HAND-GESTURE SENSING LEVERAGING RADIO AND VIBRATION SIGNALS

BY SONG YANG

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Engineering

Graduate in Electronic and Computer Engineering

Written under the direction of

Yingying Chen

And approved by

New Brunswick, New Jersey

May, 2019

ABSTRACT OF THE THESIS

Hand-gesture Sensing Leveraging Radio and Vibration Signals

by Song Yang

Thesis Director:

Yingying Chen

Gesture recognition that enriches human-computer interaction (HCI) has gained considerable attention recently. Existing solutions such as computer-vision-based approaches recognize and track human hand/body gestures using cameras or visible light. However, they all require line-of-sight and are susceptible to interference from light sources. In this thesis, an innovative approach using ambient radio and vibration signals is implemented to achieve fine-grained hand/finger gesture recognition. By sensing the influence of hand/finger gestures on the transmitted radio signals (e.g., millimeter wave signals) and physical vibrations, the position of the hand/finger can be precisely estimated through machine-learning-based techniques. Particularly, we implemented two types of solutions that work separately. (1) mmWave-based: we leverage frequency-modulated continuous-wave (FMCW) radar to track hand movements and recognize various hand gestures; and (2) vibration-based: we capture the tiny disturbance in the surface vibrations caused by user's finger touches to discriminate user's finger input on the surface. Extensive experiments demonstrate that our proposed approaches can accurately track and recognize user's hand gestures with high accuracy.

List of Figures

Acknowledgements

First of all, I am very grateful to my advisor Prof. Yingying Chen for her insightful feedback and kindly guidance of my research direction. I appreciate Prof. Richard Howard's gently help in any period of this research. I would express thanks to my mentors, Jian Liu and Chen Wang. Their keen research observation helps me specify the research methods. I also like to thank Xin Yang. As my research partner, we had great cooperation which accelerates the research a lot. Last but not least, I' d like to thank my family for their encouragement and support.

Table of Contents

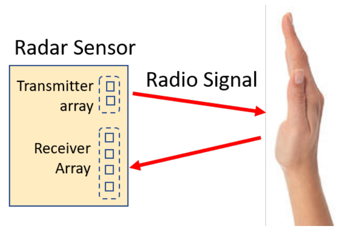
Chapter 1

Introduction

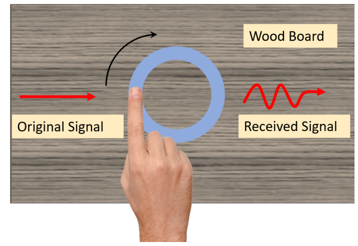
Gestures recognition pertains to recognizing meaningful expressions of motion by a human [1]. Gestures can express different human emotions. It is silent but vivid. Hand-gesture is a kind of expressive and meaningful body language involving the movement of fingers and hands. Hand-gestures can achieve the interaction with keyboards, mouse and electronic devices. Input methods like typing, clicking and scrolling are learned by acquired learning. Therefore, the gesture is a more natural and reasonable input action for humans.

There are some existing gesture recognition solutions. Some of them are based on vision systems [2] [3]. These systems complement the existing camera solution with other sensors such as a depth camera. However, vision-based systems are critical to the environment. Besides, how the background influences the prediction result is complicated. Some other gesture recognition utilizes visible low-power LED lights and their power density of different angles [4]. However, this system is fragile to ambient light like the camera solutions. And the users are required to input the touch points or gestures on a certain surface. Our work, with its resistance to the light changes and in-the-air gestures detection feature, steps forward to the natural gesture recognition. Comparing to the camera solution, the signal transmitters and receivers that signal-based system cost less than the cameras.

In this thesis, two gesture recognition methods are investigated to provide interaction between people and the environment. The user performs the hand-gesture in the air, in front of the radar antennas as shown in Figure 1.1(a). By utilizing interactions such as hand-gestures and body activities, the computer can understand and makes



(a) Radio-signal-based system.



(b) Vibration-based system.

Figure 1.1: Illustrations of hand-gesture sensing systems.

corresponding feedback to human behaviors. First, we build a radio-signal-based system. We use the radar sensor to measure the distance of each part of the human's hand, and furthermore achieve gesture recognition goal.

The second recognition system we built is vibration-based gesture recognition. The user is required to perform the gesture on a specific board. The overall idea is that, when the user performs a gesture on a media which works as a signal transmitting, we collect the signal on the receiver end. The illustration in Figure 1.1(a) shows the gesture performing process. By comparing the fluctuated and original signals, the software can learn the fluctuation pattern of each gesture. Finally, we use a threshold based classification to distinguish gestures. The overall system is similar to [6]. For the classifier of both systems, a gesture can be regarded as a series of time-related points. The similarity among the time series is quantifiable. We define the similarity with the combination of two existing distance indicators, dynamic time wrapping (DTW) and earth mover's distance (EMD).

To conclude, the objective is to achieve gesture recognition leveraging radio and vibration signals. We reduce the difference among users and build both gesture recognition systems for in-the-air hand gestures and physical touched gestures recognition. Two threshold-based systems are built to distinguish two types of gestures. Profiling the gestures is required in both recognition systems. For evaluations, we finally get 96.3% accuracy for the radio-signal-based system, in differentiating two hand-gestures and get overall 91% accuracy in the vibration-based system for differentiating three different gestures.

Chapter 2

Related Work

The main contribution would be radio-signal-based and vibration-based gesture recognition systems. They are both signals in a certain waveform with different bandwidth. By utilizing shorter wavelength and FMCW features, the radio-signal-based system provides a higher resolution.

A radio signal is a form of electromagnetic wave. These signals are already used for high-speed data transmitting such as 5G [16]. Since the normal high-frequency mmWave signal is easy to get block by the solid surfaces, a research [19] targeted to the mm-wave range, 75-110 GHz investigated the signal absorbing performance on the frequency selective surfaces. For gesture recognition, there are already several pieces of research indicate that the mmWave has an ability to detect human activity and have some advantages compared to Wi-Fi channel state information (CSI) [9]. The paper proof that the mmWave have an advantage compared with the other two signals. They choose to use a self-designed testbed X60 [10]. The testbed costs at least $400k and are customizable. Others use mmWave beamforming to recognize gesture [11]. They introduce a 94 GHz radar transceiver with on-chip antennas. The transceiver provides phase, amplitude, and time-of-flight information on echo pulses. This work contributes more on the antenna and circuit design.

Vibration-based gesture recognition is also a part of the application of the high-frequency signal. The High frequency signal is widely used for carrying information or used for information transferring [12] [13]. Since it is hard to fluctuate, and the carried data should be well-saved. Some gesture recognition system use visible light [14]. By utilizing the features of radio and vibration signals, we can regard the signal as a feature. By fluctuating the signal, the signal will hint some information about the fluctuation process.

Chapter 3

Background and Feasibility

In this section, the feasibility of recognition using radar signals and vibration is investigated. Vibration-based gesture recognition makes use of physical vibration propagation features. When the designed signal vibrates through the surface, any tiny influence to the surface can cause a relatively huge difference to the receiver side in the frequency domain. For vibration-based system setup, the following sections will present the fundamental theory of vibration attenuation, and how human's finger can effect such process. The basic gesture detection strategy in detecting movement using mmWave is introduced. There is no physical medium along the transceivers, users will perform the hand-gestures in front of the radar sensor. We can differentiate fine-grained gestures according to the related spectrogram.

