# Multi-Touch in the Air: Device-Free Finger Tracking and Gesture Recognition via COTS RFID

Chuyu Wang[†], Jian Liu[‡], Yingying Chen[‡], Hongbo Liu[*], Lei Xie[†], Wei Wang[†], Bingbing He[†], Sanglu Lu[†]
[†]State Key Laboratory for Novel Software Technology, Nanjing University, China
Email: {wangcyu217, hebb}@dislab.nju.edu.cn, {lxie, ww, sanglu}@nju.edu.cn
[‡] WINLAB, Rutgers University, New Brunswick, NJ, USA
Email: jianliu@winlab.rutgers.edu, yingche@scarletmail.rutgers.edu
[*]Indiana University-Purdue University, Indianapolis, IN, USA
Email: hl45@iupui.edu

***Abstract*—Recently, gesture recognition has gained considerable attention in emerging applications (e.g., AR/VR systems) to provide a better user experience for human-computer interaction. Existing solutions usually recognize the gestures based on wearable sensors or specialized signals (e.g., WiFi, acoustic and visible light), but they are either incurring high energy consumption or susceptible to the ambient environment, which prevents them from efficiently sensing the fine-grained finger movements. In this paper, we present *RF-finger*, a device-free system based on Commercial-Off-The-Shelf (COTS) RFID, which leverages a tag array on a letter-size paper to sense the fine-grained finger movements performed in front of the paper. Particularly, we focus on two kinds of sensing modes: *finger tracking* recovers the moving trace of finger writings; *multi-touch gesture recognition* identifies the multi-touch gestures involving multiple fingers. Specifically, we build a theoretical model to extract the fine-grained *reflection feature* from the raw RF-signal, which describes the finger influence on the tag array in $cm$-level resolution. For the finger tracking, we leverage K-Nearest Neighbors (KNN) to pinpoint the finger position relying on the fine-grained reflection features, and obtain a smoothed trace via Kalman filter. Additionally, we construct the reflection image of each multi-touch gesture from the reflection features by regarding the multiple fingers as a whole. Finally, we use a Convolutional Neural Network (CNN) to identify the multi-touch gestures based on the images. Extensive experiments validate that RF-finger can achieve as high as $88\%$ and $92\%$ accuracy for finger tracking and multi-touch gesture recognition, respectively.***

## I. Introduction

With the flourishing of ubiquitous sensing techniques, the human-computer interaction is undergoing a reform: the natural human gestures, e.g., *finger movements in the air*, is progressively replacing the traditional typing-based input devices such as keyboards to provide a better user experience. Such gesture-based interactions have promoted the development of both Virtual Reality (VR) and Argument Reality (AR) systems, where users could directly control the virtual objects via performing gestures in the air, e.g., writing words, manipulating the tellurion or playing the VR games. Toward this end, the gesture-based interaction can further enable the operations on the smart devices in the Internet-of-Things (IoT) environments, e.g., withdrawing the curtains, controlling the smart TVs.

Yingying Chen and Lei Xie are the co-corresponding authors.

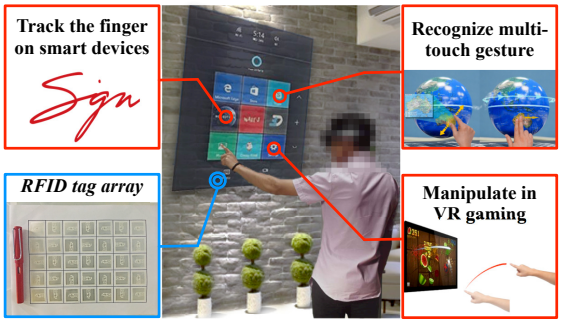


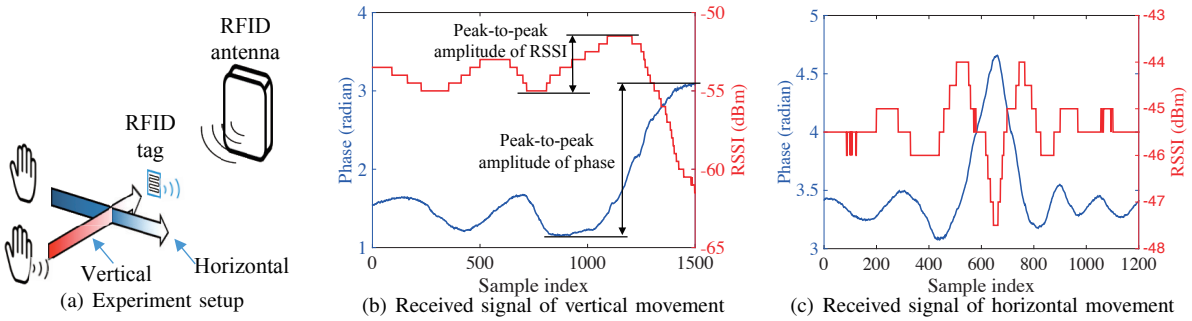

Fig. 1. Illustrations of application of RF-finger.

Therefore, accurately recognizing gestures in the air, especially fine-grained finger movements, has a great potential to provide a better user experience in emerging VR applications and IoT manipulations, which will have a market value of USD 48.56 billion by the year of 2024 [2].

Existing gesture recognition solutions can be divided into two categories: (i) *Device-based* approaches usually require the user to wear sensors, e.g., RFID tag or smartwatch, and track the motion of the sensors to recognize the gestures [15, 17]. These studies usually derive the gestures by building theoretical models to depict the signal changes received from the sensors. However, device-based approaches either suffer from the uncomfortable user experience (e.g., attaching the RFID tag on the finger) or the short life cycles due to the high energy consumption. (ii) *Device-free* approaches recognize the gestures from ambient signals through different kinds of techniques without requiring the user to wear any devices. As the most popular solutions, camera-based solutions, such as Kinect and LeapMotion, construct the body or finger structure from the video streams for accurately gesture recognition. Nevertheless, they usually involve high computation and may incur privacy concerns of the users. More recent works try to recognize the gestures based on WiFi [16], acoustic signals [18] and visible light [9]. However, these solutions are either easily affected by the environmental noise or incapable of sensing fine-grained gestures at the finger level. In this work, we are in search of a new device-free mechanism that can recognize finger-level gestures to facilitate the growing



978-1-5386-4128-6/18/$31.00 ©2018 IEEE




(a) Experiment setup (b) Received signal of vertical movement (c) Received signal of horizontal movement

Fig. 2. Preliminary study of the RF signal reflection.

VR applications and IoT operations.

