Demo: iPand: Accurate Gesture Input with Smart Acoustic Sensing on Hand

Shumin Cao*, Panlong Yang*, Xiangyang Li*, Mingshi Chen†, Peide Zhu*
* School of Computer Science and Technology, University of Science and Technology of China, Hefei, China
† Institute of Communications Engineering, Army Engineering University, Nanjing, China
Email: {shumcao, cms603421, zpeide}@gmail.com, {plyang, xiangyangli}@ustc.edu.cn

***Abstract*—Finger gesture input is emerged as an increasingly popular means of human-computer interactions. In this demo, we propose iPand, an acoustic sensing system that enables finger gesture input on the skin, which is more convenient, user-friendly and always accessible. Unlike past works, which implement gesture input with dedicated devices, our system exploits passive acoustic sensing to identify the gestures, e.g. swipe left, swipe right, pinch and spread. The intuition underlying our system is that specific gesture emits unique friction sound, which can be captured by the microphone embedded in wearable devices. We then adopt convolutional neural network to recognize the gestures. We implement and evaluate iPand using COTS smartphones and smartwatches. Results from three daily scenarios (i.e., library, lab and cafe) of 10 volunteers show that iPand can achieve the recognition accuracy of 87%, 81% and 77% respectively.**

I. Introduction

Have you ever felt helpless while thinking about using more gestures to interact more conveniently with the smartwatch or smart bracelet but limited by the screen size? Expanding wearable devices interaction is a popular problem, where operations are limited by the small screens and insufficient physical buttons. Many researches are looking to address this problem and explore many input techniques such as acoustic sensing, cameras or infrared sensing and so on[8], [1]. For example, Huawei recently patented an invention of expanding input surface, implementing a projection-vision system for wearable devices by combining infrared and ultrasound sensors.

Gesture input in human skin using acoustic transmissions has recently received much attention. The two key approaches in this domain can be broadly categorized as active and passive acoustic based systems. Some researchers adopt active mode to detect accurate finger movements by analyzing phase changes, such as LLAP[7] and FingerIO[5]. However, this solution is not customized for gestures recognition, which needs a relatively long time to monitor and consumes large amounts of energy to emit tracking signals. Moreover, for active mode, the modulated ultrasound signal will possibly lead to healthcare issues. Skinput[2] uses a dedicated bio-acoustic sensing armband to localize fingers tapping on the skin while Acoustic barcode[3] and Lamello[6] analyze a passive signal from input device with grooves. These works all use customized devices in passive acoustic sensing. WritingHacker[9] leverages embedded microphone on the phone to snoop the victim's input gesture but suffers from relatively poor accuracy.

In this demo, we design iPand, which exploits the microphone embedded in mobile devices to extract the friction sounds between fingers and the hand-back for real-time recognizing of gestures. We observe that the microphones on COTS smart devices can capture the weak skin friction while we take the hand-back as a gesture input surface. This offers an opportunity to identify finger gestures through analyzing the sound collected by the microphone. Specifically, in the first place, iPand leverages the embedded microphone in smartwatch to collect the acoustic signals as input. Then, the smartphone receives the data transmitted from the smartwatch and extracts the fragment for each gesture. As the input of the pre-trained convolutional neural network (CNN)[4] model, the spectrogram images are transformed from the sound fragments. Finally, the smartphone or smartwatch takes corresponding action according to the recognition result.

Developing such a system, however, entails the following challenges.

- **Denoising:** The skin friction can be easily affected by the ambient noise, which results the extracted effective sound fragment drowning by noise. How to detect the beginning point of the finger gesture from the received audio signals remains challenging.
- **Adaptation:** As the diversity of gesture input habits, the acoustic signals of different gestures vary with people. Prior collection of all users' gestures is impractical and computationally intensive and we need to carefully deal with the problem.

Contributions: To summarize, this demo makes the following contributions:

- It presents a prototype system that utilizes COTS mobile devices to collect audio signals which emitted by the finger gestures to recognize gesture input. It provides a skin input service for the users to interact with wearable devices without equipping heavy devices.
- High quality features are extracted and learned for the acoustic signals recorded by the smartwatch. It presents the design and implementation of iPand based on CNN, which makes high enough accuracy than other classified algorithms in the presented system.
- It demonstrates the system and evaluates it under real-world environment, i.e., lab and cafe. As a result, iPand can provide real-time gesture input of wearable devices

