Demo: Robust Contactless Gesture Recognition Using Commodity WiFi

Shujie Ren[1,2], Huaibin Wang[1,2], Bin Li [*,4], Liangyi Gong[1,2], Hao Yang[3], Chaocan Xiang[4], and Bo Li[5]

[1]Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology, Tianjin, China,300384.

[2]School of Software and BNRist, Tsinghua University, Beijing, China, 100084; with the School of Computer Science and Engineering, Tianjin University of Technology, China,300384.

[3]School of Information Science and Technology, Yancheng Teachers University, Jiangsu China,224002.

[4]the Army Logistics University of PLA, Chongqing, China, 401311.

[5]School of Mechanical Engineering,Chongqing University of Technology, Chongqing, China, 400050.

Abstract

Contactless gesture recognition is an interactive technique in ubiquitous computer and linguistics that analyzes, judges and integrates human gestures through mathematical algorithms. The traditional contactless gesture recognition in complex indoor environment suffers numerous challenges. In this paper, a robust contactless gesture recognition using commodity WiFi equipment is presented. We de-noised the collected signal in a sliding window, and extract the eigenvalues of channel phase information. Then the contactless gesture recognition is realized based on navie Bayes technique.

Keywords

Fresnel Zone Model; Channel State Information; Gesture Recognition

1 Introduction

In recent years, with the rapid development of ubiquitous computer, users can interact with the smart devices in a contactless way. Traditional interaction requires a physical touch of a user. Compared with it, gesture based interaction can provide a more convenient and natural way for users to interact with devices. In new applications, they use gestures to interact provides users with great convenience, as FIGURE 1 shows. It is possible to interact with the sensing devices when it is not convenient to physically touch the computer.

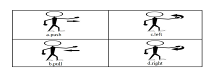


Figure 1: Gestures diagram

Traditional contactless gesture recognitions have two limitations:(1)Professional hardware equipment needs to be installed in advance [1]. (2)Gesture recognition use commercial radio frequency equipment. This method can work under non-line-of-sight conditions [3]. However these systems are realized based on machine learning. It requires a lot of training and learning in the early stage, and requires a high level of the environment in which the user is located.

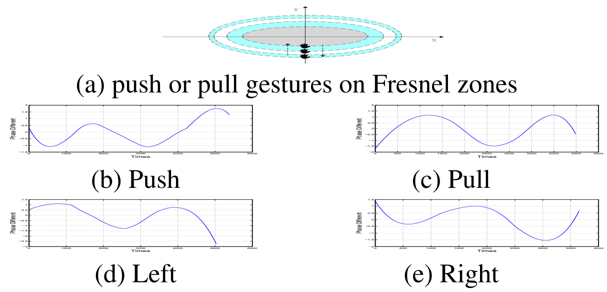
For overcoming the existing challenges in roubust contactless gesture recognition, a novel WiFi based system, FiGest. The FiGest is based on the basic theory of radio frequency sensing. It can be used to recognize the gestures made by the human body in a fine-grained and contactless way. There are two challenges in the implementation of the system. (1) Filter selection in denoising. (2) Selection of two carrier signals. The main contributions of this work are as follows:

- We propose a novel robust contactless gestures recognition system, called FiGest. The FiGest can recognize common gestures based on Fresnel zone mode and Bayes classifier.
- The FiGest system achieves complex and successive gesture recognition based on the hidden Markov model. It can be widely used in various indoor scenarios.
- We take experimental evaluation in different environments. Experimental results show that FiGest can detect human gestures in real time as the accuracy of 91%.

*Corresponding author: Bin Li, e-mail: xcc123dbs@163.com

International Conference on Embedded Wireless
Systems and Networks (EWSN) 2019
25–27 February, Beijing, China

ISBN: 978-0-9949886-3-8



(a) push or pull gestures on Fresnel zones

(b) Push (c) Pull

(d) Left (e) Right

Figure 2: Phase difference waveform of push and pull gestures

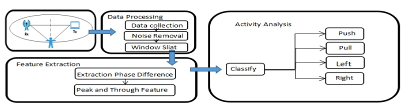


Figure 3: System model

2 OBSERVATIONS

As shown in the FIGURE2(a), if the user makes the push gesture, the palm moves closer to the center of the Fresnel zone. The peak and valley values of phase difference in Fresnel zone tend to increase. And while the user makes the pull gesture that palm moves away from the center of Fresnel, it will decrease, as the FIGURE2(b),(c) shows.

