Sign Language Estimation Scheme Employing Wi-Fi Signal

A Thesis Submitted to the Department of Computer Science and Communications Engineering, the Graduate School of Fundamental Science and Engineering of Waseda University in Partial Fulfillment of the Requirements for the Degree of Master of Engineering

Submission Date: July 16th,2021

Changhao Liu

(5119FG39-6)

Advisor: Prof. Shigeru Shimamoto

Research guidance: Research on Wireless Access Scheme

Table of content

List of Figures

List of Tables

Acknowledgments

First, I would like to express my gratitude to my supervisor, Professor Shigeru Shimamoto who supported my research and offered me many suggestions during my master period. I can not make so much progress in academic studies without his patient guidance.

Second, I sincerely acknowledge the help of Professor Jiang Liu, who gave me detailed suggestions about my research so many times, and taught me how to write academic paper. With her consistent instruction, I learned how to solve a big problem step by step.

In my past two years in Shimamoto Laboratory, I learned a lot from all members and received so much help from them, especially Pro. Pan, Dr Wang, Dr Yoshii, Ms. Megumi Saito, master Xianzhi Chu who not only helped me in my research, but brought care and love in my life.

At last, my parents and family always give me power and confidence, their love and strength let me never fear anything.

Abstract

The sign language recognition system plays an important role in the field of human-computer interaction. In the daily life of hearing-impaired people, sign language is used as the main tool to communicate with the world. Although sign language can satisfy simple conversation, it is difficult to deal with medical emergencies or educational consultation.This paper proposes a sign language recognition system based on Wi-Fi to improve the life of the disabled. The proposed system collects the Channel State Information (CSI) due to the change of hand movement. Through the analysis of all subcarriers, we decide to use the amplitude of CSI to reflect the characteristics of different sign languages, and remove the high-frequency noise in the amplitude of CSI to obtain a smoother signal Gesture feature, we propose a gesture feature extraction method based on the variance of time series, and use three different algorithms to recognize nine common Japanese Sign Languages including Dynamic Time Wrapping, Artificial Neural Network and Support Vector Machine. We set two daily conditions to test the system, and the experimental results show that the system performs well in different conditions. The main contributions are as follows:

1. A Japanese sign language gesture recognition system based on Wi-Fi is proposed in this paper for the first time. Nine dynamic gestures of Japanese sign language are analyzed and selected .

2. In the data processing part, Principle Component Analysis and Butterworth filter are used to remove high-frequency noise and make the curve smoother according to the type of gesture.

3. We analyze the characteristics of Japanese sign language and define a window for feature extraction and data segmentation, the result shows the system can locate all feature in time series.

4. We use three different classification algorithms for recognition, we also define two daily communication scenarios for testing the performance of our system, the average accuracy can reach 90% finally.

Chapter 1

Introduction

1.1 Background and Motivation

According to the World Health Organization (WHO) report in 2021, more than 1.5 billion people worldwide have some degree of hearing loss and about 430 million have a hearing loss of moderate or higher severity , And the number of people who are with hearing problem is growing rapidly every year. In Japan, there are about 300,000 deaf and hard of hearing people, and according to the studies of Japanese Association for sign language, the estimated number of Japanese Sign Language users is around 60,000 in Japan. Therefore, sign language is a common communication tool for them in daily life, it acquires a lot of communication and consulting especially in the medical, legal, education and other scenarios. However, sign language users can only communicate with another sign language user or someone who is familiar with sign language, so our original intention is to enable the deaf people to communicate with all people smoothly.

Because sign language interpreters can not provide help anytime and anywhere, and it will consume a lot of manpower and financial resources, to improve the lives of deaf people, a gesture recognition system becomes the best choice to solve this problem. At present, gesture recognition systems can be divided into three categories: Wearable sensor-based system, computer vision-based system, and radio communication-based system. However, there are some limitations in the practical

application of these three systems. For example, the system based on wearable sensors needs to wear special devices which not only increases the cost of detection but also limits the mobility of users and can not be widely used in daily life. Computer vision-based systems need to work in a favorable light condition, and can not be widely used in Japan due to personal privacy issues. The effective detection distance of the system based on radio signals is short while the cost is very high. Wi-Fi signal is the most widely used wireless signal in daily life. With the development of wireless sensor technology, more and more information in Wi-Fi signals is also used in the field of biological detection. A gesture recognition system based on Wi-Fi signals is also possible. The basic working principle of this system is that the channel state information in the Wi-Fi signal can be described in detail The transmission of the signal from transmitter to receiver. Compared with RSS, CSI is more sensitive to the propagation changes caused by hand gestures, which is used in OFDM demodulation.

In this paper, we design a device-free gesture recognition system based on Wi-Fi signal, and successfully recognize nine kinds of Japanese sign language which are frequently used in daily life as shown in Fig.1. According to people's speaking habits., we conduct experiments in two self-defined scenarios, and we use PCA to select the optimal subcarrier of thirty subcarriers of each antenna, use Butterworth filter to remove the high-frequency noise, and propose a method to extract gesture features based on the variance of CSI time series. Finally, we use DTW, Neural Network and SVM to classify all the sign languages. The results show that the accuracy of three methods are 76%, 87%, 91%.

1.2 Related work

Researchers have never stopped studying how to improve the lives of deaf people. Sign language is the most commonly used communication tool for them. However, there are many kinds of sign language and complex gestures, the previous research system has more or less many limitations. Therefore, many researchers focus on how to recognize sign language gestures accurately. As early as 1995, the system based on fuzzy control proposed by Yamaguqi et al. recognized 16 sign language words [1], with an average accuracy of about 85%. Imagawa et al. proposed a vision-based system [2], which requires users to wear colored gloves. Since most Japanese sign language words require hands to move near the face, the system needs to track the movement of hands instead of the face. Hirohiko et al. proposed to recognize sign language sentences, the system uses HMM method [3] at the first time, each sign language word is recognized and sorted correctly. The recognition accuracy of words and sentences is 86.6% and 58% respectively. Natsuki et al. Proposed a sign language recognition system based on computer vision [4], which can recognize sign language and annotate video sign language at the same time. Although the above research has achieved results, the system based on vision will expose the privacy of users, especially in countries that pay attention to the protection of privacy. Therefore, the practicability of the system is low.

To solve the above problems, some researchers focus on the way to recognize sign language by obtaining hand movements from various sensors. In the system proposed by Vasiliki et al [5]., the surface muscle sensor and acceleration sensor are

tied to the arm at the same time to recognize 60 sign language gestures. Yuichino et al. Proposed using video and data gloves to recognize words according to hand, body shape, and finger bending, and binding specific sensors with data gloves to obtain the features of hand movement changes [6], the system recognizes 20 common Japanese gesture words with an average recognition accuracy of 51%. Although the system based on the wearable sensors can obtain fine-grained gesture features, it can not be widely used because of the problems of distance and portability.

In recent years, it is possible to collect CSI from Wi-Fi signals. As a fine-grained signal, CSI can also be used in gesture recognition systems, which has been proved by many researchers. Ali et al. Found that different users in the indoor environment caused different changes in the waveform and proposed Wikey [7], this can be used as a feature to identify users. Wang et al. Identified the content of speaker according to the action of the mouth, and proposed the WiHear system [8].In [9], when there are more mobile people, the change of CSI is more dramatic. According to this discovery, they designed an electronic eye to count people in a room.. Wi-Sign [10] uses three Wi-Fi transceivers to associate gesture movement with CSI signal waveform changes, and realizes the recognition of 8 characters, commonly used sign language gestures. These systems are designed to recognize micromotion, so their solutions and system settings can not handle the motion recognition of hand and arm well. Moreover, Japanese sign language has its uniqueness, so the existing system is no longer applicable.

Chapter2

Sign Language Recognition System design

2.1 Analysis of Japanese sign language

Different from ASL, JSL has its unique hand movement and expression, which is more suitable for Japanese habits. Through the survey of the Japanese Sign Language Association, we selected nine sign languages commonly used in daily life as shown in table1, including can, good, bad, know, like, friend, sign language, thanks, sorry. These sign languages can meet the basic daily life and communication.

TABLE1 nine Japanese sign language

C1	C2	C3	C4	C5	C6	C7	C8	C9
Can	Good	Bad	Know	Like	Friend	Sign language	Thanks	Sorry

(1) 大丈夫, できる (Can)

This gesture is parallel to the right chest with the right hand from the left chest, and accompanied by nodding action, indicating that can do something or you are welcome. The detailed gesture decomposition is shown in Figure 1.

(2) ᥒᥒᥒ (good)

This gesture is the right hand clench, starting from the nose along Sliding into the air, this gesture is easy to be confused with other gestures, and the action amplitude is roughly the same, so it is difficult to recognize.

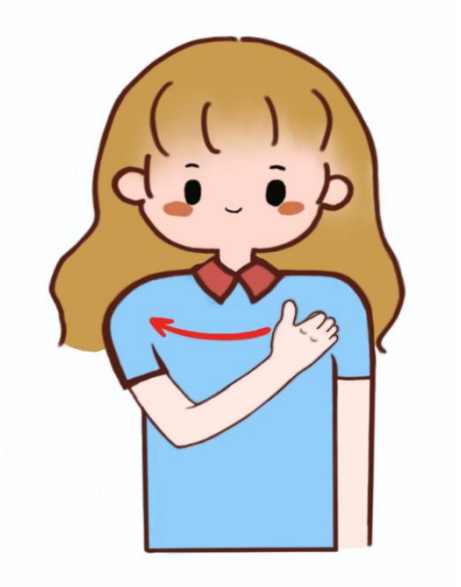


Fig1. The hand movement of gesture ‘can’

Fig2. The hand movement of gesture ‘good’

(3) 悪い (bad)

This gesture is relatively recognizable. The forefinger of the right hand extends out, and the arm slides down from the air to the chest. Generally speaking, the movement will be accompanied by the contraction of facial muscles, indicating the meaning of aversion and bad.