978-1-5386-4281-8/18/$31.00 ©2018 IEEE

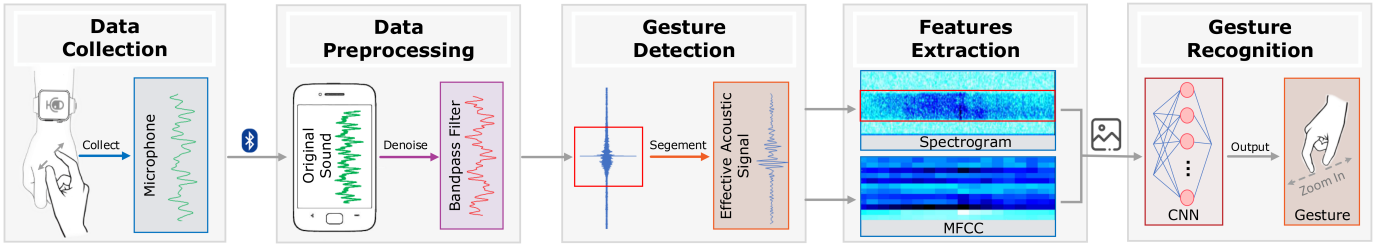


Fig. 1. iPand Working Flow

with a relative high precision. Besides, iPand is user-independent and can be implemented on wearable devices, which achieves the average recognition accuracy of 86.3% for each user. And its average recognizion accuracy in cafe achieves 80.5%.

II. Background

Works in acoustic sensing fall in two key domains. ***(i)*Active acoustic sensing**. The essential idea is to compare the d-ifferences between the original sound waves and received sound. A device emits and transmits original sound waves through modulating audio signals such as OFDM and FMCW, etc. These signals get reflected from nearby objects and can be recorded as an echo at the device's microphone. These make active acoustic sensing reach a high accuracy. ***(ii)*Passive acoustic sensing**. Compared with active acoustic, passive a-coustic sensing does not require specific devices, meaning that the modification cost will be reduced. Only the microphone embedded in the devices is needed to collect the sound waves, and then the specified analyzation will be executed. It is not difficult to find out that passive acoustic sensing consumes less energy. However, as the absence of referenced original signal, passive acoustic sensing may collect weak signals which is sensitive to ambient noise, reducing its robustness.

III. System Overview

iPand is an acoustic sensing system that recognizes finger gestures leveraging the microphone in commercial smart devices to achieve human-computer interaction. Fig. 1 shows the system architecture of iPand, mainly including five conceptual modules, namely data collection module, data preprocessing module, gesture detection module, features extraction module, and gesture recognition module.

For the first module, a smart watch and a smartphone comprise the data collection unit. The smart watch is utilized as a recorder to collect the passive acoustic signal emitted by the finger gestures. For the performance limitation of the smart watch, iPand uses the smartphone to compute the following module. The smartphone receives the acoustic signal recorded by the smart watch through Bluetooth transmission. While the system is working, the microphone embedded in the smart watch continuously listens the motion.

In the second module, iPand minimizes the interference of environmental ambient noise by applying bandpass filter to the original audio signal. Then a novel gesture detection method is used in the gesture detection module to eliminate the silent part of audio signal. In the features extraction module, we arrange gestures as separated finger gesture features by transformed the acoustic signals into spectrogram and Mel-scale frequency cepstral coefficients (MFCC) data.

In the last module, iPand provides the recognition of finger gestures by applying CNN, where the finger gesture features are transformed into visual images to feed into. Finally, the smartwatch or the smartphone calls the corresponding function of each individual application based on the output by the neural network to interact with users.

IV. Implementation

We build a prototype of iPand using HUAWEI WATCH I with Android 4.3 OS to recognize finger gestures toward hand-back. The system is evaluated in three scenarios: lab, library and cafe.

Hardware: We adopt the microphone on HUAWEI WATCH I to collect audio signals. Specifically, the microphone records a .wav audio file with the sampling rate of 44100Hz. We mainly use HUAWEI Nexus 6P smartphone (Android 7.0) to receive the acoustic signals and send the prediction of finger gesture to HUAWEI WATCH I after calculating.

Software: We adopt HUAWEI WATCH I to record sound through getting the microphone recording permission and storage authority. Then, the Android mobile platform receives the signals and sends it to an offline trained model which executes and runs on HUAWEI Nexus 6P. Note that Matlab (version R2017a 9.2.0.538062) is employed to train the neural network model on a desktop computer, which is equipped with an Intel Core i7-6700 at 3.40GHz, 16G memory and an NVIDIA Quadro P2000 graphics card. HUAWEI WATCH I receives the recognized result from mobile phone and make the corresponding action. The software is implemented using Android Studio(version 2.2.3).