If the user makes the right gesture, the peak and valley values of phase difference in Fresnel zone tend to increase. And the right gesture is slanting to cut the Fresnel boundary. The time of phase difference waveform from peak to adjacent valley is obviously longer than push gesture. In the same way, the left gesture takes longer than the pull gesture. The experimental results are shown in the FIGURE2(d),(e).

3 SYSTEM DESCRIPTION

As the FIGURE 3 shows, the system is divided into three modules.

3.1 Data Processing

1. Data Collection:CSI data are collected at receivers in the form of real-time streams and are sent to a computer to process.

2. Noise Removal:The noise signal is smoothed by filtering technique [2]. There are a lot of noise interference in the CSI data collected. We use the Savitzky-Golay filter to smooth the signal to solve the above problem. It smoothes the signal without causing much problems.

3. Window Sliat:The speed of Normal people gesture is about 0.3 to 2 meters per second. In the Fresnel zone, the peak distance is slightly greater than $\lambda/2$, which is about 3cm in the 5GHz band. As a result, CSI power fluctuates about (0.3,2) / 0.03 times per second, equivalent to 10 Hz to 70Hz. So we choose 0.1 second as window size.

3.2 Feature Extraction

1. Extraction Phase Different:If the WiFi card is configured with 40 MHz bandwidth, the CSI value from adjacent OFDM subcarriers is 1.25MHz according to the 802.11n-2009 specification. We select two CSI subcarriers, and they are intered per five indicators, such as 1 and 6, 2 and 7, etc.

2. Peak and valley Feature:The selection of eigenvalues is estimated by analyzing the delay distribution. The encode the extracted eigenvalues As the TABLE 1 shows.

3.3 Activity Analysis

1. Classify:Naive Bayes technique is used to deal with the collected feature codes. Then, the gesture action is judged according to the feature coding. When the detected feature code is (1,1), the gesture is push. If the detected feature code is (0,1), the gesture is pull. And the feature code (1,0) suggests the gesture is right. The feature code (0,0) suggests the gesture is left.

2. Complex gesture recognition:Hidden Markov Model is applied to infer the next gesture. We can define the current HMM model as $\lambda = \{\pi, A, B\}$, and use Viterbi algorithm to infer the hidden state according to the observed state.

Define a probability of reaching an intermediate state as δ, an implicit state as x, an observable output as y, a state transition probability as a, and an output probability as b.

For calculating the maximum partial probability of the first state in which the t moment can be observed as $\delta(i) = \max_j(\delta_{t-1}(j)a_{ji}b_{ik_t})$

Where a_{ji} denotes the probability of transition from state j to state I the b_{ik_t} probability that state I is observed as k_t. A backward pointer *phi* is used to record the previous state that cause the maximum local probability of a state, that is

$$\varphi_t(i) = \arg\max_j(\delta_{t-1}(j)a_{ji}) \tag{1}$$

4 CONCLUSIONS AND FUTURE WORK

This article demonstrates the ability to recognize human gesture information by a ready-made WiFi device. The core technology of this paper is based on CSI, Fresnel phase analysis theory.

Our experimental results show that the FiGest system can recognize gestures in different indoor environments, and the overall recognition rate error is less than 10%. We believe that this new theory can be used in many environment-aware devices, and it can be applied to a wider range of macro micro human activities identification applications.

Table 1: Features Coding

Peak difference \ time difference	large	small
positive number	(1,1)	(1,0)
negative number	(0,1)	(0,0)

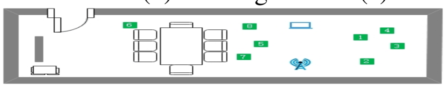
5 References

[1] G. K. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler. A real time system for robust 3d voxel reconstruction of human motions. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 714–720. IEEE, 2000.

[2] B. Sun, Q. Ma, S. Zhang, K. Liu, and Y. Liu. iself: Towards cold-start emotion labeling using transfer learning with smartphones. *ACM Transactions on Sensor Networks (TOSN)*, 13(4):30, 2017.