The recent advances demonstrate that the emerging RFID technology not only can sense the status of objects with device-based solutions [7, 10–12, 20], but also has the potential to provide device-free sensing by leveraging the multi-path effect [4, 21]. In this work, we present *RF-finger*, a device-free system based on RFID tag array, to sense the fine-grained finger movements. Unlike previous studies, which either locate the human body in a coarse-grained manner [21] or simply detect single stroke from the hand movement for letter recognition [4], RF-finger focuses on tracking the *finger trace* and recognizing the *multi-touch gestures*, which involves a smaller tracking subject and more complicated multi-touch gestures than existing problems. As shown in Figure 1, by leveraging the tag array attached on a letter-size paper, RF-finger seeks to support different applications including writing, multi-touch operations, gaming, etc.

Specifically, we deploy only one RFID antenna behind the tag array to continuously measure the signals emitted from the tag array, and recognize the gestures based on the corresponding signal changes. In designing the RF-finger system, we need to solve three main challenging problems. *i) How to track the trajectory of the finger writings?* Since the finger usually affects several adjacent tags due to the multi-path effect, it is inaccurate to locate the finger as the position of tags. In our work, we theoretically model the impact of the moving finger on the tag array to extract the *reflection features*, and then exploit the reflection feature to pinpoint the finger with a $cm$-level resolution. *ii) How to recognize the multi-touch gesture?* Multi-touch gesture indicates the RF-signals reflected from multiple fingers are mixed together in the tag array, making it even more difficult to distinguish these fingers for gesture recognition. To address this problem, we regard the multiple fingers as a whole for recognition and then extract the reflection feature of the multiple fingers as images. We then leverage a Convolutional Neural Network (CNN) to automatically classify the corresponding gestures from the image features. *iii) How to obtain stable signal quality from the tag array?* In real RFID systems, misreading is a common phenomenon due to the dynamic environments that affects the signal quality, especially when reading multiple tags simultaneously, such as a tag array. To address this problem, we utilize a signal model to depict the mutual interference between tags, which provides recommendations on tag deployment that re-arranges the adjacent tags in a perpendicular way to reduce the interference.

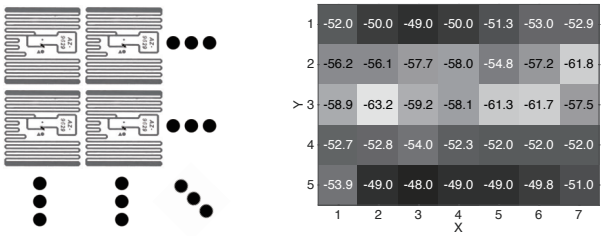
The contributions of RF-finger are summarized as follows: *i*) We design a new device-free solution based on Commercial-Off-The-Shelf (COTS) RFID for both finger tracking and multi-touch gesture recognition. To the best of our knowledge, we are the first to recognize the multi-touch gestures based on a RFID system through a device-free approach. *ii*) We build a theoretical model to depict the reflection relationship between the tag array and the fingers caused by the multi-path effect. The theoretical model provides guidelines to develop two algorithms to track the finger trajectories and recognize the multi-touch gestures. *iii*) We experimentally investigate the impact of tag array deployment on the signal quality. We analyze the mutual interference between tags via a signal model and provide recommendations on tag deployment to reduce the interference. *iv*) We implement a system prototype, *RF-finger*, for finger tacking and gesture recognition. Experiments show that RF-finger can achieve the average accuracy of $88\%$ and $92\%$ for finger tracking and gesture recognition, respectively.

## II. Preliminaries & Challenges

In order to design a system to track the fine-grained finger movements, we first conduct several preliminary studies on the impact of finger movement on the RF-signals, and the feasibility to use RFID tag array for gesture recognition. Based on the observations, we summary three challenges for designing our system.

### A. Preliminaries

**Impact of Finger Movement on RF-Signals.** RFID technique has been widely used in locating and sensing system based on the physical modalities on RF-signal [20], i.e., phase and Received Signal Strength Indicator (RSSI). Moreover, when a human moves around the tag, both the phase and RSSI are changing accordingly due to the multi-path environment variance [21]. Therefore, we first investigate the impact of finger movement on RF-signals, which is much smaller than human body. As shown in Figure 2(a), a typical finger movement can be decomposed into two basic directions: *horizontal movement* (i.e., swipe in front of the tag) and *vertical movement* (i.e., approach/departure the tag). Hence, we conduct two experiments to investigate the influence of these two finger movements. Figure 2(b) presents the signal's phase and RSSI readings when the finger is moving towards (i.e., vertically) the tag from $20cm$ away. We find that both the phase and RSSI readings change in a wavy pattern, and




(a) Universal deployment (b) RSSI distribution of universal deployment

Fig. 3. Preliminary study of the tag array deployment.

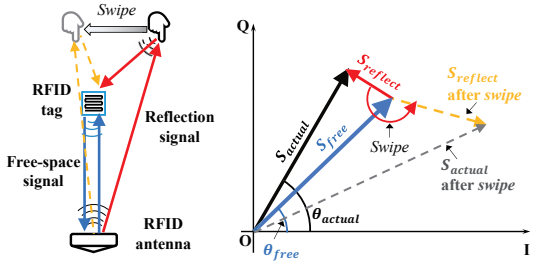



Fig. 4. Reflection model of a tag.

the *peak-to-peak amplitude* [1] increases slowly with the approaching finger. This indicates that the approaching finger leads to larger reflection effect.

Additionally, when we swipe the fingers $40cm$ along the horizontal direction as shown in Figure 2(a), we observe similar phenomenon in Figure 2(c). The peak-to-peak amplitude first increases and then decreases as the fingers swipe across the tag. The results indicate that the peak-to-peak amplitude correlates with the distance between the finger and the tag, which is later analyzed in Section III. Since the peak-to-peak amplitude indicates the linear distance between finger and tag, we can deploy a tag array to track the moving finger.

**Signal Interference within a Tag Array.** When we deploy the tag array to capture the finger movement, the density of the array is a fundamental factor on understanding the granularity of the gestures. For example, a sparse tag array can only recognize the coarse-grained strokes based on the detected tags affected by the whole hand [4]. Therefore, to recognize the finger-level gestures, we should exploit a dense tag array deployment to serve better recognition capability.

In this work, we use the small RFID tag *AZ-9629*, whose size is only $2.25cm \times 2.25cm$, so that the tags can arrange tightly. Specifically, we deploy a $5 \times 7$ tag array into $15cm \times 21cm$ rectangular space, while each tag only occupies $3cm \times 3cm$ space. A simple deployment is to universally deploy all tags with the same orientation as shown in Figure 3(a). Under this deployment, Figure 3(b) shows the RSSI distribution of 35 tags in the unit of $dBm$ when there is no finger around. We observe that the RSSI readings vary greatly across different tags due to the electromagnetic induction between the dense tags [8]. In particular, larger RSSI values are captured from the marginal tags than those from the tags in the center. Therefore, a new deployment is proposed in Section IV-B to provide stable and uniform RF-signals.

### B. Challenges

To develop the finger-level gesture tracking system under realistic settings, a number of challenges need to be addressed.