3.1 Millimeter Wave Object Detection

Millimeter wave (mmWave) is the signal that the band of its spectrum is between 30 gigahertz (GHz) and 300 GHz. Millimeter means that the wavelength of the wave is in the millimeter level. A 30 GHz millimeter's wavelength is about 9.99 mm. Normally, the thickness of the human's hand is 2-3 cm. Such band allows the signal containing more information including the distance, angle, and velocity of the object. The Figure 3.1 shows the 1D trace of the swiping right gesture. The mmWave utilize frequency-modulated continuous wave (FMCW) to detect the object. In FMCW, the transmitted signal is a linear frequency modulated chirp signal with a short duration for each chirp signal. By calculating the frequency and phase differences between the original and reflected chirp signals, the radar could even detect multiple objects.

The basic algorithm of detecting the distance, angle, and velocity of the object relies

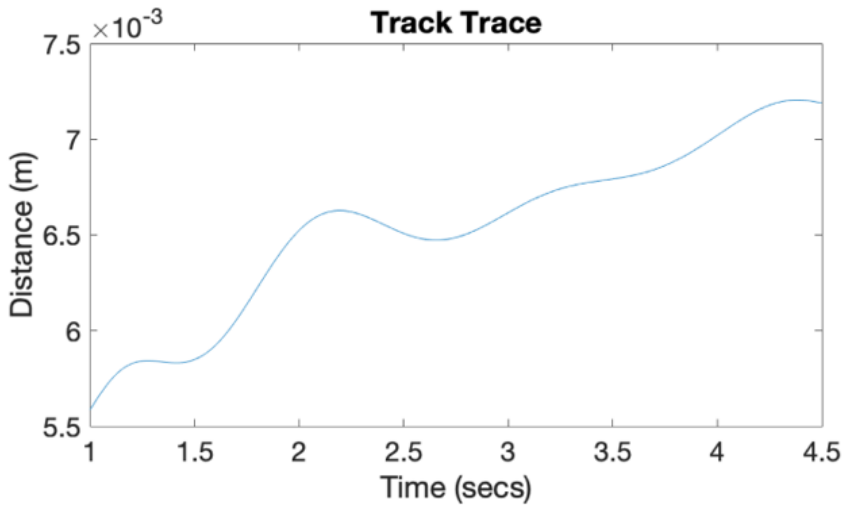


Figure 3.1: Feasibility experiment swiping right tracing.

on the frequency and phase difference. The fundamental concept is that the antenna array sends the chirp signal and receive the reflected signal wave. The increasing speed of the frequency of a signal is called the slope. The difference between transmitted and received a single chirp signal when overlapping is defined as the intermediate frequency (IF) signal as we can see the solid part in Figure 3.1. The IF signal is computed by the hardware circuit. The initial phase (ϕ_0 in Figure 3.1) of the IF signal is the difference between the phase of transmitted and received chirp which is known. Besides, it can be derived to Formula 3.1. f_c is the start frequency of the chirp signal. τ is the time delay which is 40 microsecond in this Figure.

$$\phi_0 = 2\pi f_c \tau \tag{3.1}$$

The time delay (τ) can be calculated. Then, the distance can also be calculated. Distance estimating is simply measuring the frequency difference between transmitted and received signals. With the distance to the object, we calculate the time delay (τ) between transmitted and received signal with the following formula 3.2. c is the speed of the light [15].

$$\tau = \frac{2d}{c} \tag{3.2}$$

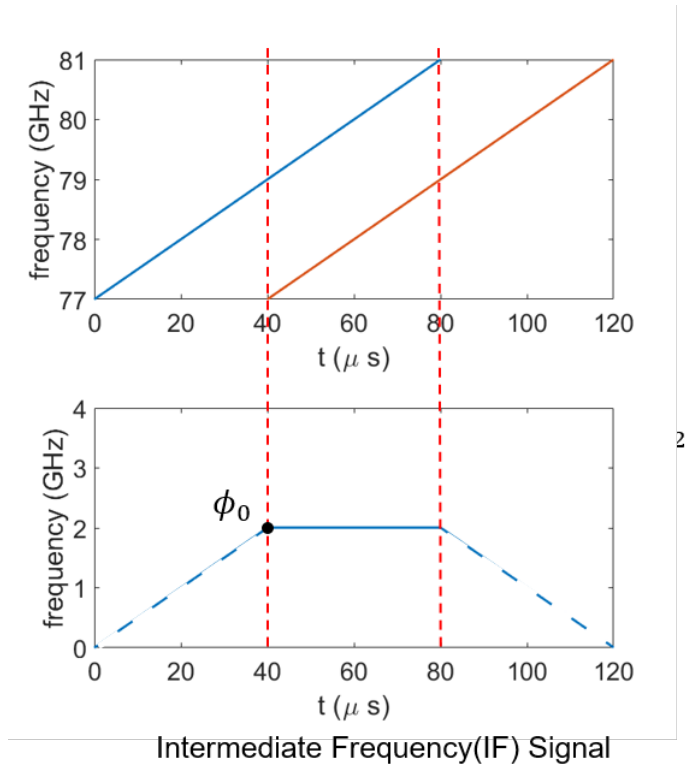


Figure 3.2: Chirp and intermediate frequency signal in FMCW.

The velocity and angle measurement are both based on the phase difference. According to the Doppler Shift, an object moving to the wave source would cause the change in frequency or the wavelength of a wave. The frequency and phase difference is caused by the movement of the object. When there are two chirps in transmitted to a moving object, the tiny phase difference ($\Delta\Phi$) in each chirps difference can be calculated. Since the absolute distance can be estimated through range measurement, with two distances and intermediate transmitting time different, the velocity (v) of the object can also be calculated with this following formula 3.3. T_c is the length in time of a single chirp. λ is the wavelength of the radar signal [15].

$$v = \frac{\lambda \Delta\Phi}{4\pi T_c} \tag{3.3}$$

In the velocity calculating algorithm, the velocity resolution is closely related to the difference among transmitted times. This algorithm is actually calculating the average speed of two points based on range detection. Assuming that the target object is moving

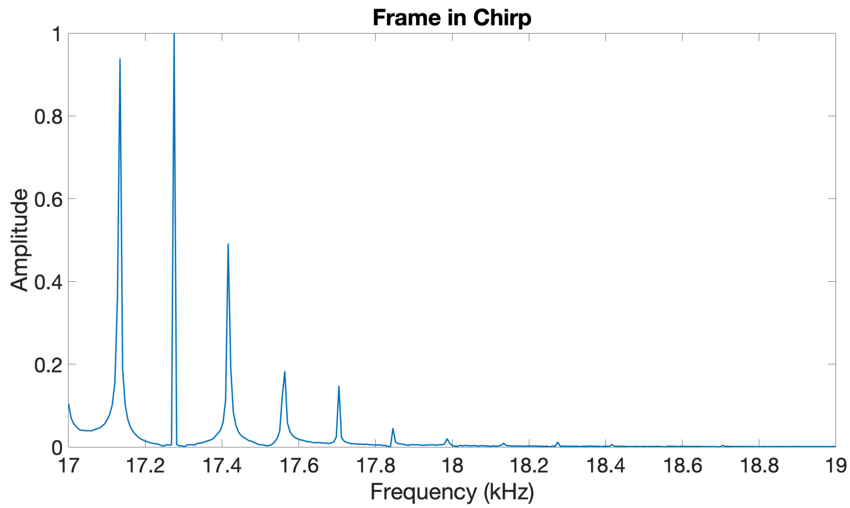


Figure 3.3: Received 17kHz to 19kHz chirp signal.

along the centric line of the mmWave radar. The actual resolution is blurred but it can still find out the direction with range and velocity measurement.

