IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE IN CYBERSECURITY

Received June 17, 2019, accepted July 8, 2019, date of publication July 11, 2019, date of current version July 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928318

In-Air Gesture Interaction: Real Time Hand Posture Recognition Using Passive RFID Tags

KANG CHENG[1], NING YE[1,2], (Member, IEEE), REZA MALEKIAN[3,4], (Senior Member, IEEE), AND RUCHUAN WANG[1,2,5]

[1]College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[2]Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[3]Department of Computer Science and Media Technology, Malmö University, 20506 Malmö, Sweden
[4]Internet of Things and People Research Center, Malmö University, 20506 Malmö, Sweden
[5]Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Post and Telecommunications, Nanjing 210003, China

Corresponding authors: Ning Ye (yening@njupt.edu.cn) and Reza Malekian (reza.malekian@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572260, Grant 61872196, and Grant 61872194, and in part by the Scientific and Technological Support Project of Jiangsu Province under Grant BE2017166.

ABSTRACT In-air gesture interaction enables a natural communication between a man and a machine with its clear semantics and humane mode of operation. In this paper, we propose a real-time recognition system on multiple gestures in the air. It uses the commodity off-the-shelf (COTS) reader with three antennas to detect the radio frequency (RF) signals of the passive radio frequency identification (RFID) Tags attached to the fingers. The idea derives from the crucial insight that the sequential phase profile of the backscatter RF signals is a reliable and well-regulated indicator insinuating space-time situation of the tagged object, which presents a close interdependency with tag's movements and positions. The KL divergence is utilized to extract the dynamic gesture segment by confirming the endpoints of the data flow. To achieve the template matching and classification, we bring in the dynamic time warping (DTW) and k-nearest neighbors (KNN) for similarity scores calculation and appropriate gesture recognition. The experiment results show that the recognition rates for static and dynamic gestures can reach 85% and 90%, respectively. Moreover, it can maintain satisfying performance under different situations, such as diverse antenna-to-user distances and being hidden from view by nonconducting obstacles.

INDEX TERMS Gesture recognition, radio frequency identification (RFID), phase.

I. INTRODUCTION

As wireless sensor technology develops, the human-computer interactions (HCI) have transformed from traditional ways of keyboards and touchscreen to in-air gesture interaction. Gesture is a natural and intuitive way to express human ideas, and the user can interact with the smart devices around using pre-defined air gestures. Therefore, the research on precise and flexible gesture recognition has shown a broad and beautiful application prospect. It can be used in smart home to control appliances at home [1], which reduces the dependence on remote controllers and mobile terminals. And for sign language recognition, gestures can help deaf people or other inconvenient crowds improve their standard of living [2]. Another common application is the Remote Control Robot [3], in which robots can be controlled by gestures to perform the tasks in dangerous scenes such as fire scene, flood rescue front line, underground operations and other situations not conducive to personal safety.

All existing gesture recognition systems can be generally divided into several major categories. The first category is based on the wearable sensors, which achieves the recognition effect by utilizing the specific information returned from the sensor [4]–[7]. Especially, in [8], sensors are deployed to capture arm gesture for implicit authentication. Compared to sensor-based gesture recognition mechanism, the computer vision-based systems such as Kinect [9], PlayStation Eye [10] and application [11] are well developed after several years of trials. The fundamental techniques of vision-based multi-touch technique is the high performance image processing and template matching. Otherwise, with the increase of wireless technology and the expansion of wireless network coverage area, the RF-based gesture recognition developed gradually [12]–[14]. However, in the above mentioned

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Quan.

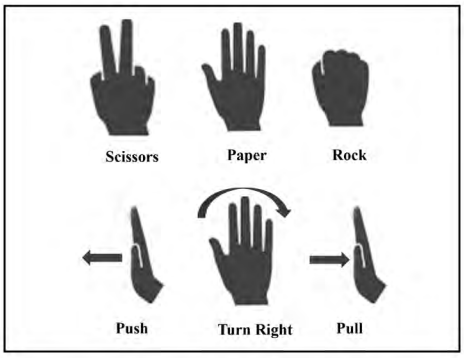


FIGURE 1. Example gestures used for the experiments in our system.

systems, there are some internal defects listed as follows: 1) Poor portability: in order to maintain reliable system performance, battery-based sensors are lack of space for portability improvement; 2) Low robustness: for vision-based recognition systems, high dependence on light conditions limits general use of them in certain environments. Because the illumination changes, mutual occlusion will bring low system stability; 3) High cost: specialized or modified hardware for wireless signal processing, which is unfavorable for our daily usage.

As the core support technology of Internet of things (IOT), RFID provides a low-cost means of information exchange to achieve effective communication between nodes. Meanwhile, the passive RFID tag which is a battery-free node that can send data to the UHF reader autonomously. RFID has been achieved success in many case, such as the application on Warehouse Management System (WMS) [15], the tracking of target object [16]–[18] and indoor location [19]–[23]. Specifically, research on RFID-based gesture recognition is also on the rise during recent years and obtains positive breakthroughs in [1], [24], [25]. However, they have not taken both static and dynamic gestures into consideration and the system performance is limited by the complexity of the environment. To make up for the deficiencies, we combine static and dynamic gesture features to establish a relational model of phase trend and relative position of tags.

In this paper, we propose a real time static and dynamic gesture recognition system for in-air interaction, which using the backscatter communication between the battery-free passive tags and the COTS RFID reader. We summarize the advantages of this system over other existing systems as follows:1) Non-line of sight (NLOS) identification, since the backscatter signal is capable of high penetration; 2) Energy-free sensing, since passive RFID tags have no independent power source and must be activated by RFID reader; 3) low cost and lightly carrying, since the passive RFID tags are light and cheap while the COTS RFID reader release the user from any hardware modifications.

The main idea is based on the observation that the movement of a tag will cause remarkable and special variations of the reflected phase trend and the tag's stationary position has a strong correlation with the real-time phase values. Thus, by analyzing and measuring the phase trend of the backscatter signals, both static and dynamic gesture recognition can be achieved naturally.

However, developing such a system is not so straightforward. The design of our system involves the following challenges:

(1) How to eliminate the influence of phase wrapping? The raw phase is a periodic function and will jump at each adjoining propagation cycle. Specially, when the phase value decreases to 0, it will jump to 2π, and then increase as usual. we term this as a phase wrapping, which may be mistakenly considered as the fluctuation of the phase, thus degrading the system performance.

We address this problem by using phase unwrapping method. The general idea is to add or subtract 2π on the phase value when the phase jump occurs.

(2) How to extract the feature template of each gesture? Since static gestures appear in the time range of stable data while dynamic gestures take place in period of data fluctuation. For static case, the key to the problem is to make use of the relative position measure between tags. And for dynamic gesture extraction, the major task is to determine the start and end times of each gesture from the continuous phase stream.

In our system, we screen out three remarkable features properly for distinguishing static gestures. Simultaneously, based on KL divergence, we comparing the discrete probability distribution function (PDF) of phase values within two adjacent sliding windows to extract dynamic gestures.

(3) How to recognize the predefined gestures? Traditional recognition system takes advantages of wearable sensors e.g. accelerators and gyroscopes to detect changes in gesture posture. However, this simple data collection method has relatively high false detection rate.

We deal with this problem by measuring the matching degree between the current gesture sequences and the feature template. Based on this, the k-nearest algorithm is used to achieve gesture classification and recognition.

We accomplish our system by using the COTS Impinj RFID reader for collecting the phase trend reflected from the passive RFID tags. And after the collected data preprocessing, we come to the gesture feature extraction and sample training. The recognition mechanism is based on the matching methods for gesture classification. We deploy the evaluation experiments upon six gestures (vary in state) shown in Figure 1, namely ''Scissors,'' ''Paper,'' ''Rock,'' ''Push,'' ''Turn Right'' and ''Pull.''

