GRfid: A Device-Free RFID-Based Gesture Recognition System

Yongpan Zou, *Student Member, IEEE*, Jiang Xiao, *Member, IEEE*, Jinsong Han, *Member, IEEE*, Kaishun Wu, *Member, IEEE*, Yun Li, and Lionel M. Ni, *Fellow, IEEE*

Abstract—Gesture recognition has emerged recently as a promising application in our daily lives. Owing to low cost, prevalent availability, and structural simplicity, RFID shall become a popular technology for gesture recognition. However, the performance of existing RFID-based gesture recognition systems is constrained by unfavorable intrusiveness to users, requiring users to attach tags on their bodies. To overcome this, we propose GRfid, a novel device-free gesture recognition system based on phase information output by COTS RFID devices. Our work stems from the key insight that the RFID phase information is capable of capturing the spatial features of various gestures with low-cost commodity hardware. In GRfid, after data are collected by hardware, we process the data by a sequence of functional blocks, namely data preprocessing, gesture detection, profiles training, and gesture recognition, all of which are well-designed to achieve high performance in gesture recognition. We have implemented GRfid with a commercial RFID reader and multiple tags, and conducted extensive experiments in different scenarios to evaluate its performance. The results demonstrate that GRfid can achieve an average recognition accuracy of 96.5 and 92.8 percent in the identical-position and diverse-positions scenario, respectively. Moreover, experiment results show that GRfid is robust against environmental interference and tag orientations.

Index Terms—Gesture Recognition, COTS RFID, phase

1 Introduction

Gesture recognition has gained substantial attention recently for reshaping the Human-Machine Interface (HMI). It enables convenient interaction between human and devices via various kinds of body gestures, and has played an increasingly significant role in our daily lives. For example, in a smart home, gesture-based HMI provides device-free access to remote control of various buildings subsystems such as heating, ventilation, air-conditioning and lighting. Another typical application of gesture-based HMI are those interactive devices for mobile gaming, in which users play games by simply performing gestures in front of devices and enjoy extrication from control handles.

Existing gesture recognition systems can be categorized by their supporting technologies. A few prominent examples include computer vision-based systems such as Xbox [1], Leap [2], PlayStation Eye [3], PointGrab [4] and [5], [6], sensor-based systems [7], [8], [9], [10], [11], [12], RF-based systems [13], [14], [15], etc. However, they need to prevail over some of the following limitations: 1) intrusiveness: the demand for wearing sensors [7], [8], [9], [10], [11], [12] on the body, which undermines quality of user experience; 2) low scalability: high dependence on Line of Sight (LOS) and light conditions for vision-based systems [1], [2], [3], [4], which prohibits their pervasive use in certain environments; 3) high cost: specialized or modified hardware [13], [14], [15] for wireless signal processing, which is unfavorable for our daily usage.

Fortunately, Radio Frequency IDentification (RFID) promises an alternative prospect to satisfy our demand with its structural simplicity and widespread availability in smart homes, museums and art galleries [16], [17]. Except for conventional functions, RFID has been witnessed to succeed in a wide variety of applications such as find-grained motion tracking [18], [19], [20], indoor localization/navigation [21], [22], [23], [24], and gesture recognition [25], [26], [27], [28]. Among these, RFID-based gesture recognition is relatively a sunrise area which draws increasing attention from academia and industry. Although several pioneering work has been conducted in this area, they all either employ complicated implementation or require users to wear tags. To avoid these, our objective is to provide an effective, reliable and non-intrusive RFID-based gesture recognition system using off-the-shelf devices, and meanwhile without instrumenting environments too much.

In this paper, we propose GRfid, a device-free gesture recognition system by exploiting phase information of RFID signals with Commercial Off-The-Shelf (COTS) devices. We argue that such a gesture recognition system possesses several advantages over other existing systems: 1) low cost and simplicity,[1] since GRfid is based on COTS

- *Y. Zou, J. Xiao, and K. Wu are with the College of Computer Science and Software Engineering, Shenzhen University, P.R. China, and the CSE Department, Hong Kong University of Science and Technology (HKUST), Kowloon, Hong Kong. E-mail: {yzouad, jxiao, kwinson}@cse.ust.hk.*
- *J. Han and Y. Li are with Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an, Shaanxi 710049, P.R. China. E-mail: {hanjinsong, yunli}@mail.xjtu.edu.cn.*
- *L.M. Ni is with the University of Macau (UM), Avenida da Universidade, Taipa, Macau, China. E-mail: ni@umac.mo.*

Manuscript received 9 Apr. 2015; revised 29 Feb. 2016; accepted 25 Mar. 2016. Date of publication 31 Mar. 2016; date of current version 5 Jan. 2017. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TMC.2016.2549518

1. Considering the wide use of RFID systems, we make a reasonable assumption that a reader is equipped in environment as a basic infrastructure, as done in [20].

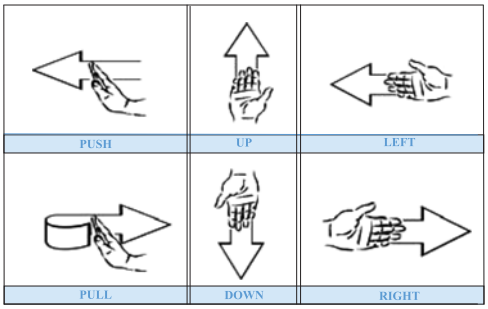


Fig. 1. Gestures that can be recognized by GRfid.

devices and doesn't require any hardware modifications; 2) non-intrusiveness, since users don't need to wear tags on their bodies any more; 3) scalability, since tags have unique IDs and can be easily attached to devices, we can conveniently extend the whole system by adding subsystems if needed.

The key idea of GRfid is to utilize the RFID signal phase changes to recognize different gestures. The inspiring insights for this lie on the following two key observations. First, when gestures are performed in front of tags deployed in an environment, the multipath of each tag's signal propogation will be changed along with hand movements, which can be captured by the RFID phase information. Second, RFID phase information is demonstrated to be more robust to environmental interference, tagged objects' properties, and also tag orientations, compared with Received Signal Strength Indicator (RSSI) [29], [30]. Since different gestures go through different spatial paths and cause distinct multipath effects, as a result, phase profiles of gestures differ from one another. Moreover, current commodity RFID products can accurately output detailed phase information without any modification. Therefore, we envision phase information as a potential metric for developing a gesture recognition system.

However, developing such a system is not so straightforward. The first challenge is how to construct suitable phase-based profiles of diverse gestures. Our extensive experiments have shown that when users stand at different locations, phase changes of one tag caused by even the same gesture will differ so much that it fails to recognize the gestures correctly. Second, it is non-trivial to recognize a gesture using prestored profiles. Naive matching algorithms based on simple metrics such as similarity always fail to work, without considering high position dependency between phase changes and gestures.

We design our system with COTS hardwares for data collection and well-designed software for data analysis. Specifically, we utilize a commercial reader and multiple tags to collect data. Then, received data packets are sequentially passed onto several functional blocks, namely, data preprocessing, gesture detection, gesture profiles training and gesture recognition. Each of them is carefully designed in order for favorable performance of gesture recognition. We conduct comprehensive experiments under various settings to evaluate GRfid's performance of recognizing six gestures as shown in Fig. 1, namely, 'PUSH', 'PULL', 'UP', 'DOWN', 'LEFT' and 'RIGHT'.

Contributions. In summary, our main contributions can be concluded as follows:

- To the best of our knowledge, this is the first attempt to design a device-free gesture recognition system based on RFID phase information, which is available from COTS devices and proved to be accurate and robust against external interference such as multipath, environmental dynamics and tag orientation.
- Second, we propose two algorithms for gesture profiles training and gesture recognition, respectively. Conjunctive use of these two algorithms can effectively handle diversities in gestures performed at different positions. As a result, gestures can be accurately recognized even at diverse sites different from where the profiles are constructed.
- Third, we implement GRfid system with commodity devices in three diverse indoor scenarios. Experimental results demonstrate that GRfid can effectively recognize six predefined gestures with average accuracy of 96.5 percent in the identical-position scenario and 92.8 percent in the diverse-positions scenario. Experiments also demonstrate that GRfid possesses robustness to both environmental interference and tag orientations.