Fig3. The hand movement of gesture ‘bad’

(4) 知ってる (know)

Right hand clasp, gently bit chest twice, on behalf of I know, I understand. The amplitude of this gesture is very small, and the smaller the amplitude, the more difficult the gesture recognition, because it can not cause large-scale signal fluctuations.

(5) 好き (like)

This gesture is similar to gesture 2. Hold your chin with your right thumb and index finger, then slide into the air and let your two fingers touch.

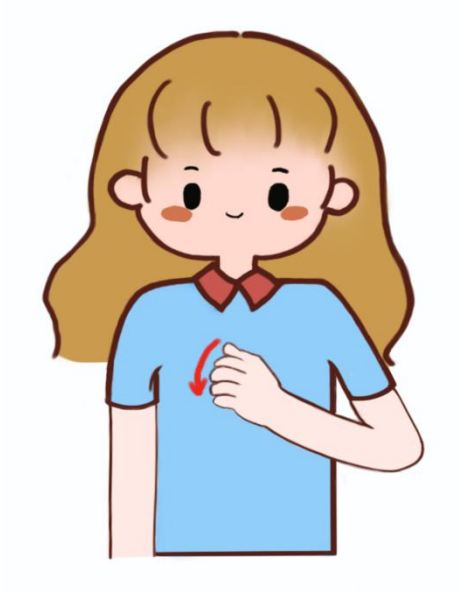


Fig4. The hand movement of gesture ‘know’

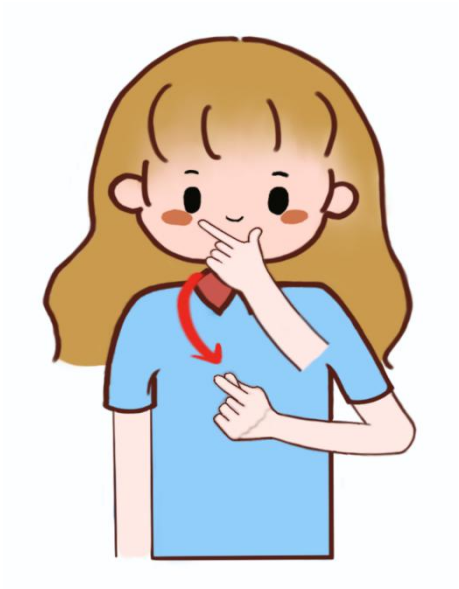


Fig5. The hand movement of gesture ‘like’

(6) 友達 **(friend)**

The range of this gesture is larger, and the time of the gesture is longer than the other two gestures. Clench your hands tightly, and do three circular movements in front of your chest.

Fig6. The hand movement of gesture ‘friend’

(7) 手話 (Sign language)

This gesture is also very complex. The forefinger of both hands is stretched out and crossed around the chest for several times. The reason for choosing this gesture is that first of all, it is almost the same as a friend's action. It can test the robustness of the system and add some difficulty to gesture classification. Moreover, this gesture is also very common in life.

(8) ありがとう (Thanks)

This gesture can be said to be a very common one, with the left hand flat and the right hand unfolded, sliding slowly from the left hand to the forehead to express thanks.

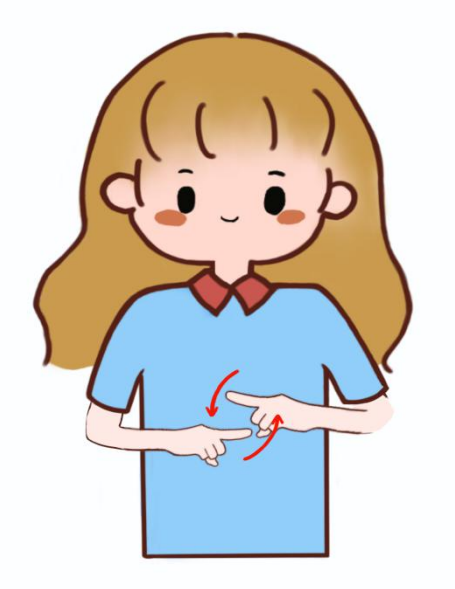


Fig7. The hand movement of gesture ‘Sign language’

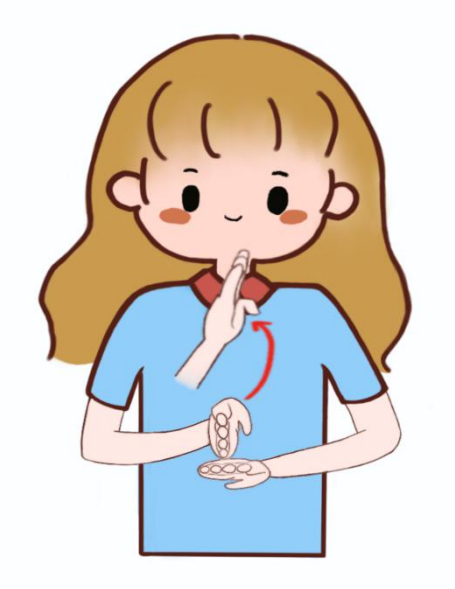


Fig8. The hand movement of gesture ‘Thanks’

(9) すみません (Sorry)

This gesture is one of the most commonly used gestures, whether it is to apologize or to attract the attention of the other party to start a conversation, it is the most used in life. Make a fist with your right hand, slide down from the top of your head, then stretch out three fingers, and finally spread out your palm and slide to your chest.

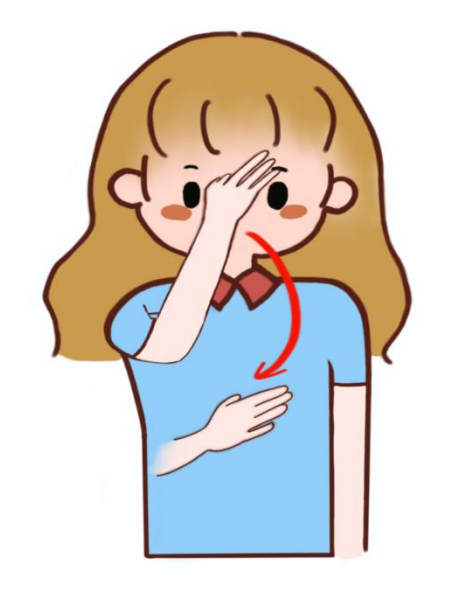


Fig9. The hand movement of gesture ‘sorry’

2.2 Preliminary of Channel State Information

2.2.1 Received Signal Strength Indication (RSSI)

RSSI is an optional part of the wireless transmission layer, which is used to determine the link quality and whether to increase the broadcast transmission strength. Because the wireless signal is mostly MW level, it is polarized and converted into DBM. Through the analysis of 802.11 underlying protocol frame, we know that PHY is the interface between MAC and wireless media, which transmits and receives data frames on shared wireless media. At present, all wireless network cards provide PMD according to the protocol_ RSSI. Indicate service primitive, and we can get the RSSI value of the wireless signal through the program interface.

2.2.2 Channel State Information

Channel state information. In the field of wireless communication, the so-called CSI is the channel attribute of communication link. It describes the attenuation factor of the signal in each transmission path, that is, the value of each element in the channel gain matrix H, such as signal scattering, multipath fading or shadowing fading, power decay of distance and so on. CSI can make the communication system adapt to the current channel conditions, and provide guarantee for high reliability and high rate communication in multi antenna system. Many IEEE802.11 standards use OFDM modulation signals to transmit them through multiple positive and alternating subcarriers, each of which has different signal strength and phase. Recently, some

common IEEE 802.11n standard commercial wireless network cards (such as Intel 5300) can provide detailed amplitude and phase information of different subcarriers in the form of CSI.

Specifically, with the ready-made Intel 5300 network card and the fine-tuning driver, a sample version of CFR within the Wi-Fi bandwidth range can be output in CSI form.

The fine-grained information of CSI includes how the Wi-Fi signal propagates and reflects in the indoor environment. In the process of transmission from TX to Rx, different sign language gestures will cause signal amplitude and phase changes. In the subcarrier, different from some simple basic gestures, most sign language gestures will make the arm and finger movement at the same time, and the duration of gestures will be longer than other gestures. In the narrow indoor environment, the receiver not only receives the transmitted signal directly but also receives the reflection from the walls, tables, and chairs, which is called multi-path reflection. Therefore, we need to describe these reflections through CSI. *i, j* refer to the number of subcarriers and the streams, the amplitude and phase of CSI can be summarized as follows:

$$H_{i,j} = \left\| H_{i,j} f_t \right\| e^{j\angle H_{i,j} f_t}$$

CSI signals can be extracted from commodity wireless network interface controllers (NICS). In our system, the open source CSI tool can obtain the CSI signal in Intel 5300 NIC. We use an access point (AP) with two antennas as the transmitter

and a detection point (DP) with three antennas as the receiver. The 2x3 CSI data streams can be described as a matrix of as follows:

$$CSI_{i,j} = \begin{Bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,30} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,30} \\ \vdots & \vdots & \vdots & \vdots \\ H_{6,1} & H_{6,2} & \cdots & H_{6,30} \end{Bmatrix}$$

2.3 System Design

The system framework is shown in the Fig.10 below. Due to the multipath reflection in the indoor environment, the collected data need to be filtered and noise removed. Then, after extracting the feature matrix, the training samples are obtained, and through the training, the human behaviors and activities are preliminarily identified.