To conclude, the mmWave equipped with its resolution in range and angle, have the ability to detection fine-grained hand and finger gestures. But the limitation is that the horizontal field of view (FOV) it 60 degrees maximum and the vertical FOV is just about 5 degrees. Thus it limits our gesture pattern design.

3.2 Physical Vibration Propagation and Attenuation on Common Boards

As we discussed in Section 3.1, object detection is based on the frequency change of each chirp signal transmitting process. And the vibration propagation is different. The vibration transmits as a wave in the air and the environment can easily get blocked and then affect the procedure. By transmitting the signal into a solid board, the solid board works as a low pass filter in this situation. By choosing a proper method to combine the surface and the signal transmitter sensor, it is possible to revert the frequency peaks.

However, the signal quality may change after transmitting through the board. According to the different texture and pattern of the board, the vibration wave may lose some part of the amplitude which is called signal attenuation. Figure 3.3 is the received 17kHz to 19kHz chirp signal. The frequency peaks show at the same frequency points as long as the transmitter signal is fixed, no matter how we touch the transmitting media. The only change is the amplitude. Thus it is possible to measure the amplitude differences in each frame along with time for gesture recording.

Since in our setup the signal is not actually saved information on any carriers, the loss of parts of amplitude is acceptable. But the recognition effect can be varied. In theory, a lighter board is easier to drive and vibrate. We should comprehensively consider all aspect to performance a better result.

Chapter 4

System Design

4.1 Radio-signal-based System

4.1.1 Radio-signal-based System Design

In this section, different from the previous system, a new system utilizing mmWave is built for testing the radio-signal-based hand-gesture recognition. The system architecture is showed in Figure 4.1. We utilize an off-the-shelf millimeter wave product from Texas Instrument (TI) [20] to detect the hand-gesture. As we discussed in Section 3.1, the mmWave is theoretically available to detect hand-gesture. Figure 4.1 is the illustration of system design. In the data collecting part, when the user performs a hand-gesture in 5 seconds in front of the radar antennas and the mmWave radar is triggered as recording, the data capture card will record the signal. The recorded signal from antenna transfers to the data handling block. By the fundamental knowledge of FMCW, we can extract range, velocity, and angle estimation. Furthermore, we can plot 1D trace with these data. More importantly, we analyze these data and build profiles for gestures. In profile building, existing indicators are used to calculating the similarity of hand-gestures as time sequences. The final decision is made by majority vote algorithm. With the similarities, we can distinguish hand-gestures by with a threshold-based classifier.

4.1.2 Hardware Design

The antenna on the mmWave board is aligned as Figure 4.2(a) which is the front view of the antenna setup. A mmWave is set right in front of the user. Since the discussion above in Section 3.1, the FOV is limited in the vertical direction as it shows in Figure

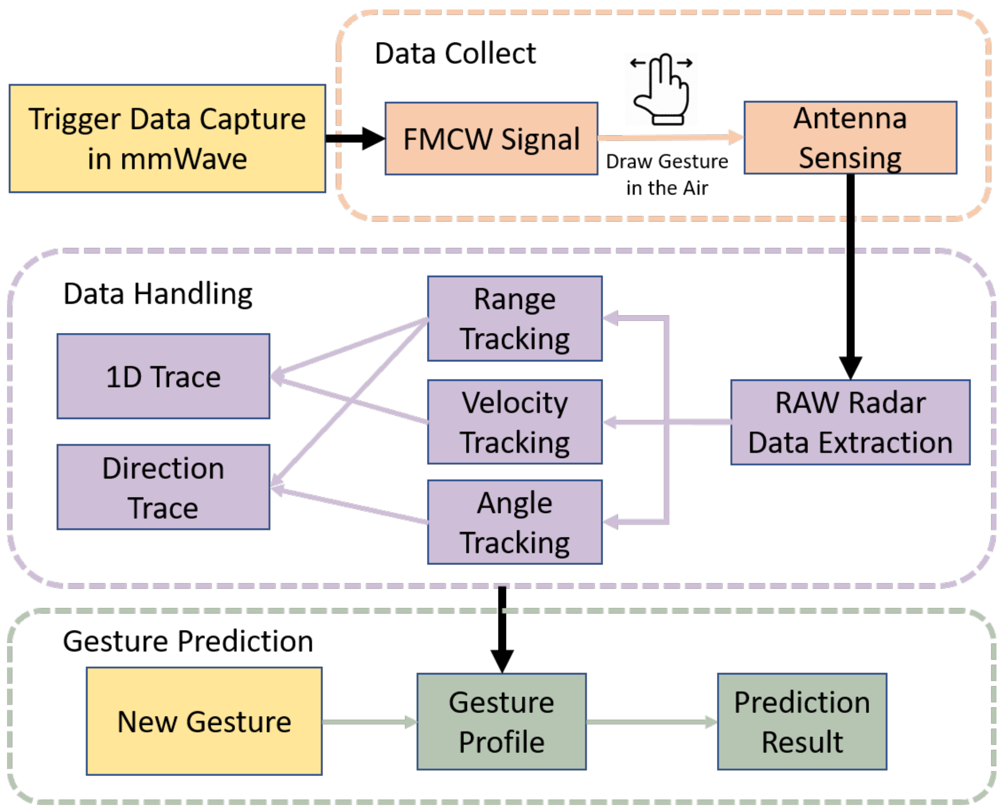


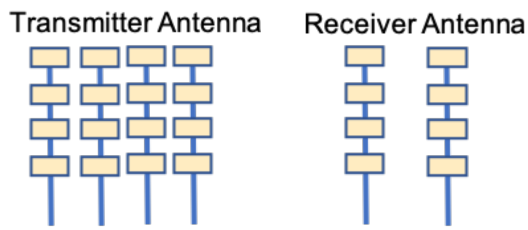
Figure 4.1: System architecture for radio-signal-based gesture recognition.

4.2(b). So we perform the gesture not exceed the FOV along the horizontal surface. We setup the mmWave to capture mode. By connecting the Ethernet cable on the board to the computer, we can stream data directly from the radar sensor.

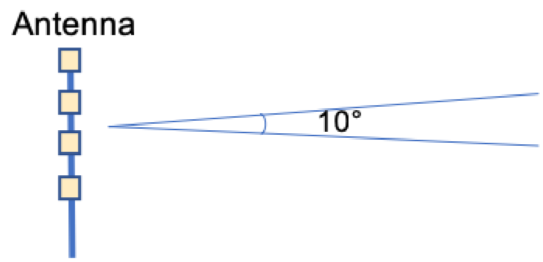
4.1.3 mmWave Radar Sensor

We utilize the mmWave Radar board AWR1642 to achieve our requirements for transmitting signal. The board can produce the signal starting from 76GHz to 81 GHz. It coverages 4 GHz available bandwidth. The transmitter transmits with the power of 12.5 dBm.

TI also provides a series of radar control programs. These programs can customize the radar in different configurations. We capture data in complex and real formats. After extracting the five-second data from the radar sensor, we extract and analysis features with our Matlab code. We build gesture profiles, track the moving trace of the object in 1 dimension, and find out the difference among gestures.



(a) Frontview of antennas.



