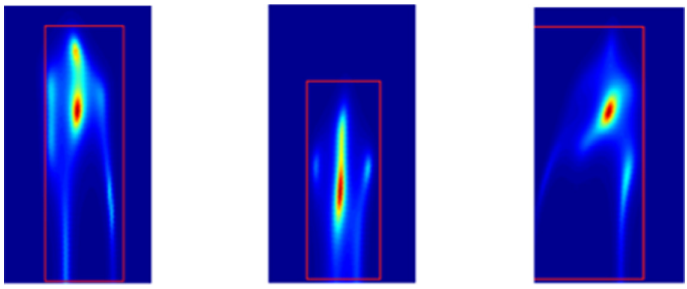

Fig. 2. Variation of different action areas

In this paper, the ambient noise has been removed in the previous background processing, and a single background is obtained through normalization processing. Therefore, by detecting the value of human body area and background area, select a

single threshold a to distinguish the background and human image, and then use a rectangular box to surround the human image as close as possible. The rectangular box is represented by, where t is time, (x,y) represents the coordinates of the upper left corner of the rectangular area box, and W and H are the width and height of the rectangular area. Let represent the rectangular area over time. As shown in Fig. 2 when the human body stands, bends and kicks, the area function will change, and the start and end of each action will cause the maximum and minimum points of the area function. However, due to the instability of human action, the change of area is not stable, and there are many maximum and minimum points of interference, which makes it difficult to find the real action segmentation points. Therefore, smoothing method is used to remove the unstable points.

3.2 Action Feature Extraction

1) Sample dimension alignment
   The image sequence of each action can be obtained through continuous action segmentation. Each sequence represents the imaging result of an action, which is called a sample. During action segmentation, various steps are processed to make the number of sequences in each action sample inconsistent. By averaging the number of sequences of different samples, the reference frame number T is determined, which is less than the supplement of the reference, and a certain number is deleted if it is greater than the reference. The dimension of the samples is aligned to facilitate subsequent processing.
2) Get the eigenvector
   Through grid division, feature points are intensively sampled at multiple scales of the picture in the action clip. These feature points are tracked over time and the eigenvalues of the autocorrelation matrix of all pixels are obtained. Some feature points lower than the threshold are removed through the threshold.

Calculate the feature point coordinates of the next frame from the feature point coordinates of the previous frame through the formula:

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * \omega_t) * x_t, y_t \tag{10}$$

Use formula (10) to obtain the optical flow median in the neighborhood of the feature point, so as to see the motion vector of the feature point. From the motion vector, the feature point coordinates of the subsequent frame can be calculated. These coordinates form trajectories over time, and the characteristics of actions are extracted from these trajectories. The above track shape can be expressed by $\Delta P_t, \ldots, \Delta P_{t+L-1}$, and the obtained trajectory features are:

$$T = \frac{(\Delta P_t, \ldots, \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \Delta P_j} \tag{11}$$

The features outside the track are further extracted, and the surrounding n existing as track feature points in each frame heat map in a certain track. The region of N

constitutes a time-space body for grid division, the space is divided into parts, and the parts are evenly selected for time division. Finally, feature extraction is performed in two regions.

The histogram of gray image gradient is calculated to obtain the hog feature, with a length of 96 (2 * 2 * 3 * 8). The histogram of optical flow direction and amplitude information is calculated to obtain the HOF feature, and the feature length is 108 (2 * 2 * 3 * 9).Calculate the histogram of optical flow image gradient to obtain MBH feature, and the feature length is 192 (2 * 96).

3) Principal component analysis dimensionality reduction
The feature length used for training is 426 dimensions, which is the sum of the above feature dimensions. In this way, there is too much redundant information of feature vector, and the dimension needs to be reduced by principal component analysis.