[3] J. Yang, Y. Ge, H. Xiong, Y. Chen, and H. Liu. Performing joint learning for passive intrusion detection in pervasive wireless environments. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

(a) Classroom (b) Meeting room (c) labouratory



(d) Position diagram

Figure 4: Experimental environment

6 Experimental

We will evaluate the performance of our contactless gesture recognition system. First, we describe experimental setup and experimental environment. Then the detailed experimental results are given, and the experimental results are analyzed comprehensively.

6.1 Experimental setup

The FiGest system requires a WiFi access point and two computers with wireless network cards. In this experiment, we use the 3G BX3H-5010 microcomputer and Intel 5300 wireless card. Install CSI tools developed by Halperin on micro PC to collect CSI samples of each received packet. Then we use MATLAB in Ubuntu14.04LTS to process CSI data in real time.

The transmitter drops some packets in a predefined mode every 10 seconds as a synchronization signal. To capture finer Fresnel phase differences between subcarriers, gestures can be identified more accurately. We chose the 5.32 GHz band to experiment with a bandwidth of 40 MHz. To capture the signal fluctuations generated by human gesture recognition better, we set the sampling frequency to 500 packets per second.

6.2 Experimental environment

We designed experiments in three test environments, as shown in FIGURE 4. (1)A classroom of 6m*6m, there are multiple sets of tables and chairs in the classroom, but it is relatively empty; (2)A laboratory of 4m*3m, which has many sets of desks, chairs and desktops; (3)A meeting room of 4m*5m, There are two sets of tables and chairs in it.

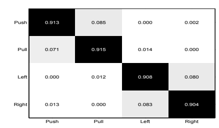
We put the devices roughly with the same position in the three experimental environments. WiFi transmitters and receiving devices are placed on walls and they are 1.4 metres above the ground. The distance between the receiver and the transmitter is 4m. The microcomputer was placed on the lectern not far away from devides.

We had four volunteers doing the experiment, and each volunteer was assigned four gestures: push, pull, left, and right at a given test point (5 test points). Each group of gestures was repeated 5 times. There will be 500 groups of experiments in each environment.

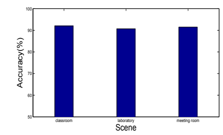
6.3 Evaluation

- 1. FIGURE5(a) shows the confusion matrix for all four gestures at all locations. Each row represents the actual gesture performed by the user and each column represents the classified gesture. Each element in the matrix corresponds to the fraction of gestures in the row that were classified as the gesture in the column.

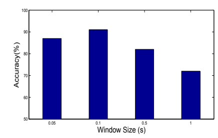
  When classifying the four gestures, the average accuracy is 91%. Through the comparison of the four gesture recognition details, it is found that the recognition accuracy is similar. Among them, the recognition accuracy of push and pull is relatively high, while the accuracy of left and right is relatively low. This accuracy shows the robustness of the FiGeat system.



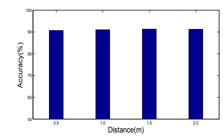
(a) Confusion matrix graph



(b) Environments



(c) Window sizes



(d) Distance

Figure 5: Detection rate in different experimental conditions

- 2. As shown in FIGURE5(b), the average rate of detection when the volunteer stayed in different experimental environments. The accuracy of gesture detection was 91.7% when the volunteer was in classroom, and 90.2% in laboratory, 91.1% in meetingroom.

  It can be seen that the accuracy of the FiGest system will not change in different experimental environments. It is stable at about 91%, so the environment will not affect the FiGest system.

- 3. From FIGURE5(c) we can observe that when we choose the sliding window size to be 0.1s, the detection rate of the system is the highest. With accuracy rate up to 91%. Therefore, the sliding window size of 0.1s is the best choice for the system.

- 4. FIGURE5(d) shows the average detection rate at different distances between the body and the center of the Fresnel zone model within the specified range. When the distance is very close, the accuracy of gesture detection is about 90.7%. About 91.3% in moderate distance, and 91% in far distance.

  Therefore, when the distance between volunteer and the center of Fresnel zone model is different, the accuracy of FiGest system will not change basically. It will be stable at about 91%, so the distance between the user and the center of Fresnel zone model will not affect the FiGest system in a specified range.