**Tracking Fine-grained Finger-writing.** Given the area size $3cm \times 3cm$ of each tag, it can only achieve a coarse-grained resolution of the finger moving trace by detecting the significantly disturbed tag. Moreover, the dense tag deployment may also lead to the detecting errors due to the mutual tag interference as shown in Figure 3(b). Therefore, we should have an in-depth understanding about the signals from the tag array during the finger movement and then develop the system to track finger trace in fine granularity.

**Recognizing Multi-touch Gestures.** Unlike the finger-writing, multi-touch gesture indicates several parts of the tag array are affected by different fingers. However, the distance between adjacent fingers is similar to the size of the tag, and the finger may affect the tags even though it is $10cm$ away as shown in Figure 2(b) and Figure 2(c). Hence, it is difficult to distinguish these fingers from the coarse-grained tag information. To address the challenge, we treat multiple fingers as a whole without distinguishing each finger and design a novel solution to recognize the multi-touch gestures from the whole of the multiple fingers.

**Reducing the Mutual Interference of Tag Array.** The received signal of the RFID tag can be easily affected by the adjacent tags, as shown in Figure 3(b). Such interference may lead to large tracking error, we thus need to find a way to obtain the uniform signal across all tags by reducing the mutual interference effect of the tag array.

## III. Modeling Finger Tracking upon a Tag Array

In this section, we introduce the reflection effect of RFID tag array with a theoretical wireless model. Particularly, we start from the reflection of a single tag, which explains the experimental results in Section III and introduces to extract the reflection feature in our system. Then, we move forward to the reflection of a tag array, which integrates the reflection features of nearby tags to facilitate the perception of the fine-grained finger movement and the multi-touch gestures.

### A. Impact of Finger Movement on a Single Tag

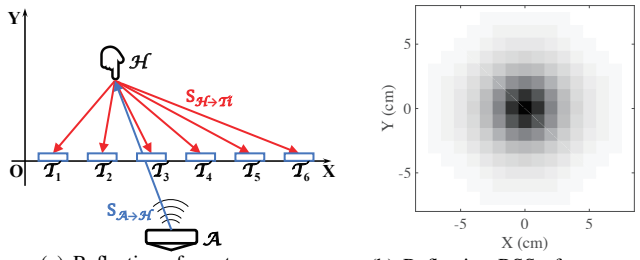
The signal received from the tag is typically represented as a stream of complex numbers. In theory, it can be expressed as:

$$S = X \cdot S_h, \tag{1}$$

where $X$ is the stream of *binary bits* modulated by the tag, and $S_h = \alpha e^{J\theta}$ is the *channel parameter* of the received signal. In RFID system, we can obtain the channel related information, including both the RSS in the unit of dBm as $\mathcal{R}$ and the phase value as $\theta$, thus the channel parameter $S_h$ can be calculated as:

$$S_h = \sqrt{\frac{10^{\frac{\mathcal{R}}{10}}}{1000}} e^{J\theta} = \sqrt{10^{\mathcal{R}/10-3}} e^{J\theta}. \tag{2}$$

Figure 4 illustrates the reflections in RFID system with a simple case, where the finger swipes across a tag. Besides the *free-space signals* directly sent from the RFID antenna, the tag would also receive the signals reflected by the moving finger. In the corresponding I-Q plane, two received signals can be




(a) Reflections from tag array (b) Reflection RSS of tag array

Fig. 5. Reflection model of a tag array.

represented as $S_{free}$ and $S_{reflect}$, respectively. Therefore, the actual signal received by the reader can be represented as:

$$S_{actual} = S_{free} + S_{reflect}. \tag{3}$$

Here, the finger movement affects $S_{reflect}$ due to the change of *reflection path*, thus both the RSS and phase of $S_{actual}$ also vary accordingly. In order to track the finger movements, we need to separate $S_{reflect}$ from the received signals to roughly describe the distance of the reflection path. Specifically, we can estimate $S_{reflect}$ by subtracting $S_{actual}$ by $S_{free}$, where $S_{free}$ can be measured without the reflection object.

### B. Impact of Finger Movement on a Tag Array

The single tag model depicts the signal change on one tag caused by the finger movement, but the tag array involving multiple tags, meaning the finger affects several adjacent tags at the same time. To better understand the reflected signals from the finger, we derive the theoretical model of tag array as follows. In Figure 5(a), we use a one-dimension tag array to illustrate the finger impact on the tag array for simplicity. Specifically, the antenna $\mathcal{A}$ interrogates six tags $\mathcal{T}_1$ to $\mathcal{T}_6$, while the finger $\mathcal{H}$ is hanging upon the tag array.

According to the single tag model, we can derive the reflection feature $S_{reflect}$ for each tag. Additionally, $S_{reflect}$ can be further divided into two parts based on the reflection path in Figure 5(a):

$$S_{reflect} = S_{\mathcal{A}\to\mathcal{H}} S_{\mathcal{H}\to\mathcal{T}_i}. \tag{4}$$

where $S_{\mathcal{A}\to\mathcal{H}}$ represents the signal from $\mathcal{A}$ to $\mathcal{H}$. $S_{\mathcal{H}\to\mathcal{T}_i}$ represents the signal reflected from $\mathcal{H}$ to $\mathcal{T}_i$, and varies based on the tag's position. In an ideal channel model [5], $S_{\mathcal{H}\to\mathcal{T}_i}$ is defined as:

$$S_{\mathcal{H}\to\mathcal{T}_i} = \frac{1}{d_{\mathcal{H}\mathcal{T}_i}^2} e^{J\theta_{\mathcal{H}\mathcal{T}_i}}, \tag{5}$$

where $d_{\mathcal{H}\mathcal{T}_i}$ is the distance between $\mathcal{H}$ and $\mathcal{T}_i$. $\theta_{\mathcal{H}\mathcal{T}_i}$ is the phase shift over the distance $d_{\mathcal{H}\mathcal{T}_i}$. Formally, the phase shift can be calculated from the wave length $\lambda$ as:

$$\theta_{\mathcal{H}\mathcal{T}_i} = 2\pi \frac{d_{\mathcal{H}\mathcal{T}_i}}{\lambda} \mod 2\pi. \tag{6}$$

For each tag $\mathcal{T}_i$, we can combine Eq. (4) and Eq. (5) to calculate the power of the $S_{reflect}$ [5] as:

$$P_{reflect} = |S_{reflect}|^2 = C * \frac{1}{d_{\mathcal{H}\mathcal{T}_i}^4}, \tag{7}$$

where $|\cdot|$ denotes the module of the complex parameter to get the power and $C = |S_{\mathcal{A}\to\mathcal{H}}|^2$ is a constant power. Therefore, the magnitude of $P_{reflect}$ is determined by $d_{\mathcal{H}\mathcal{T}_i}$, meaning the finger leads to larger reflection power to the close tags.