The remainders of this paper are organized as follows. Section 2 presents the related work. Section 3 describes the overview of the system. Section 4 shows the methodology of GRfid. In Section 5, we present the system implementation and performance evaluation. Section 7 conducts a discussion of this work and future directions. In Section 8, we conclude this paper.

2 Related Work

This section surveys the state-of-the-art work that deals with subjects of RF-based gesture recognition and RFID-based indoor localization, which are most relevant to our work.

RF-based gesture recognition. Recently, RF-based gesture recognition systems have attracted intensive research interest. Most of them are based on Wi-Fi or RFID, owing to the shared merits of non-intrusiveness and pervasiveness. The examples of Wi-Fi-based gesture recognition mainly include Wisee [14], WiVi [13], Witrack [15].[2] They succeed in distinguishing various kinds of gestures at the cost of sophisticated system implementation, and/or specialized and modified hardware. Another example is proposed in [31], which leverages Channel State Information (CSI) and signal phase as metrics to distinguish fine-grained gestures. However, the system works in a relatively short range (within 80 cm) and is highly position-dependent because of the utilization of directional antennas. RFID is an alternative promising technology for gesture recognition. For example, Allsee [32] leverages RFID backscatter signals to recognize signals. However, the tag used by Allsee should be connected to a logical computation unit which is unavailable on most low-cost commodity tags. Authors in [26] leverage the gestures recognized by RFID readers to identify a pair of RFID toys, yet requiring much computational processing. RFIDraw [19], Tagoram [18] and Tagball [20] can trace a tag's

2. To be strict, WiTrack utilizes RF signals that partially cover Wi-Fi frequency band (5G) in a different modulation scheme. We classify it into Wi-Fi-based systems for the sake of clearness in the recital.

spatial trajectory at a granularity of centimeters, while they all require attaching a tag(s) to the tracked object. Consequently, our work is different from all those above in that we concentrate on building up a device-free gesture recognition system with RFID phase information, which owns the capability of capturing the diverse features of gestures for distinction.

RFID-based indoor localization. There have been a large amount of research on RFID-based indoor localization. LANDMARC [21] is a pioneering positioning system relying on RFID with RSSI which suffers from severe multipath problems in indoor environments. PinIt [23] resolves both multipath and NLOS by employing adaptive dynamic time warping technique. VIRE [22] introduces the concept of virtual tags to improve precision as well as to reduce deployment cost. Our design is inspired by the most recent work [18] that utilizes phase value of the backscattered signal to build a virtual antenna array for real-time localization. Instead of acquiring the absolute location like the above systems, GRfid aims to detect and identify the more subtle gestures.

3 An Overview of GRfid System

In this section, we first briefly present the background knowledge of an Ultra-High-Frequency (UHF) RFID system and its signal propagation, with a special focus on the phase property, which lays the foundation of our work. Next, we conduct two sets of preliminary experiments using a COTS ImpinJ reader and commercial tags to demonstrate key issues relevant to our system. Finally, we introduce the architectural framework of the GRfid system, consisting of four functional blocks: 1) signal processing, 2) gesture detection, 3) profiles training and 4) gesture recognition.

3.1 Background

A UHF RFID system typically operates on frequency band between 902 and 928 MHz using backscatter links. During a whole communication process, a reader transmits a high-power RF signal to interrogate tags. Then, nearby tags within range make responses by modulating the transmitted signal from the the reader, using On-Off keying modulation by changing the impedance on their antennas. By doing so, passive tags can accomplish data transmission purely by harvesting energy from the backscattered signals. A COTS RFID system usually consists of commercial UHF readers, which can interrogate tags within a range from several to a few dozen meters, depending on the power of reader's antennas. Present COTS RFID readers such as Impinj Speedway can output rich signal information such as signal phase, RSSI and Doppler shifts [33], which provide potential for gesture recognition.

For an RFID communication system, when taking into account the Multipath Components (MPCs), the receieved signal can be modeled as:

$$s(t) = \sum_{i=1}^{N} \beta_i p(t - \tau_i) e^{j\phi_i}, \tag{1}$$

where $p(t)$ is the pulse from the transmitter. β_i, τ_i and ϕ_i stand for the complex magnitude, the Time of Arrival (ToA) and phase of the ith arriving path, respectively. This signal propagation model considers the multipath effect, in which MPCs arrive at the receiever via different propagation mechanisms (i.e., reflection, transmission or scattering).

We expand Eq. (1) by applying Euler formula so as to obtain the following:

$$s(t) = \sum_{i=1}^{N} \beta_i p(t - \tau_i) \cos\phi_i + j \sum_{i=1}^{N} \beta_i p(t - \tau_i) \sin\phi_i. \tag{2}$$

The phase Φ of received signal can therefore be presented by:

$$\Phi(t) = \arctan\left(\frac{\sum_{i=1}^{N} \beta_i p(t - \tau_i) \cos\phi_i}{\sum_{i=1}^{N} \beta_i p(t - \tau_i) \sin\phi_i}\right). \tag{3}$$

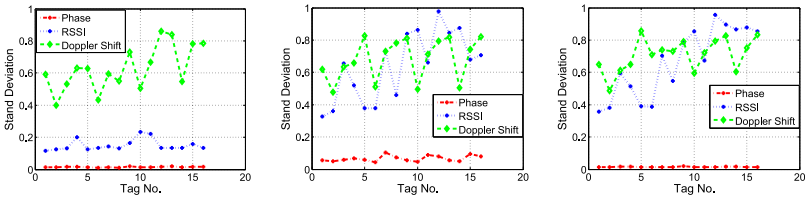
From Eq. (3), it is obvious that the received phase information is a complex combination of all MCPs. As for RFID phase-based localization or motion tracking systems, suppressing the phase shifts caused by multipath needs to be emphasized in order to make the LOS component the dominant part. However, this requirement is always violated because of the rich multipath in indoor environments. Fortunately, in our case, LOS is no longer the necessary condition. On the contrary, multipath which is intentionally produced by our gestures are fully utilized for gesture recognition. As a hand moves in front of tags, the propagation paths of each tag's backscatter link will be altered, which brings in phase changes for each tag's backscattered signal. Diverse gestures induce unique profiles of phase changes, which can be used for gesture recognition.

3.2 Preliminary Experiments

In this part, we conduct two sets of preliminary experiments using a COTS Impinj reader and 16 commercial tags, to demonstrate the following two observations before designing our system, covering what information and how many tags we will utilize for our application.

Observation 1. Phase information is a more favorable metric for gesture recognition, compared with two other kinds of outputs, namely, RSSI and Doppler shifts.

As previously mentioned, present COTS RFID readers can output RSSI, phase and Doppler shifts. In order to select the best metric for gesture recognition, we conduct experiments to examine their characteristics in three different scenarios. In the first experiment, all 16 tags are placed in a relatively static and open space, that is, less multipath and no dynamic interference. In the second scenario, environmental interference are produced by placing a metallic board nearby and volunteers' walking around. In the third experiment, we place each tag in diverse orientations at the same position. For each setting, we collect phase, RSSI and Doppler shifts and calculate their standard deviation, respectively. The results are shown in Fig. 2, illustrating that the standard deviation of phase is much smaller than that of RSSI and Doppler shift in all of the above scenarios. It demonstrates that phase information is more stable and robust than RSSI and Doppler shifts, especially under environmental interference and diverse orientations. As a result, we select signal phase as indicator of gesture recognition in our system.



(a) Static (b) Dynamic Environment (c) Diverse Orientation

Fig. 2. Standard deviation of phase, RSSI, and doppler shifts in three different scenarios.

Observation 2. In one-tag scheme, signal profiles of gestures are highly positon-dependent, which puts strong constraints on this scheme and makes it fail to work practically.