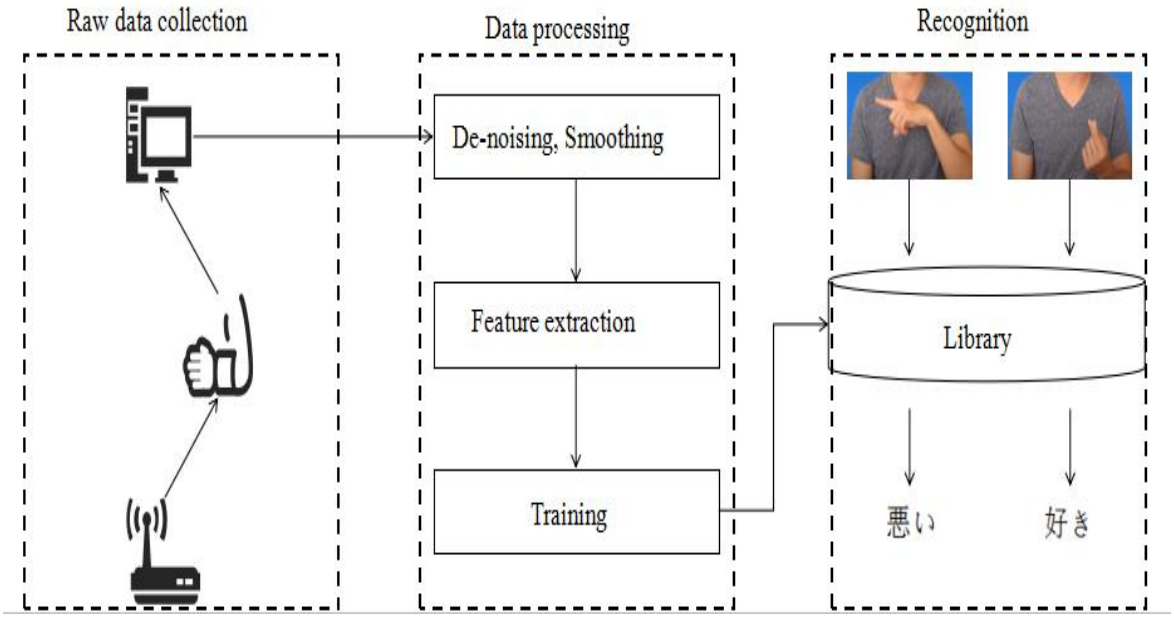


Fig.10 Framework of proposed system

our proposed system can be divided into three parts: CSI raw data collection, data processing, and recognition. First, CSI is collected at the receiver or so-called detection point (DP) which is a laptop installed Ubuntu system with open access CSI-Tool [11]. The collected data preprocessed to select the appropriate CSI data and to remove noise by applying Principle Component Analysis (PCA) and low-pass filter. Thereafter, DTW is applied to classify the proposed actions. In what follows, each

stage will be described in detail.

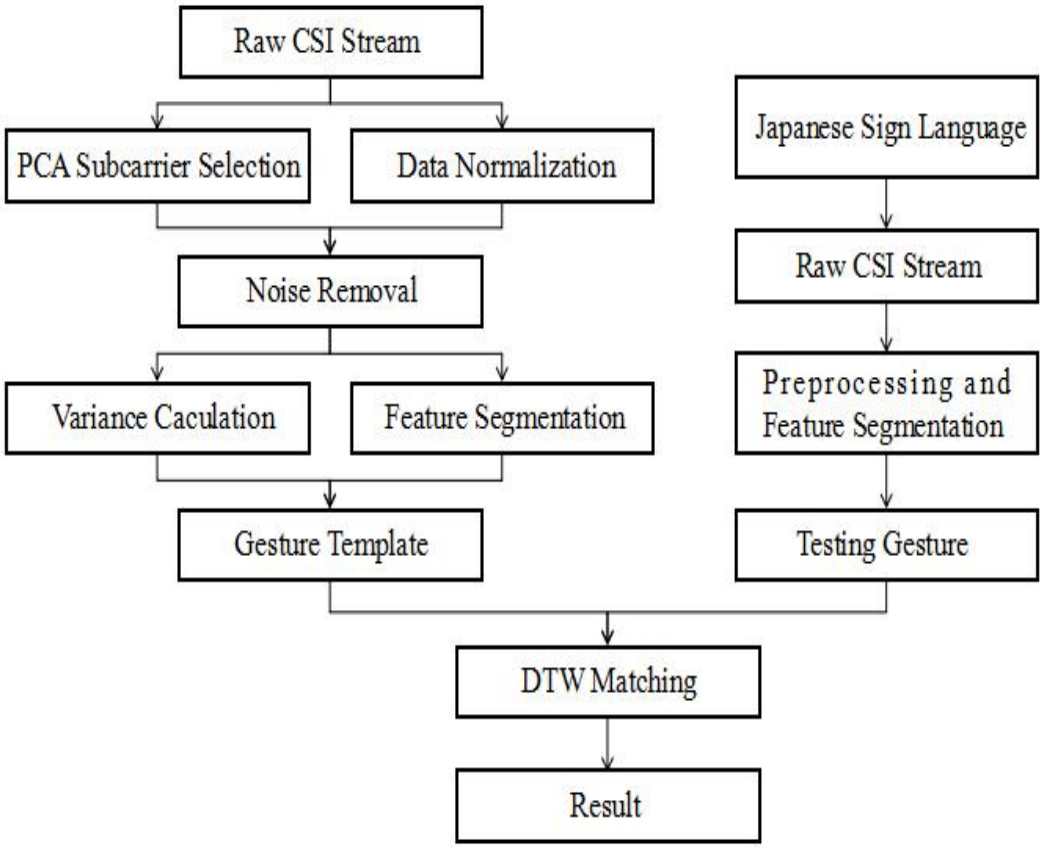


Fig.11 Details of data processing and recognition

The recognition part can be divided into two, when we get a new time stream, we consider it as a Raw CSI stream, then there are some data preprocessing steps. When get nine gesture templates, we use DTW algorithm for matching testing gesture and template. In the second half of the paper, we will also introduce how to use neural network and support vector machine to classify gestures.

Chapter 3

Data processing and Normalization

3.1 The description of CSI

The CSI raw data detected by the system is complex. Firstly, we use MATLAB to distinguish the amplitude and phase of CSI raw data, and extract all useful information from data file. Figure 12 (a), (b), (c) shows three ways to describe a CSI signal: SNR, CSI amplitude and CSI phase.

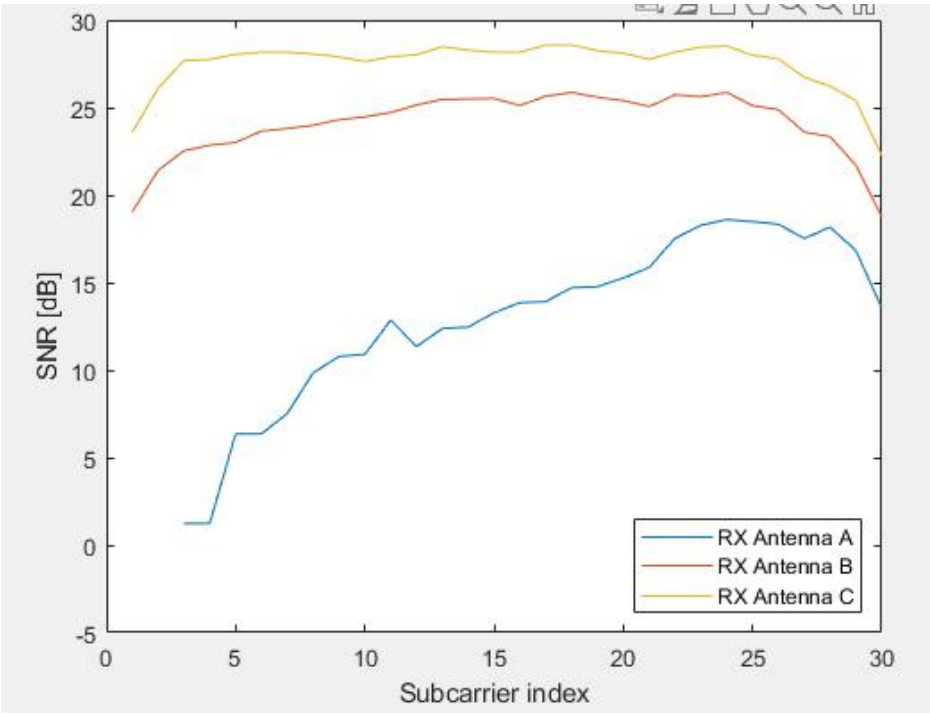
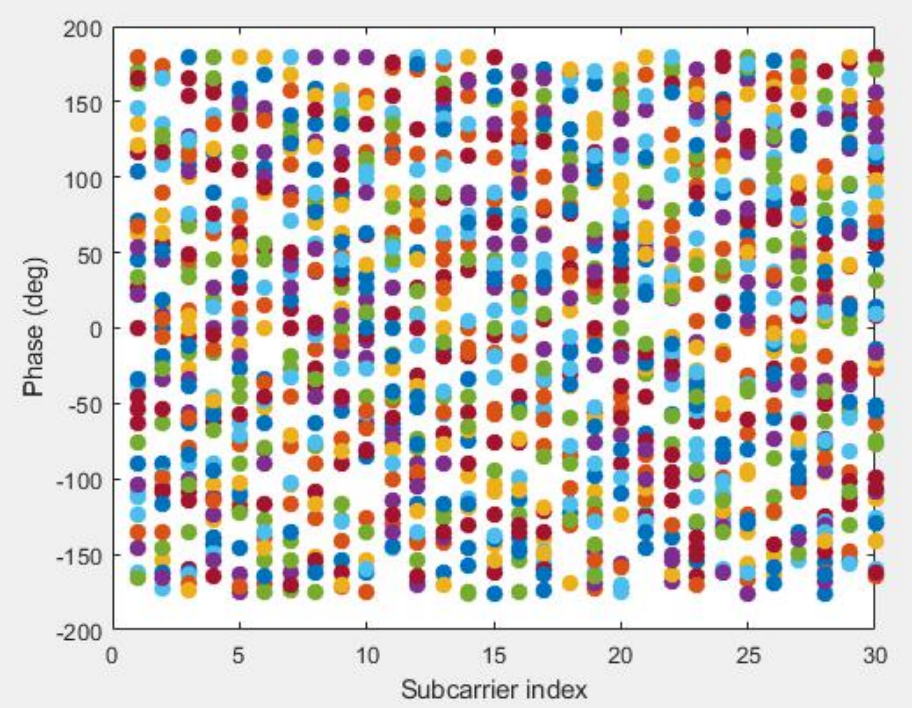


Fig.12 (a) SNR of a CSI signal

From the figure, we can see that SNR signal can effectively describe the strength of the reflected signal received by the receiver, but this parameter is too rough to

accurately describe the signal changes caused by small gestures.