Given the position of $\mathcal{H}$, we can calculate the distribution of reflection power $P_{reflect}$ in the 2D space from Eq. (7).

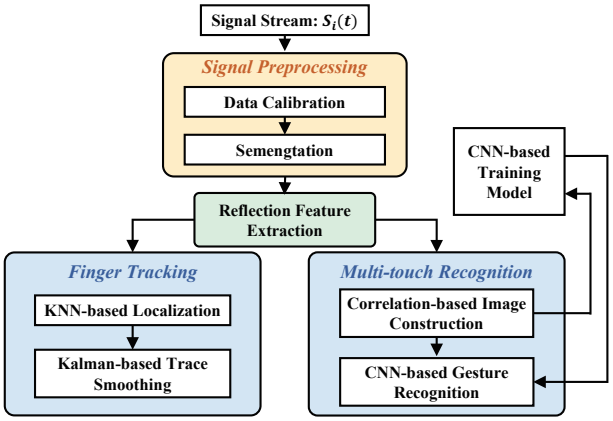



Fig. 6. System framework.

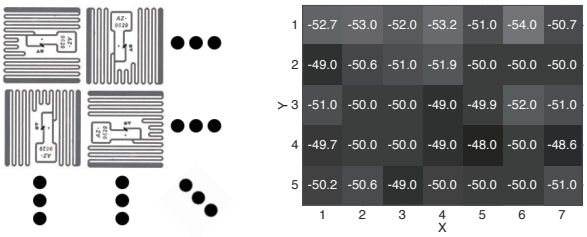
Figure 5(b) illustrates the case where the finger is at $(0, 0)$ coordinate with $3cm$ height and each $1cm \times 1cm$ grid is supposed to deploy a tag. We set $C$ to 1 for simplicity in this figure. We note that the power highly concentrates at the position of the finger. Therefore, we can use the theoretical power distribution as a pattern to estimate the finger position from the measured power distribution of the whole tag array. By computing the theoretical power distribution in a fine-resolution manner, we are able to refine the recognition resolution of the tag array with the correlation-based interpolation. In Section V-B, we will show the effectiveness of the tag array model by extracting the *reflection feature* from the reflection power distribution.

## IV. System Overview

### A. System Architecture

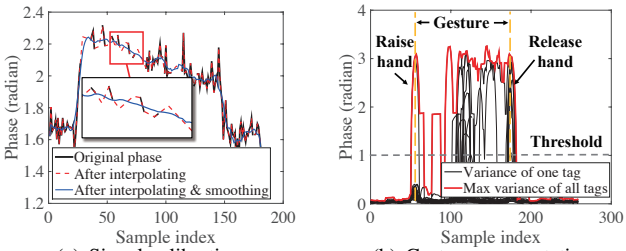
The major objective of our work is to recognize the fine-grained finger gestures via a device-free approach. Towards this end, we design an RFID-based system, *RF-finger*, which captures the signal changes on the tag array for gesture recognition. As shown in Figure 6, RF-finger consists of four main components: two core modules *Signal Pre-processing* and *Reflection Feature Extraction*, followed by two functionality modules *Finger Tracking* and *Multi-touch Recognition*. Specifically, RF-finger takes as input the time-series signal $s_i(t)$ received from each tag $i$ of the tag array, including both the RSSI and phase information. The *Signal Pre-processing* module first calibrates the measured signal by interpolating the misreading signal and smoothing the signal. Next, we divide the smoothed signals into separated gestures by analyzing the signal variance of all tags, which accurately estimates the starting and ending point of a gesture. Then, the *Reflection Feature Extraction* module extracts the reflection features of the gesture based on our reflection model in Section III.

After extracting the reflection features from RF signal, two main functionality modules are followed for finger tracking and multi-touch gesture recognition. For the finger-writing, the *Finger Tracking* module locates the finger from the reflection features in each time stamp based on the K-Nearest Neighbors (KNN) algorithm. Locations in consecutive time stamps are connected together and smoothed via Kalman filter to obtain a fine-grained trace. For the multi-touch gestures,




(a) Shuffled deployment (b) RSSI distribution of shuffled deployment

Fig. 7. Shuffled deployment of the dense tag array.




(a) Signal calibration (b) Gesture segmentation

Fig. 8. Illustration of signal preprocessing in RF-finger.

the *Multi-touch Recognition* module leverages a Convolutional Neural Network (CNN) to automatically classify each gesture from the visual features. Particularly, it constructs a 3-frame image of the gesture from the reflection features, which describes the influence range of the multiple fingers in the starting/middle/ending period of the gestures. Then we learn the neural network model from the 3-frame image for gesture classification. Finally, we can recognize the gestures by analyzing the classification scores based on CNN.

### B. Dense Tag Array Deployment

As illustrated in Section II-A, we observe that the adjacent tags in the dense tag array have great impacts on the signal quality of other tags due to the electromagnetic interference [8, 19]. The principle behind such influence is the electromagnetic interference between the two tags [8]. As a result, the parallel deployed tags will affect the nearby tags due to the mutual interference. To eliminate such mutual interference, we shuffle the directions of part tags as shown in Figure 7(a) by *making the nearby tags perpendicular to each other*. In this way, we can minimize the interference between nearby tags by making the electromagnetic interference perpendicular to each tag. As a result, we can then achieve a stable RSSI measurement across all tags, which is shown in Figure 7(b). Therefore, we adopt the *perpendicular* deployment of the tag array in our system.

## V. RF-Finger System Design

In this section, we will talk about the detailed design of the proposed RF-finger system. Specifically, we first preprocess the raw RF-signals and then extract the reflection features to depict the finger influence on the tag array. Finally, we track the finger trace and recognize the multi-touch gestures from these reflection features.

### A. Signal Preprocessing

Given the received RF-signals, which involve some inherent measurement defects such as misreading tags and noise, the data calibration process is developed to improve the reliability of the RF-signals by interpolating the misreading tags and smoothing the signal. In RFID system, the misreading tags are usually caused by the highly dynamic environment during the finger movement. Therefore, we can interpolate the misreading RF-signals from adjacent sampling rounds based on the continuous movement of the finger. Take a *phase stream* $\theta(t)$ as an example, which is time-series phase values from one tag. If there is a misreading phase $\theta(t_i)$, we calculate the interpolation value from other phase reading as:

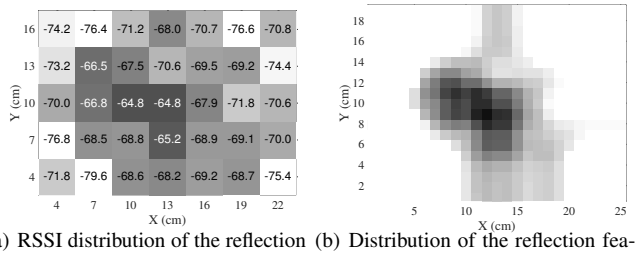
$$\hat{\theta}(t_i) = \theta(t_{i-1}) + (\theta(t_{i+1}) - \theta(t_{i-1}))\frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}, \quad (8)$$

where $\theta(t_{i-1})$ and $\theta(t_{i+1})$ are two adjacent phase readings before and after time $t_i$. After interpolation, a moving average filter is applied to smooth the signal, which further removes high-frequency noise. Figure 8(a) illustrates the effectiveness of our data calibration by comparing the phase stream before and after data calibration. The phase stream shown in the figure is from one tag in the array, when the user is performing right rotate gesture. From the enlarged figure, we could clearly see the misreadings are well interpolated. Moreover, after smoothing, the high-frequency serrated waves are removed.