(b) Sideview of antennas.

Figure 4.2: Views of antennas.

4.1.4 Gesture Design

As shown in Figure 4.3, 3 gestures are designed. The ability to distinguish each other will be tested. These gestures are performed in front of the mmWave radar sensor. The symmetry gestures such as swiping right are not given since we could easily differentiate them with the moving direction in the same profile. Hand gestures are chosen rather than finger gestures like fisting because the radar sensor receives the reflected signal mostly from the hand palm whose size is large enough. The circle gesture is chosen because this gesture has a moving range in all distance. We'd like to find the FOV in vertical. Other two gestures are all learned from daily life.

4.2 Vibration-based System

The section describes the theoretical system overview and practical system design in details. The structure and functions of the prototype will be explained including material and method selections, approach trade-offs, and other considerations that may affect the system.

(a) Push and pull. (b) Swipe right. (c) Circle.

Figure 4.3: Illustration of radio-signal-based gestures.

4.2.1 Vibration-based System Overview

The overview of the system is described as follows. Regarding the background influence indicators of all investigated results, it is feasible to detect the frequency difference comparing to the stable status in performing gestures situations. The theory is that, when the finger is touching on the surface, due to the physical change of the material, the receiver receives the significantly different signal. By utilizing the divergence using learning algorithms, the smartphone could sense the touched position. Furthermore, owing to the different strength and size of the fingertip in different gestures, the smartphone could differentiate gestures.

The prototype contains the following steps to achieve prediction goal as it presented in Figure 4.4. Generally, the specified signal trough sensor broadcasting on the surface, and picked up by the other sensor which works as a receiver. All kinds of the signal from all directions will be collected by a smartphone through the headphone jack. Then we process collected data and compare with other gestures. Finally, output a predicted result. A separate introduction of all hardware parts, together with findings during experiments, is provided in the next following subsections.

4.2.2 Sensors Selection

In this prototype, sensors are required for both transmitter and receiver sides. The piezoelectric sensor is an ideal choice since the principle of the piezoelectric sensor which is a physical dimension, transformed into a force, acts on two opposing faces

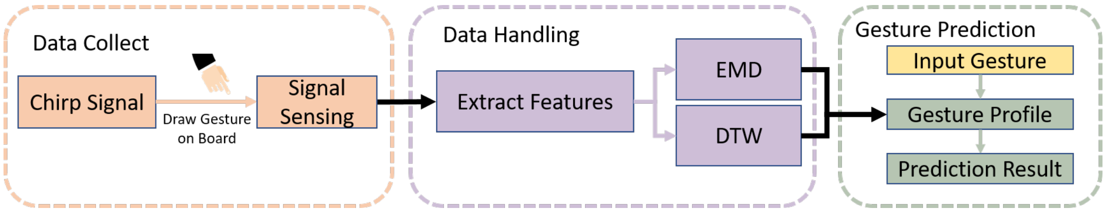


Figure 4.4: System architecture for vibration-based gesture recognition system.

of the sensing element.To ensure most of the signal wave produced by the sensor is transmitted into the physical surface rather than in the air, sensors should firmly be attached to the board. When transmitting the signal, leveraging the sensor to convert the signal wave to physical vibration. Higher amplitude of the signal can produce a robust signal. We looked for a certain piezoelectric sensor whose resonant frequency is near 20 kHz. Finally, we found an 18kHz sensors sold by Newark, 18kHz is also used as the centric frequency of the sweep signal.

4.2.3 Design of Transmitting Signal

As mentioned in the previous paragraph and according to the system design, the source signal is not only required to transmit in 20-centimeter's range but also needs to be in the scope of microphone frequency response range. In addition, the receiver expects to receive a clear and relatively stable signal with significant frequency peaks. The frequency peaks are short Fourier transform of the received signal as is showed in Figure 3.3. To get the ideal signal, solutions are provided below.

The signal needs to balance among aspects. The commonly stated range of human hearing is 20 Hz to 20 kHz. To reach the inaudible goal and considering the maximum frequency that the smartphone can handle with, the upper bound of the frequency is 22 kHz. Since the recognition mechanism is based on a training algorithm, the number of valid samples or features is a dominant factor in predicting the result. In contrast to the signal with a single frequency peak, it contains more valid sample if multiple frequency peaks show at the same time. Sweep signal and linear superimposed are two methods to achieve this goal. The linear superimposed signal is so stable comparing with sweep signal that causes the over-fitting problem in the test. An over-fitting training model makes the same gesture difficult to pass the recognition. As explained, the sweep signal with little fluctuation is an ideal choice. Higher amplitude can fight against the background noise while traveling through the surface so that a 5W amplifier is applied at very front. Another way to amplify the signal is trying to use the signal around the resonant frequency since the sensor will produce the strongest vibration feedback at the resonant frequency. The resonant frequency of selected sensors is 18 kHz which fairly meets the requirements. Because the project is not a data transferring application and does not need much bandwidth, 17k-19kHz signal is enough for feature extraction and training. The sample rate is set at the upper bound of smartphone acceptance which is 192kHz. As a result, 17k-19kHz is selected to be used in future tests, and that fully satisfy our needs.

As we tested and learned from the vibwrite [6] before. The chirp signal with 0.004 sweep period performs well on the PIN Pad authentication scenario. However, reducing differences between users is the objective of this recognition system. We reduce the sweep time to 0.003 and get a better result in the vibration-base gesture recognition.

4.2.4 Gesture Design

The gesture is expected to work as a switch with more flexible and reliable features since there is no physical limitation. We designed 3 patterns in Figure 4.5. Since in the previous work, the system is confirmed that can authenticate users by touching the PIN number on the surface [6]. The two-line and triangle gestures evaluate the system in multi-points features. The circle is a closed graphic, hinted the continuous recognition possibility in future work. With a clear gap in different steps, the triangle is also selected as the tested gesture.

4.3 Profiling

4.3.1 Similarity Definition

Once a series of data is received we need to build a profile for each gesture. Profile samples are connected by the similarity between each other. We define three kinds of

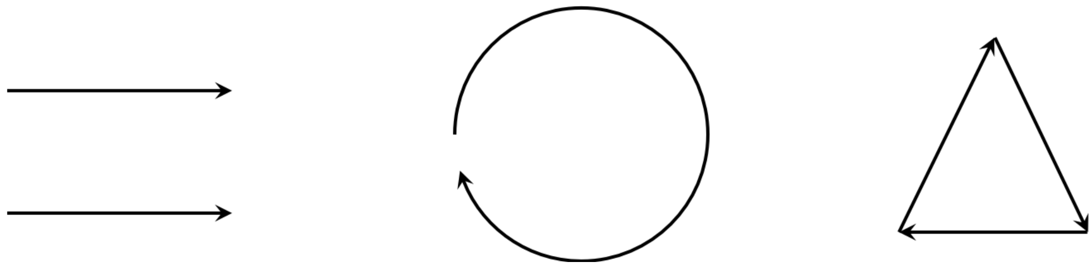
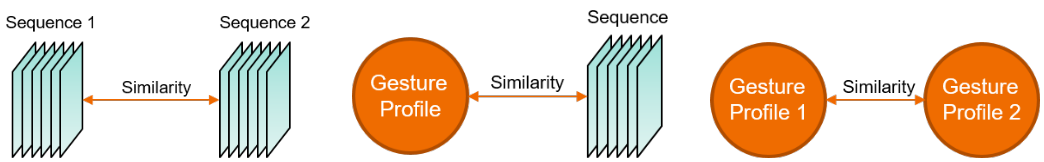


Figure 4.5: Illustration of target gestures.