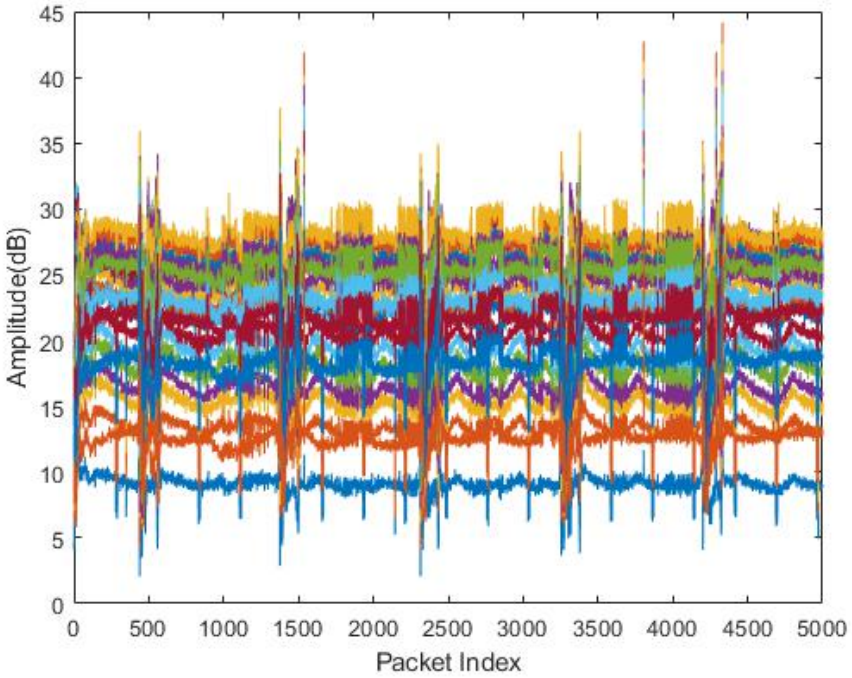


(b) The phase of a CSI signal

It seems that the phase information can be used as an effective parameter to describe the change of gesture. Many researchers extract good features and get good classification results through the processing of CSI phase. However, it is difficult to capture the change of phase characteristics. As shown in figure (b), if we need a more intuitive and easy to understand parameter, the amplitude which requires a lot of mathematical processing parameters is obviously not suitable.

Therefore, CSI amplitude is selected as the feature to describe gesture. As shown in figure (c), it is the amplitude feature caused by gesture action. When it is regarded as free propagation, any small change will be captured and reflected in the change of

CSI amplitude.



(c) The amplitude of CSI

3.2 Optimal subcarrier selection

3.2.1 The meaning of selection

At present, most commercial routers are based on OFDM modulation and demodulation, so the CSI amplitude obtained has multiple subcarriers, and the number of subcarriers varies according to different commercial network cards. For example, the number of subcarriers obtained by Atheros series wireless network cards is about 110, and the number of subcarriers obtained by Intel series wireless network cards is 30.

30 subcarriers describe the same gesture action will be different, and the amplitude changes of each subcarrier caused by different gestures are also different. But the general trend of all subcarriers is the same. Therefore, we need to reduce the dimension of data to get an optimal subcarrier, which can replace other subcarriers as classification features.

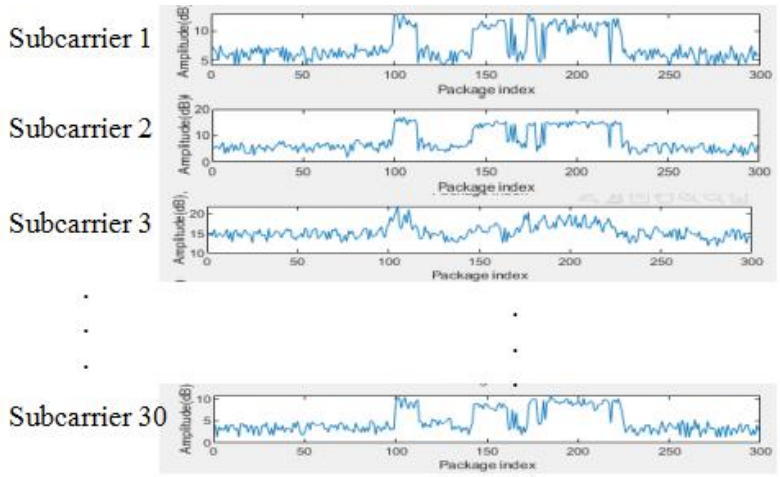


Fig.13 The curve trend of 30 subcarriers

3.2.2 Principle Component Analysis

Generally speaking, the mainstream methods of data dimension reduction are projection and popular learning. In most problems, features are not evenly distributed in all dimensions, and many of them are almost unchangeable, while other features are highly correlated.

Therefore, projection is the best choice for our system, and principal component analysis is the most popular dimension reduction algorithm so far. First, it identifies the hyperplane closest to the data, and then projects the data onto it.

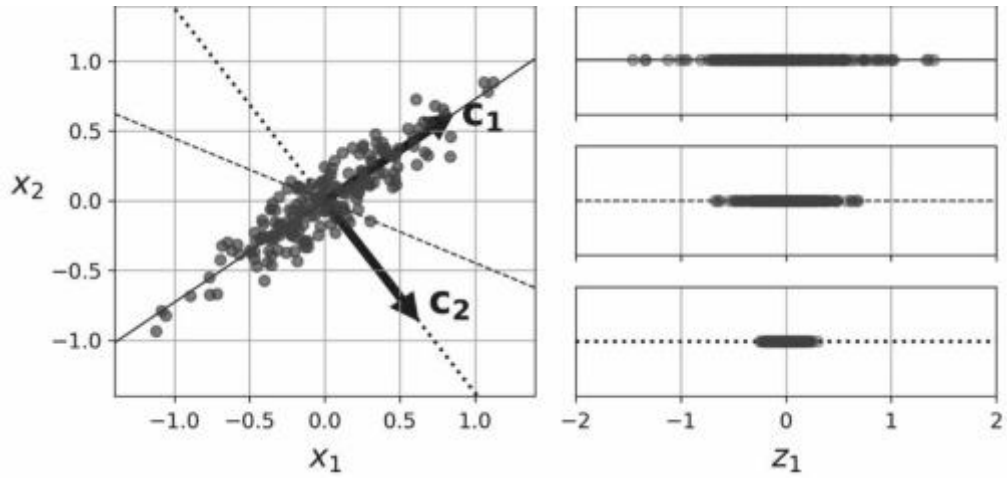


Fig14. The basic theory of PCA

When choosing different hyperplane, the reserved feature is different. We need to find the hyperplane with the most reserved feature, so the hyperplane may be much less than other lost information. PCA can identify which plane projected to the highest contribution to the difference. The I axis is called the first major component of the data. The main component matrix obtained by decomposition is as follows：

$$V = \begin{pmatrix} | & | & & | \\ c_1 & c_2 & \cdots & c_n \\ | & | & & | \end{pmatrix}$$

We usually choose the first principal component C_1 as the optimal subcarrier. Through theoretical analysis and practice, it can be proved that the first principal component contains the most gesture features. As shown in Figure 15, (a) is the principal component of the original CSI matrix after PCA, (b) shows the principal component with the highest score.

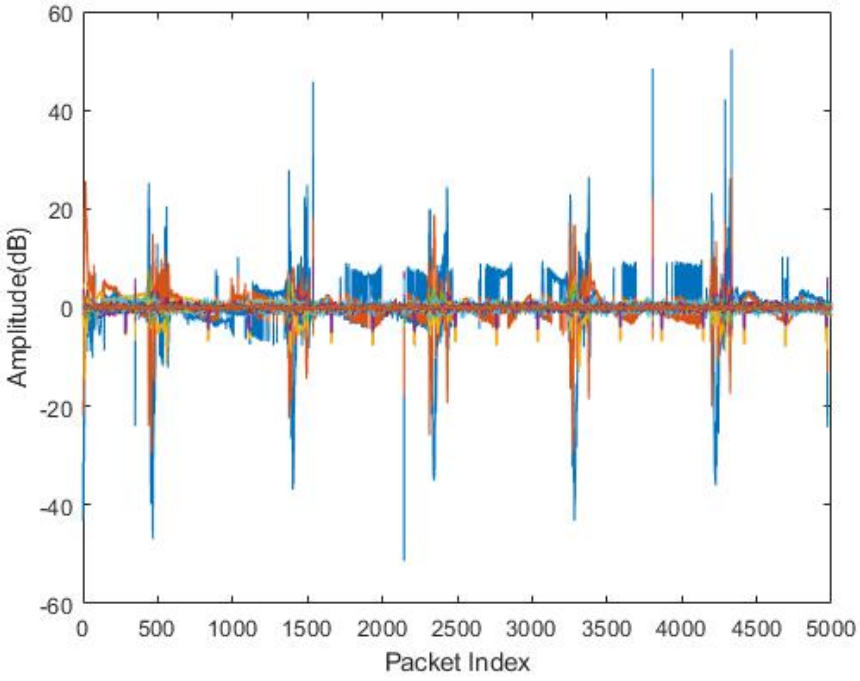
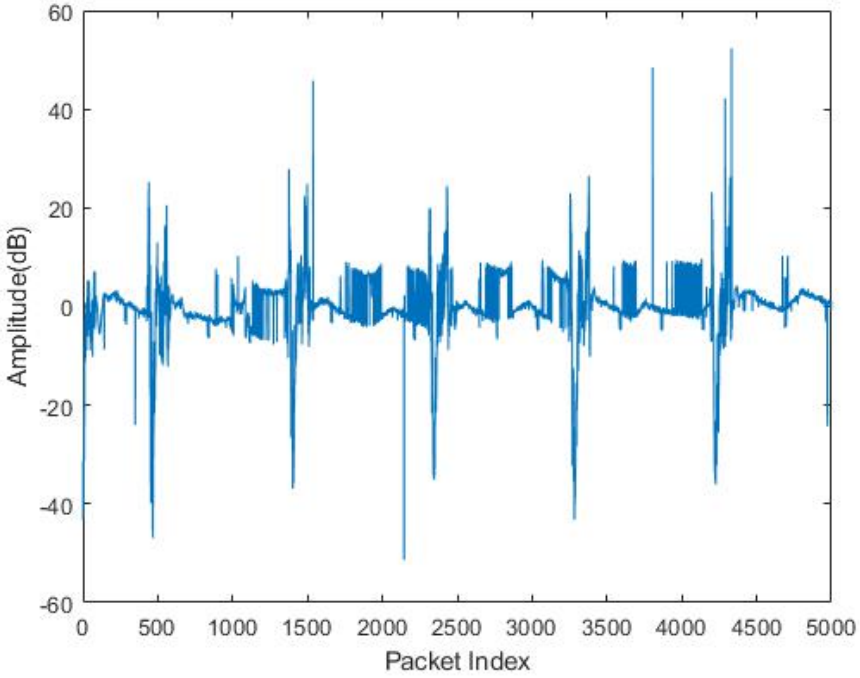