If there are m samples, there are m features. If the feature dimension is set to N, the feature matrix X of M * n can be formed. The feature normalization of each column of matrix X is that the mean value of each column is 0 and the variance is 1. The covariance matrix C of matrix X is calculated

$$\mathrm{C} = \frac{1}{M} X^T X \tag{12}$$

where C is the covariance matrix, and its diagonal elements represent the variance of each feature. The elements of row I, column J and row J, column I are the same, representing the covariance of the I and j features. In order to achieve the purpose of maximum information retention and minimum data repetition, in this matrix, it is the transformation matrix to maximize the diagonal elements and minimize the other elements to achieve orthogonality. Therefore, we need to diagonalize matrix C. Eigenvalue decomposition covariance matrix to obtain the eigenvector and eigenvalue of the matrix. From small to large, the eigenvectors corresponding to each eigenvalue are combined into a matrix, so that the matrix composed of the front row is called linear algebra. It has been proved that $P * C * P^T = \Lambda$, Λ is a diagonal matrix. The covariance matrix C is diagonalized, which is the linear transformation required by PCA. Y = XP, the original data X is changed into k-dimensional data y through the linear change P of k-dimension.

It should be noted that the size of dimension k of the data after dimension reduction is selected, as shown in Figs. 3 and 4. It is necessary to make a trade-off between dimension k and how many data features are retained, and select an appropriate value. In this study, when the reduced dimension is 213, 98% of the information is retained, and K is 213. In this way, the original feature dimension is reduced by half, and then the feature group of each action segment is encoded.

3.3 SVM Classification Based on GA Algorithm Optimization

In this paper, SVM is used to classify and identify the movements of waving, leg raising, bending and squatting. The original purpose of SVM is linear binary classification. Here, There is a multi classification problem and the data distribution is

complex, so SVM needs to be modified, so we can maximize the gap between two different class and minimize probability of misclassification. the RBF kernel function be used to get new decision function, and we use SVM in each pair of action. By above method, the problems of non-linear sample space and multi-classification are solved.

Samples are obtained by using 2.4 GHz and 5 GHz Wi-Fi signals respectively. Some of these samples have good effects and some have poor effects, but they all contain features that represent actions, and their corresponding optimal SVM parameters are also different. However, the original SVM uses a deterministic search method to find the optimal parameters with fixed directions and relationships, The search speed is slow and there is a great possibility that the global optimization cannot be obtained. In view of this situation, genetic algorithm is used to optimize SVM, which is mainly to optimize the parameters and improve the adaptability of the model under various conditions.

On the one hand, the stochastic characteristics of genetic algorithm are used to accelerate the speed of SVM in the parameter seeking process; On the other hand, the fitness function is used as the optimization standard to improve the accuracy of action recognition under different conditions. The comparison of fitness function between SVM optimized by genetic algorithm and original SVM is shown in the figure below. Aiming at the weakness of SVM, genetic algorithm is introduced to improve SVM from two aspects: efficient adjustment of parameters and recognition accuracy under different conditions. The training time is about 1/3 of that when it is not optimized, and the accuracy is improved by 8%–10%.

4 Experiment and Result Analysis

In this section, we will introduce the experimental configuration and experimental design, compare different experimental parameters to test the system performance, and finally display and analyze the experimental results.

4.1 Experimental Configuration

In this paper, two 9300 network cards are selected as the transmitter and receiver of the signal. Three antennas are led out from the transmitting end, and six omnidirectional antennas are used at the receiving end to form an antenna array to realize three transmitting and four receiving devices. Six omnidirectional antennas are used to form a uniform circular array to receive reflected signals. The height of the transmitter and receiver is 0.7 m, which is basically located in the middle of the human body and can receive all information about the human body. The experimenter acted in the area where the signal could reach, and collected CSI data using PicoScenes platform. In each case, 10 experimental acquisitions were carried out for different continuous actions, each acquisition time was 5 min, the contracting interval was set to 5 ms, the signal transmission frequency was set to 5800 MHz, the transmission energy was set to 30 dBm, and the subcarrier bandwidth was set to 20 MBps. 60000 data packets can be obtained in each experiment, and each data packet contains CSI data of all subcarriers corresponding to links between antennas.