The number of tags used in GRfid system is sure to have an effect on the final performance of gesture recognition and a multiple-tags system is expected to outperform a single-tag one. The insight behind this is similar to a multi-antennas RF system, in which multiple antennas provide spatial diversity and is beneficial for gesture recognition [14]. However, our preliminary experiments are merely conducted to justify the failure of single-tag scheme due to its high sensitivity to the positions of performing gestures. Specifically, we perform the same gesture ('Down' as shown in Fig. 1) at three different positions, namely in the directly front, left diagonal front and right diagonal front of a tag as is shown in Fig. 3. Collected phase data are processed to extract corresponding signal segments. For simplicity, we only show typical signal segments when 'Down' is performed in the directly front and right diagonal front of a tag respectively in Fig. 4. Similarity calculation via cross-correlation indicates that these two segments share little similarity. This observation implies that when designing a gesture recognition system, a single-tag scheme is not a preferable choice. Inspired by the principle of a multi-antennas system design, we utilize multiple tags for the exploration of an accurate and robust gesture recognition system. Detailed evaluation of choosing the number of tags is conducted in Section 5.

3.3 System Overview

From the application level perspective, GRfid is a wireless system that enables a device-free interactive interface for people to interact with other devices using gestures. Overall speaking, GRfid consists of hardware acting as gesture sensing and softwares performing data analysis. As for hardware components, as is shown in Fig. 3, GRfid mainly contains one PC, multiple tags and an RFID reader with a directional antenna, all of which are COTS devices. The tags are attached to objects and spatially separated. The directional antenna is mounted with a motor and fixed at a certain distance from tags. The antenna is connected to the reader, which in turn is connected with the PC. When a person performs gestures in front of tags, the reader will receive data packets and transmit them to the PC for further analysis.

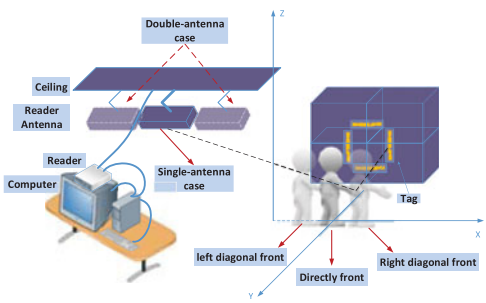


Fig. 3. The hardware components of GRfid.

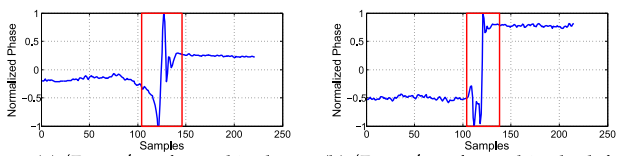
Fig. 5 demonstrates the schematic digram of GRfid from a perspective of data flow. Upon receiving data packets, the reader feeds them into the software part for data analysis, including four functional blocks, i.e., signal preprocessing, gesture detection, gesture profiles training and gesture recognition. The signal preprocessing block fulfills phase unwrapping, smoothing and normalization in sequence. Next, GRfid performs automatic gesture detection so as to extract signal segments corresponding to gestures and ignore those perturbed signals induced by environmental noise. In the following, the obtained segments are transfered into gesture profiles training block, which calibrates segments, trains gesture profiles and builds gesture files farm. Finally, an weighted matching algorithm is applied for an unknown gesture to be recognized, to find a optimal match within the predefined gestures. In the following section, we focus on exploring specific methodologies of each part.

4 Methodology

In this section, we give a detailed demonstration of the methodology of GRfid, based on the aforementioned four basic functional blocks: signal preprocessing, gesture detection, profiles training and gesture recognizing, as shown in Fig. 5. For each functional block, we give an introduction to specific techniques we applied and state the insights behind them.

4.1 Signal Preprocessing

When raw phase data has been collected, the primary task is to unwrap the phase. This is because the signal phase is a periodic function, which is also termed as a wrapped phase. To be specific, in order to differentiate between each packet, we need to consider all possible times of 2π, known as phase unwrapping. To achieve accurate phase differentiation, we adopt the commonly used One-Dimensional Phase Unwrapping method [34].



(a) 'Down' performed in the directly front of a tag. (b) 'Down' performed in the left diagonal front of a tag.

Fig. 4. One-tag scheme may fail to work at different locations.

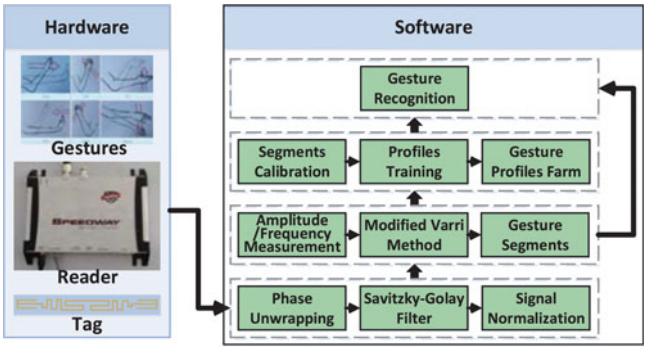


Fig. 5. GRfid system architecture.

After the phase-unwrapping stage, we apply a Savitzky-Golay filter to remove any random noise and smooth the data. The Savitzky-Golay filter is defined as a weighted moving average with weights given as a polynomial of specific degree, which serves as a powerful tool for smoothing signals and tends to preserve features of the distribution such as relative maxima and minima [35], making it more suitable than other filtering methods for our case.

After filtering, we then proceed to the normalization of phase data, which rules out unwanted effects resulting from the variance of absolute magnitude while keeping track of patterns of the signals. This will ultimately benefit the gesture recognition at a later stage, as gestures sharing similar patterns could differ in magnitude when performed by different people. In our work, we accomplish normalization by:

$$\widetilde{x_i} \begin{cases} \frac{x_i - \frac{\sum_{k=1}^{N} x_k}{N}}{\arg\max x_{k,k\in 1,\ldots,N}}, & x_k \geq 0 \\ \frac{x_i - \frac{\sum_{k=1}^{N} x_k}{N}}{-\arg\min x_{k,k\in 1,\ldots,N}}, & x_k < 0, \end{cases} \tag{4}$$

where $\widetilde{x_i}$, x_i and N represent the normalized signal, the original signal and the total number of signals, respectively.

4.2 Gesture Detection

Once cleaned signal sequences are obtained, we come to the gesture detection stage at which segments corresponding to gestures need to be extracted. This task is rather important as well as challenging considering the following two points. First, it is required to extract segments corresponding to gestures exactly. Otherwise, noise will be introduced or useful information shall be lost, both degrading the final performance of gesture recognition. Another fundamental challenge is that patterns of gestures are not so stable, when taking diversities of gestures, variance for performing even the same gesture into consideration. For example, signal sequences of the same gesture present much difference, when performed at different speeds even by the same person.

To respond to the challenges, we apply a segmentation method called the Modified Varri Method, combined with Savitzky-Golay filter described above [36]. This method is based on the combination of a frequency measure and an amplitude measure of the signal in the relevant windows. Specifically, when applying this segmentation method to a signal, there are two windows moving along the signal in charge of calculating two parameters: 1) amplitude measure A_{diff}; and 2) frequency measure F_{diff} that are defined as follows, respectively:

$$A_{diff} = \sum_{k=1}^{l} |x_k| \tag{5}$$

$$F_{diff} = \sum_{k=1}^{l} |x_k - x_{k-1}|, \tag{6}$$

where l and x_k represent the length of the sliding windows and the kth data point, respectively. Based on Eqs. (5) and (6), the measure difference function (G) is defined to determine the boundaries of segments:

$$G_m = A_1 \left| A_{diff_{m+1}} - A_{diff_m} \right| + F_1 \left| F_{diff_{m+1}} - F_{diff_m} \right|, \tag{7}$$

where m is the index of sliding window.

Those local maximums in the G function being above a threshold, which is empirically defined, indicate the boundaries of each segment. It should be noted that parameters of A_1, F_1 and the window length l should be precisely tuned in order for better accuracy of segmentation. However, automatic segmentation utilizing the Modified Varri Method without tuning parameters manually can be accomplished by applying a Genetic Algorithm, which also achieves better performance [37]. Since genetic algorithms (GAs) are meta-heuristic and as such there is no general complexity analysis that applies to all genetic algorithms at once. As a result, the term 'convergence time' is usually used for evaluating the complexity of GAs which depends on the number of iterations and population size. In our experiments, the average convergence time in different settings is about 100 ms. For more details about the tuning process, please refer to the corresponding references.