Fig. 2. Experiment Setup

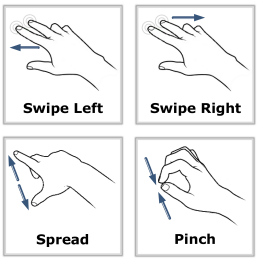
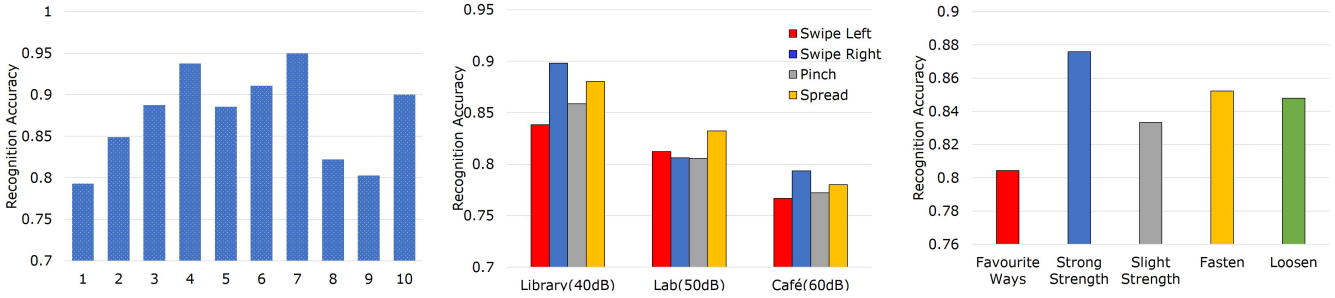


Fig. 3. The Basic Gestures We De-fined



(a) The recognition accuracy per user (b) Impact of scenarios (c) Imapct of user behaviours

Fig. 4. The Recognition Accuracy

V. Evaluation

A. Experiment Setup

We conducted experiments on HUAWEI WATCH I using its rear microphone with the watch on person's wrist as shown in Fig. 2 and HUAWEI smartphone using its calculation for finger gesture recognition. We defined 4 basic finger gestures: swipe left, swipe right, pinch (zoom out) and spread (zoom in), as shown in Fig. 3. Further, we extend those basic finger gestures to 12 with multiple fingers, i.e., swipe left and swipe right can be performed with 2 to 4 fingers while the next two are performed with 3 to 5 fingers. Indeed, data set is a necessary part of training operation for classification, which is collected from ten human users (7 males, 3 females). Each volunteer takes each finger gesture in their used way for 20 times. Fig. 4 shows the recognition accuracy of our system.

Performance of users: As the human anatomy affects the performance of iPand, we recruit ten volunteers of age 20-60 at our university. The total recognition accuracy on the known and the unseen test set is 86.3% and 70.2% respectively. Furthermore, we explore whether the sex of volunteers impact the performance. And the average recognition on the male and female test set is 85.7% and 85% respectively.

Performance of different scenarios: iPand is robust to ambient noise and achieves an average recognition accuracy of 81.99% under noise interferences. The experiments are under three normal environments, i.e., the library, lab and cafe. The tone pressure levels measured at these three environments are 40 dB, 50 dB, and 60 dB respectively. We find that iPand has slightly lower accuracy under noise interferences.

Performance of users behaviors: To analyze the performance of characteristic behaviors, we conduct three types of experiments. First, we let users take finger gestures in their favorite ways, and the detection accuracy is around 80.4%. Then, we ask the users to perform the finger gesture in a strong and a slight strength, where the accuracy achieves 87.6% and 83.3% respectively. Finally, the users are required to wear the smart watch in a fasten and loosen way, achieving the mean accuracy both around 85%. Note that although the friction sound can be easily affected by users' behaviours, the embedded microphone in smartwatch still have the ability to capture it in normal scenarios.

References

[1] K.-Y. Chen, K. Lyons, S. White, and S. Patel. utrack: 3d input using two magnetic sensors. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 237–244. ACM, 2013.

[2] C. Harrison, D. Tan, and D. Morris. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–462. ACM, 2010.

[3] C. Harrison, R. Xiao, and S. Hudson. Acoustic barcodes: passive, durable and inexpensive notched identification tags. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pages 563–568. ACM, 2012.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[5] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1515–1525. ACM, 2016.

[6] V. Savage, A. Head, B. Hartmann, D. B. Goldman, G. Mysore, and W. Li. Lamello: Passive acoustic sensing for tangible input components. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1277–1280. ACM, 2015.

[7] W. Wang, A. X. Liu, and K. Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 82–94. ACM, 2016.

[8] M. Weigel, T. Lu, G. Bailly, A. Oulasvirta, C. Majidi, and J. Steimle. Iskin: flexible, stretchable and visually customizable on-body touch sensors for mobile computing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2991–3000. ACM, 2015.

[9] T. Yu, H. Jin, and K. Nahrstedt. Writinghacker: Audio based eavesdropping of handwriting via mobile devices. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 463–473. ACM, 2016.