The experimenter selected two men and two women of different height and body type, with height and weight of [183 cm, 126 kg], number a, [176 cm, 143 kg] number B, [164 cm, 115 kg] number C and [152 cm, 104 kg] number d.

4.2 Human Imaging and Motion Recognition

The action sequence is segmented by using the action segmentation method proposed in Sect. 4, in which the size of the selection matrix is determined based on the threshold A. for the heat map value in this paper is between 0 and 1, and the background is normalized close to 0, take the kick action as an example, select the threshold a as 0.01, 0.05, 0.1 and 0.15 respectively, and the effect is shown in the following picture.

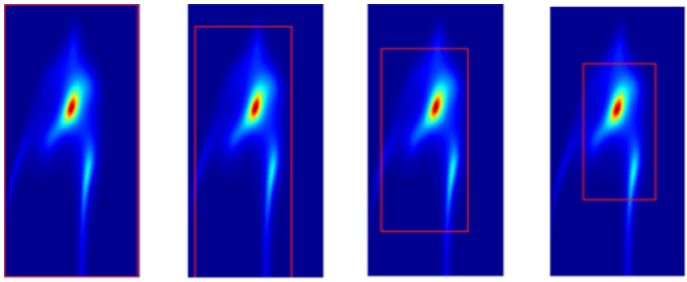

Fig. 3. Threshold selection (a) 0.01, (b) 0.05, (c) 0.1, and (d) 0.15

It can be seen from the above figure that when the threshold is too small, the matrix is large and contains a lot of background. With the increase of the threshold, the bounding box gradually decreases. When the threshold is too large, the bounding box does not contain the weak reflection caused by limbs. The appropriate threshold should be between 0.05 and 0.1, which effectively includes the reflection of limbs, and the change of bounding box with action is obvious. In the subsequent experiments, the threshold is 0.06. In this paper, we need to distinguish five kinds of actions, and the experimenter segments them after continuous different actions. After smoothing, similar to the method of selecting threshold a, select =1 *s*, =4 *s*, area difference =100, =300. the SVM will extract a series of features of each action under the stable time sequence.

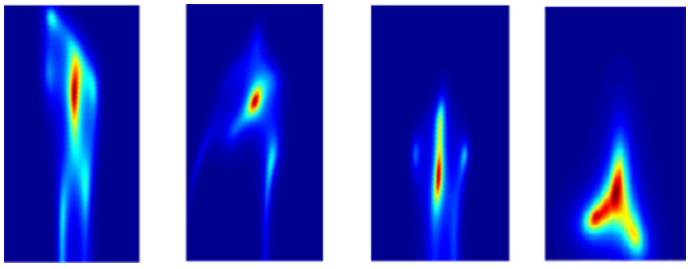

Fig. 4. Imaging results of different actions (a) wave, (b) kick, (c) bend, and (d) squat

4.3 Result Analysis

Figure 5 shows the action recognition results. Generally speaking, the recognition rate of different actions has reached more than 90%, but there is still a gap in the recognition rate between different actions, which is related to the recognition effect of different actions.