We summarize the contributions of this paper as follows:

- To the best of our knowledge, this is the first attempt to design a gesture recognition system which utilizes available phase from COTs devices to support both static and dynamic gesture recognition.
- We discover unique features differentiating each gesture type. And for the feature extraction, we leverage the observation that static gestures tend to appear within the time period of phase data stabilization while dynamic gestures occur during the periods of fluctuation.

- For superior recognition, we carry out different normalization and classification schemes on static and dynamic gestures.
- We design and implement our system with COTS RFID devices. Experiments demonstrate the system feasibility under varying deployments such as visual field occlusion and different antenna-to-user distances.

This paper is organized as follows. In section 2 the related work is presented. After that, we introduce the background and empirical studies in section 3. The main design of our system is explained in section 4. We conduct our implementation and evaluation in section 5. Section 6 gives the limitations of our system and a foreword on future work. Finally, our work is concluded.

II. RELATED WORK

The related work we focus on can be broadly divided into a few categories.

A. RFID INDOOR LOCATION

RFID-based
solutions for location has relatively large number of research results. There are several well-known systems work well in performance, such as TagOram [16], LANDMARC [19] and RF-IDraw [23]. LANDMARC introduces reference tags at a known location to correct the uncertainty of the tracking tag. The system improves the positioning accuracy of the system with fewer readers and greatly reduces the system cost. Otherwise, both RF-IDraw and TagOram utilize the phase characteristics in RFID signals to improve the robustness and accuracy of location, which is the basis of our work that the phase characteristic is a periodic function that is closely related to the communication distance and often has a strong anti-interference ability against complicated indoor environments. However, instead of obtaining the tag's track or accurate position, we focus on the detection and recognition of gestures.

B. ACTIVITY RECOGNITION

There are plenty of research on the human activity recognition. Based on its implementation, it can be divided into two categories: device-free and device-carried. Most of these research on object behaviors are using RFID or sensors. The examples of RFID-based behavior recognition such as GRfid [24] and FEMO [26], having achieved a convincing success in distinguishing different human behaviors. The GRifd system utilizes the phase changes of RFID signal to detect dynamic gestures, which is a device-free system and brings great convenience to users. However, Non-contact recognition using the RFID tags will also sacrifice the signal stability and lead to unwilling diversity of phase waveforms, which will increase tasks of gesture segmentation and matching. FEMO proposes a platform for bodybuilder to assess their performance after free-weight exercise. It works by analyzing the Doppler shifts extracted from the backscatter signals between the tags attached on the dumbbells and the RFID reader equipped with antennas, and it also use the UI module to illustrate the results of detection, recognition and exercise performance. To meet the real-time performance of HCI, unlike the FEMO triggering results after activity finished, the recognition mechanism of our design is adaptive and suitable for most interactive applications.

III. PRELIMINARIES

In this section, we introduce the backscatter communication in RFID systems and RF phase. We then take two sets of empirical experiments on the reflected backscatter signals and finally confirmed that phase is a favorable feature for real-time gesture recognition.

A. BACKSCATTER COMMUNICATION IN RFID SYSTEMS

Passive RFID systems use tags with no internal power source and instead are powered by the electromagnetic energy transmitted from an RFID reader [27]. Figure 2 provides a conceptual diagram of the radio wave propagation between an RFID reader and a passive RFID tag. The reader sends energy to an antenna which converts that energy into an RF wave that is sent into the read zone. Once the tag is read within the read zone, the RFID tag's internal antenna draws in energy from the RF waves. The energy moves from the tag's antenna to the Integrated Circuit (IC) and powers the chip which generates a signal back to the RF system. This is called backscatter. The backscatter, or change in the electromagnetic or RF wave, is detected by the reader (via the antenna), which interprets the information.

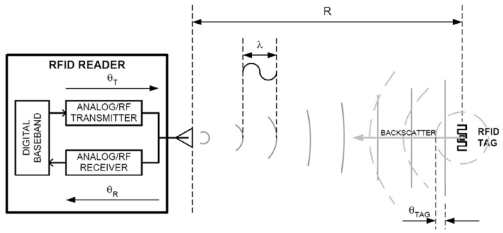


FIGURE 2. Backscatter communication.

B. RF PHASE

For an RF carrier wave at frequency f(Hz), the relation between frequency and wavelength is given by $\lambda = c/f$, where c is the speed of the EM wave in the communication medium which, in air, is equal to the speed of light($\approx 3\times10^8$). As shown in Figure 2, the total distance traversed by the signal will be $2R$. In addition to the RF phase rotation over distance, the reader's transmit circuits, the tag's reflection characteristic, and the reader's receiver circuits will all introduce some additional phase rotation θ_T, θ_{TAG} and θ_R respectively [28]. The total phase rotation can be expressed as:

$$\varphi + 2k\pi = 2\pi\frac{2R}{\lambda} + \theta_T + \theta_{TAG} + \theta_R \quad (1)$$

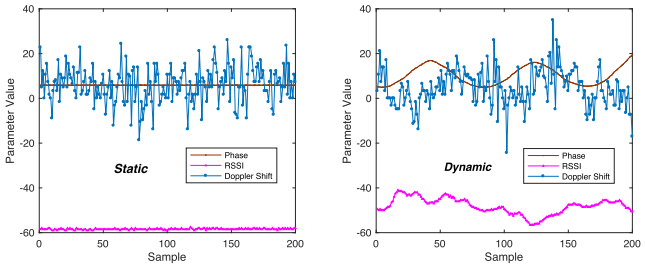


FIGURE 3. The parameter value of Phase, RSSI and Doppler shifts in both the static and dynamic cases.

where φ, an output parameter from the RFID reader, is the measured phase within $[0, 2\pi]$. The unknown parameter k is an integer guaranteeing the phase periodicity.

C. EMPIRICAL EXPERIMENTS

In this subsection, we conduct a serial of empirical experiments on the reflected backscatter signals using COTS ImpinJ reader and a passive tag. These experiments are deployed to verify the following three observations before our design.

Observation 1: The phase is more reliable and well-regulated than other output parameters from the reader, such as RSSI and Doppler shift.

In the first experiment, we focus on the three outputs from COTS Impinj reader, namely Phase, RSSI and Doppler shift. In order to select the most suitable indicator for our design, we compare the value of the three parameters in the static and dynamic condition of the tag. The result is shown in Figure 3, we can see that RSSI and Doppler shifts tend to be easily influenced and irregular while the phase is stable in static and well-regulated in dynamic.

Observation 2: Once the tag position is fixed, the raw phase obeys the Gaussian distribution and is barely influenced by the tag orientation.

As positions of wearable devices tend to be multivariate, it is crucial to consider the phase noise. Since its fixed position with respect to the antenna, we rotate the tag around its long axis with a step of 15°. Figure 4(a) shows the raw phase collected by UHF reader under different tag orientation. We can see that owing to the inherent channel noise, the raw phase values seem to be steady around same level. Then we use *Q-Q* plot method to examine whether the values obey the Gaussian distribution (shown in Figure 4(b)). The corresponding points in the *Q-Q* plot approximately lie on the line $y = x$, which means that the phase values are normally distributed.

Observation 3: The phase has a linear relation to the distance within an intra-wave and a stable periodicity at inter-wave.

This observation is the main basis for the identification model establishment. We conduct this experiment using a toy car attached by the target tag, and we control the car to move away from the antenna at a constant speed. The results is plotted in Figure 4(c). We can see that the measured cycle is about 161.8*mm* and the periodicity presented by phase waveform well matches the theory according to Equation (1), that is the phase will clearly repeat from 0 to 2π at distances separated by integer multiples of half wavelength ($\frac{\lambda}{2} \approx 162.9mm$).