Fig. 6 verifies the feasibility and efficiency of gesture detection, in which each subgraph corresponds to a predefined gesture experiment. In such an experiment, volunteers are required to perform gestures in the experimental scenarios shown in Fig. 3 and received phase data (with timestamp) are then fed into the signal preprocessing block. Considering the limited space here, only part of results are shown in this figure. In each subplot, the blue line represents preprocessed phase, and signals encompassed by the red rectangles represent segments extracted. Experimental results of six gestures indicate that this method can work well in different scenarios for different gestures.

After gestures can be correctly detected, i.e., signal segments corresponding to gestures processes can be extracted, they will be delivered to the next stage, namely, profile training stage when the system is under training, or gesture recognizing stage when gestures need to be recognized.

4.3 Gesture Profiles Training

In order to recognize various types of gestures, GRfid requires the user to perform each predefined gesture as shown in Fig. 1 for multiple times (assume N), providing inputs for the gesture profiles training process. It is noted that since GRfid utilizes M tags, the total number of obtained signal sequences should be M for each gesture. Then, the system perform signal preprocessing and gesture

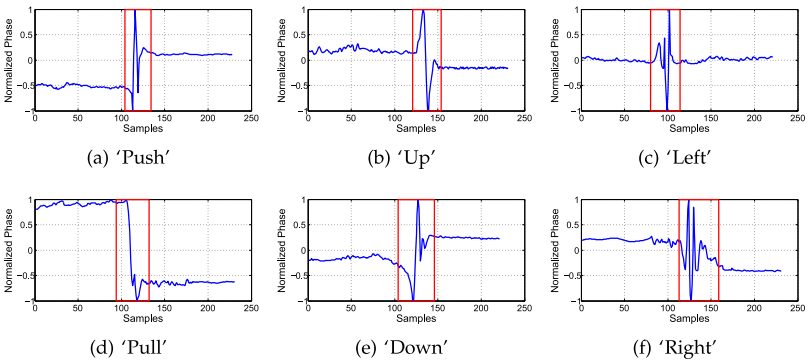


Fig. 6. Segmentation accomplished by modified varri method: each subplots show phase change of a single tag caused by the predefined gesture. The blue curve represents normalized phase and the red rectangles represent the segments corresponding to various gestures.

detection described above on raw signal sequences. For the sake of making each gesture profile reliable and self-consistent, we propose to carry out calibrations on segments set, that is, we select a subset of representative segments (exemplars) to be signal profiles of each gesture, instead of utilizing all of them.

To do this, we borrow the concept of proximity-based outlier detection [38], in which an object (a point or vector) is considered to be an outlier if it is far away from the remaining objects. Specifically, for the N segments, i.e., N vectors $T_1, T_2, \ldots, T_N$ of each gesture from each tag to be calibrated, a proximity matrix $A_{N\times N}$ is constructed in which element $A_{i,j}$ represents the normalized DTW distance (NDTW) which is defined in Section 4.4 between any pair of signal segments T_i and T_j. For any segment T_i, its proximity score can be denoted by the summation of NDTWDs to other $N-1$ segments. Finally, we pick out the top K segments with the lowest proximity scores as exemplars. Similarly, for other tags, the exemplars can be selected in the same way. As a result, for each gesture, we can get a total number of $K \times M$ signal segments and each tag occupies K segments as gesture profiles. For each gesture, the selected exemplars are attached with the gesture's label, i.e., the index among $\{1, 2, \ldots, P\}$, where P is the total number of gestures to be recognized. Selected exemplars of all gestures are stored as a gesture profiles pool.

4.4 Gesture Recognition

The training process provides a set of reference templates for each gesture. When an unknown gesture is performed and the corresponding signal sequences are obtained, it is required to find the most likely gesture match among the predefined gestures, which is essentially a time series matching problem. In this section, we first give an brief introduction to the metric for measuring the similarity between two time series, i.e., normalized dynamic time wrapping (DTW) distance denoted by NDTW. Based on this, we then propose an adaptive weighting algorithm focusing on finding the optimal match for gestures to be recognized.

4.4.1 Normalized Dynamic Time Wrapping Distance

Dynamic time wrapping is an efficient technique for measuring the similarity between two time series by finding the optimal alignment, which allows for elastic shifting in the time domain and matches sequences that are similar but out of phase. A well-known application of DTW has been automatic speech recognition, to cope with varied speeds of speech. Formally, in the context of gesture matching for our application, given two gesture profiles $P_i = P_i(1), P_i(2), \ldots, P_i(\alpha), \ldots, P_i(M)$ and $P_j = P_j(1), P_j(2), \ldots, P_j(\beta), \ldots, P_j(N)$, the cost of mapping any pair of points $P_i(\alpha)$ and $P_j(\beta)$ to each other is defined as the euclidean distance between them, that is:

$$d_{\alpha,\beta} = \|P_i(\alpha) - P_j(\beta)\|, \tag{8}$$

where α and β are the indexes in the two sequences P_i and P_j. Based on this, DTW searches a wrapping path $W = w_1, w_2, \ldots, w_K$, where $w_k = (\alpha, \beta)$, between P_i and P_j that can minimize the total cost, i.e.,

$$DTW(P_i, P_j) = \underset{left}{\text{minimize}} \sum_{k=1}^{L} d_{w_k}, \tag{9}$$

where D is a matrix of size $M \times N$ whose elements are $w_i, i = 1, 2, \ldots, M \times N$, namely, (α_1, β_1), $(\alpha_1, \beta_2), \ldots, (\alpha_M, \beta_N)$. $DTW(P_i, P_j)$ represents the dynamic time wrapping distance, i.e., the optimal total cost of wrapping between gesture sequence P_i and P_j, which is a metric of similarity of any two time series. However, it should be pointed out that directly using DTW as a similarity metric will be biased by two factors: differences in amplitudes of time series and variable lengths of comparing sequences [39]. The former problem can be effectively handled by amplitude normalization described in Section 4.1. As for the latter, longer sequences are obviously expected to be further apart than short sequences. As a result, we apply another normalization method as:

$$NDTW(P_i, P_j) = \frac{DTW(P_i, P_j)}{M + N}. \tag{10}$$

Compared with the widely used distance measurement such as Euclidean distance, Cosine similarity, DTW has taken into account the stretching or shrinking effect of data sequences along the time axis, making it a powerful tool to cope with varying speeds even for the same gesture. It is noted that the problem of finding the optimal wrapping path can be solved by dynamic programming in $\mathcal{O}(n)$ time [40].

TABLE 1
Notations and Corresponding Meaning

Notations	Meaning
N	the number of repetitions performed for each gesture in profiles training stage
P	the number of gestures to be recognized by GRfid
M	the number of tags used in GRfid
K	the number of exemplars selected for each gesture from each tag
$T^i_{k,j}$	the kth representative segment for the ith gesture from the jth tag, where $k = 1, \ldots, K$, $i = 1, \ldots, P$, $j = 1, \ldots, M$
$\mathbf{T}^i_k = \langle T^i_{k,1}, T^i_{k,2}, \ldots, T^i_{k,M} \rangle$	the kth exemplars of the ith gesture
S^l_j	a signal segment corresponding to an unknown gesture from the jth tag, where $l \in \{1, 2, \ldots, P\}$ to be identified
$\mathbf{S}^l = \langle S^l_1, S^l_2, \ldots, S^l_M \rangle$	tuple of signal segments representing a gesture
NDTW(S^l_j,$T^i_{k,j}$)	normalized dynamic time wrapping distance between two sequences $T^i_{k,j}$ and S^l_j

4.4.2 Weighted Matching Algorithm

Gesture profiles training process provides signal templates for each gesture and similarity metric-NDTW for time series matching has been introduced in Section 4.4.1. Then, when a user performs gestures intending to interact with other devices, GRfid processes collected data via preprocessing and segmentation, and then is required to compare those segments with the prestored templates and determine the optimal match, which is essentially an exemplar-based matching process.

Before formally formulating the matching problem in GRfid, for the sake of convenience, we concludes the notations in Table 1, which will be adopted in our later problem formulation and algorithm design.