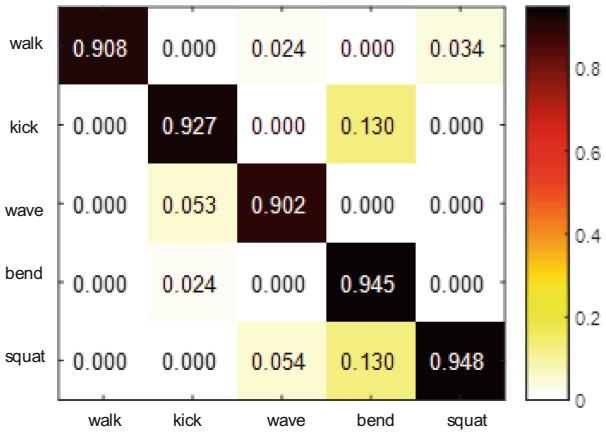

Fig. 5. Different action recognition result

When standing upright, the limbs are closer to the trunk, which is difficult to distinguish. With the extension of the action, the distance between the limbs and the trunk is farther and farther, and it is easier to be distinguished. For squatting and bending movements, the images of the imaging heat map and the real experimental scene are displayed. It can be found that there is a high corresponding relationship. With the experimenter squatting and bending down, the main reflection point is also moving downward, and the pitch angle relative to the receiving antenna array is gradually reduced. For the wave movement, the arm is close to the chest, the arm is thin, and the reflection signal is weak. The wave movement cannot be clearly observed, but when the arm is far away, the change of the azimuth of the reflection point can still be observed. For kicking, the distance between the leg and the chest is far, and the cylindrical structure is easy to generate reflection signals, so the imaging results are better than the waving action, so the recognition rate should be higher.

4.4 Model Test

Individual Factors. Among the experimenters numbered a, B, C and D, the action samples of experimenters numbered B and C are used for training, and the resampled experimental data numbered a, B, C and D are used for testing in the training stage. The recognition effect is shown in Table 1.

It can be found that there is a certain relationship between the recognition rate and the height and body shape of the experimenter. The influence of different height and body shape is more obvious in the bending and leg lifting actions. The larger the reflection area caused by the height and weight of the experimenter, the higher the action recognition rate. When the experimenter carries out the leg lifting action, the leg action amplitude caused by height is quite different. Similarly, Height and body shape also have a great influence when bending. However, the influence of this factor is not obvious when the experimenters squat, walk and wave. The reasons for the same phenomenon of the two movements are also different. For the squatting movement, when the human body squats, the change range of human motion imaging trajectory is the largest, and after the squatting movement is completed, the reflection area of the experimenter is the least affected by height and body shape factors. For walking and waving movements, the changes of human body parts on the heat map during walking can be basically ignored. Therefore, different body shapes will not affect the recognition of this movement, and waving movements are the same reason. In the overall trend, the recognition rate of men is higher than that of women, mainly because the overall reflection area of men becomes larger due to larger height and body shape, and the chest reflection area of women is much larger than that of other parts due to body shape factors, which affects the classification.

Table 1. Effects of different volunteers

Experimenter accuracy/%	A	B	C	D
Walk	90.24	92.33	88.67	89.57
Lift your legs	92.45	91.87	91.34	88.79
Wave	88.21	89.92	89.43	87.94
Bend	94.56	92.74	92.14	90.32
Squat down	94.36	95.48	93.67	92.96

Distance Factor. Since there is walking in the set recognition action, the recognition rate of different actions is also different when the experimenter walks to different distances from the equipment. In order to understand the role of distance factor in model recognition, the experimenter can walk to a distance of 2 m, 4 m, 6 m and 8 m from the equipment to perform other actions. The results are shown in Fig. 6. We observed that the distance factor was roughly negatively correlated with the recognition accuracy. The decrease of recognition rate with distance is mainly because the farther the distance, the higher the degree of signal attenuation and the less useful information in the reflected signal. The experiment here removes the influence of the wall. Therefore, it can be seen that compared with the influence of the wall on the recognition rate, the distance has little effect on the recognition rate. There is no wall at the same distance, even if the distance is far, The reduction of recognition rate is also very gentle, and the signal attenuation with distance is mainly related to the signal propagation frequency. Generally, the maximum space without walls in indoor space will not exceed 8 m * 8 m, and the recognition accuracy of the model for actions at a distance

of 8 m is more than 80%, which fully meets the requirements. Therefore, it can be said that the algorithm has high adaptability to distance.