Obviously, all of the above experiments verify our observations. Which indicates that backscatter communication is appropriate for gesture detection and recognition.

IV. SYSTEM DESIGN

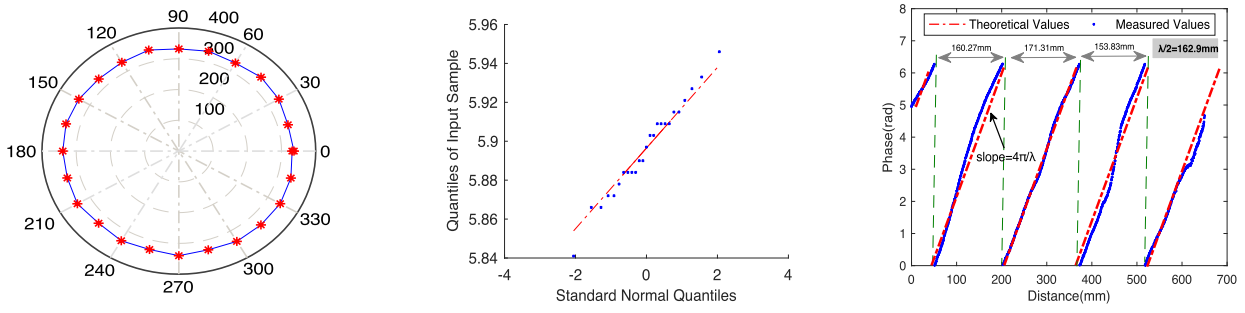
In this section, we give details on specific approaches to our design, based on three major modules: Preprocess, Gesture Feature Extraction and Training, and Gesture Recognition, as shown in Figure 5. For each module, we give an introduction to its methods and procedures in next few subsections.

A. PREPROCESSING

As mentioned before, the raw phase date we collected by the reader is a periodic function and obeys the Gaussian distribution. In order to obtain accurate phase values, both phase unwrapping and smoothing are required.

1) PHASE UNWRAPPING

Since the raw phase is a periodic function, the phase will jump at each adjoining propagation cycle, we term it as a phase jump. i.e., when the phase value decreases to 0, it will jump to 2π (shown in Figure 6). In addition, we observed that the read speed of the RFID reader is extremely fast, resulting in a



(a) (b) (c)

FIGURE 4. (a) Phase vs. Orientation (b) Q-Q Plot of Sample Data versus Standard Normal (c) Phase vs. Distance.

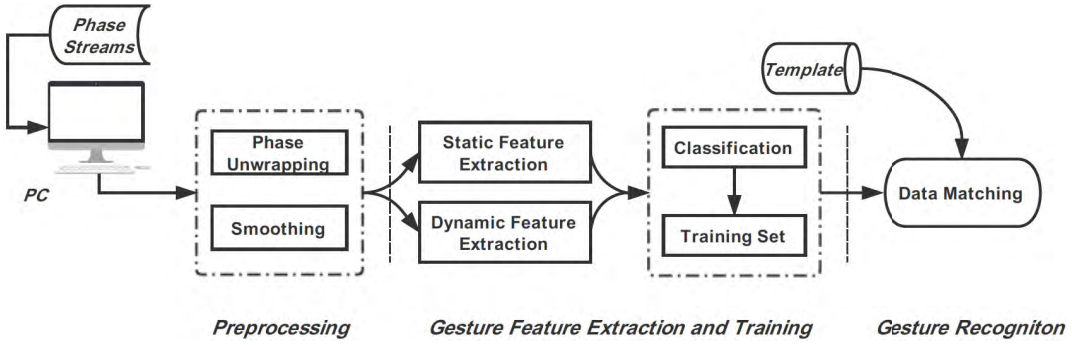


FIGURE 5. An overview of system work-flow.

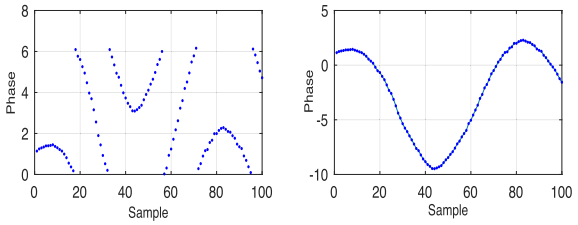


FIGURE 6. Phase trend before /after the de-periodicity algorithm is executed.

very small displacement d between two consecutive readings of a single tag. We utilize this observation and assume $d \leq \lambda/4$ ($\approx 8.1cm$). Based on the triangle constraint, we have $0 < |\Delta R| < d \leq \lambda/4$, where ΔR is the difference of antenna-to-tag distance by two consecutive readings. Further, the constraints of the two consecutive phase values can be obtained using Equation (1):

$$\left|\frac{4\pi}{\lambda}\Delta R\right| = |\Delta\varphi_{i,i-1} + 2\pi \cdot \Delta k_{i,i-1}| < \pi \tag{2}$$

Since k is an integer, we can calculate $\Delta k_{i,i-1}$ based on the constraints, i.e.,

$$\Delta k_{i,i-1} = \begin{cases} 0 & |\Delta\varphi_{i,i-1}| < \pi \\ -1 & \pi \leq \Delta\varphi_{i,i-1} \leq 2\pi \\ 1 & -2\pi \leq \Delta\varphi_{i,i-1} \leq -\pi \end{cases} \tag{3}$$

Therefore, similar to the de-periodicity method described in TagBooth system [29], we set the first reading φ_0 as a reference and start the accumulating compensation from the second sample of the phase streams, i.e., $\varphi'_0 = \varphi_0$; $\varphi'_i = \varphi_i + 2\pi \cdot \sum_{i=1}^{i} \Delta k_{i,i-1}$, $i \geq 1$, where $\Delta k_{i,i-1}$ is derived from Equation (3) using $\Delta\varphi_{i,i-1} = \varphi_i - \varphi_{i-1}$. Figure 6 gives the phase trend before/after measurement, which illustrates that the algorithm we adopted is well performed.

2) SMOOTHING

Based on Observation 2, we find that the raw phase obeys the Gaussian distribution and is barely influenced by the tag orientation when the distance from antenna to tag is unchanged. It is prerequisite, therefore, to smooth the primary phase noises. In our design, we use the moving average(MA) filter with its window size set as 10. Figure 7 shows the filtering effect in both the static and movement cases. We can see the fluctuation of filtered phase tends to fade down in the static case. In the dynamic casdase, however, the features of distribution such as maxima and minima seems to be flattened by MA filter.

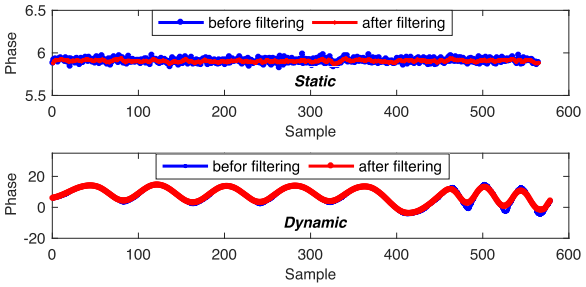


FIGURE 7. Phase measurements before/after filtering under static/dynamic cases.