Formally, for gesture recognition in GRfid, the problem formulation can be given as follows:

Problem Formulation. *Given profiles of P predefined gestures, namely, $T^i_{k,j}$, where $k = 1, \ldots, K$; $i = 1, \ldots, P$; $j = 1, \ldots, M$. For an unknown signal segment S^l_j, where $j = 1, \ldots, M$ and $l \in \{1, 2, \ldots, P\}$, it is required to identify the exact value of l, such that the cost for classifying this unknown gesture to a certain class is minimized.*

The key point of this problem is to define the overall cost of classifying a gesture into a certain class. A straightforward idea is to minimize the overall NDTW when matching with templates in this gesture class, that is:

$$l = \underset{l \in \{1,\ldots,P\}}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{j=1}^{M} NDTW(T^l_{k,j}, S^l_j). \tag{11}$$

Essentially, the key idea indicated by Eq. (11) is the same as the Nearest Neighbors algorithm, which is a naive classification method in data mining. Ideally, it will work well when we ignore the noise caused by environmental variance, gesture diversity performed even by the same person, and especially position diversities even for the same gesture. However, this naive matching algorithm fails when these considerations are taken into account in practice. The reasons behind this can be concluded in two aspects: 1) noise caused by environmental variance presents distinct interference on RF links between tags and the reader; 2) different weights should be assigned to tag-reader links according to the overall similarity calculation.

Different from the above, we propose a gesture recognition scheme in which different scores are assigned to each gesture class, according to the corresponding distance. Specifically, for a certain gesture to be recognized represented by $\mathbf{S}^l$, the distance between it and the kth exemplars of ith gesture will be calculated as:

$$NDTW(\mathbf{S}^l, \mathbf{T}^i_k) = \sqrt{\sum_{j=1}^{M} w(j) \times NDTW(S^l_j, T^i_{k,j})}, \tag{12}$$

where $w(j)$ is a normalized Variance-based Weighting function, that is:

$$w(j) = \frac{var(NDTW(S^l_j, T^i_{k,j}))_{(k=1,2,\ldots,K)}}{\sum_{j=1}^{M} var(NDTW(S^l_j, T^i_{k,j}))_{(k=1,2,\ldots,K)}}. \tag{13}$$

The key insight of this weighting scheme is that a tag possessing higher variance indicates less certainty in recognizing a gesture. After this step, for each gesture class, we can obtain K distance values and thus a total number of $K \times P$ distance values can be obtained when an unknown gesture $\mathbf{S}^l$ is matched with all predefined gestures.

Next, we propose such a weighted voting scheme for making the final decision. We select the top k nearest neighbors, that is, select the top k smallest values among the obtained $K \times P$ values, suppose to be sorted as $d_1 < d_2 < \cdots < d_k$, with each labeled by a corresponding gesture index $l_i \in \{1, 2, \ldots, P\}$. By assigning weight to each of them as follows:

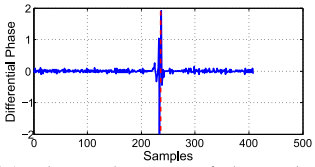
$$w_{d_i} = \begin{cases} \frac{d_k - d_i}{d_k - d_1}, & d_k \neq d_1 \\ 1, & d_k = d_1. \end{cases} \tag{14}$$

Based on this, the final gesture recognition result can be acquired by the majority weighted voting:

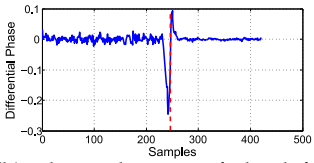
$$l = \underset{l}{\operatorname{argmin}} \sum_{T} w_{d_i} \times \delta(l = l_i), \tag{15}$$

where $T = \{l_1, l_2, \ldots, l_k\}$ and $\delta(l = l_i)$ represents the Dirac delta function. This weighted matching algorithm for gesture recognition can be concluded as Algorithm 1.

Another subtle information that can be utilized to assist gesture recognition is timestamps contained in signal sequences, which implies moving directions of certain gestures (like 'Left', 'Right', 'Up' and 'Down'). Since multiple tags are placed on a plane with certain distance between them, it is understandable that different tag-reader links will encounter obvious interference, when certain gestures are performed. For example, when a 'Right' gesture is performed, tags on the left side will be influenced earlier, compared with the right-side ones, as shown in Fig. 7. By pinpointing the corresponding timestamps, we can estimate the direction of a gesture.



(a) Phase changes of the right tag.



(b) Phase changes of the left tag.

Fig. 7. Phase changes of two tags when performing a 'Left' gesture.

Algorithm 1. Weighted Matching Algorithm

Input:
Gesture templates:
$\{\mathbf{T}_k^i = \langle T_{k,1}^i, T_{k,2}^i, \ldots, T_{k,M}^i \rangle, i = 1, \ldots, M; k = 1, \ldots, K\}$;
Unkonwn gesture:
$\mathbf{S}^l = \langle S_1^l, S_2^l, \ldots, S_M^l \rangle$;

Output: Identified gesture index: $l \in \{1, \ldots, P\}$

1: **for** each gesture $i = 1$ to P **do**
2: **for** each tag $j = 1$ to M **do**
3: Calculate each $NDTW(S_j^l, T_{k,j}^i)$ for $k = 1, 2, \ldots, K$;
4: **end for**
5: Calculate $w(j)$ by Eq. (13);
6: Calculate $NDTW(\mathbf{S}^l, \mathbf{T}_k^i)$ for $k = 1, 2, \ldots, K$, by Eq. (12);
7: **end for**
8: Select top k smallest value among $K \times P$ and assign weight w_{d_i} to each of them;
9: Output the value of l by weighted majority voting, by Eq. (15);

5 System Implementation and Experiment Settings

In this section, we first present the prototype implementation of GRfid, which includes both hardware and software implementations. Afterwards, we describe several basic experiment scenarios, through which the performance of GRfid will be evaluated.

5.1 System Implementation

1) *Hardware.* As for hardwares in our system, GRfid mainly includes a commodity RFID reader (Impinj R420) and off-the-shelf UHF passive tags from two mainstream manufactures, i.e., Alien and Impinj. The Impinj reader, equipped with a directional antenna (Laird $A9028R30NF$ with 8dbi gain), operates at a fixed working frequency of 924.375 MHz. The antenna is mounted on the cell of a lab by a rotating motor, such that we can control the antenna to circularly spin around for direction adjustment. The tags are attached to paper cartons, with a center distance of 20∼30 cm between each other. The distance between geometric centers of tags and the reader antenna is about 200 cm.

2) *Software.* On the other hand, to accomplish the task of data collection and analysis, we implement software on the PC tailored for two purposes. one of the purposes is to extract low-level signal information from the reader. For this purpose, we integrate Octane SDK, an extension of the LLRP Toolkit with our software, which supports the collection and analysis of signal phase, RSSI and Doppler shifts. The other one is for data analysis as described above in the form of functional blocks, which are implemented with Matlab programming.

5.2 Experiment Settings

5.2.1 Identical-Position Scenario

In this setting, gestures are trained and recognized at the same position. Specifically, volunteers first perform predefined gestures at a certain position. The corresponding data segments can be obtained by feeding the raw data through functional blocks sequentially. Later we apply K_f-fold Cross Validation ($K_f = 5$ in our experiments) to test the accuracy of gesture recognition. Three different positions are tested, namely, directly front (DF, about 75∼90 cm from geometric center of tags), left diagonal front (LDF, about 20 cm at the left of DF) and right diagonal front (RDF, about 20 cm at the right of DF).

5.2.2 Diverse-Positions Scenario

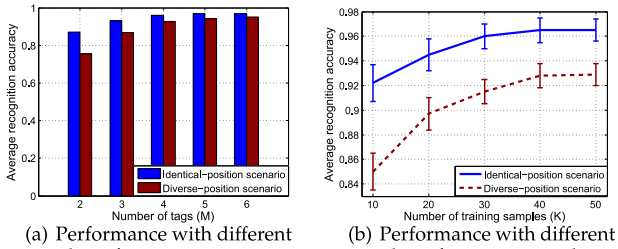
In this scenario, gestures are trained and recognized at separate positions. To be specific, we take data collected at one position (like DF) as training set and data of another position (like LDF) for testing. The recognition accuracy of each gesture can be obtained from each set of experiments and the average is defined as the final result. It is noted that the number of tags when conducting experiments are varied from 2 to 6 in both identical-position and diverse-positions scenarios, in order to explore the effect of the number of tags on the system's performance.