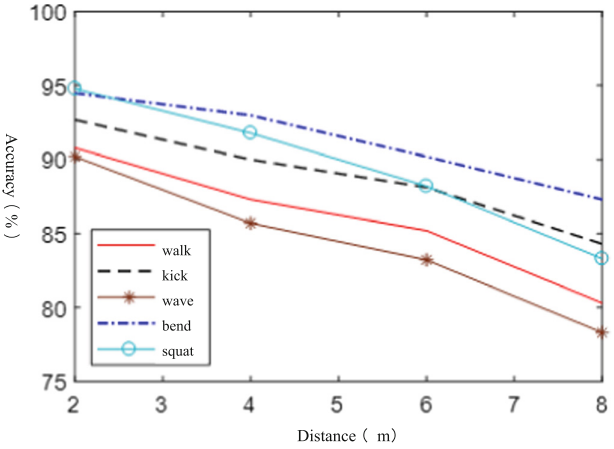

Fig. 6. Influence of different measurement distance

Wi-Fi Signal Frequency Factor. The above experiments lead to the information that the Wi-Fi signal changes with distance and the attenuation degree of wall blocking is related to frequency. In the current Wi-Fi frequency standard, 2.4 GHz is a general frequency, and 5 GHz has been widely popularized. with the development of technology, higher frequency Wi-Fi signal frequencies are also being studied, and high-frequency and low-frequency signals have their own advantages in propagation, therefore, experiments are conducted on the Wi-Fi signals of the above two frequencies. When there is no wall blocking between man and machine, the recognition rates of different Wi-Fi signals are shown in Fig. 7. It can be seen that when the attenuation degree of Wi-Fi signal is small, the recognition rates of the two Wi-Fi signals are very high and similar, and the accuracy of 5g signal model is slightly greater than 2.4G signal, This is mainly because 5g signal has fast transmission rate and contains more information, so the accuracy is slightly higher.

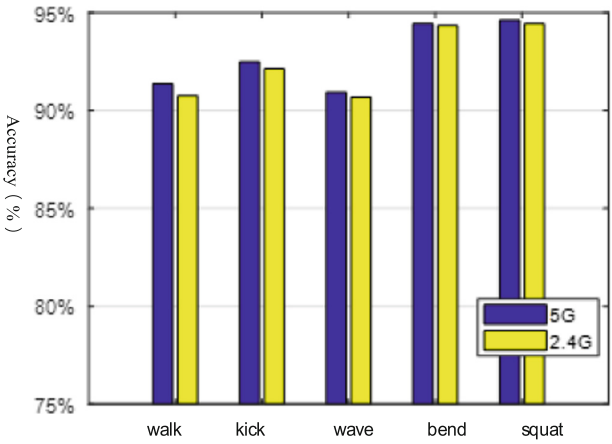

Fig. 7. Propagation path without wall

4.5 Comparison of Different Models

There are other models in the direction of action recognition, such as CARM (an activity recognition scheme), rtfall (a fall detection scheme) and falldefi (several fall action recognition schemes). In order to compare with the model in this paper, the accuracy, recall and F1 score, referred to as PRF, are used to evaluate the performance of the model. The results are shown in the table below (Table 2):

Table 2. Comparison of effects of various models

	Paper model	FallDeFi	CARM	RTFall
F1 score	90.24	86.34	76.41	82.45
Precision	86.65	84.12	73.60	80.01
Recall/TPR	92.33	90.48	80.32	85.20

It can be seen from the table that the effect of the model in this paper is better than other models. On the one hand, the imaging accuracy of the imaging algorithm in this paper is higher, and the information of human limbs is obtained, so the accuracy is higher in some movements about limbs. On the other hand, GA is used to optimize SVM, which has higher resistance to different factors affecting classification, so a certain classification accuracy is improved.

5 Conclusion

For the Wi-Fi imaging part, this paper solves several difficulties of Wi-Fi imaging, transforms the imaging problem into an angle measurement problem, extends the one-dimensional MUSIC algorithm to two-dimensional, and realizes coarse-grained

imaging. However, because the number of antennas is too small, it cannot be applied to human height recognition. Directly increasing the number of physical antennas is a direct way to improve the accuracy. However, limited to the implementation conditions and application scenarios, this method is not applicable. Therefore, using the characteristics of Wi-Fi signal transmission using subcarriers, this paper demonstrates the equivalent role of subcarriers and physical antennas in angle measurement theory. Adding subcarriers is actually equivalent to increasing the number of antennas and further improving the imaging accuracy. Combined with the characteristics of various decoherence algorithms, a decoherence algorithm is designed, which first decoherence locally and then decoherence globally. Finally, the effects of these improvements are verified by simulation experiments. Through these improvements, the imaging algorithm can be used in human motion imaging.