B. STATIC FEATURE EXTRACTION

1) RELATIVE POSITION CALCULATION

For a single tag, phase unwrapping is an effective way to eliminate periodic disturbance so that the tag's displacement of two adjacent samples for antenna can be computed accordingly. However, static gesture feature extraction utilizes multiple tags whose positions are fixed within a short period. As a result, we can assume that these positions are on the movement trajectory of a single tag and keep the distance between two tags less than or equal to quarter-wave (i.e., $d_{tag_{i,j}} \leq \lambda/4$). Further, Based on the idea of *Phase unwrapping* in *subsection A*, the absolute value of the distance difference from any two tags to antenna $|\Delta R_{tag_{i,j}}|$ can be approximated using Equation (2) and (3):

$$|\Delta R_{tag_{i,j}}| = \begin{cases} \dfrac{|\Delta\varphi_{i,j}|}{4\pi} \times \lambda & |\Delta\varphi_{i,j}| < \pi \\ \dfrac{2\pi - |\Delta\varphi_{i,j}|}{4\pi} \times \lambda & other \end{cases} \tag{4}$$

where $\Delta\varphi_{i,j}$ is the phase difference between the reflected signals from tag_i and $tag - j$, and λ is the wavelength.

2) FEATURE SELECTION AND EXTRACTION

To determine the gesture recognition model, appropriate selection and extraction of static gesture features are consistently required. We test various relative positions of tags properly on the static gestures available in our implementation, ending up with three remarkable features denoted as x, y and z. There are some insights here:

- Antenna array in our system is placed in three dimensions (i.e., $A_1 \rightarrow$ X-axis, $A_2 \rightarrow$ Y-axis, $A_3 \rightarrow$ Z-axis), guaranteeing full range of gesture reading.
- Selecting a feature set is contingent on the predefined static gestures for reason of the diversity of tag locations.
- The sub-feature x is extracted based on the observation that gesture "Scissors" causes obvious differences in RF phase between forefinger (tag_2) and middle finger (tag_3). Similarly, gesture "Paper" gives the exclusiveness of sub-feature y.
- With regard to the sub-feature z, we conducted several sets of data tests on gesture "Rock." The results showed subtle data difference. Considering that the positional relationship between tag_1 and tag_5, which is read by antenna A_1 and A_3, appears to outstanding in gesture "Rock." We set $\left|\frac{\Delta R^{A_1}_{tag_{1,5}}}{\Delta R^{A_3}_{tag_{1,5}}}\right|$ as the weight and multiply by the sum of relative position data of each tag detected from A_2 to obtain a significant distinction sign.

As a result, the definition of x, y and z is as follows, respectively:

$$x = \left|\Delta R^{A_2}_{tag_{2,3}}\right| \tag{5}$$

$$y = \left|\Delta R^{A_3}_{tag_{1,2}}\right| \tag{6}$$

$$z = \left|\frac{\Delta R^{A_1}_{tag_{1,5}}}{\Delta R^{A_3}_{tag_{1,5}}}\right| \cdot \sum_{k=2}^{5} \left|\Delta R^{A_2}_{tag_{k,k-1}}\right| \tag{7}$$

where $\left|\Delta R^{A_i}_{tag_{k,k-1}}\right|$ represents the difference between the two distances from antenna i to tag k and $k-1$ respectively.

It should be noted that the values of x, y and z fluctuate around the specific threshold identified by each gesture accordingly. Thus, our basic idea is to infer the distinctions between multiple gestures using the features calculated by above equations.

C. DYNAMIC FEATURE EXTRACTION

Unlike the static case, dynamic gesture recognition in our system only utilizes the phase values from tag_3 read by A_1. Because dynamic gestures can incur significant phase waveform changes, which is sufficient for recognition. However, multi-tag readings will increase matching complexity and reduce recognition accuracy accordingly. Once preprocessed phase sequences are obtained, we need to extract useful dynamic features to promote our design. The challenge comes mainly from two aspects. First, the data streams update constantly and quickly, which brings biggish difficulty to determine the start and end points of each gesture. Otherwise, the patterns of gestures are not stable due to inappropriate device deployment. Therefore, there is a pressing need for accurate and efficient segmentation method.

1) SEGMENTATION-BASED EXTRACTION

In our implementation, the phase streams segmentation scheme is based on the fact that the dynamic gestures usually take place in the period of continuous phase changing, we denote the period as *waking gap*. The remaining time period that be denoted as *sleeping gap* naturally separates the dynamic gestures and is used for static gestures detection and recognition. As a result, the actual deployment of segmentation scheme to detect whether a dynamic gesture occurs is based on the KL divergence. Denote the phase stream as $S = (s_i) \in R^{1\times N}$, where N represents the sampling point. For each ω consecutive phases, we group them into a window. We categorize phase values into multiple bins, and get the PDF of phase values within each window. Give two consecutive windows ω_i and ω_j, let I and J be their PDFs. Thus, the KL divergence from J to I is defined as follows:

$$D_{KL}(I||J) = \sum_i I(i) log \frac{I(i)}{J(i)} \tag{8}$$

The KL divergence measures the distance between the distributions within two consecutive windows to determine the similarity of them. Based on this, we can conclude that $D_{KL}(I||J)$ should be small (close to zero) when ω_i and ω_j are both within the sleeping gap. Conversely, if one or two of the two consecutive windows are both within the waking gap, there will be significant differences between their PDFs which will lead to a large value of $D_{KL}(I||J)$. Hence, we can compare $D_{KL}(I||J)$ with a threshold ρ to determine whether the current window is within the sleeping gap. The dynamic features can be extracted accordingly by determining the boundaries of segments after finding all the available windows within the sleeping gap.

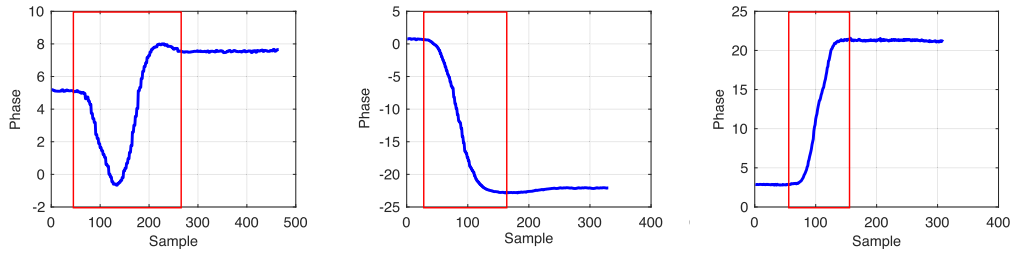
The segmentation result over the phase streams is shown in Figure 8, in which each subgraph corresponds to a dynamic gesture that is defined before and the blue curve represents preprocessed phase streams while the red rectangle marks the boundaries of the segments. We can see that all of these gestures are correctly identified.

D. GESTURE PROFILES TRAINING

Due to individual differences, the gestures performed by users are diverse. As a result, we add gesture profiles training module to reduce the impact of individual differences in our implementation.

1) STATIC GESTURES PROFILES TRAINING

As mentioned above, each static gesture can be distinguished by its corresponding feature set. To ensure the system functions are realized and errors minimized, we recruit M volunteers (vary in type of figure) to perform each predefined gesture for n times per person. Further, we define the feature



(a) "Turn Right" (b) "Push" (c) "Pull"

FIGURE 8. Segmentation result of the three dynamic gestures.

set obtained each time as $t_i = \{x_i^{(w)}, y_i^{(w)}, z_i^{(w)}\}$, where w ($w = 1, \ldots, W$) represents the corresponding gesture and i is the index of training set (i.e., $i = 1, \ldots, M \times W \times n$ and $i \geq w$). Therefore, after collecting all the sample data, the training set can be expressed as: $T = \{t_1, t_2, \ldots, t_{M\times W\times n}\} \in R^{(M\times W\times n)\times 3}$.

2) DYNAMIC GESTURES PROFILES TRAINING

Since the recognition of static gestures is equipped with bulk of sample data, similarly, in our system, dynamic gestures recognition also requires users to input feature data for n times by performing each predefined gesture. Once the gesture performs, the system start carrying on data preprocessing and feature extraction over the phase streams to achieve the corresponding segments set.