5.2.3 Environmental Interference

Environmental interference that we consider in this paper covers static and dynamic multipath interference, which is be created by placing a highly reflective board nearby and people's walking around. Specifically, we place a metallic board of 1.5 m × 1.0 m with high reflectivity at a distance (around 30 cm) close to the position where gestures are performed. In this way, complex static external multipath is produced. Meanwhile, diverse numbers of volunteers walk randomly around the system (about 30 cm from the prototype) to produce dynamic environmental interference.

5.2.4 Amplitudes of Gestures

An additional point to be noted is about the amplitudes of gestures. In the context of this paper, the amplitude of a gesture is defined as the distance between start position and end position. It is noted that the six predefined gestures in experiments are performed with a amplitude of about 20∼30 centimeters for each. For example, when performing a PULL gesture, the volunteer drags back his/her straight arm with a distance of 20∼30 centimeters. According to our experiments, we find that such an amplitude can guarantee GRfid high accuracy of recognizing gestures and doesn't cause unnatural feelings to users.



(a) Performance with different number of tags.

(b) Performance with different number of training samples.

Fig. 8. GRfid's performance under different system parameters.

6 Performance Evaluation

In this section, we evaluate the performance of GRfid in various scenarios, considering the practical application of GRfid. The main metric we adopt for performance evaluation is confusion matrix. In a confusion matrix, elements in each row represent the ratios that a performed gesture is recognized as other gestures (including itself) respectively. One advantage of this metric is that other evaluation metrics such as *false positive (FP)* and *false negative (FN)* can be directly derived by definitions.

6.1 Performance Evaluation under Different System Parameters

6.1.1 Performance with Different Number of Tags

When designing GRfid, the number of tags (M as described in Table 1) should be carefully considered, as there is a trade-off between recognition accuracy and computational overhead. As a result, we vary the number of tags and conduct experiments in identical-position and diverse-position scenarios. Fig. 8b shows GRfid's average accuracy under different number of tags in both scenarios. It indicates that the effect of number of tags is more obvious in diverse-positions scenario than that in identical-position scenario. This is reasonable since more tags provide richer spatial diversity which is more significant for GRfid in the diverse-positions scenario. Another point is that increasing the number of tags indeed improves the performance, but the gain becomes continually smaller, especially when M exceeds four (with an increment rate less than 3 percent). Taking the aforementioned tradeoff into account, we only evaluate the performance of GRfid in the case where four tags are utilized in the following experiments.

6.1.2 Performance with Different Number of Training Samples

Another parameter needs to be considered when designing GRfid is the number of training samples, specifically, the size of K as described in Table 1. Similar to the above, we consider both identical-position and diverse-position scenarios, respectively. We scale the size of K from 10 to 50, and test the average accuracy of gesture recognition in each case. Fig. 8 shows the results. As we can see, the overall trends in both scenarios are the same, that is, the accuracy increases with the size of K in the beginning, and then converges to a constant after K exceeds a certain threshold. In both scenarios, when the number of training samples exceeds 40, the accuracies remain consistent with an increment rate less than 1 percent. As a result, taking the tradeoff between accuracy and computational overhead into consideration, we set K to be 40 in the following evaluation experiments.

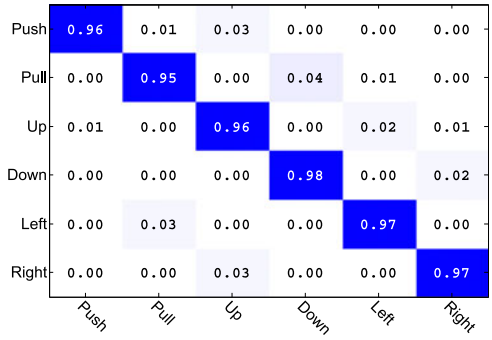


Fig. 9. Confusion matrix for the identical-position scenario.

6.2 Performance in Identical-Position Scenario

In this session, we investigate the performance of GRfid in the identical-position scenario. We conduct experiments and perform training-testing analysis on collected data at three positions, namely, DF, LDF and RDF as described in Section 5.2. The accuracy of recognizing a gesture is defined as the average value at the above three positions. Fig. 9 shows the gesture recognition results in the form of a confusion matrix. As we can see, the average recognition accuracy of six gestures is up to 96.5 percent, which is slightly higher than the accuracy of WiSee [14] (94 percent) and comparable with that of AllSee [32] (97 percent). This indicates that all predefined gestures can be recognized accurately in the identical-position scenario. In other words, it also reveals that gestures present stable and unique patterns which can be extracted by phase changes.

6.3 Performance in Diverse-Positions Scenario

We also evaluate the performance in the diverse-positions scenario. Since there are several subsets of experiments, we average the results to obtain the accuracy of GRfid system. Fig. 10 shows the confusion matrix of recognizing gestures over varying sites. It can be clearly shown that the average accuracy of gesture recognition for predefined gestures even in such a diverse-positions scenario can remain relatively high, with a mean value of 92.8 percent. In addition, compared to the identical-position case, there is a slight performance degradation of 3.7 percent. These results indicate that by applying the four-tag scheme and the adaptive weighting matching algorithm, dependency on a user's position can be effectively alleviated, which ensures the robustness of this system.

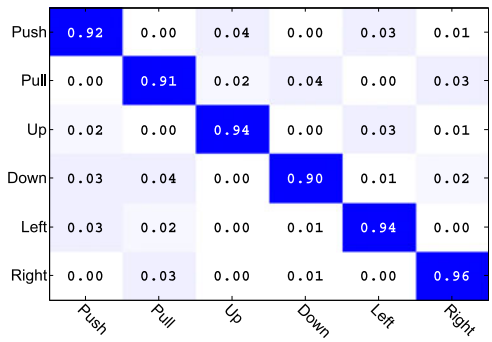


Fig. 10. Confusion matrix for the diverse-positions scenario.

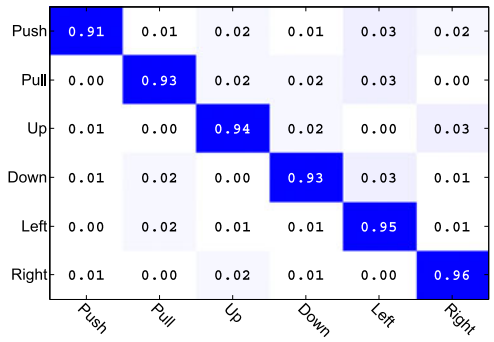


Fig. 11. Confusion matrix for multipath scenario.

6.4 Performance in the Presence of Environmental Interference

6.4.1 Impact of Multipath

We first test the accuracy of our system when the metallic board is placed nearby and a single volunteer is roaming around. Since performance in both identical-position and diverse-positions scenarios does not differ so much when multiple tags are utilized, we only focus on performance evaluation under different conditions in the former scenario in the following. Fig. 11 shows the average accuracy of gesture recognition in this case is about 93.7 percent, with a slight decrease of 2.8 percent compared with that in the identical-position scenario without much environmental interference. The results indicate that GRfid can retain robustness even under severe environmental interference.

6.4.2 Impact of Multiple Entities

To further explore the robustness of this system against dynamic interference, we conduct a set of experiments with two to four volunteers walking around in each experiment, respectively. Fig. 12 shows the performance varying with the number of roaming entities. It is obvious that within three persons walking around while performing gestures, the recognition accuracy is acceptable and when this number further increases, the accuracy drops a lot. Through this experiment, we can conclude that the distance between walking persons and the prototype significantly influences the overall system performance. In particular, when the distance is larger than a certain threshold, interference caused by entities' movement can be greatly mitigated.

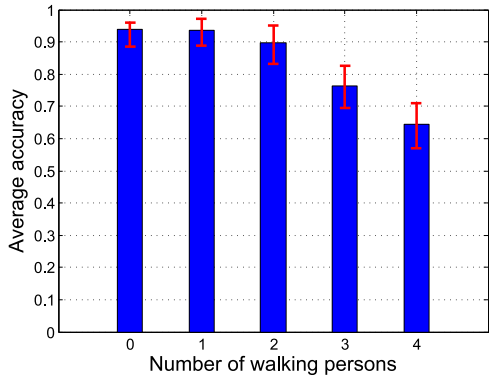


Fig. 12. Impact of multiple entities.