To capture the signal pattern of a specific finger movement, we need to identify its starting and ending point, which correspond to the gesture people tend to raise the hand up and drop the hand down. Therefore, a segmentation method based on the detection of the calibrated RF-signals variance is developed to detect the actions of rasing/releasing hand to segment gestures. Intuitively, we observe that the signal should be stable when people drop the hand down, and the signal of some tags experiences distinct variations when the user performs the gestures. Therefore, we further leverage a sliding window to calculate the *variance stream* of each tag from the calibrated RF-signals, and the starting/ending points should have large variance values. Figure 8(b) illustrates the variance stream of all the 35 tags, which takes as input the calibrated phase stream. We find only part of the tags have large signal variance at the same time, because the finger only affects several tags close to the finger. Thus, we continuously calculate the maximum variance of each sliding window for the *maximum variance stream*. Based on the first and last peak of the max variance stream, we can detect the action of raising/releasing hand and then take the signal stream between them as the gesture signal.

### B. Reflection Feature Extraction

After signal processing, we have the segmented and noiseless signal of each individual gesture, so we first leverage the reflection model in Section III-A to derive the reflection signals $S_{reflect}$ of each tag. Then we extract the *reflection features* from the $S_{reflect}$ as the likelihood distribution inside the tag array zone, where the likelihood of each position depicts the probability that the finger locates at the position. Before defining the likelihood, we derive the reflection signal of each tag by removing the free-space signal as $S_{reflect} = S_{actual} - S_{free}$. Particularly, $S_{actual}$ is collected during the gesture period and $S_{free}$ is collected before the gesture.




(a) RSSI distribution of the reflection signal (b) Distribution of the reflection feature

Fig. 9. Illustration of the extracted reflection features.

Therefore, $S_{reflect}$ demonstrates the reflection signal caused by the finger movements. Figure 9(a) illustrates the RSSI distribution of $S_{reflect}$ when the finger is at $(10, 10)$. We find that the finger affects several tags around $(10, 10)$ and the adjacent tags even have the same RSSI value. The reason is that both the finger and the palm reflect the RF-signal, making the reflection signal mixed together.

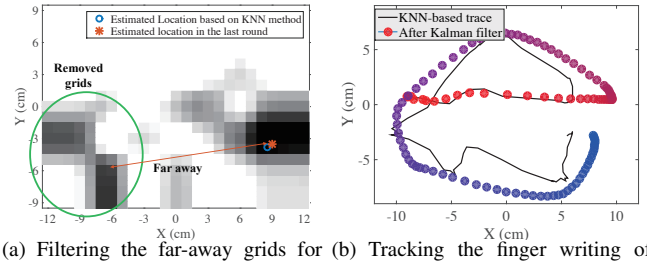
We further use the reflection model of tag array in Section III-B to extract the reflection features from $S_{reflect}$. Specifically, we partition the reflection range of our tag array into $cm$-level *grids*. Suppose the finger is right upon the grid $(x, y)$, then we can derive the theoretical reflection power $P_i$ for each tag $i$ according to Eq. (7). Given the measured RSSI values $R_i$ for tag $i$, we define the likelihood $I_{x,y}$ of grid $(x, y)$ from the Pearson correlation coefficient [13] as:

$$I_{x,y} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{P_i - \mu_P}{\sigma_P}\right)\left(\frac{R_i - \mu_R}{\sigma_R}\right), \tag{9}$$

which indicates the probability that the finger locates at $(x, y)$. $N$ is the size of the tag array. $\mu$ and $\sigma$ are the corresponding mean and standard deviation value of $P$ and $R$, respectively. All the probabilities $I_{x,y}$ form a new likelihood matrix as the *reflection feature* in our work. Figure 9(b) illustrates the reflection feature extracted from the RSSI distribution of $S_{reflect}$. We can observe a peak on the probability distribution around $(10, 10)$, representing the estimated location range of the finger.

### C. Finger Trace Tracking

Based on the extracted reflection features, we next demonstrate how to track the finger writing by locating the finger continuously at each sampling round. The basic idea is to use the *K*-Nearest-Neighbor (KNN) method to track the tendency of finger movement on the whole and leverage Kalman filter to smooth the trace for better recognition. The intuition of KNN method is that the reflection features concentrate on the position of finger as shown in Figure 9(b), so grid $I_{x,y}$ with larger value is closer to the finger. However, noisy reflection features may deviate the localization result away from the groundtruth, because traditional *KNN* method just weight averages the $K$ grids without considering the position of them. Therefore, we first filter the grids based on the fact that the finger always moves continuously, which removes the grids far away from the finger location in the last sampling round. Then, we estimate the location of the finger $F(t)$ at time $t$ from the $K$ grids with the largest likelihood as:




(a) Filtering the far-away grids for KNN-based localization (b) Tracking the finger writing of letter "e"

Fig. 10. Illustration of tracking the finger trace from reflection features.

$$F(t) = \frac{\sum_{i=1}^{K} I_i \times (x_i, y_i)}{\sum_{i=1}^{K} I_i}, \tag{10}$$

where $I_i$ is the $i$th largest grid and $(x_i, y_i)$ is the corresponding coordinate. The concatenation of the estimated locations $F(t)$ is the trace of the finger. At last, we use the Kalman filter to smooth the trace of finger-writing trace based on the fact that the finger is continuously moving for writing. Due to the space limitation, we only present the state transition function based on a velocity model as:

$$F(t) = F(t-1) + v(t-1) * \Delta t, \tag{11}$$

where $v(t)$ is the moving speed and $\Delta t$ is the sampling gap. Based on the Kalman filter, we are able to migrate the errors in KNN localization to provide a smooth trace from the velocity model. Figure 10 uses a sample case to illustrate the effectiveness of our tracking method. Figure 10(a) presents the mechanism of filtering the grids for KNN localization. By removing the grids that are far away from the estimated location in the last round, we can reduce the interference of the reading errors from some tags. Besides, Figure 10(b) illustrates the effectiveness of tracking the finger-writing of letter "e" using KNN method and Kalman filter.