Fig.15 (a) the amplitude of CSI after applying PCA



(b) The first PC of CSI

After the original data is projected by PCA, the new features will be distributed on the positive and negative half axes of the X axis. This is because the data in the original coordinate system is projected into the new coordinate system, and the data in the original coordinate system is scaled to the new coordinate system. Therefore, this step is also to regularize the data.

3.3 Butterworth filter

3.3.1 Basis of Butterworth filter

Butterworth filter is a kind of electronic filter, which is also called maximum flat filter. The characteristic of Butterworth filter is that the frequency response curve in the pass band is flat to the maximum extent, without ripple, while it gradually drops to zero in the stop band. On the potter diagram of logarithmic diagonal frequency of amplitude, the amplitude decreases with the increase of the angle frequency, and tends to be negative infinity from the angle frequency of one side.

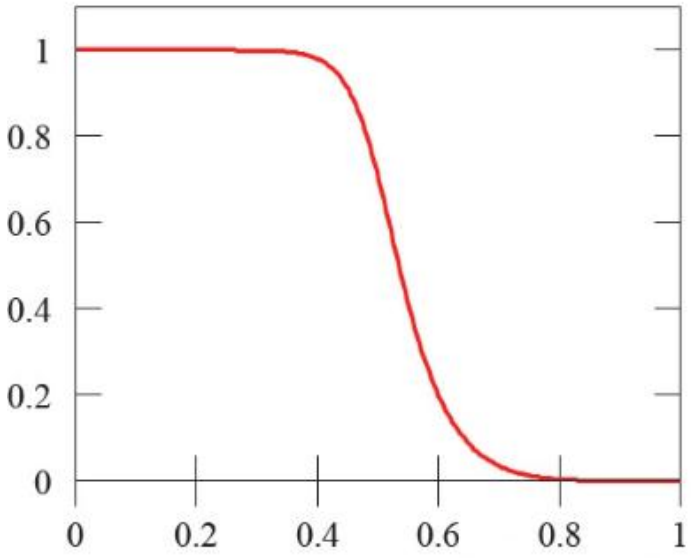


Fig.16 Butterworth filter

The attenuation rate of the first order Butterworth filter is 6 dB per second harmonic generation and 20 dB per tenth harmonic generation. The second order Butterworth filter's attenuation rate is 12 dB per second harmonic generation, the third order Butterworth filter's attenuation rate is 18 dB per second harmonic generation,

and so on. The amplitude diagonal frequency of Butterworth filter decreases monotonously, and it is the only filter whose amplitude diagonal frequency curve keeps the same shape regardless of order. However, the higher the order of the filter, the faster the amplitude attenuation in the stopband. The higher-order amplitude diagonal frequency of other filters has different shapes from the lower order amplitude diagonal frequency. Butterworth low pass filter can be expressed by the formula of square to frequency of amplitude as follows:

$$|H(\omega)|^2 = \frac{1}{1+\left(\frac{\omega}{\omega_c}\right)^{2n}} = \frac{1}{1+\epsilon^2\left(\frac{\omega}{\omega_p}\right)^{2n}}$$

Where n is the order of the filter, ω_c is the cut-off frequency.

Butterworth filter is a kind of filter design classification, similar to Chebyshev filter, it has high pass, low pass, band pass, high pass, band stop and other filters. It has stable amplitude frequency characteristics both inside and outside the passband, but it has a long transition band, which is easy to cause distortion in the transition band. When calling Butterworth filter in MATLAB for simulation, the signal will always be slightly distorted in the first cycle. But in the future, the amplitude frequency characteristics will be very good.

3.3.2 The application of Butterworth filter

According to the data collected in our experiment, the first component of CSI amplitude are approximately distributed in the range of [-50dB, 50dB], so we choose the fifth order Butterworth filter, whose frequency doubling attenuation rate is 30 dB, and its pass band edge frequency is set to 0.2, which can retain gesture features to the maximum extent and filter noise, The features after applying the filter are shown in Figure 17.

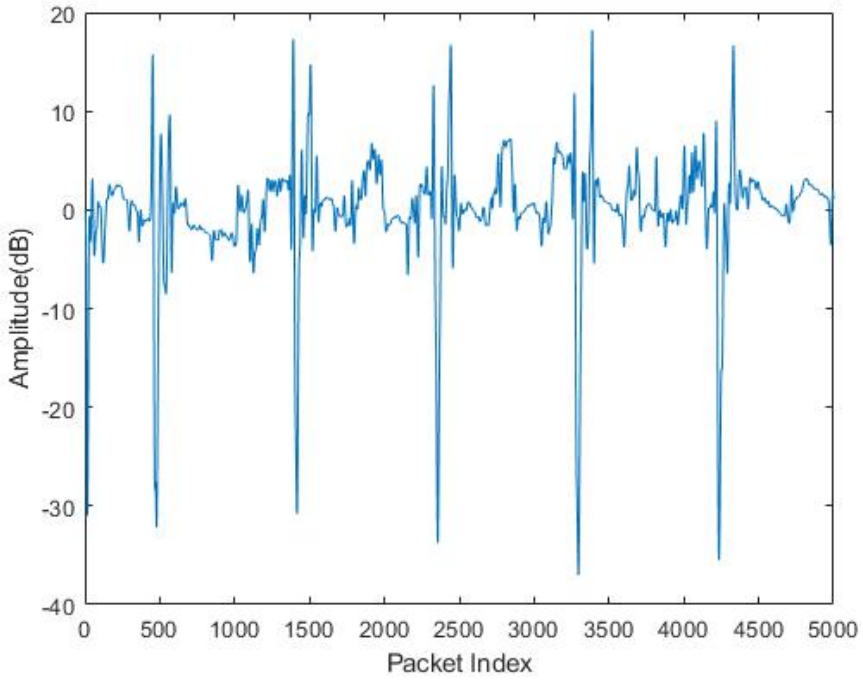


Fig.17 CSI after applying Butterworth filter

We can compare with Figure 15 (b) and find that the amplitude after filtering is much less high-frequency noise. Due to the change of CSI amplitude affected by the external environment, we can clearly see that the area of each frequency drop in the

figure represents a sign language. After passing PCA and Butterworth filter, we can clearly find the position of gesture features, But there is still a lack of a feature extraction method.

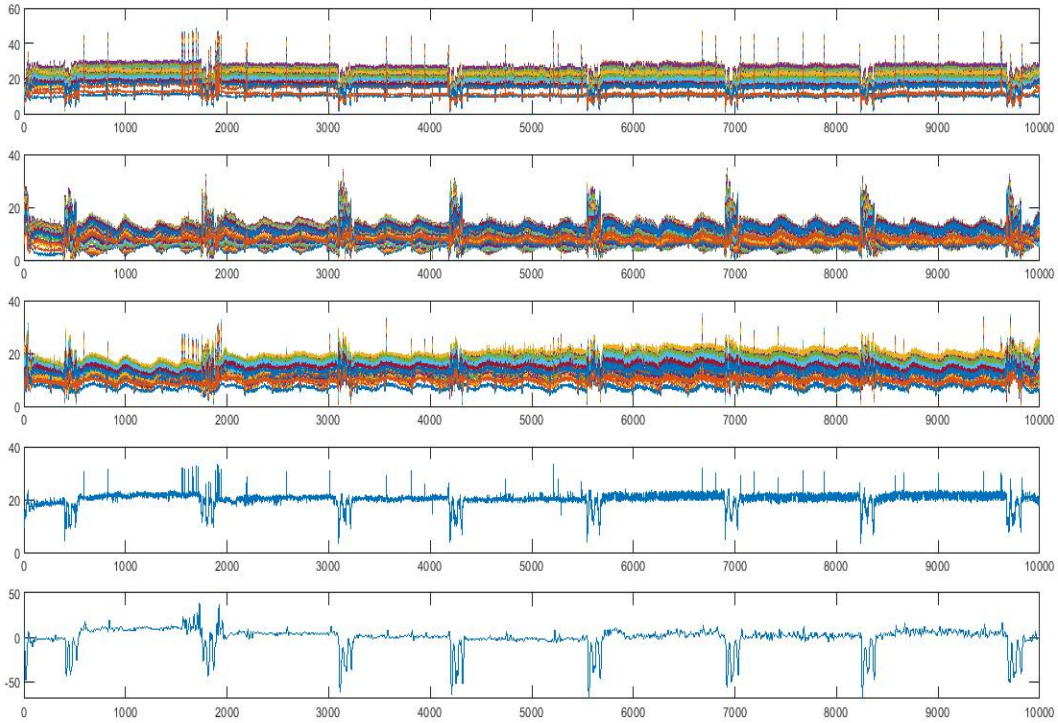


Fig.18 Summary of data processing

3.4 Feature Segmentation

In Figure 17, we can see that in 5000 packets, there are about 5 gesture samples, so we can't complete the classification of gestures in a time series. Therefore, feature segmentation is also a very important step. We propose a window based feature segmentation method.