In the image processing part, the background interference is removed first, the continuous action is segmented according to the characteristics of imaging and action, the trajectory features and other features of the special action are obtained through intensive sampling, and the dimension of the sample data is reduced by PCA to retain 96% of the information. The classification accuracy of SVM is improved and optimized by genetic algorithm.

Experimental phase, at 9300 The algorithm prototype is implemented on the network card and PicoScenes software platform, and then the imaging and motion recognition are evaluated. In the imaging part, the human imaging effect in CARM is compared to verify the effectiveness of the imaging algorithm and decoherence algorithm in this paper. Then different actions are imaged and displayed. In the classification part, SVM is used to train the model because of the amount of data. For the trained model, several groups of comparative experiments are designed to simulate the impact of factors that may affect the recognition in reality, such as scene, propagation path, human clothing and recognition distance, so as to illustrate the robustness of the system. It is proved that the motion recognition algorithm based on Wi-Fi imaging proposed in this paper is practical and feasible.

Acknowledgment. This work was supported by NSFC 62072367, 61772413, 61802299, 62002284, and Natural Science Foundation of Shaanxi Province 2021JM-025.

References

1. Forbrig, P., Paternó, F., Pejtersen, A.M.: Human-computer interaction. Encycl. Creativity Invention Innovation Entrepreneurship **19**(2), 43–50 (2017)
2. Zesheng, Y.: Based on Kinect Research and Design of Smart Home System Based on Gesture Recognition. Liaoning University of Science and Technology, Liaoning (2017)
3. Yibo, L., Yulin, D.: Indoor abnormal behavior detection of the elderly living alone based on intelligent monitoring. Comput. Appl. Softw. **31**(02), 188–190 (2014)
4. Kwapisz, J.R., Weiss, G.M., Moore, S.: Activity recognition using cell phone accelerometers. ACM SIGKDD Explor. Newsl. **12**(2), 74–82 (2011)
5. Lv, M., et al.: Bi-view semi-supervised learning based semantic human activity recognition using accelerometers. IEEE Trans. Mobile Comput. **17**(9), 1991–2001 (2018)
6. http://www.36kr.com/p/212759.html