### D. Multi-touch Gesture Recognition

In this work, we consider to recognize 6 multi-touch gestures as shown in Figure 13(b). When we track the finger trace, the RF signals received from the tag array are only affected by one main moving finger. In regard to the multi-touch gestures, the signals affected by different fingers are mixed together, making it hard to distinguish each finger. Intuitively, each multi-touch gesture usually has a unique motion pattern within the tag array zone. In order to effectively discriminate different multi-touch gestures, we evenly separate the gestures period into 3 *frames* of equal length, which represent the starting/middle/ending period of the gestures, respectively. For each frame, we accumulate the reflection features $I_{x,y}(t)$ of time $t$ to generate the statistic feature $\mathbb{I}_{x,y}$ as:

$$\mathbb{I}_{x,y} = \sum_{t \in T} I_{x,y}(t), \tag{12}$$

where $T$ is the duration of a frame. The statistic feature $\mathbb{I}_{x,y}$ thus constructs an *image* about the unique pattern of gesture during this frame. Then the *3-frame image* is used as the basic feature representation for gesture recognition. Figure 11 illustrates the 3-frame image of "left rotation", while the gesture is shown in Figure 13(b). We can roughly detect the rotation pattern from this 3-frame image, which reflects the physical movement of the hand.

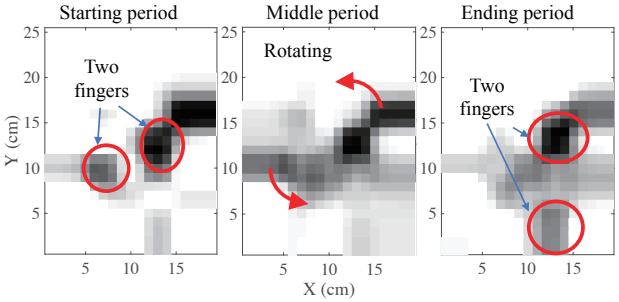



Fig. 11. 3-frame images of the "left rotation" for gesture recognition.

Given the feature representation (i.e., 3-frame image) of each multi-touch gesture, we leverage Convolutional Neural Network (CNN) to recognize the target gestures, which provides better performance in image classification mission. Figure 12 presents the structure of our CNN model, which takes as input the 3-frame image and produces the classification scores of each gesture for recognition. Particularly, our CNN model contains five hidden layers, including two convolutional (Conv) layers and two pool layers followed by a Fully Connected (FC) layer. Conv layer is the core building block, which leverages a set of learnable filters to extract the *local properties* of the image. For example, in Figure 11, the two fingers of starting period are placed horizontally, and then rotate to vertical direction at the ending period. Therefore, based on these well learned filters, CNN can automatically detect these local properties for gesture recognition, even though the gestures are not performed at the same place.

During the training process, we learn the model by collecting the 3-frame images for each gesture with manually labels. The model automatically learns the properties from the 3-frame images, which can accurately character each gesture from the view of images. In the validation process, we construct the 3-frame image from the reflection features of testing gestures and use the trained model to classify them. Finally, we recognize the testing gestures based on the CNN model.

## VI. Performance Evaluation

### A. Experimental Setup & Metrics

In order to validate the effectiveness of the proposed RF-finger system, we conduct the experiments on both the finger tracking and multi-touch gesture recognition in realistic settings. The experimental setup of RF-finger system consists of a $5\times7$ tag array of *AZ-9629* RFID tags and an ImpinJ Speedway *R420* RFID reader integrated with a *S9028PCL* directional antenna as shown in Figure 13(a). The tag array is deployed using the shuffled deployment as shown in Figure 7(a) and the average sampling rate is $13Hz$. The RFID antenna is placed $50cm$ behind the tag array to interrogate the tags, while the user performs finger gestures in front of the tag array. A LeapMotion is also deployed under the tag array to collect video stream for comparison.

The experiments are carried out in a typical indoor environment involving 10 participants in total (8 males and 2 females). Before performing each gesture, the user is required to drop

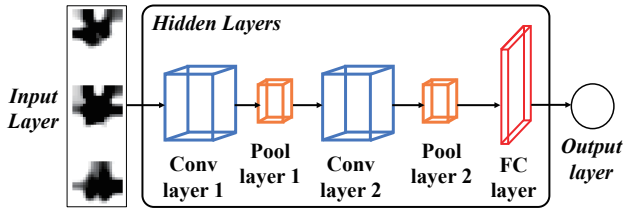
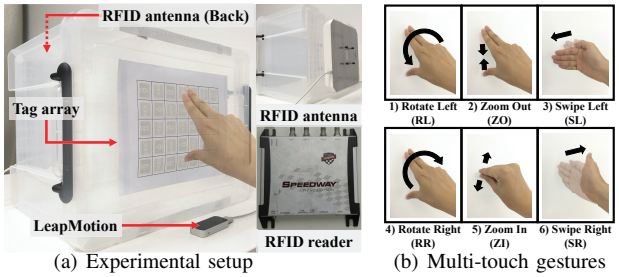



Fig. 12. Illustration of CNN structure for gesture recognition.

his/her hand down to collect the free-space signal $S_{free}$, which is used for reflection feature extraction. For the finger tracking, we ask 4 participants to write the 26 letters and 4 shapes (i.e., □,△,◯,♡) 10 times. In the KNN method, $K$ is set to 5 as default. For the multi-touch gesture recognition, we ask all the 10 participants to perform each of the 6 gestures as shown in Figure 13(b) 30 times. Particularly, 80% of the gesture related RF dataset (i.e., 1440 gestures) are used to train the CNN model, and the other 20% are used to evaluate the trained model. Only one CNN model is trained for all the users.




(a) Experimental setup (b) Multi-touch gestures

Fig. 13. Evaluation setup & multi-touch gestures.

We define three different metrics to evaluate the performance of the finger movements.