In a certain time series, it contains some gesture features and a large number of useless blank values. First, we estimate the duration of gesture about 1.5s according to the time volunteers do each Japanese sign language. We know that the contract time interval is 0.01s from the system, so we can estimate the characteristics of gesture in 150 data packets.

According to the above inference, we first define a window with a length of 150. From the beginning to the end of the time series, we calculate the variance in 150 windows in turn, and get a series based on variance, as shown in Figure 19. Because when there is gesture feature, the fluctuation range of curve is very large, so we can preliminarily determine that the local maximum of this series is the position of gesture feature. So we only need to arrange variance sequence from big to small, and take the position of the first n maxima, which is the position of N gesture features in a period of time series. After applying this algorithm, we get nine Japanese gesture features as shown in Figure 20. We can see that the variance based feature segmentation algorithm successfully extracts gesture features from time series.

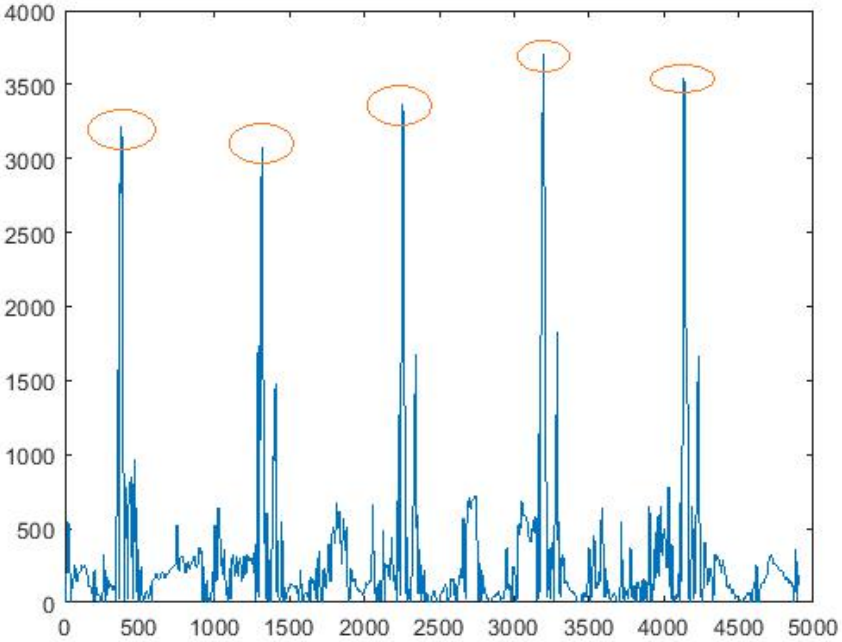


Fig.19 Variance based feature segmentation

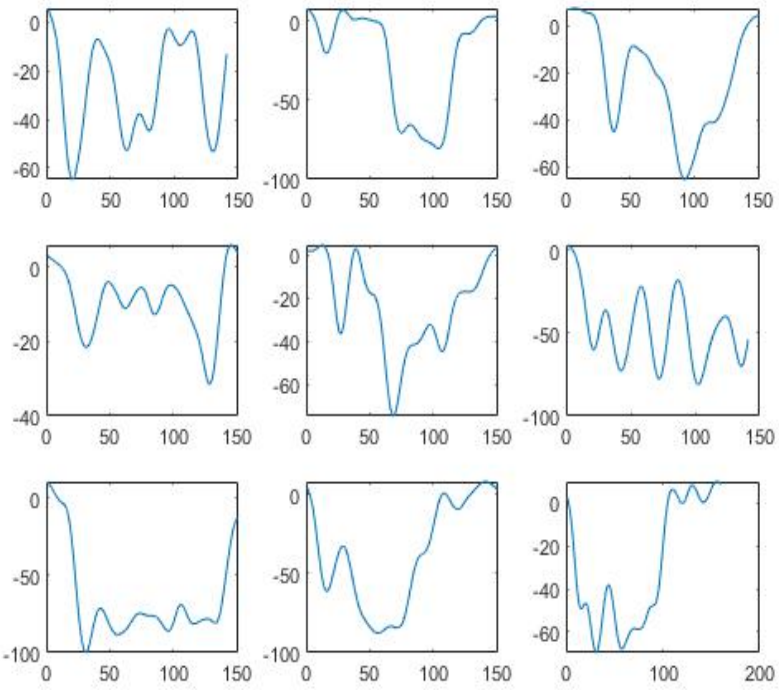


Fig.20 Features of nine Japanese sign language

Chapter 4

Classification of Gesture features

4.1 Experiments

The system consists of an ELECOM WRC-1167GS2-B router as the transmitting node, a Dell VOSTRO 3902 desktop with Ubuntu 14.04 system as the receiving end, and the distance between AP and DP is set to 1.5m. In order to make the experiment more practical, we simulated two kinds of communication conditions in daily life of deaf people. In the first condition, we assume that the experimenter is talking face to face with other people in the conference room, as shown in Fig.21 and Fig.22, a volunteer who is familiar with Japanese sign language sits between AP and DP and talks in sign language. In the second condition, we assume that a volunteer is consulting in front of a window. The distance of two speakers is 1m, and one volunteer makes sign language gestures between the AP and DP on the table.

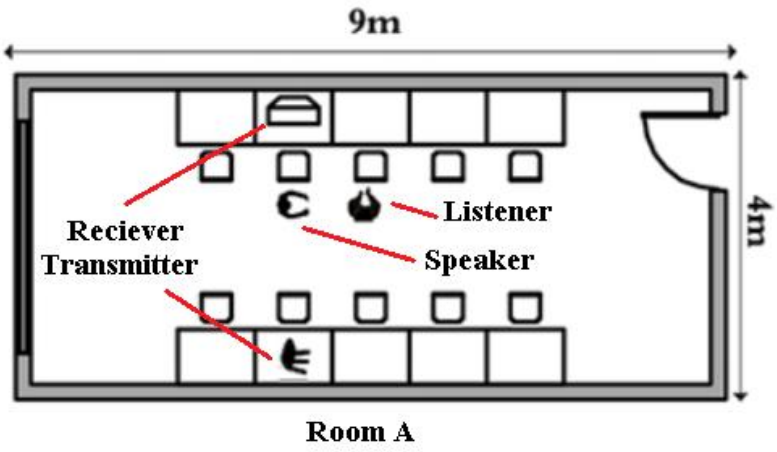


Fig.21 Room A: a face-to-face scenario

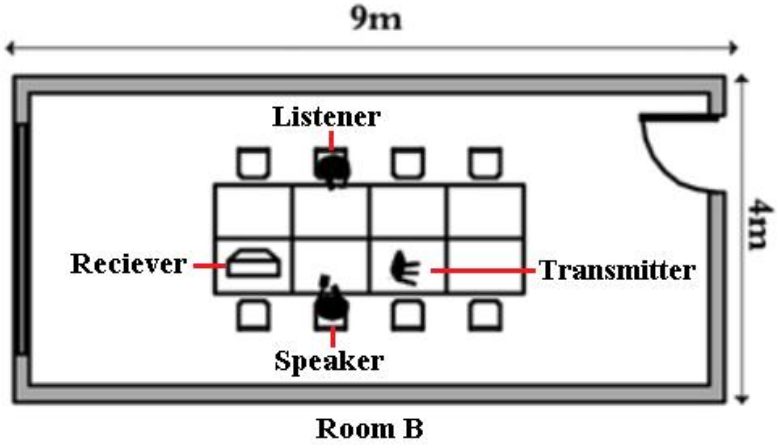


Fig.22 Room B: a consulting window scenario

We asked a volunteer who is familiar with Japanese gestures to make nine gestures at first, and take them as template data. Then we repeated nine gestures, with 100 test data for each gesture and each experiment environment. For each gesture, we repeated five times in a group of 20.

TABLE 2 Experiments details

Factor	Value
AP	ElECOM WRC-1167GS2-B
DP	DELL VOSTRO 3902 with Ubuntu 14.04
Distance of AP and DP	1.5m
Collected Samples	100 Samples for each Gestures
Packet Time Interval	0.01s for each packet
Type of Japanese Sign Language	C1: Can, C2: Good, C3: Bad, C4:I know, C5: Like, C6: Friend, C7: Sign Language, C8: Thanks, C9: Sorry

4.2 Dynamic Time Wrapping

4.2.1 Preliminary of DTW

In most disciplines, time series is a common representation of data. For time series processing, a common task is to compare the similarity of two series.

In time series, we usually need to compare the differences between the two ends of the audio. And most of the length of these two audio segments is not equal. In the field of speech processing, different people speak at different speeds. Even if the same person utters the same sound at different times, it is impossible to have exactly the same length of time. And everyone's pronunciation speed of different phonemes of the same word is also different. Some people will drag "a" a little longer, or "I" a little shorter. In this complex situation, the traditional Euclidean distance can not be used to obtain the effective similarity between the two time series.