(a) Similairty between gestures. (b) Similairty between a gesture and a profile. (c) Similairty between profiles.

Figure 4.6: The similarity definitions.

similarity, including similarity between gestures, the similarity between a gesture and a profile, and the similarity between profiles as the illustration shows in Figure 4.6.

Similarity between gestures. We selected the Earth Mover's Distance (EMD). EMD is a measure of the distance between two probability distributions. Assuming that we have two gesture $P = \{(P_1, w_{P_1}), (P_2, w_{P_2}), (P_3, w_{P_3}), ..., (P_M, w_{P_M})\}$ and $Q = \{(Q_1, w_{Q_1}), (Q_2, w_{Q_2}), (Q_3, w_{Q_3}), ..., (Q_N, w_{Q_N})\}$. P_i/Q_j is the certain feature of P/Q which is each frame of the time sequence. We define a D as the distance matrix describing the distance between each pair of features. Thus the dimension of D will be $M \times N$. We are looking for a matrix flow F that can minimize the cost function $\sum_{i=1}^{m}\sum_{j=1}^{n} f_{i,j}d_{i,j}$ which is an dynamic programming program. The EMD is calculated with following formula.

$$EMD(P, Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{i,j}d_{i,j}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{i,j}} \tag{4.1}$$

Besides, we also need an indicator to calculate the D in EMD calculation. This indicator should match the spectrogram pattern in each frame. The dynamic time warping (DTW) is selected. As we can see in the received signal in Figure 3.3, frequency peaks are fixed. So DTW will only calculate the Euclidean distance. For the rest of the remaining, in case of some certain frequency have a usual fluctuation DTW is necessary. We use the DTW similarity sequence as the weight of EMD and calculate the similarity between two gestures . Thus, the EMD with weighted as DTW in each frame is regarded as the similarity between gestures.

Similarity between a gesture and a profile. This similarity is used when validating a gesture to a profile. There is simply two steps for calculating the similarity between a gesture and a profile. First, calculate the similarity between the test gesture and each profile gestures. We will get a distance matrix after that. Then the validating decision is made by over 60% majority vote. The definition of threshold is $threshold = MEAN(Profile\,Sim\,Matrix) + STDEV(Profile\,Sim\,Matrix)$.

Similarity between profiles. This similarity is used to evaluate the profile building result. We simply define it as the average similarity of the training data in each profile.

4.3.2 Profiling

In either system, the gesture profiles are required to distinguish with other gestures. The profile of a gesture is mainly depending on the 'distance'. The distance is defined by the summation of DTW weighted EMD. The processing is described below. Starting with the profiling samples, the new input gesture whose distance to the profile is smaller than a threshold is valid to such profile. The decision is made by the majority vote algorithm. With profiling, the distance to a gesture is defined. The database of a gesture can update to fit variety gesture inputs.

Chapter 5

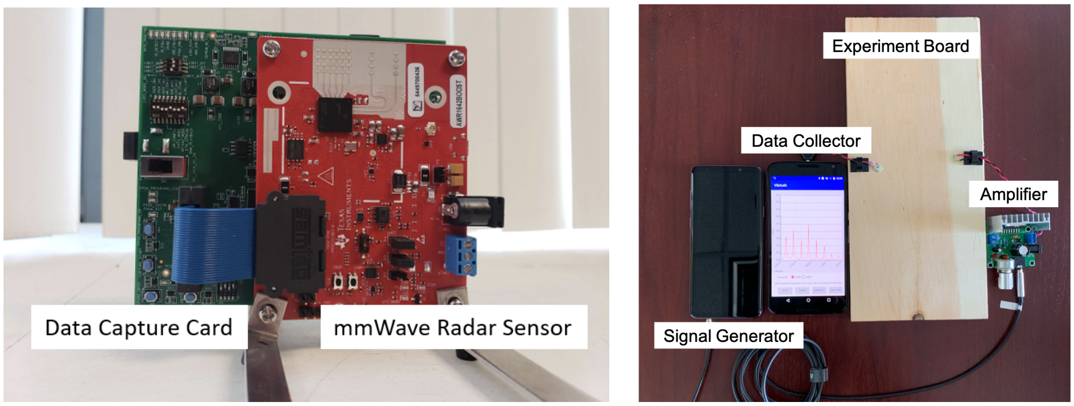
Performance Evaluation

The performance of the prototype is given in the following section. The prototype is tested in different aspects to evaluate its performance in real situations. In addition, the motivation for all experiments will also be presented.

5.1 Data Collection

For the vibration-based gesture detection, the selected transmitting signal is a sweep signal in the frequency range from 17kHz to 19kHz in a 0.003 seconds period. We collect signals from the headphone jack and cut signal with such frequency range to decrease the influence of noise on other frequency bands. There is about 8 frequency peeks showed in the frequency domain graph as Figure 3.3. These peaks are the key features to predict the position of the finger. The user is required to perform the gesture on the board in 5 seconds. The Android application will capture these data sampled with a 200ms sliding window which means the program sample the data with 5 frames per second. As a result, a series of points for each frame is captured. For each gesture, there is a data matrix including frames and signal points in each frame.

In the evaluation experiment, for the radio-signal-based system, we perform each gesture 10 times and calculate the accuracy for each gesture in all 30 gesture samples. We repeat the experiment 10 times to get the overall accuracy. For the vibration based system, the user performs each gesture 10 times and calculate the true positive and false positive accuracy for each gesture in all 30 gesture samples. We repeat the experiment 10 times to get the overall average accuracy.



(a) Vibration-based gesture recognition setup

(b) Radio-signal-based gesture recognition setup

Figure 5.1: The experimental setups

5.2 Prototyping and Experimental Setup

Both systems identify the user's input trace in frames. The system should resist common obstruction such as slight finger moving and noise vibration from other sources on the table. With that said, we need information for each status such as speed, and angle.

The experiment setup is showed in Figure 5.1. Note that in Figure 5.1(a), other than the wood board, acrylic board is also tested. Amount of experiments is finished on the board. There are two sensors at each side of the board in the figure. They can be either transmitter or receiver depending on the transmitted signal and wire connection to the smartphone. The transmitter transmits the signal from a smartphone as we discussed in the previous section. The receiver sensor's wires go to the headphone jack of another smartphone for data collecting and training. The smartphone in receiver and transmitter side can either be split or combine in the same phone since the transceivers utilize different channel of the headphone jack features. The user is expected to perform the gesture in the same position. The dimension of the sensor is shown in Figure 5.2. With 18kHz resonant frequency, the receiver can receive a clean and stable signal.