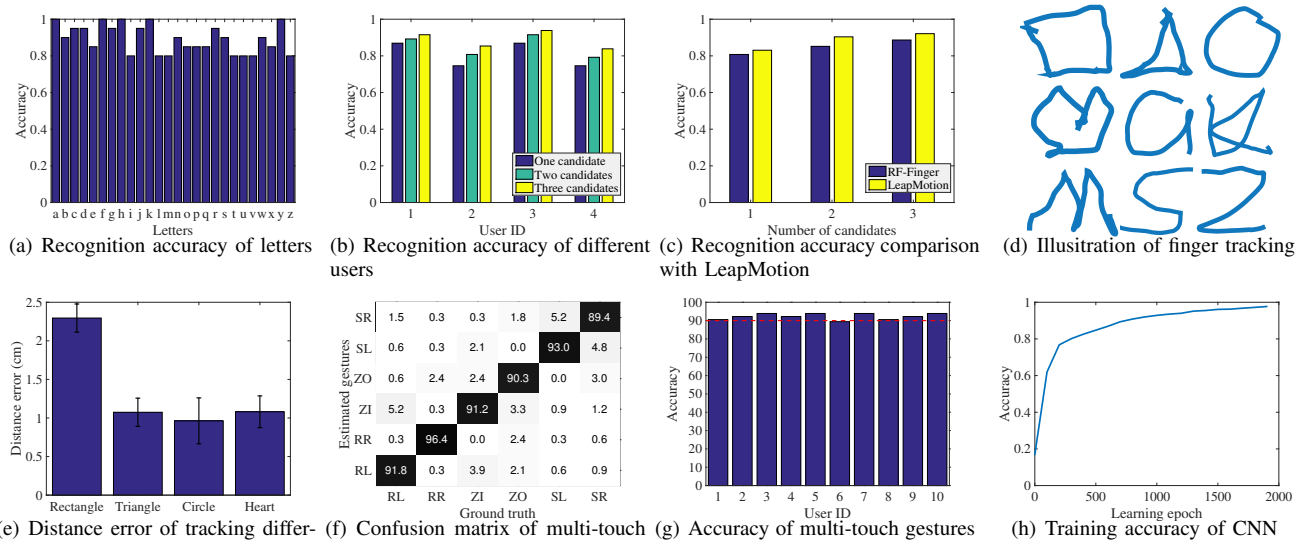
**Recognition accuracy:** For the finger writing letters, we recover the writing trace and then use LipiTk [3] to recognize the trace, which provides a candidate letter set $\mathbb{C}$ with different confidences. Given a test set $\mathbb{T}_x$ of the traces for letter $x$, the recognition accuracy of $x$ is defined as $\frac{\sum||\{x\}\cap\mathbb{C}||}{||\mathbb{T}_x||}$, where $||\cdot||$ measures the set size.

**Distance error:** For the shapes in finger tracking, the distance error is defined as $\frac{DTW(F,F_G)}{\max(L(F),L(F_G))}$, which indicates the average Dynamic Time Warping (DTW) distance between the tracking trace $F$ from RF-finger and the groundtruth shape $F_G$. $L()$ calculates the number of points in the trace.

**Classification accuracy:** For the multi-touch gestures, the classification accuracy is defined as $\frac{G_c}{G_a}$, where $G_c$ and $G_a$ are the numbers of correctly classified gestures and performed gestures, respectively. Particularly, we first train a general CNN model and then use the model to classify all the multi-touch gestures.

### B. Finger Tracking of Letters

We first evaluate the accuracy of recognizing the finger writing letters based on LipiTk. Since LipiTk produces several candidate letters with different confidences, we use the first three candidates with the larger confidence as the recognition result. As shown in Figure 14(a), RF-finger achieves an average recognition accuracy of 88%. For all the letters, the recognition accuracies are all above 80%, while 14 of 26 letters achieve more than 90% recognition accuracy. Particularly,




(a) Recognition accuracy of letters (b) Recognition accuracy of different users (c) Recognition accuracy comparison with LeapMotion (d) Illustration of finger tracking

(e) Distance error of tracking different shapes (f) Confusion matrix of multi-touch gestures (g) Accuracy of multi-touch gestures with users (h) Training accuracy of CNN

Fig. 14. Evaluation results.

letter "a", "f", "h", "k" and "y" are correctly recognized with $100\%$ due to their distinct shapes.

Moreover, we evaluate the robustness of RF-finger by comparing the recognition accuracy across different users. We also vary the size of candidate set produced by LipiTk for comparison. As shown in Figure 14(b), all the users achieve more than $75\%$ the accuracy based on the first candidate. As we increase the number of candidates to three, the accuracy increases to more than $85\%$, meaning we can correctly recognize the letters from the first three candidates with more than $85\%$ probability. Particularly, user 3 achieves the highest recognition accuracy as $94\%$, while the lowest accuracy is $84\%$ for user 4. Therefore, RF-finger is robust to recognize the letters from the finger writings of different users.

Additionally, we compare the letter recognition accuracy of RF-finger with that of LeapMotion by varying the number of candidates. As shown in Figure 14(c), it is encouraging to find that the accuracy of RF-finger is only $3\%$ to $6\%$ lower than the LeapMotion, which validates the accuracy of RF-finger. Particularly, RF-finger achieves about $89\%$ recognition accuracy when we use 3 candidates, and LeapMotion achieves $92\%$ accuracy. Therefore, *RF-finger achieves comparable accuracy for the recognition of finger writings with the video-based technique (i.e., LeapMotion).*

### C. Finger Tracking of Shapes

Next, we evaluate the accuracy of finger tracking by comparing the shapes of RF-finger with the shapes of groundtruth drawn on the paper. Particularly, we test 4 basic shapes, i.e., rectangle, triangle, circle and heart (□, △, ◯, ♡), respectively. Figure 14(d) illustrates the traces of RF-finger, which include □, △, ◯, ♡ and letter "a", "k", "m", "s", "z". All the finger traces can be easily recognized with little distortion. Besides, all the traces are written in a $15cm \times 15cm$ square, indicating RF-finger can track the trajectory with fine-grained resolution.

Furthermore, we compare the trace of RF-finger with the groundtruth on the paper. Particularly, we use DTW to map each location in the trace of RF-finger to the groundtruth on the paper. We use the average DTW distance to characterize the tracking accuracy of RF-finger as shown in Figure 14(e). We find three of the shapes have the average error as low as $1cm$, while the error for rectangles is about $2.3cm$. Through in-depth investigating, we find all the tracked rectangles are easily recognized (similar to Figure 14(d)), but they are distorted with some rotations, leading to a little bit higher tracking error than the other shapes. Overall, *RF-finger is able to accurately track the finger trace with small error.*

### D. Multi-touch Gesture Recognition

Finally, we evaluate the performance of multi-touch recognition using the CNN based classification algorithm. Figure 14(f) presents the confusion matrix of classifying the 6 gestures. We find 5 of the 6 gestures achieve over $90\%$ accuracy for gesture recognition. Even though these gestures are not performed at exactly the same position over the tag array, CNN model can still correctly classify them via the local property of the images, e.g., the relative positions of fingers in different periods. The average accuracy of the all gestures achieves as high as $92\%$, indicating RF-finger can be used to accurately recognize the multi-touch gesture.

We also show the robustness of the CNN model by comparing the recognition accuracy across different users. All the 10 users perform the 6 gestures in front of the tag array, while the users randomly choose the position over the tag array to perform. As shown in Figure 14(g), the proposed method achieves around $90\%$ accuracy for most of the users. Particularly, the lowest accuracy is as high as $89\%$, while the highest accuracy is $94\%$. Therefore, *RF-finger can accurately classify the multi-touch gestures based on the properties extracted from the CNN model.*

Besides, we also present the learning rate of our CNN model as shown in Figure 14(h). We randomly choose 1440 gestures from all the 1800 to train our CNN model. All the parameters in each CNN layer automatically update in each epoch to improve the recognition accuracy of the training dataset. Particularly, we find the CNN model achieves as high

as 90% accuracy when training dataset exceeds 800 learning epochs, while the training accuracy reaches 98% after 2000 learning epochs. The result indicates that our CNN model can converge quickly to about 90% accuracy with fewer epochs and reasonable time.