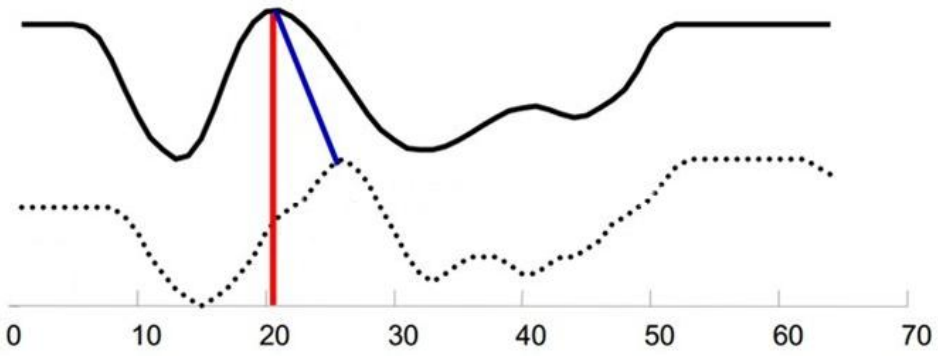


Fig.23 Dynamic Time Wrapping

In most cases, the two sequences have very similar shapes as a whole, but they are not aligned on the x-axis. So before we compare their similarity, we need to warp

one (or two) sequences under the timeline to achieve better alignment. DTW is an effective way to realize this warping distortion. DTW calculates the similarity between the two time series by extending and shortening the time series.

4.2.2 DTW based classification

In the first experimental environment, CSI propagates approximately freely at the transmitter and receiver, so there is not too much interference. The average recognition accuracy of gestures is 88%. In the second experimental environment, due to the reflection of the desktop, there will be some interference signals, and the average recognition accuracy is 86%. The results of comprehensive analysis of the two scenes are shown in Fig.26, and the overall recognition rate of the system is 87%.

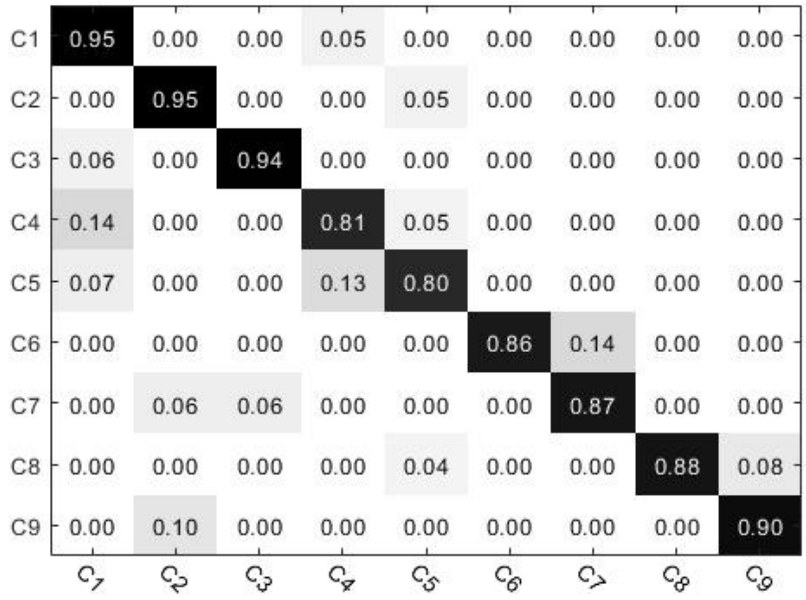


Fig.24 Confusion matrix of the classification in room A

DTW is used to calculate the matching degree between the test samples and the templates. The results are shown in the confusion matrix in Fig.24 and Fig.25. The accuracy of most gestures is about 90%, while the accuracy of some gestures, such as C4 and C5, is lower. In different environments, the accuracy of the nine sign languages is roughly the same.

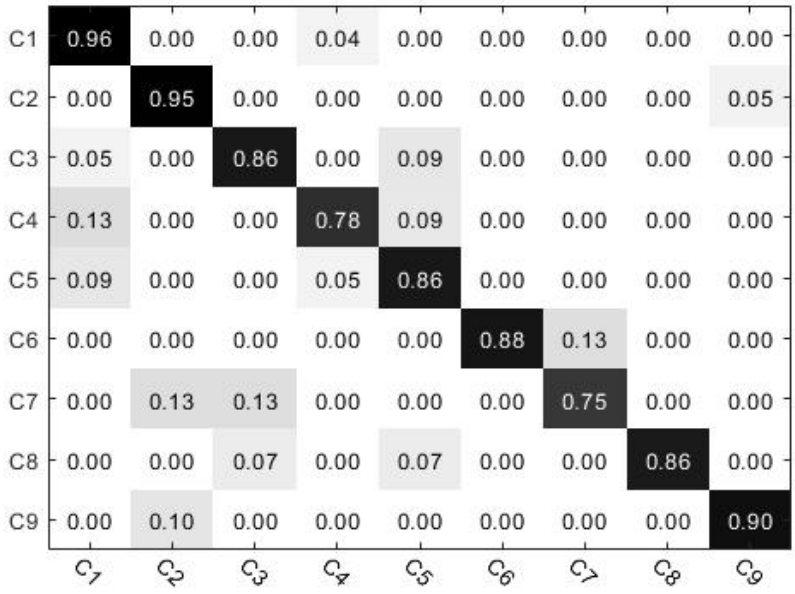


Fig.25 Confusion matrix of the classification in room B

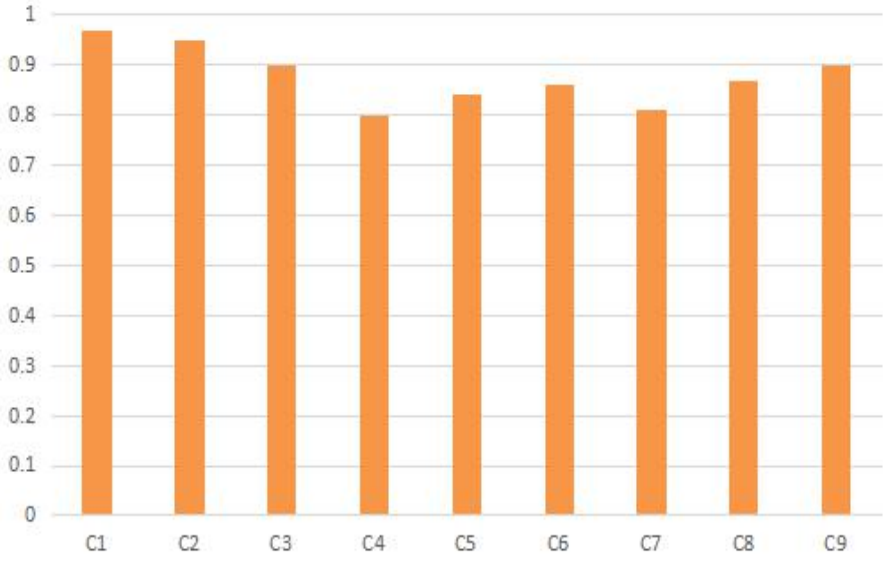


Fig.26 Total accuracy of nine Japanese sign language

There are two reasons: first, DTW algorithm is used to find the best matching gesture, but no threshold is set in our system, This is because even if you make the same gesture, different people's gesture range is very different. Second, some sign

language actions are difficult to recognize. If we increase the types of sign language, it will lead to the accuracy lower.

4.3 Artificial Neural Network

4.3.1 Preliminary of ANN

We got inspiration from birds, learned to fly, got inspiration from burdock, invented Velcro, and many other inventions were inspired by nature. So it's logical to look at the composition of the brain and expect to be inspired to build intelligent machines. This is also the basic source of artificial neural network (ANN). Artificial neural network is the core of deep learning. They are versatile, powerful and scalable, making them very suitable for large and highly complex machine learning tasks, such as classifying billions of images (such as Google images), supporting voice recognition services (such as Apple's Siri), and recommending the best videos to thousands of users (such as YouTube) every day, Or learn to beat the world champion in deepmind's alphago.

Artificial neural networks have been around for a long time: they were first proposed by neurophysiologist Warren McCulloch and mathematician Walter Pitts in 1943. They proposed a simplified computational model, which calculated how biological neurons work together in the animal brain, and used propositional logic to carry out complex calculations. This is the first artificial neural network architecture.

The characteristics and advantages of artificial neural network are mainly shown in three aspects: First, it has the function of self-learning. For example, when realizing image recognition, many different image templates and corresponding recognition

results are input into the artificial neural network first, and the network will learn to recognize similar images slowly through the self-learning function. Self learning is very important for prediction. It is expected that the future artificial neural network computer will provide economic forecast, market forecast and benefit forecast for human beings, and its application prospect is very broad.

Second, it has associative storage function. This association can be realized by the feedback network of artificial neural network.

Thirdly, it has the ability to find the optimal solution at high speed. Finding the optimal solution of a complex problem often requires a lot of calculation. By using a feedback artificial neural network designed for a certain problem and giving full play to the high-speed computing ability of the computer, the optimal solution can be found quickly.