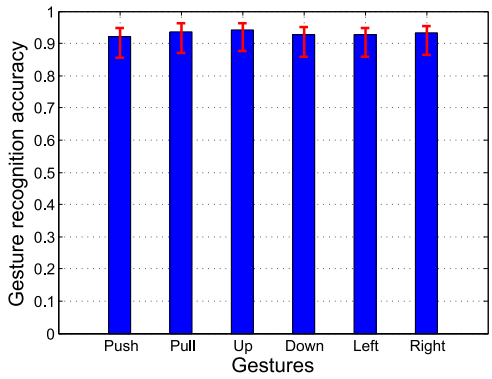


Fig. 13. Exploration of tags' orientations.

6.5 Exploration of GRfid Tag's Orientation

Thus far, GRfid has shown its success in gesture recognition in various scenarios, even in the presence of environmental interference. Yet it remains unclear that whether GRfid will retain its benefits when the tags are placed with varying orientations. To evaluate the effect of tag orientations on the performance of GRfid, we rotate tags around their own centers by 30 degree each time and repeat experiments as done in the identical-position scenario. We compute the average recognition accuracy of each gesture at orientations between 0 to 360 degree. Fig. 13 presents the results of this set of experiments. The accuracy of recognizing gestures in such a setting is relatively high, with a minimum value of 92.0 percent and an average of 93.1 percent. Compared with the accuracy in identical-position scenario (an average of 96.5 percent), the performance slightly degrades, which indicates tag orientation has slight effect on the performance of GRfid.

6.6 An Extension of GRfid-the Double-Antenna Case

In this set of experiments, we evaluate the performance of a kind of extension of GRfid, with a pair of Laird antennas connected to the same reader. Two antennas are placed at the top-left and top-right corner of tags respectively, as shown in Fig. 3. Intuitively, more antennas are beneficial to reduce position dependency when recognizing gestures in the diverse-positions scenario. As a result, we solely evaluate the performance of GRfid with multiple antennas in diverse-position experiment scenario. For the purposes of comparison, we show the results of double-antenna case (represented by dark red bars) together with results of single-antenna case (represented by blue bars) in Fig. 14. The results

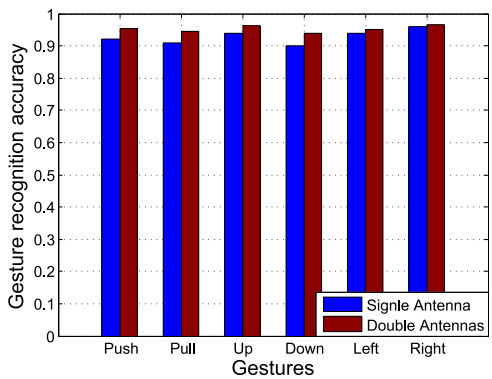
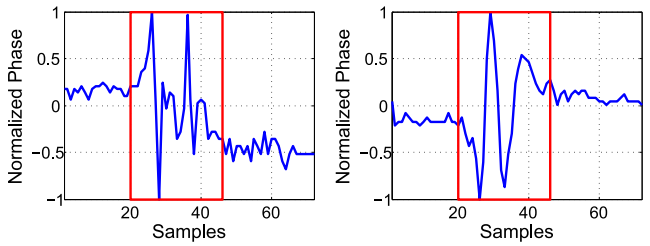
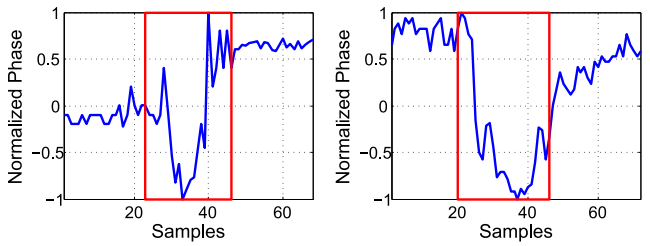


Fig. 14. Performance with multiple antennas.



(a) Signal of the top-left tag (b) Signal of the top-right tag



(c) Signal of the bottom-left tag (d) Signal of the bottom-right tag

Fig. 15. Phase Profile of 'Knock' gesture.

show that with two antennas, GRfid can achieve an average recognition accuracy of 95.9 percent, which achieves a 3.1 percent performance gain compared with the 92.8 percent accuracy in single-antenna case. Similar to the utilization of multiple tags, the rational of performance gain is that adding an antenna can enhance the capability of capturing hand movements, thus can improve gesture recognition accuracy.

7 Discussion and Future Work

In this section, we discuss several final thoughts on the future of "GRfid".

First, in this paper, we just take six basic representatives of them to verify the performance of our system, namely, 'Push', 'Pull', 'Up', 'Down', 'Left' and 'Right'. Fortunately, GRfid can be easily extended to recognize C_6^2 synthetic gestures, which are the arbitrary combinations of six gestures such as 'Pull-Push', 'Up-Left' and etc. To be specific, when a composite gesture is conducted, the corresponding signal sequence can be segmented to multiple su-segments, each of which corresponds to a basic gesture included. However, it should be emphasized that gestures that can be recognized by GRfid are not limited to the above ones and their combinations. Actually, more gestures with the same level of granularity can also be recognized by this system. We take a 'Knock' (that is, knocking a door or a wall) for example. The profile constructed for this gesture is shown in Fig. 15, which is different from profiles of the six gestures even performed in the same position. As a result, we claim that by various combinations of basic gestures, GRfid can perform a large number of gesture-based commands.

Second, we have to admit that gestures recognized by GRfid are at a relatively coarse granularity until now, compared with the dramatically fine-grained gestures tracing in RF-IDraw [19], which utilizes two readers and eight antennas and requires users to attach a tag on the finger. However, since phase provides rich knowledge about location, we envision it potentially to be an alternative choice for device-free gesture tracing with a fine-grained granularity by more complex system designs such as deploying multiple antennas or readers. We leave this more interesting and challenging topic as our future work.

8 Conclusion

In this paper, we demonstrate that phase information output by a COTS RFID reader unfolds new possibilities for device-free gesture recognition, without hardware modifications and instrumenting the environment much. Based on the observation that gesture is closely correlated with RFID signal phase, we propose a device-free gesture recognition system GRfid. We describe the key motivation, objectives, and design methodologies of GRfid. Extensive experiments have been conducted in various settings to evaluate its performance from different aspects, mainly covering the accuracy and robustness. Our experimental results show that GRfid can recognize gestures accurately in both identical-position and diverse-positions scenarios, with an average accuracy of 96.5 and 92.8 percent respectively. Moreover, we demonstrate that its overall performance can retain high even with environmental interference, and for diverse tag's orientations.

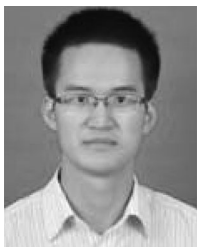
Acknowledgments

This research is supported in part by Guangdong Natural Science Funds for Distinguished Young Scholar (No. S20120011468), the Shenzhen Science and Technology Foundation (No. JCYJ20140509172719309, KQCX2015032416 0536457), China NSFC Grant 61472259, Guangdong Young Talent Project 2014TQ01X238, Hong Kong RGC Grant HKUST16207714, the University of Macau Grant SRG2015-00050-FST, by Guangdong S&T Project Grant 2013B051000003, and GDUPS (2015). This work is also supported partially by National Basic Research Program of China (973 Program) under Grant No. 2015CB351705, NSFC under Grant Nos. 61572396 and 61325013. This work is conducted when Lionel M. Ni was with HKUST. Kaishun Wu is the corresponding author.