To make sure each gesture profile is exclusive and reliable, we then select some representative samples from obtained segments set using the conception of an outlier detection method [30]. The main idea is based on the notion of object's (a point or vector) proximity and this notion reflects the closeness of a vector to other considered vectors. That is, the higher the proximity's degree of a vector, the greater the likelihood that it is outlier. Suppose there are n segments, i.e., n vectors L_1,L_2,...,L_n of each gesture available, and for each segment L_i, its proximity's degree is acquired by calculating the summation of the normalized DTW distance between L_i and other $n-1$ segments. The top δ segments with lowest proximity's degree, therefore, can be determined by comparing the proximity's degree of each segment. Denote the selected top δ segments of each gesture as $p_w = \{L_{min}^1, L_{min}^2, \ldots, L_{min}^\delta\}$, where w ($w = 1, \ldots, W$) is the index of dynamic gestures to be recognized. Finally, after picking up all the representative segments of available gestures, the training set can be expressed as: $P = \{p_1, p_2, \ldots, p_w\} \in R^{W\times\delta}$.

E. RANGE NORMALIZATION

Since the feature sequence we acquire every time is a vector and our recognition system utilizes the distance-based matching method, both the individual feature whose value is far larger than others in the relevant sequence and the amplitude diversity of the time series templates (segment set) performed by different users may entail ambiguous experimental results. Therefore, it is imperative for us to normalize these vectors to the same range of values before gesture recognition. In our application, range normalization is divide into global and local, and is respectively applied to feature set for static recognition and time series segment set for dynamic recognition.

1) GLOBAL NORMALIZATION

The global normalization requires considering all the alternatives of the training set. As the feature set of static gestures is defined above as $t_i = \{x_i^{(w)}, y_i^{(w)}, z_i^{(w)}\}$, where $w = 1, \ldots, W$; $i = 1, \ldots, M \times W \times n$; $i \geq w$, we have to normalize $M \times W \times n + 1$ sequences after adding the operational feature set to be recognized. To do this, we create three one-dimensional arrays according to the three notable features $x_i^{(w)}$, $y_i^{(w)}$ and $z_i^{(w)}$ respectively. i.e., we have $X = \{x_1, \ldots, x_{M\times W\times n+1}\}$, $Y = \{y_1, \ldots, y_{M\times W\times n+1}\}$ and $Z = \{z_1, \ldots, z_{M\times W\times n+1}\}$. Thus, in our experimental deployment, the three arrays are normalized basically via:

$$newValue = \frac{oldValue - min}{max - min} \tag{9}$$

to the [0, 1] range. In the normalization procedure, the variables *min* and *max* are the smallest and largest values in the array.

2) LOCAL NORMALIZATION

In the local normalization, contrarily, each sequence template is normalized separately. Given a time series template $L = (l_1, \ldots, l_M)$, the local normalization is defined as follows using Equation 9:

$$\widehat{L} = \frac{L - min(L)}{max(L) - min(L)} \tag{10}$$

3) SHORTCOMINGS DEALING

It should be noted that linear normalization does have a drawback. The normalized results are influenced by the stability of *max* and *min*, which making the subsequent recognition ineffective. In actual implement, we handle this problem by boundary detection and replace *max* and *min* by empirical constants, i.e., each phase value crossing the boundary will be re-assigned.

F. STATIC GESTURE RECOGNITION BASED ON K-NN

As the feature set of each static gestures constitutes a feature space, which is the premise of pattern recognition. For the sake of classifying the real-time samples to be recognized diametrically, in our application, we adopt the k-Nearest Neighbors (k-NN) algorithm [31]. k-NN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions) and has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique.

It works like this: we have an existing set of example data (i.e., the training set $T = \{t_1, t_2, \ldots, t_{M\times W\times n}\}$) and labels w for all of this data (i.e., we know what class each piece of the data should fall into). And we first perform the global normalization for each piece of data to be utilized when given a new piece of data $t = x, y, z$ without a label. Otherwise, we compare that new piece of data to the existing data, every piece of existing data, using the Euclidean distance, that is:

$$d_i = \sqrt{(\hat{x} - \hat{x}_i^{(w)})^2 + (\hat{y} - \hat{y}_i^{(w)})^2 + (\hat{z} - \hat{z}_i^{(w)})^2} \tag{11}$$

where d_i is the Euclidean distance between normalized t and t_i. Note that w is actually a label used to mark different gestures, and can be calculated by combining i and the storage structure of the feature space. We then take the top k most similar pieces of data (the nearest neighbors) from our known dataset and look at their labels. Finally, the static gesture with feature set t can be recognized naturally by a majority vote of these k neighbors. In our implementation, k is empirically set as 3.

G. DYNAMIC GESTURE RECOGNITION BASED ON DTW

In designing the metric for dynamic gesture recognition, we have a challenge here: the phase trend of each segment is inconsistent, which means that segments are with different length due to the individual differences of the same gestures, fragile backscatter links and multi-path effect (shown in Figure 9). To address this, we bring in DTW [32], to measure the distance of different segments.

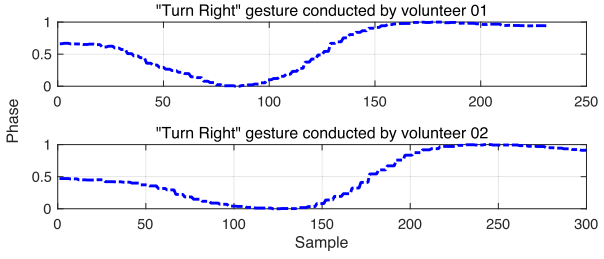


FIGURE 9. Normalized segment of "Turn Right" from different volunteers.

DTW is a time series alignment algorithm developed originally for speech recognition. It aims at aligning two sequences of feature vectors by warping the time axis iteratively until an optimal match (according to a suitable metrics) between the two sequences is found. Consider two segments of discrete time series, whose length are M and N, respectively: $A = a_1, a_2, \ldots, a_M$, $B = b_1, b_2, \ldots, b_N$, and let $d(A, B) \in R^{M\times N}$ be the pairwise distance matrix, where $d(m, n)$ is the Euclidean distance between points a_m and b_n, i.e., $d(m, n) = \sqrt{(a_m - b_m)^2}$. To align these two sequences, an (N, M)-warping path $v = (v_1, v_2, \ldots, v_k, \ldots, v_K)$ defining an alignment between A and B is obtained by assigning the element a_{m_k} of A to the element b_{n_k} of B, i.e., $v_k = (m_k, n_k)$. The total cost $c_v(A, B)$ of a warping path v between A and B is defined as:

$$c_v(A, B) = \sum_{k=1}^{K} d(v_k) \tag{12}$$

Furthermore, the optimal alignment path v^* having minimal total cost among all possible warping paths can be solved by use of dynamic programming algorithm and additional constraints. The DTW distance $DTW(A, B)$ between A and B is then defined as the total cost of v^*:

$$\begin{aligned} DTW(A, B) &= c_{v^*}(A, B) \\ &= min\{c_v(A, B) | v \text{ is an (N, M)-warping path}\} \end{aligned} \tag{13}$$

In our implementation, we use the similarity metric (DTW distance) mentioned above to ensure the attainment of dynamic gesture recognition. Given the normalized training set matrix $\widehat{P}$ with labels w and the operational segment $\widehat{Q}$ to be classified, the predefined class that minimize the DTW score between $\widehat{P}$ and $\widehat{Q}$ can be found using Equation 13, that is:

$$w = \underset{w\in\{1,\ldots,W\}}{\operatorname{argmin}} \sum_{\tau=1}^{\delta} DTW(\widehat{P}_\tau^w, \widehat{Q}) \tag{14}$$

where $\tau = 1, \ldots, \delta$ indexes the training segments of related gesture class w.