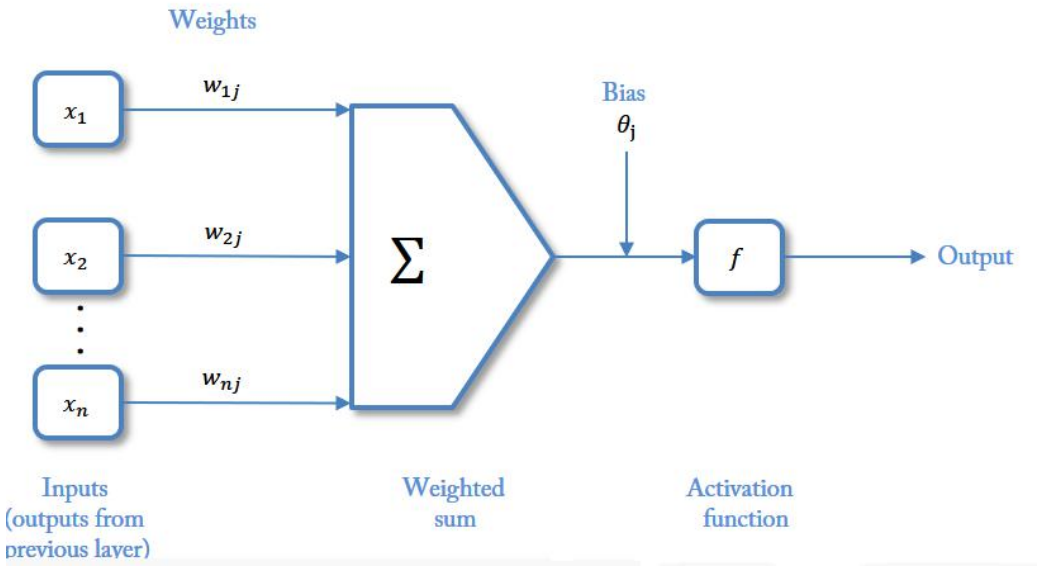


Fig.27 A simple Artificial Neural Network

4.3.2 ANN based classification

According to the previous experiment, we have 150 input features, so we set the number of input layers as 150, hidden layer 1 has 300 neurons, and hidden layer 2 has 100 neurons. Both hidden layers use relu activation function, and the number of model training is 20. Each gesture has 100 sets of data, 80 of which are used for training and 20 for testing. The results are shown in Figure 26. The average accuracy is 76%

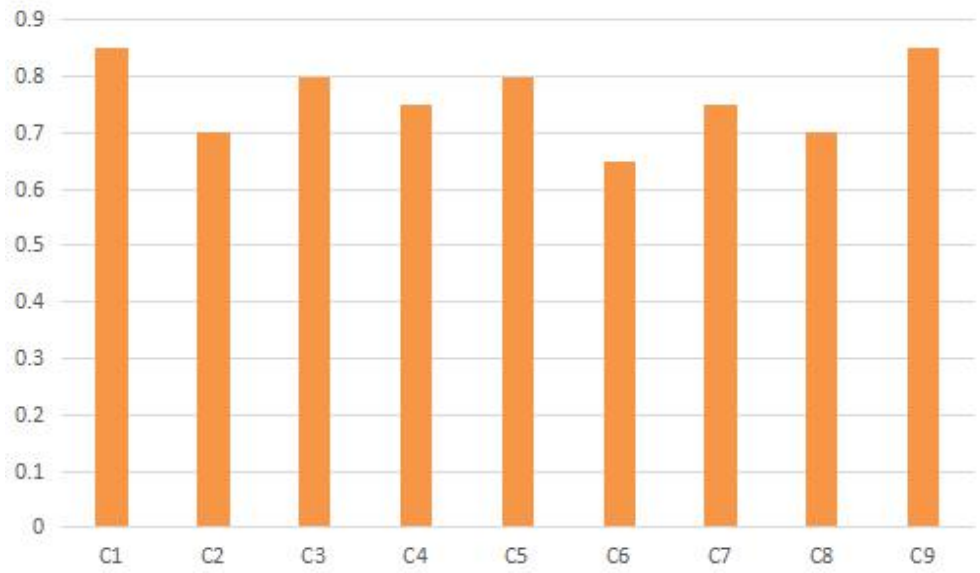


Fig.28 Total accuracy of ANN classification

We can see that the classification result is not as good as that of DTW algorithm, because the amplitude of CSI is not in a fixed interval, and it can not be fixed in the same position when collecting samples. Although the feature change trend of different samples of the same gesture is roughly the same, the amplitude may be very different, which has a great impact on the training of neural network.

4.4 Support Vector Machine

4.4.1 Preliminary of SVM

Support vector machine (SVM) is a powerful and comprehensive machine learning model, which can perform linear or nonlinear classification, regression, and even outlier detection tasks. It is one of the most popular models in the field of machine learning. Support vector machines (SVM) is a binary classification model. Its basic model is the linear classifier with the largest interval defined in the feature space, which makes it different from perceptron; SVM also includes kernel techniques, which makes it essentially a nonlinear classifier. The learning strategy of SVM is interval maximization, which can be formalized as a convex quadratic programming problem, and also equivalent to the regularized hinge loss function minimization problem. The learning algorithm of SVM is the optimization algorithm for convex quadratic programming.

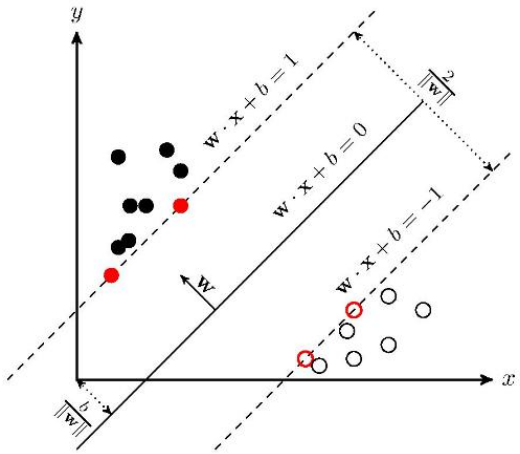


Fig.29 Basis of SVM

4.4.2 SVM based classification

In this section, we firstly build a model of SVC and initialize it, the classification model is one-vs-rest, we use the same training data and testing data as used in ANN, we pick Gaussian kernel function as the main kernel, and the classification results are shown below.

TABLE 3 Accuracy of SVM based classification

C1	C2	C3	C4	C5	C6	C7	C8	C9
75%	100%	94%	100%	100%	100%	100%	50%	100%

The average recognition rate of SVM classification is 91%. This classification algorithm seems to perform best, but for a certain gesture such as C8, it can be said that it can not be classified correctly. SVM system sacrifices part of the recognition rate of gesture and improves the overall recognition rate. If we want to consider the practicability and stability of the system, the performance of DTW algorithm is more balanced.

Chapter 5

Conclusion and Future work

5.1 Conclusion

We propose a device-free Japanese sign language recognition system based on Wi-Fi by using CSI, according to the characteristics of Japanese sign language actions, we use time series based data processing methods, including subcarriers selection, denoising, data segmentation and gesture classification. The proposed system is used to recognize nine commonly used Japanese sign language gestures in two daily life scenes, the results show that the system achieves high accuracy. Although our system can effectively recognize a single Japanese sign language gesture, it can not complete the recognition of continuous Japanese gesture combination, that is, a complete Japanese sign language sentence, which will be the focus of our future work.

5.2 Future work

Although our system can recognize nine Japanese sign language gestures, there are still many problems if it is applied to daily life.

1. Real time recognition: at present, we only test our system by collecting samples of certain time series. But if we can't achieve real-time input and output, the system can't be widely used.

2. Classification algorithm: Although the performance of ANN is not as good as the other two classification algorithms, it does not mean that it is not suitable for sign language recognition. In the future, we can get better classification results through the improved neural network algorithm.

3. Types of Japanese sign language: in my research, we only tried to identify nine kinds of Japanese sign language, and the number of sign languages to be identified is far more than nine, because the future work should focus on how to identify more types of sign language, while maintaining a high recognition rate.

References

[1] T. Yamaguchi, M. Yoshihara, M. Akiba, M. Kuga, N. Kanazawa and K. Kamata, "Japanese sign language recognition system using information infrastructure," *Proceedings of 1995 IEEE International Conference on Fuzzy Systems.*, Yokohama, Japan, 1995, pp. 65-66 vol.5.

[2] K. Imagawa, Shan Lu and S. Igi, "Color-based hands tracking system for sign language recognition," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, pp. 462-467.

[3] H. Sagawa and M. Takeuchi, "A method for recognizing a sequence of sign language words represented in a Japanese sign language sentence," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, Grenoble, France, 2000, pp. 434-439.

[4] N. Takayama and H. Takahashi, "Sign Words Annotation Assistance Using Japanese Sign Language Words Recognition," *2018 International Conference on Cyberworlds (CW)*, Singapore, 2018, pp. 221-228.

[5] V. E. Kosmidou and L. J. Hadjileontiadis, "Sign Language Recognition Using Intrinsic-Mode Sample Entropy on sEMG and Accelerometer Data," in *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 12, pp. 2879-2890, Dec. 2009.

[6] Y. Mori and M. Toyonaga, "Data-Glove for Japanese Sign Language Training System with Gyro-Sensor," *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, Toyama, Japan, 2018, pp. 1354-1357.

[7] K. Ali, A. X. Liu, W. Wang and M. Shahzad, "Recognizing Keystrokes Using WiFi Devices," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1175-1190, May 2017.

[8] G. Wang, Y. Zou, Z. Zhou, K. Wu and L. M. Ni, "We Can Hear You with Wi-Fi!," in *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2907-2920, 1 Nov. 2016.

[9] W.Xi, J.Zhao, Y.Li, K.Zhao, S.Tang, X.Liu, and Z.Jiang, Electronic frog eye: Counting crowd using wifi. In *Proceedings* of the *INFOCOM,* pages 361-369. IEEE, 2014.

[10] J.Cheng, J.Wu, A Robust Sign Language Recognition System with Multiple Wi-Fi Devices. In *Proceedings* of the Workshop on Mobility in the Evolving Internet Architecture, ACM, pp. 19-24, 2017.

[11] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, ”Tool release: gathering 802.11 n traces with channel state information,” ACM SIGCOMM Computer Communication Review, vol. 41, no. 1, pp. 53–53, 2011.

Research Achievement

Changhao Liu, Jiang Liu, Shigeru Shimamoto, “Sign Language Estimation Scheme Employing Wi-Fi Signal” in IEEE Sensors Applications Symposium, Stockholm city, Sweden. August 23th 2021. Accepted

Changhao Liu, Jiang Liu, Shigeru Shimamoto, “Machine Learning based Sign Language Recognition System” in IEICE Conference, September 17th 2021.