We collect the real and complex data on the mmWave by a data capture card produced by TI company. These data are processed and converted to temporal sequences. We also collect data from the Android application. And they are also in the format of

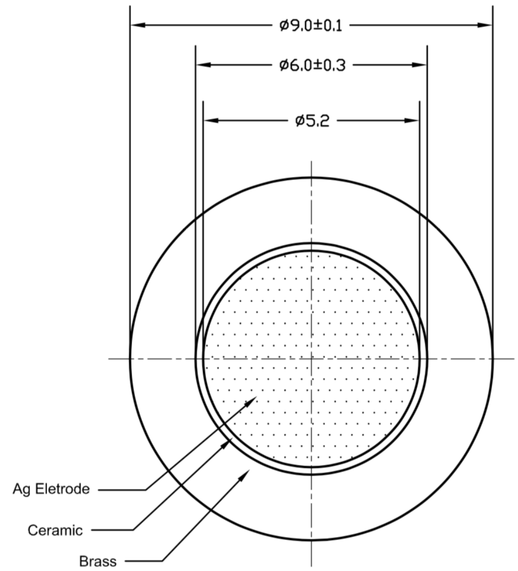
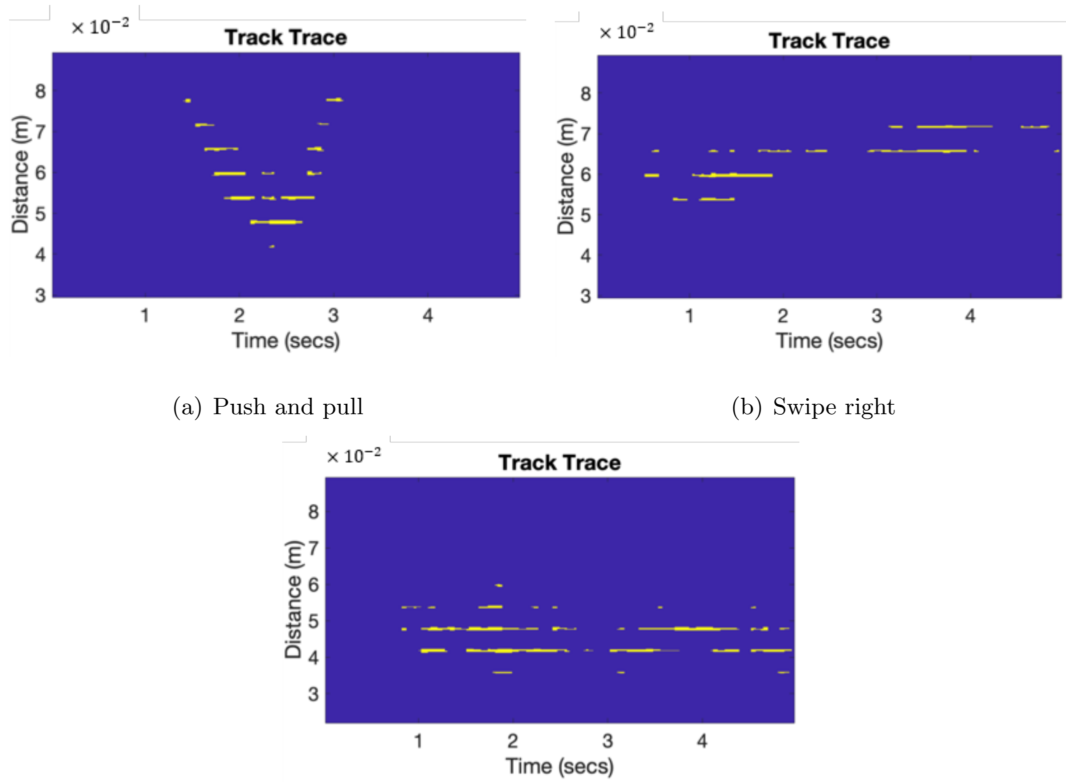


Figure 5.2: Selected piezoelectric sensor.

time sequences.

In the evaluation of the radio-signal-based gesture recognition, the mmWave is set 50 centimeters away from the user. As we already know, the gesture is performed in the range of 1 meter, we manually cut the range of 1 meter to gain a pure data. Thus we do not need to worry about the size of the room during the experiments. The background of the test environment is clear and stable in order to block the multi-path noise of the radar.

For the physical touched gesture, the texture and material of the surface influence the transmitting efficient. For instance, light wood is tested at the first. The experiments indicate that the vibration travels faster along the texture of the wood which leads to low sensitivity when user touching places that are vertical to the texture lines. The receiving sensor receives the signal mostly transmitted by the central texture line. This finding guides us to choose materials that have a uniform density and is completely integral. As a result, the experiment are designed on both 1" $\times$ 6" $\times$ 12" wood board and 12" $\times$ 12" $\times$ 1/4" acrylic board. Note that, since we use 17kHz to 19kHz signal in vibration-based gesture recognition, the environment is confirmed without any noise lower than 20kHz.



(a) Push and pull

(b) Swipe right

(c) Circle

Figure 5.3: Preliminary traces of gestures.

5.3 Evaluation of Radio-signal-based System

As we could observe from the moving traces of 3 gestures in Figure 5.3, the mmWave radar can trace the moving trend by using only 1 transmitter and receiver. We can profile these gestures with more data and differentiate with others. The scattered points are grouped in a frame and we can track the moving trace with the group of each frame. The trace of swiping right gesture is shown in the previous preliminary experiment result in Figure 3.1. For the comparison between pushing/pulling and swiping right, the numerical similarity result is shown in 5.4. Gesture 1 indicates pushing/pulling and gesture 2 indicates swiping right. Pushing/pulling is a more complex gesture comparing to swiping right. The mean value of profiling similarity of pushing/pulling is 0.059 which is higher than the mean value of profiling similarity of swiping right which is 0.03. It is normally harder for a complex gesture to build a profile. However, the similarity between different gesture is still above the mean of pushing/pulling profiling similarity.

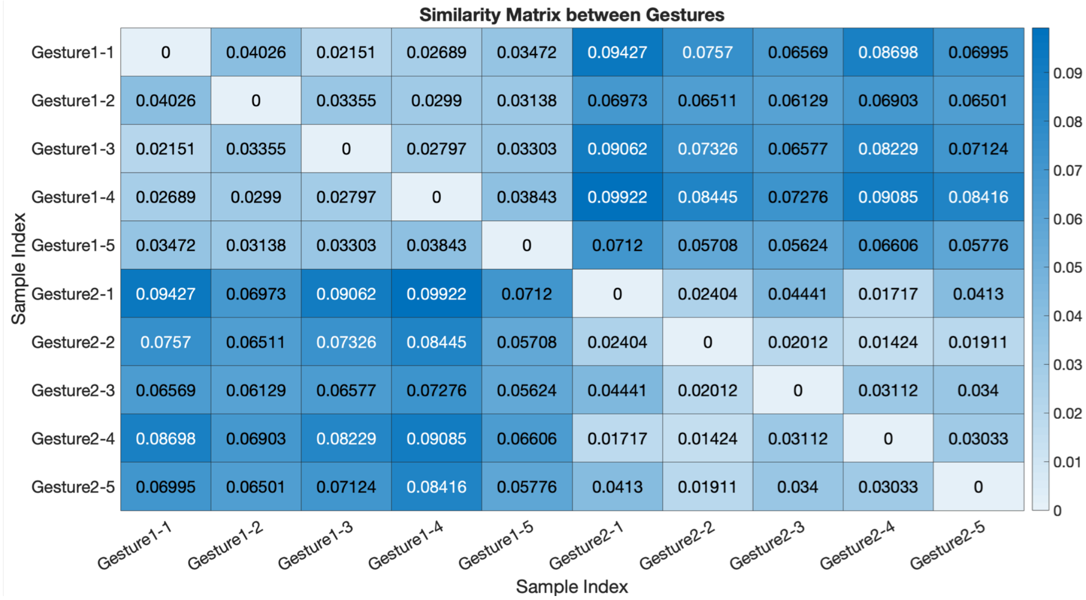


Figure 5.4: Similarity between 'push and pull' gesture and 'swipe right' gesture.