V. IMPLEMENTATION AND EVALUATION

In this section, we present the details of system implementation, and conduct extensive experiments to evaluate its performance of accuracy and effectiveness.

A. IMPLEMENTATION

1) HARDWARE

We implement our gesture recognition system using UHF COTS RFID products, including an Impinj R420 reader which is configured as the fastest RF mode (*Max Throughput*, *Dualtarget* and *Session 0*) for high read rate, three 8*dBi* directional antennas with the size of $280mm \times 280mm \times 40mm$ and Impinj E42 RFID tags. The connection between reader and PC is utilizing an Ethernet cable with the most common interface RJ 45. The whole system's operating frequency is set as 920.625 *MHz*, while in China, UHF RFID operating frequency band is $920.625 \sim 924.375$ *MHz*. The scope of tag's activity is $1 \sim 2m$ from the antenna.

2) SOFTWARE

We create a C# Windows Forms Application with the guide of Low Level Reader Protocol (LLRP) [33]. The reader operation and data collection in our implementation is based on the Octane SDK, which is the extension of LLRP protocol. Meanwhile, the system performance is evaluated in MATLAB R2017a. we runs the software on a Lenovo Thinkpad E450c PC equipped with a dual-core i5-4210U 1.7GHz CPU and 4GB RAM.

B. EXPERIMENT SETUPS

Figure 10 gives the experimental scene and information of COTS RFID devices. We set the antenna arrays to fit in the three-dimensional coordinate space and perform gestures in a certain area where is about 1.5m away from the antennas. The tag placement is shown in Figure 11, we attach five E42 RFID tags on the glove and make the tag IDs specify five fingers.

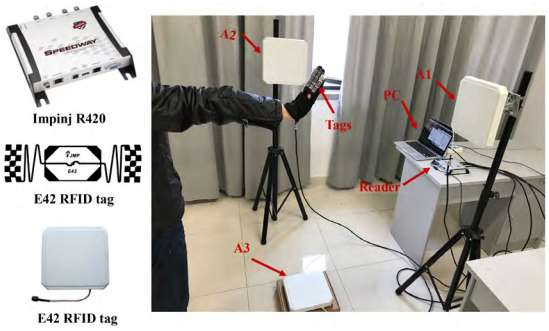


FIGURE 10. The experimental scene of our system.

FIGURE 11. The tag placement in our system.

Note that the E42 passive tag can also work well when embedded in the glove, but it is a non-metal mount tag and gives poor performance when attached on metal surfaces. In order to achieve a comprehensive evaluation on the accuracy and influencing factors of our system, we recruit 20 volunteers to perform the predefined gestures for 5 times per person and label the segmented data for further database storage.

C. THE EVALUATION OF DYNAMIC GESTURE SEGMENTATION

In terms of dynamic gesture recognition, we need to evaluate the accuracy of feature segmentation for each gesture. We use the *Offset* (from the theoretical value) of each boundary as the evaluation indicator, i.e. we have $Offset = i_{diff} + j_{diff}$, where i and j are the start sampling point and end sampling point of each segment, respectively. We first assess the impact of two parameter configurations on the *Offset*, and then compare the segmentation effect of the three dynamic gestures on the whole. Finally we give a comparison of the segmentation performance between our system and GRifd.

1) OFFSET WITH REGARD TO THE DIVERSITY OF WINDOW SIZE ω AND THRESHOLD ρ

The window size and threshold are both the pivotal parameters in boundaries detection, and they need to be comprehensively considered according to the settings of UHF reader (reading rate) and the operating characteristics of gesture (speed and magnitude). Therefore, appropriate parameter configuration is required for better segmentation effect. We vary ω from 6 to 10, and calculate the different *Offset* of the same phase trend respectively. The *Offset* represents error distance between real split points and theoretical split points, and we plot the calculated *Offset* under various threshold settings in Figure 12(a). We can see that the segmentation method reaches the lowest *Offset* value when $\rho = 0.2$ and $\omega = 8$, we thus use them as our parameter settings in the following experiments. It should be noted that the *Offset* values shown in Figure 12(a) is within [10, 30], and it is affordable considering the rapid reading speed and calculation errors.

2) OFFSET WITH REGARD TO THE DIVERSITY OF GESTURES

To examine the performance of segmentation, we repeat each gesture for 10 times and get the overall *Offset* distribution. Figure 12(b) gives the result of total segmentation effect with respect to different gestures. The operation on gesture ''Turn Right'' tends to have more offset points than that on gesture ''Push'' and ''Pull.'' In line with the theory, the median *Offset* calculated by the segments extracted from the gesture ''Push'' and ''Pull'' is almost the same. The result shows that, despite the existing offset points, the phase segment caused by dynamic gestures can also be differentiated properly.

3) OFFSET COMPARED WITH GRFID

We use the *Offset* to measure the performance of dynamic gesture segmentation algorithm between our system and GRfid. In GRfid, they apply the Modified Varri Method [34] for gesture segmentation. This method is based on the combination of a frequency measure and an amplitude measure of the signal in the relevant windows. Changes of the amplitude or frequency of the signal is calculated by Modified Varri method which is an acceptable algorithm for segmenting a signal. We perform each dynamic gesture for 10 times and

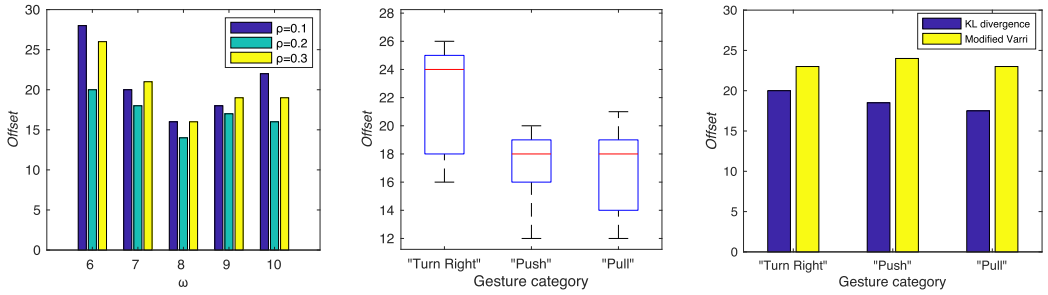


FIGURE 12. (a) *Offset* over different window sizes and threshold (b) *Offset* over three gestures (c) *Offset* over three gestures compared with Modified Varri Method in GRfid.

employ these two algorithms to obtain data segments. In order to ensure the reliability of the experiment, we use the same data stream to implement data segmentation concurrently. The window size is set as 8. Figure 12(c) gives the comparison of the average *Offset* between KL divergence and Modified Varri Method. We can see that the Modified Varri Method has the average *Offset* of 23, 24 and 23 while the KL divergence based segmentation method has the average *Offset* of 20, 18.5 and 17.5. Moreover, the phase stream in our system is more accurate due to the direct manipulation of tags, and in GRfid, they utilize the effect of external factors on fixed tag which results in visible data noise. So the KL divergence can significantly improve the accuracy of signal segmentation.

D. THE EVALUATION OF OVERALL GESTURE RECOGNITION

1) OVERALL ACCURACY OF RECOGNITION

Based on the adequate experiments that each gesture is performed 10 times by 20 volunteers, we conduct some analysis on the collected result data items to calculate the overall accuracy. Figure 13 plots the confusion matrix of gesture recognition, which shows the recognition accuracy on each gesture. As we can see, this result clearly states that our system is able to maintain satisfying performance with the average accuracy 89%. In other words, it also shows that our matching algorithms work well under the multi-mode recognition mechanism.