7. Jalal, A., Kamal, S., Kim, D.: Shape and motion features approach for activity tracking and recognition from kinect video camera. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, pp. 445–450. IEEE (2015)
8. Tang, S., Andres, B., Andriluka, M., Schiele, B.: Multi-person tracking by multicut and deep matching. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9914, pp. 100–111. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48881-3_8
9. Yoshida, T., Taniguchi, Y.: Estimating the number of people using existing WiFi access point in indoor environment. In: Proceedings of the 6th European Conference of Computer Science (ECCS '15), pp. 46–53 (2015)
10. Weng, J., Weng, C., Yuan, J.: Spatio-temporal naive-Bayes nearest-neighbor (ST-NBNN) for skeleton-based action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4171–4180 (2017)
11. Zou, Y., et al.: Wi-Fi radar: recognizing human behavior with commodity Wi-Fi. IEEE Commun. Mag. **55**(10), 105–111 (2017)
12. Wilson, J., Patwari, N.: Radio tomographic imaging with wireless networks. IEEE Trans. Mob. Comput. **9**(5), 621–632 (2010)
13. Sigg, S., Blanke, U., Tröster, G.: The telepathic phone: frictionless activity recognition from WiFi-RSSI. In: 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 148–155. IEEE (2014)
14. Han, J., et al.: GenePrint: generic and accurate physical-layer identification for UHF RFID tags. IEEE/ACM Trans. Netw. **24**(2), 846–858 (2016)
15. Xia, S., et al.: Indoor fingerprint positioning based on Wi-Fi: an overview. ISPRS Int. J. Geo Inf. **6**(5), 135 (2017)
16. Sameera, P., et al.: FallDeFi: ubiquitous fall detection using commodity Wi-Fi devices. IEEE Trans. Mob. Comput. **15**(11), 2474–9567 (2019)
17. Shangguan, L., et al.: STPP: spatial-temporal phase profiling-based method for relative RFID tag localization. IEEE/ACM Trans. Netw. **25**(1), 596–609 (2016)
18. Kotaru, M., et al.: Spotfi: decimeter level localization using WiFi. In: Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, pp. 269-282 (2015)
19. Yuw, Z., et al.: ZeroEffort cross-domain gesture recognition with Wi-Fi. In: Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services, 978-1-4503-6661-8. ACM (2020)
20. Stamnes, K., Stamnes, J.: Scattering of Electromagnetic Waves (2001)
21. Liu, Y., et al.: Channel estimation for OFDM. Commun. Surv. Tutorials IEEE **16**(4), 1891–1908 (2014)
22. Farhang-Boroujeny, B., Moradi, H.: OFDM inspired waveforms for 5G. IEEE Commun. Surv. Tutorials **18**(4), 2474–2492 (2016)
23. Schmidt, R.: Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag. **34**(3), 276–280 (1986)
24. Fugui, L., Mingzhen, L.: Structural optimization of depth CNN model based on convolution kernel decomposition and its application in small image recognition. J. Jinggangshan Univ. (Natural Sci. Edition) **39**(02), 31–39 (2018)
25. Agarwal, M., et al.: Face recognition using principle component analysis, eigenface and neural network. In: 2010 International Conference on Signal Acquisition and Processing, pp. (310–314). IEEE (2010)
26. Roska, T., Pazienza, G.: Cellular Neural Network (2000)
27. Cortes, C., Vapnik, V.N.: Support vector networks. Mach. Learn. **20**(3), 273–297 (1995)
28. Vapnik, V.N., Lerner, A.: Pattern recognition using generalized portrait method. Autom. Remote. Control. **24**(6), 774–780 (1963)

29. Kimeldorf, G., Wahba, G.: Some results on Tchebycheffian spline functions. J. Math. Anal. Appl. **33**(1), 82–95 (1971)
30. Huang, D., Nandakumar, R., Gollakota, S.: Feasibility and limits of Wi-Fi imaging. In: Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems, pp. 266–279. ACM (2014)
31. Xiong, J., Jamieson, K.: ArrayTrack: a fine-grained indoor location system. In: Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation. USENIX Association (2013)
32. Adib, F., et al.: Capturing the human figure through a wall. ACM Trans. Graphics (TOG) **34** (6), 219 (2015)
33. Lei, C., et al.: Principle and architecture analysis of Wi-Fi technology. Telecommun. Sci. **31** (9), 175–181 (2015)
34. Ting, W.: Research on Two-Dimensional DOA Estimation of Coherent Signals. Harbin Engineering University, Harbin (2010)
35. Kang, L., Youbao, X., Zhi, L.: Signal correlation and correction musicTwo dimensional DOA estimation algorithm. Comput. Appl. **32**(02), 592–594 (2012)
36. Ma, Y., Zhou, G., Wang, S.: WiFi sensing with channel state information: a survey. ACM Comput. Surv. (CSUR) **52**(3), 1–36 (2019)
37. Jiang, Z., et al.: PicoScenes: enabling UWB sensing array on COTS Wi-Fi platform. In: Proceedings of the 2019 International Conference on Embedded Wireless Systems and Networks, pp. 264−266 (2019)
38. Mingxing, X.: Human feature extraction and measurement based on image. Zhejiang Univ. Technol (2018)
39. Hang, L.: Statistical Learning Methods. Tsinghua University Press, Beijing (2012)
40. Gao, R.X., Yan, R. Discrete wavelet transform. In: Proceeding of the International Conference on Imaging Science, pp. 193–197 (2016)
41. Kay, S.M.: Modem Spectral Estimation: Theory and Application. Prentice-Hall, Englewood, NJ (1988)