We also collect data to distinguish all three different gestures. We perform 10 times for each gesture to test the error rate. The result shows in Figure 5.5. The label 1, 2 and 3 in the figure indicates 'push and pull', 'swipe right', and 'draw a circle' separately. Label 4 indicates 'unknown gesture'. It is hard to recognize the circle as a gesture since the pattern of the signal is complicated. On the other hand, the poor profile result did not influence the recognition of other profiles. The pushing/pulling and swiping right still remains a high accuracy. For the first two gestures, we can get 96.3% true-positive accuracy and 5% false-positive accuracy. And 100% with an adjusted threshold/fault tolerance.

The circle gesture performs poor accuracy. As we learn from Figure 5.3. In theory in the 1D trace, when performing the circle gesture, the distance should remain the same. However, in the trace, we see clear trace breaks, which indicates the radar lost the reflected signal. During the breaks, the hand is actually out of the FOV.

5.4 Evaluation of Vibration-based System

By performing the different gesture 5 times by the same person, we learn the similarity between gestures. Starting from this preliminary result, we evaluated the average

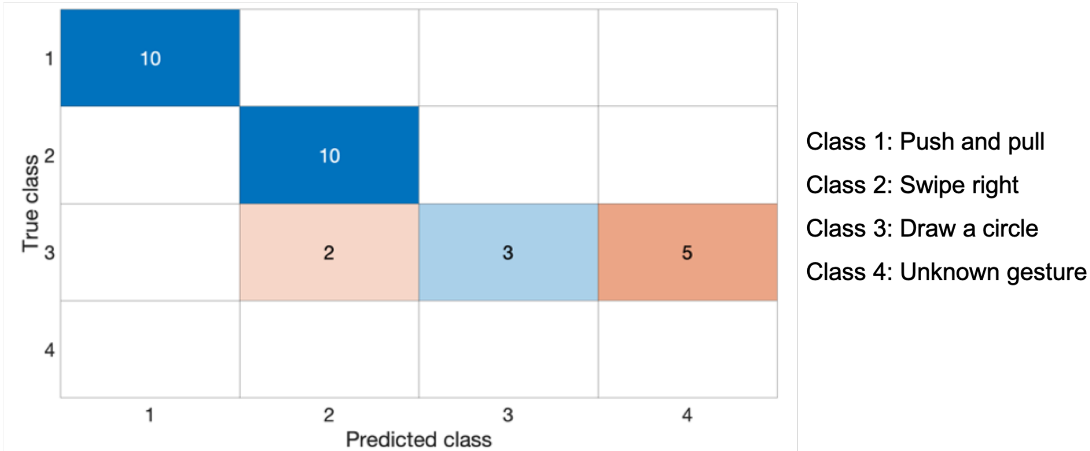


Figure 5.5: Gestures confusion matrix.

similarity of the profile of the two line gesture is as we can see in Figure 5.6. However, DTW results depend on segment accuracy. The heavy calculating cost also makes it unsuitable for real-time reorganization. Thus both EMD and DTW should be utilized to reduce error.

Figure 5.7 shows a similar trend when using DTW weighted EMD. The DTW weighted EMD evaluation method is presented in 4.3.1. There is a significant difference gap between user 1 and user 2. For 2 line gesture pattern, the mean value of the EMD among same gesture samples for user 1 and user 2 are 0.0688 and 0.0570 separately. For two line gesture, the true-positive accuracy is 93% and the false-positive accuracy is 2%. For circle and triangle gesture, since they are similar and the accuracy varies, the overall accuracy is 89% and 90% for true-positive accuracy for each gesture and 8% and 4% false-positive accuracy separately. By setting an adjusted threshold, higher accuracy is possible.

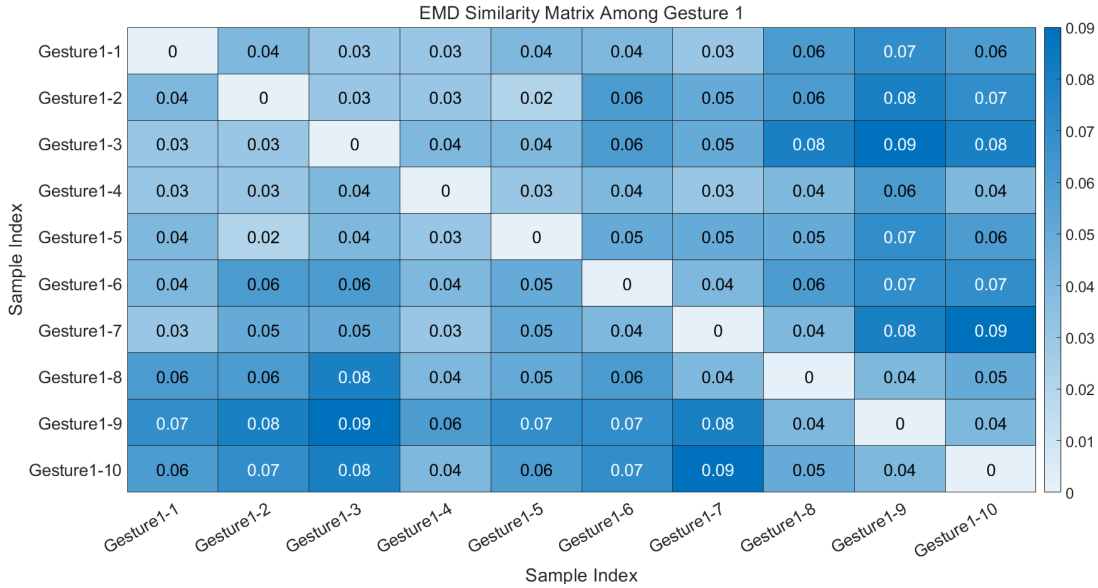


Figure 5.6: Similarity among drawing two line2 gesture samples.

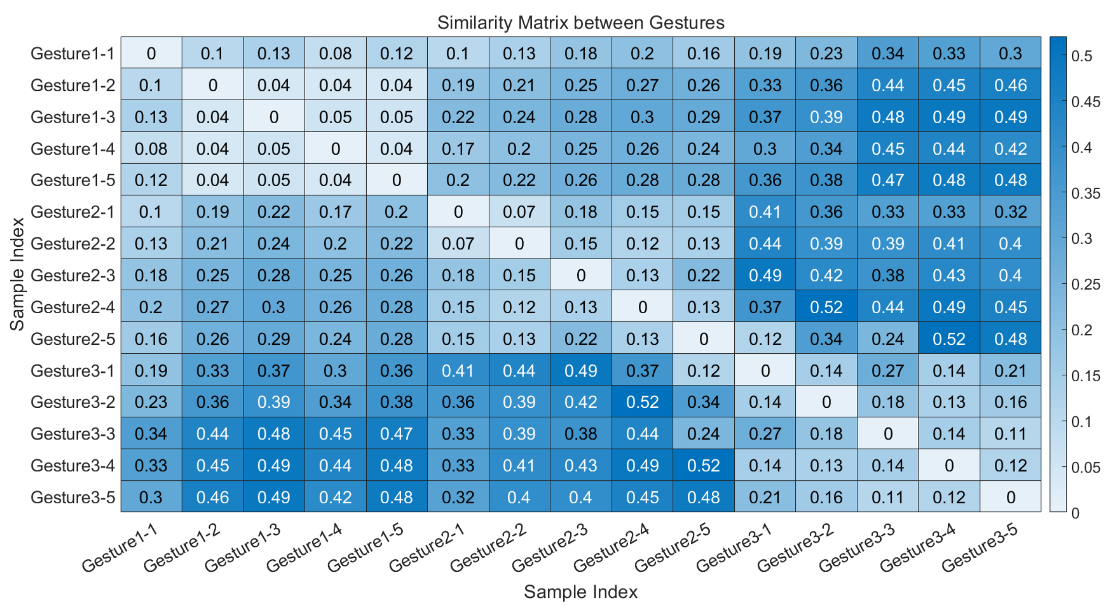


Figure 5.7: Similarity among three gestures including drawing two lines, drawing a circle and drawing a triangle.