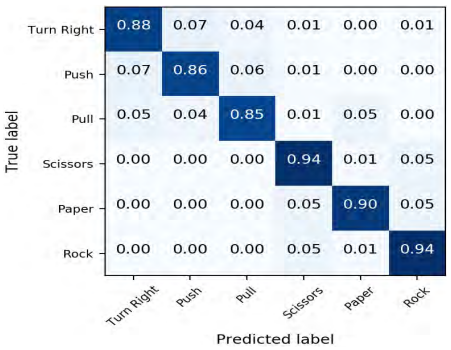


FIGURE 13. Confusion matrix of gesture recognition.

2) IMPACT OF DISTANCE BETWEEN THE USER AND ANTENNA

As is mentioned above, the scope of tag's activity is $1 \sim 2m$ from the antenna A_1, which means the user need to perform their gesture in the specific area. In this set of experiments, we vary the distance between the user and the antenna from $1m$ to $2m$ (i.e. the centers of operating areas are deployed $1m$, $1.5m$ and $2m$ away from the antennas), and explore the impact of this distance on the recognition accuracy. For each distance, we dispose different gestures and calculate the corresponding accuracy respectively. Figure 14 gives the result, indicating that the accuracy of gesture recognition in our design shows a downward trend but still maintain a high rate when the distance increases. Specifically, when the distance is $1.5m$, the accuracy tends to be relatively stable and insensitive to different gestures. However, due to the impact of fragile backscatter links and read range of the tag, recognition of long distance will show the unsatisfactory effects.

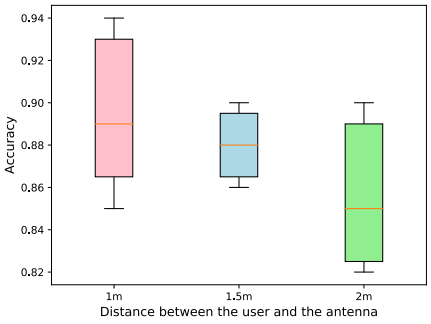


FIGURE 14. Accuracy over different antenna-to-user distance.

3) IMPACT OF THE TRAINING SET SIZE

Unlike the deep learning, which requires a large data set to improve its performance. Considering the matching method in our system is similar to machine learning based classification, the size of training set may have substantial impact on the accuracy. To examine the actual recognition effect under different training set, we carry out the corresponding experiments and select six sizes of data sets from 50 to 300.

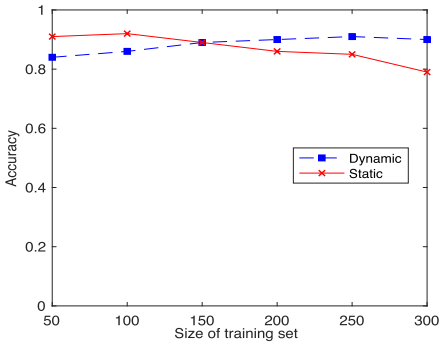


FIGURE 15. Accuracy of static&dynamic gesture recognition under different sizes of training set.

Figure 15 shows the average accuracy of gesture recognition under different sizes of training set. The results indicate that the accuracy of dynamic gesture recognition is approximately proportional to the size of the training set and tends to be steady when the scale is large enough. In contrast, the result of static gesture recognition manifests a poor performance under the big data set ($>$ 100), which means that the lack of characteristics corresponding to static gestures and the interference between multiple tags may reduce the average accuracy.

4) IMPACT OF THE STATIC OBSTRUCTIONS

Electromagnetic interference destroys the wireless channel transmission, which becomes the main reasons that affect the performance of RFID applications. However, in order to study the effect of ordinary insulators on our system, we deploy cardboard (thickness of 2mm) covering the antenna and complete the overall accuracy experiments. Due to the recognition in our system is based on the relative phase values, the overall numerical change on phase streams is allowed to a certain extent. Figure 16 gives the confusion matrix for gesture recognition when antenna is occluded. It is obviously that our system can also work well when the wireless propagation path is blocked by static insulators. The average accuracy of recognition is 89%, which is weakly affected.

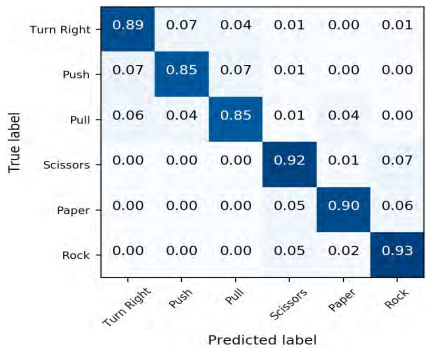


FIGURE 16. Accuracy of static&dynamic gesture recognition under different sizes of training set.

5) RECOGNITION LATENCY

It is critical for our system to achieve the real time interaction such that the user gestures can be displayed promptly. In order to evaluate this performance indicator, we measured the recognition latency, which is the duration from the time point that volunteer finishes this gesture to the time point the recognition result is shown in the Windows Forms Application. For each gesture, we randomly choose 10 repetitions from 20 volunteers and record their recognition latency. It should be noted that the size of static training set is 100 while the dynamic training set is 150. The distribution of the recognition latency for each kind of gestures is shown in Figure 17. It shows that we achieve a recognition latency of 0.27s on average for these 6 gestures. Therefore, we can conclude that our system can provide a real time gesture recognition.

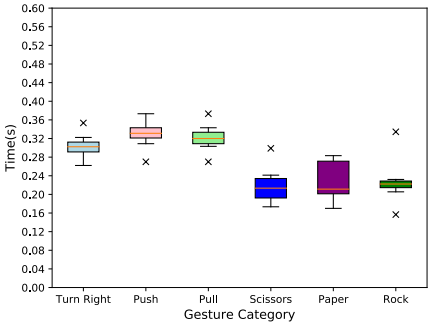


FIGURE 17. Computational latency.

VI. LIMITATIONS AND FUTURE WORK

Despite the relatively positive results we have got in previous experiments, there are still some limitations in our system.

Limitations: 1) Limited recognition distance. As is mentioned in *section V* that the distance between the user and antenna plays an important role in gesture recognition. Considering the tag diversity, the recognition distance is generally limited less than 4m. Therefore, our system can hardly achieve reliable recognition effect in long-distance HCI applications; 2) Accuracy is positively correlated with complexity. Precise dynamic gesture recognition require more training sets, and increases the number of template matching calculation, thus affecting the computational efficiency of classification. 3) Lack of corresponding applications. Applications of HCI are the most direct way to show the effectiveness of a system, which will be a future work for us. The software we developed is only a visual interface intending to manifest the feasibility of our idea.

Future Work: Just as the limitations have described above, there is still some work acquired to be done in the future. First of all, we need to improve our accuracy of recognition by more standardized deployment and deep conflict exclusion research, e.g., establishing a query table to store part of the constraint information, avoiding duplication of calculation and improving the efficiency of matching degree calculation. Otherwise, we plan to extend our system to more available

gestures by valid feature modeling. Finally, it must come to the real application, that something like mechanical arm or Unmanned Aerial Vehicle (UAV) can be controlled based on the gesture combination from our system.

VII. CONCLUSION

In this paper, we present the design, implementation and evaluation of a concurrent gesture recognition system using passive RFID tags. It uses one COTS RFID reader with 3 antennas and five passive tags attached to the five fingers. Based on the physical-layer information retrieved from the backscatter communication, we propose series of methods for data preprocess, gesture feature extraction and gesture recognition. Results from our implementation show that it could achieve high accuracy and efficiency. Our future work includes modeling of more complicated gestures and designing more robust algorithms.

REFERENCES

[1] K. Bouchard, A. Bouzouane, and B. Bouchard, "Gesture recognition in smart home using passive RFID technology," in *Proc. 7th Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, 2014, Art. no. 12.