References

[1] (2015, Apr. 6). Xbox Kinect [Online]. Available: http://www.xbox.com/en-US/kinect
[2] (2015, Apr. 6). Leap Motion [Online]. Available: https://www.leapmotion.com/
[3] (2015, Apr. 6). PlayStation Eye [Online]. Available: http://asia.playstation.com/hk/en/ps4
[4] (2015, Apr. 6). PointGrab [Online]. Available: http://www.pointgrab.com/
[5] M. Van den Bergh and L. Van Gool, "Combining rgb and tof cameras for real-time 3d hand gesture interaction," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2011, pp. 66–72.
[6] P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," *World Acad. Sci., Eng. Technol.*, vol. 49, no. 1, pp. 972–977, 2009.
[7] T. Schlömer, B. Poppinga, N. Henze, and S. Boll, "Gesture recognition with a wii controller," in *Proc. 2nd Int. Conf. Tangible Embedded Interaction*, 2008, pp. 11–14.
[8] V.-M. Mantyla, J. Mantyjarvi, T. Seppanen, and E. Tuulari, "Hand gesture recognition of a mobile device user," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2000, vol. 1, pp. 281–284.
[9] J. Rekimoto, "Smartskin: An infrastructure for freehand manipulation on interactive surfaces," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2002, pp. 113–120.
[10] J. Wu, G. Pan, D. Zhang, G. Qi, and S. Li, "Gesture recognition with a 3-d accelerometer," in *Proc. 6th Int. Conf. Ubiquitous Intell. Comput.*, 2009, pp. 25–38.

[11] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uwave: Accelerometer-based personalized gesture recognition and its applications," *Pervasive Mobile Comput.*, vol. 5, no. 6, pp. 657–675, 2009.
[12] (2015, Apr. 6). Myo [Online]. Available: https://www.thalmic.com/en/myo/
[13] F. Adib and D. Katabi, "See through walls with Wi-Fi!" in *Proc. ACM SIGCOMM Conf.*, 2013, pp. 75–86.
[14] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 27–38.
[15] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *Proc. 11th USENIX Conf. Netw. Syst. Design Implementation*, 2013, vol. 14, pp. 317–329.
[16] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.
[17] G. Borriello, "Rfid: Tagging the world," *Commun. ACM*, vol. 48, no. 9, pp. 34–37, 2005.
[18] 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.
[19] J. Wang, D. Vasisht, and D. Katabi, "Rf-idraw: Virtual touch screen in the air using rf signals," in *Proc. ACM SIGCOMM*, 2014, pp. 235–246.
[20] Q. Lin, L. Yang, Y. Sun, T. Liu, X.-Y. Li, and Y. Liu, "Beyond one-dollar mouse: A battery-free device for 3d human-computer interaction via rfid tags," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 1661–1669.
[21] 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.
[22] Y. Zhao, Y. Liu, and L. M. Ni, "Vire: Active Rfid-based localization using virtual reference elimination," in *Proc. Int. Conf. Parallel Process.*, 2007, pp. 56–56.
[23] J. Wang and D. Katabi, "Dude, where's my card?: Rfid positioning that works with multipath and non-line of sight," in *Proc. ACM SIGCOMM*, 2013, pp. 51–62.
[24] J. Wang, F. Adib, R. Knepper, D. Katabi, and D. Rus, "Rf-compass: Robot object manipulation using Rfids," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 3–14.
[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] L. Kriara, M. Alsup, G. Corbellini, M. Trotter, J. D. Griffin, and S. Mangold, "Rfid shakables: Pairing radio-frequency identification tags with the help of gesture recognition," in *Proc. 9th ACM Conf. Emerging Netw. Exp. Technol.*, 2013, pp. 327–332.
[27] M. Buettner, R. Prasad, M. Philipose, and D. Wetherall, "Recognizing daily activities with Rfid-based sensors," in *Proc. 11th Int. Conf. Ubiquitous Comput.*, 2009, pp. 51–60.
[28] R. Bainbridge and J. A. Paradiso, "Wireless hand gesture capture through wearable passive tag sensing," in *Proc. Int. Conf. Body Sens. Netw.*, 2011, pp. 200–204.
[29] J. Huiting, H. Flisijn, A. B. Kokkeler, and G. J. Smit, "Exploiting phase measurements of epc gen2 Rfid tags," in *Proc. IEEE Int. Conf. RFID-Technol. Appl.*, 2013, pp. 1–6.
[30] P. V. Nikitin, R. Martinez, S. Ramamurthy, H. Leland, G. Spiess, and K. Rao, "Phase based spatial identification of uhf Rfid tags," in *Proc. IEEE Int. Conf. RFID*, 2010, pp. 102–109.
[31] P. Melgarejo, X. Zhang, P. Ramanathan, and D. Chu, "Leveraging directional antenna capabilities for fine-grained gesture recognition," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 541–551.
[32] B. Kellogg, V. Talla, and S. Gollakota, "Bringing gesture recognition to all devices," in *Proc. 11th USENIX Conf. Netw. Syst. Des. Implementation*, 2014, pp. 303–316.
[33] Impinj, "Speedway revolution reader application note - low level user data support," 2013.
[34] K. Itoh, "Analysis of the phase unwrapping algorithm," *Appl. Opt.*, vol. 21, no. 14, pp. 2470–2470, 1982.
[35] M. U. Bromba and H. Ziegler, "Application hints for Savitzky-Golay digital smoothing filters," *Analytical Chemistry*, vol. 53, no. 11, pp. 1583–1586, 1981.
[36] H. Azami, K. Mohammadi, and B. Bozorgtabar, "An improved signal segmentation using moving average and Savitzky-Golay filter," *Int. J. Comput. Appl.*, vol. 3, pp. 39–44, 2012.
[37] H. Azami, K. Mohammadi, and H. Hassanpour, "An improved signal segmentation method using genetic algorithm," *Int. J. Comput. Appl.*, vol. 29, no. 8, pp. 5–9, 2011.
[38] A. Dik, K. Jebari, A. Bouroumi, and A. Ettouhami, "Similarity-based approach for outlier detection," *arXiv Preprint arXiv:1411.6850*, 2014.
[39] C. A. Ratanamahatana and E. Keogh, "Everything you know about dynamic time warping is wrong," in *Proc. 3rd Workshop Mining Temporal Sequential Data*, 2004, pp. 22–25.
[40] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intell. Data Anal.*, vol. 11, no. 5, pp. 561–580, 2007.



Yongpan Zou received the BEng degree in chemical machinery from Xi'an Jiaotong University, Xi'an, China. Since 2013, he has been working toward the PhD degree in the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology (HKUST). His current research interests mainly include: mobile computing, embedded systems, and wireless communication. He is a student member of the IEEE.

Jiang Xiao received the PhD degree in computer science and engineering from the Hong Kong University of Science and Technology (HKUST) in 2014. She is currently a research assistant professor in the Fok Ying Tung Graduate School at the Hong Kong University of Science and Technology. Her research interests include wireless communication, mobile computing, indoor localization, and big data. She is a member of the IEEE.

Jinsong Han received the PhD degree in computer science and engineering from the Hong Kong University of Science and Technology in 2007. He is currently an associate professor in the Department of Computer Science and Technology in Xian Jiaotong University. His research interests focus on mobile computing, RFID, and wireless network. He is a member of the IEEE, the IEEE Computer Society, and the ACM.

Kaishun Wu received the PhD degree in computer science and engineering from the Hong Kong University of Science and Technology, in 2011. After that, he worked as a research assistant professor in the Hong Kong University of Science and Technology. From 2013, he joined Shenzhen University as a distinguished professor. He is the inventor of 6 US and 43 Chinese pending patents (13 are issued). He received the best paper awards in IEEE Globecom 2012, IEEE ICPADS 2012, and IEEE MASS 2014. He received the 2014 IEEE ComSoc Asia-Pacic Outstanding Young Researcher Award and was selected as 1,000 Talent Plan for Young Researchers. He is a member of the IEEE.

Yun Li received the PhD degree in power engineering and engineering thermo-physics from Xi'an Jiaotong University in 1999. She is currently a professor in the Department of Chemical Machinery and Equipment Engineering at Xian Jiaotong University. Her research interests include: efficient chemical fluid mechanical, energy-saving heat pump system, and energy integration technology in process systems. She is a member of the Chemical Industry and Engineering Society of Shaanxi Province and is a senior member of the Chinese Association of Refrigeration.

Lionel M. Ni received the PhD degree in electrical and computer engineering from Purdue University in 1980. He is a chair professor in the Department of Computer and Information Science and a vice rector of academic affairs at the University of Macau. Previously, he was a chair professor of computer science and engineering at the Hong Kong University of Science and Technology. He has chaired more than 30 professional conferences and has received eight awards for authoring outstanding papers. He a fellow of the IEEE and the Hong Kong Academy of Engineering Science.