Chapter 6

Discussion

Along with the goal-oriented experiments, we did find out some interesting facts and features that may lead to some inspiration in research. Some of these discover did accelerate and change the original plans.

6.1 The Influence Factors in Vibration-based Gesture System

The original selected signal is 17k-19kHz chirp in 0.003 seconds as explained in Section 4.2.3. The strategy for selection is designed for detecting PIN number input rather than gesture detection. Thus it performed really bad when utilizing the original chirp signal in gesture segmentation. Relativity slower sweeping speed of a chirp signal could make fewer peak features in the spectrogram. The more peak features to the receiver side, the better the system can recover the signal and is more sensitive to the physical change of the median. Since the user's finger is unstable when drawing a gesture, the system should keep the peak features and resist the instability at the same time.

6.2 Authentication and The Sweeping Speed of Chirp Signal

For the vibration-based system, according to our previous, with sweep chirp signal in 0.004 seconds per chirp, the system has a good performance in PIN Pad authentication. Also, 16 points authentication result is also acceptable. Intuitively, shorter sweep time brings fewer peak features in the receiver plot. The number of the feature is directly related to the sensitivity to the features of input. With visible sweeping time, such as 1 second, any tiny fluctuation by mistake will be recognized as an important feature. As a result, we choose 0.003 as the speed of sweeping which takes approximately four frequency peaks on the plot and also have an ability to distinguish different gestures.

6.3 Limitations

There are some limitations to this project. First, more fine-grained hand gesture should also be tested to see the potential of mmWave. Then, there is an alternative method to against the limited FOV in vertical. We can simply put the mmWave 10 degrees heads down to get a 3-D like detection field. However, we did not make an experiment in that situation. These works can also be included in future works. Besides, the potential possibility of mmWave is still under development.

Chapter 7

Conclusion

For the physical touched gestures recognition, the recognition data is from the signal peaks fluctuation, therefore the gestures are not sensitive to the complexity of the signal. However, in the radio-signal-based gesture recognition detection, since the mmWave radar sensor tracks the moving object by sensing and limited by the poor vertical FOV, it is hard for the complex gestures like the circle to build a profile. Although lots of experiments have already done for investigating the possibility of gesture recognition, there is still more potential abilities that the mmWave may have. Future work could be people counting or fine-grained gesture recognition. We would like to do more experiments using our setups. We need to find a quantized standard that how mmWave can do in gesture field.

References

[1] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 3, pp. 311-324, May 2007.

[2] Garg, P., Aggarwal, N., & Sofat, S. (2009). Vision based hand gesture recognition. World Academy of Science, Engineering and Technology, 49(1), 972-977.

[3] Liu, X., & Fujimura, K. (2004, May). Hand gesture recognition using depth data. In Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. (pp. 529-534). IEEE.

[4] Chi Zhang, Josh Tabor, Jialiang Zhang, and Xinyu Zhang. 2015. Extending Mobile Interaction Through Near-Field Visible Light Sensing. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom '15). ACM, New York, NY, USA, 345-357.

[5] Sohrabi, F., & Yu, W. (2017). Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays. IEEE Journal on Selected Areas in Communications, 35(7), 1432-1443.

[6] Liu, J., Wang, C., Chen, Y., and Saxena, N. (2017, October). VibWrite: Towards Finger-input Authentication on Ubiquitous Surfaces via Physical Vibration. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 73-87). ACM.

[7] Liu, J., Chen, Y., Gruteser, M., and Wang, Y. (2017, June). VibSense: Sensing Touches on Ubiquitous Surfaces through Vibration. In Sensing, Communication, and Networking (SECON), 2017 14th Annual IEEE International Conference on (pp. 1-9). IEEE.

[8] Chen, W., Guan, M., Huang, Y., Wang, L., Ruby, R., Hu, W., and Wu, K. (2018, June). ViType: A Cost Efficient On-Body Typing System through Vibration. In 2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON) (pp. 1-9). IEEE.

[9] Jiang, W., Miao, C., Ma, F., Yao, S., Wang, Y., Yuan, Y., ... & Xu, W. (2018, October). Towards Environment Independent Device Free Human Activity Recognition. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (pp. 289-304). ACM.

[10] Saha, S. K., Ghasempour, Y., Haider, M. K., Siddiqui, T., De Melo, P., Somanchi, N., ... & Uvaydov, D. (2019). X60: A programmable testbed for wideband 60 ghz wlans with phased arrays. Computer Communications, 133, 77-88.

[11] A. Arbabian, S. Callender, S. Kang, M. Rangwala and A. M. Niknejad, "A 94 GHz mm-Wave-to-Baseband Pulsed-Radar Transceiver with Applications in Imaging and Gesture Recognition," in IEEE Journal of Solid-State Circuits, vol. 48, no. 4, pp. 1055-1071, April 2013.

[12] Peter A. Iannucci, Ravi Netravali, Ameesh K. Goyal, and Hari Balakrishnan. 2015. Room-Area Networks. In Proceedings of the 14th ACM Workshop on Hot Topics in Networks (HotNets-XIV). ACM, New York, NY, USA, Article 9, 7 pages.

[13] Qian Wang, Kui Ren, Man Zhou, Tao Lei, Dimitrios Koutsonikolas, and Lu Su. 2016. Messages behind the sound: real-time hidden acoustic signal capture with smartphones. In Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom '16). ACM, New York, NY, USA, 29-41.

[14] S. Berman and H. Stern, "Sensors for Gesture Recognition Systems," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 3, pp. 277-290, May 2012.

[15] C. Iovescu S. Rao "The fundamentals of millimeter wave sensors" pp. 1-8 2017 [online] Available: http://www.ti.com/lit/wp/spyy005/spyy005.pdf.

[16] Niu, Y., Li, Y., Jin, D. et al. Wireless Netw (2015) 21: 2657.

[17] Hong, W., Baek, K. H., Lee, Y., Kim, Y., & Ko, S. T. (2014). Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices. IEEE Communications Magazine, 52(9), 63-69.

[18] Musa, A., Murakami, R., Sato, T., Chaivipas, W., Okada, K., & Matsuzawa, A. (2011). A low phase noise quadrature injection locked frequency synthesizer for mm-wave applications. IEEE Journal of Solid-State Circuits, 46(11), 2635-2649.

[19] Sakran, F., Neve-Oz, Y., Ron, A., Golosovsky, M., Davidov, D., & Frenkel, A. (2008). Absorbing frequency-selective-surface for the mm-wave range. IEEE Transactions on Antennas and Propagation, 56(8), 2649-2655.

[20] AWR1642 single-chip 76-GHz to 81-GHz automotive radar sensor evaluation module, from http://www.ti.com/tool/awr1642boost