[2] T.-H. S. Li, M.-C. Kao, and P.-H. Kuo, "Recognition system for home-service-related sign language using entropy-based *k*-means algorithm and ABC-based HMM," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 1, pp. 150–162, Jan. 2017.

[3] G.-C. Luh, Y.-H. Ma, C.-J. Yen, and H.-A. Lin, "Muscle-gesture robot hand control based on sEMG signals with wavelet transform features and neural network classifier," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2017, pp. 627–632.

[4] J. Wu, G. Pan, D. Zhang, G. Qi, and S. Li, "Gesture recognition with a 3-D accelerometer," in *Proc. Int. Conf. Ubiquitous Intell. Comput.* Berlin, Germany: Springer, 2009, pp. 25–38.

[5] A. Pandit, D. Dand, S. Mehta, S. Sabesan, and A. Daftery, "A simple wearable hand gesture recognition device using iMEMS," in *Proc. Int. Conf. Soft Comput. Pattern Recognit. (SOCPAR)*, 2009, pp. 592–597.

[6] N. Siddiqui and R. H. M. Chan, "A wearable hand gesture recognition device based on acoustic measurements at wrist," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 4443–4446.

[7] A. Ferrone, X. Jiang, L. Maiolo, A. Pecora, L. Colace, and C. Menon, "A fabric-based wearable band for hand gesture recognition based on filament strain sensors: A preliminary investigation," in *Proc. IEEE Healthcare Innov. Point-Care Technol. Conf. (HI-POCT)*, Nov. 2016, pp. 113–116.

[8] A. F. Abate, M. Nappi, and S. Ricciardi, "I-Am: Implicitly authenticate me—Person authentication on mobile devices through ear shape and arm gesture," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 3, pp. 469–481, Mar. 2017.

[9] X. Wu, C. Yang, Y. Wang, H. Li, and S. Xu, "An intelligent interactive system based on hand gesture recognition algorithm and kinect," in *Proc. 5th Int. Symp. Comput. Intell. Design*, vol. 2, 2012, pp. 294–298.

[10] (2017). *Playstation Eye*. Accessed: Oct. 19, 2017. [Online]. Available: http://www.playstation.com.cn/ps4/ps4_camera.htm

[11] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," *Commun. ACM*, vol. 54, no. 2, pp. 60–71, 2011.

[12] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 27–38.

[13] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. 11th USENIX Conf. Netw. Syst. Design Implement.*, 2013, pp. 317–329.

[14] F. Adib and D. Katabi, "See through walls with WiFi!" *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 75–86, 2013.

[15] H. Wang, S. Chen, and Y. Xie, "An RFID-based digital warehouse management system in the tobacco industry: A case study," *Int. J. Prod. Res.*, vol. 48, no. 9, pp. 2513–2548, 2010.

[16] 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. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 237–248.

[17] Z. Wang, N. Ye, R. Malekian, F. Xiao, and R. Wang, "TrackT: Accurate tracking of RFID tags with mm-level accuracy using first-order Taylor series approximation," *Ad Hoc Netw.*, vol. 53, pp. 132–144, Dec. 2016.

[18] Q. Cao, D. R. Jones, and H. Sheng, "Contained nomadic information environments: Technology, organization, and environment influences on adoption of hospital RFID patient tracking," *Inf. Manage.*, vol. 51, no. 2, pp. 225–239, 2014.

[19] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," *Wireless Netw.*, vol. 10, no. 6, pp. 701–710, 2004.

[20] D. Zhang, L. T. Yang, M. Chen, S. Zhao, M. Guo, and Y. Zhang, "Real-time locating systems using active RFID for Internet of Things," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1226–1235, Sep. 2016.

[21] A. Montaser and O. Moselhi, "RFID indoor location identification for construction projects," *Automat. Construct.*, vol. 39, no. 4, pp. 167–179, 2014.

[22] C.-H. Huang, L.-H. Lee, C. C. Ho, L.-L. Wu, and Z.-H. Lai, "Real-time RFID indoor positioning system based on Kalman-filter drift removal and Heron-bilateration location estimation," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 3, pp. 728–739, Mar. 2015.

[23] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: Virtual touch screen in the air using RF signals," in *Proc. ACM Conf. SIGCOMM*, 2014, pp. 235–246.

[24] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "GRfid: A device-free RFID-based gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 381–393, Feb. 2017.

[25] P. Asadzadeh, L. Kulik, and E. Tanin, "Gesture recognition using RFID technology," *Pers. Ubiquitous Comput.*, vol. 16, no. 3, pp. 225–234, 2012.

[26] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao, "FEMO: A platform for free-weight exercise monitoring with RFIDs," in *Proc. ACM Conf. Embedded Netw. Sensor Syst.*, 2015, pp. 141–154.

[27] D. M. Dobkin, *The RF in RFID: UHF RFID in Practice*. Boston, MA, USA: Newnes, 2012.

[28] *Impinj, Speedway Revolution Reader Application Note: Low Level User Data Support*. Accessed: Mar. 1, 2019. [Online]. Available: https://support.Impinj.com

[29] T. Liu, L. Yang, X.-Y. Li, H. Huang, and Y. Liu, " TagBooth: Deep shopping data acquisition powered by RFID tags," in *Proc. IEEE Conf. Comput. Commun.*, Apr./May 2015, pp. 1670–1678.

[30] A. Dik, K. Jebari, A. Bouroumi, and A. Ettouhami, "Similarity- based approach for outlier detection," *Int. J. Comput. Sci.*, vol. 11, no. 5, pp. 1–5, 2014.

[31] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[32] J. Zhao, Z. Xi, and L. Itti, "metricDTW: Local distance metric learning in dynamic time warping," 2016, *arXiv:1606.03628*. [Online]. Available: https://arxiv.org/abs/1606.03628

[33] *Low Level Reader Protocol*. Accessed: Mar. 1, 2019. [Online]. Available: https://www.gs1.org/standards/epc-rfid/llrp/1-0-1

[34] H. Azami, K. Mohammadi, and B. Bozorgtabar, "An improved signal segmentation using moving average and Savitzky-Golay filter," *J. Signal Inf. Process.*, vol. 3, no. 1, p. 39, 2012.

KANG CHENG received the B.S. degree from the School of Electronic Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China, in 2016, where he is currently pursuing the master's degree. His research interest includes radio frequency identification (RFID) and fancy human interaction.

NING YE received the B.S. degree in computer science from Nanjing University, in 1994, the M.S. degree from the School of Computer and Engineering, Southeast University, in 2004, and the Ph.D. degree from the Institute of Computer Science, Nanjing University of Posts and Telecommunications, in 2009, where she is currently a Professor. In 2010, she was a Visiting Scholar and a Research Assistant with the Department of Computer Science, University of Victoria, Canada. Her research interests include wireless networks and the Internet of Things.

REZA MALEKIAN is currently with the Department of Computer Science and Media Technology, Malmö University, Sweden. His current research focuses on advanced sensor networks and the Internet of Things. He is a member of the IEEE Signal Processing Society Chapters Committee. He is registered as a Chartered Engineer with the Engineering Council UK and a Fellow of the British Computer Society. He is an Associate Editor of the IEEE INTERNET OF THINGS JOURNAL and the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and an Editor of *IET Networks*.

RUCHUAN WANG was born in Hefei, China. He was a Researcher on graphic processing with the University of Bremen and on program design theory with Ludwig-Maximilians-Universität München, from 1984 to 1992. He has been a Professor and a Supervisor of Ph.D. candidates with the Nanjing University of Posts and Telecommunications, since 1992. His major research interests include wireless sensor networks and information security.