## VII. Related work

There have been active research efforts in gesture recognition, which can be broadly divided into two main categories:

**Device-based Approaches.** Previous research has shown that both the built-in motion sensors on wearable devices and the wearable RFID tags attached on human body can be utilized for gesture recognition [6, 14, 17]. For example, ArmTrack [15] proposes to track the entire arm solely relying on the smartwatch. FitCoach [6] assesses dynamic postures in workouts by recognizing the exercise gestures from wearable sensors. However, these methods suffer from the short life cycles due to high energy computation. RF-IDraw [17] and Pantomime [14] track the motion pattern of RFID tags for gesture recognition. These approaches, however, require the tags to be attached to the finger or the passive object held by the user. It will reduce the user experience with the attached RFID tags on human body, especially for the manipulation in the VR applications. Different from previous studies, we propose a device-free approach with a RFID tag array, which indicates the user can perform each gesture naturally without wearing any specialized device.

**Device-free Approaches.** As an emerging solution for gesture recognition, device-free approaches gain significant attentions in recent years. As a mature technique, camera-based approaches, e.g., Microsoft Kinect and LeapMotion, are able to extract the body or finger structure based on the computer version techniques. However, reconstructing the body or finger structure from video streams usually incurs high computation and unexpected privacy leakage. Nowadays, several studies try to recognize the gestures leveraging specialized signals, e.g., WiFi [16], acoustic signal [18] and visible light [9]. However, these solutions are either easily affected by the ambient noise or incapable of sensing fine-grained gestures. Yang *et al.* propose to locate the human body based on COTS RFID technique via a device-free approach [21], which shows the potential of device-free sensing in RFID system. More recently, RF-IPad [4], another device-free approach based on RFID, is proposed to recognize the human writing by detecting the stroke. However, we focus on tracking the finger trace, which is a finger-level and fine-grained tracking problem. Moreover, we are able to recognize the multi-touch gestures with a device-free approach based on RFID, which still remains open so far.

## VIII. Conclusion

In this paper, we propose RF-finger, a device-free system to track the finger writings and recognize the multi-touch gestures based on COTS RFID system. RF-finger provides a practical solution to precisely track the fine-grained finger trace and recognize multi-touch gestures, which facilitates the in-the-air operations in many smart applications (e.g., VR/AR and IoT systems). Our key innovations lie in modeling the reflection of the finger on the tag array and extracting the reflection features of the finger based on the model. Through the reflection features, we leverage the KNN method to track the finger trace and the CNN model to recognize the multi-touch gestures. The experimental results confirm the effectiveness of RF-finger on both finger writing tracking and multi-touch gesture recognition, which achieves over 88% and 92% accuracy.

## Acknowledgment

This work is partially supported by National Natural Science Foundation of China under Grant Nos. 61472185, 61373129, 61321491, 61502224; JiangSu Natural Science Foundation, No. BK20151390. This work is partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization. This work is partially supported by the program A for Outstanding PhD candidate of Nanjing University. This work is partially supported by the US National Science Foundation Grants CNS-1514436, CNS-1716500, CNS-1717356 and Army Office Research Grant W911NF-17-1-0467.

## References

[1] Amplitude. https://en.wikipedia.org/wiki/Amplitude.
[2] Gesture recognition market. http://www.transparencymarketresearch.com/gesture-recognition-market.html.
[3] LipiTk. http://lipitk.sourceforge.net/.
[4] H. Ding, C. Qian, J. Han, G. Wang, W. Xi, K. Zhao, and J. Zhao. Rfipad: Enabling cost-efficient and device-free in-air handwriting using passive tags. In *Proc. of IEEE ICDCS*, 2017.
[5] D. M. Dobkin. *The RF in RFID: Passive UHF RFID in Practice*. Newnes, 2007.
[6] X. Guo, J. Liu, and Y. Chen. Fitcoach: Virtual fitness coach empowered by wearable mobile devices. In *Proc. of IEEE INFOCOM*, 2017.
[7] J. Han, H. Ding, C. Qian, W. Xi, Z. Wang, Z. Jiang, L. Shangguan, and J. Zhao. A customer behavior identification system using passive tags. *IEEE/ACM Transactions on Networking*, 2016.
[8] J. Han, C. Qian, X. Wang, D. Ma, J. Zhao, W. Xi, Z. Jiang, and Z. Wang. Twins: Device-free object tracking using passive tags. *IEEE/ACM Transactions on Networking*, 2016.
[9] T. Li, C. An, Z. Tian, A. T. Campbell, and X. Zhou. Human sensing using visible light communication. In *Proc. of ACM MobiCom*, 2015.
[10] J. Liu, M. Chen, S. Chen, Q. Pan, and L. Chen. Tag-Compass: Determining the spatial direction of an object with small dimensions. In *Proc. of IEEE INFOCOM*, 2017.
[11] J. Liu, F. Zhu, Y. Wang, X. Wang, Q. Pan, and L. Chen. RF-Scanner: Shelf scanning with robot-assisted RFID systems. In *Proc. of IEEE INFOCOM*, 2017.
[12] X. Liu, X. Xie, K. Li, B. Xiao, J. Wu, H. Qi, and D. Lu. Fast Tracking the Population of Key Tags in Large-scale Anonymous RFID Systems. *IEEE/ACM Transactions on Networking*, 2017.
[13] K. Pearson. Notes on regression and inheritance in the case of two parents. In *Proc. of the Royal Society of London*, 1895.
[14] L. Shangguan, Z. Zhou, and K. Jamieson. Enabling gesture-based interactions with object. In *Proc. of ACM Mobisys*, 2017.
[15] S. Shen, H. Wang, and R. R. Choudhury. I am a smartwatch and i can track my users arm. In *Proc. of ACM MobiSys*, 2016.
[16] S. Tan and J. Yang. Wifinger: Leveraging commodity wifi for fine-grained finger gesture recognition. In *Proc. of ACM MobiHoc*, 2016.
[17] J. Wang, D. Vasisht, and D. Katabi. Rf-idraw: virtual touch screen in the air using rf signals. In *Proc. of ACM SIGCOMM*, 2015.
[18] W. Wang, A. X. Liu, and K. Sun. Device-free gesture tracking using acoustic signals. In *Proc. of ACM MobiCom*, 2016.
[19] R. K. Wangsness. *Electromagnetic Fields*. New York, NY, USA: Wiley-VCH, 1986.
[20] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu. Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices. In *Proc. of ACM MobiCom*, 2014.
[21] L. Yang, Q. Lin, X. Li, T. Liu, and Y. Liu. See through walls with cots rfid system! In *Proc. of ACM Mobicom*, 2015.