WiG: WiFi-based Gesture Recognition System

Wenfeng He*, Kaishun Wu*, Yongpan Zou†, Zhong Ming*
*Guangdong Province Key Laboratory of Popular High Performance Computers,
College of Computer Science and Software Engineering, Shenzhen University, China
†Department of Computer Science and Engineering, Hong Kong University of Science and Technology
he_wenfeng@foxmail.com, yzouad@cse.ust.hk, {wu,mingz}@szu.edu.cn

***Abstract*—Most recently, gesture recognition has increasingly attracted intense academic and industrial interest due to its various applications in daily life, such as home automation, mobile games. Present approaches for gesture recognition, mainly including vision-based, sensor-based and RF-based, all have certain limitations which hinder their practical use in some scenarios. For example, the vision-based approaches fail to work well in poor light conditions and the sensor-based ones require users to wear devices. To address these, we propose WiG in this paper, a device-free gesture recognition system based solely on Commercial Off-The-Shelf (COTS) WiFi infrastructures and devices. Compared with existing Radio Frequency (RF)-based systems, WiG stands out for its systematic simplicity, extremely low cost and high practicability. We implemented WiG in indoor environment and conducted experiments to evaluate its performance in two typical scenarios. The results demonstrate that WiG can achieve an average recognition accuracy of 92% in line-of-sight scenario and average accuracy of 88% in the none-line-of-sight scenario.**

I. Introduction

Gesture recognition has recently been a hot topic in academia and industry for greatly promoting Human-Computer Interface (HCI). It enables users to convey commands to interact with devices conveniently just by performing gestures, and thus is broadly applied in our daily life. For example, in a smart house, people can remotely control household equipments such as TV, air conditioner, refrigerator, by doing simple gestures without any extral controller. Another typical application of HCI based on gesture recognition is interactive devices for mobile games, which set users' hands free from control handles.

In response to this promising trend, trendous systems have been proposed in papers or released in the market as commerical products. Overall speaking, these systems can be classified into three main categories according to their designing principles, namely, vision-based systems [4], [2], [13], [23], [32], [14], sensor-based systems [28], [19], [26], [33], [18], [22] and RF-based systems [8], [9], [24]. Vision-based gesture recognition systems make use of cameras and computer vision techniques to recognize gestures. However, they are restricted by their high dependence on Line-of-Sight (LOS) and light conditions, as well as accompanying privacy issues, which makes them unbefitting in certain application scenarios. Sensor-based systems employ various kinds of sensors as gestures input interface, but require users to wear devices with them.

RF-based systems open up new perspectives of gesture recognition, which utilize radio frequency signals as medium for interpreting gestures, without the requirements of LOS or wearing any devices. In this kind of systems, indicators such as Received Signal Strength Indicator (RSSI) [16], Doppler shifts [24], Time of Fly (ToF) [8] have been used for distinguishing different gestures. However, these existing systems either rely on specialized RF instruments/devices [9], [8], or need modifications of commercial devices [24], or are overly susceptible to interference [16], which consequently limit their pervasiveness to a great extent.

To step further, we ask such a question: *Is it possible to design a system that can achieve robust performance based solely on existing COTS RF infrastructures (namely, WiFi) and devices without any modifications?* In this paper, we give a positive response to this question by proposing WiG, satisfying the above requirements. In a high level, the design intuition of WiG is stimulated by the observation that Channel State Information (CSI) is a more fine-grained indication of a single signal stream, thanks to the Orthogonal Frequency Division Multiplexing (OFDM) modulation scheme [30]. As a result, CSI is demonstrated to be a more robust indicator even with interference for indoor localization [34]. We envision that such a property can be extendedly applied to gesture recognition.

The idea is straightforward; nevertheless, two key challenges remain to be dealt with in order to realize it.

First, how to extract valuable fearures of different gesrures carried by CSI values? Essentially, a gesture recognition process is a mapping process from the physical world to the digital world. Thus, extracting digital features of each gesture, represented by CSI values, is a prerequisite for this mapping process. Although CSI is expected to present different changes to gestures, it is non-trivial to extract the corresponding features from these noisy CSI changes to portray gestures in the physical world.

Second, how to classify various gestures based on their corresponding features extracted already? Although features extracted from CSI reflect characteristics of gestures, they can not be applied directly as unique ID to distinguish them. It is because, for a certain gesture, any single feature can not represent it completely. As a result, a remaining challenging task is to devise a proper method which can jointly integrate features to distinguish gestures reliably.

To sum up, in this paper, we make the following main contributions.

- To the best of our knowledge, this is the first attempt to achieve a fine-grained gesture recognition only by leveraging wireless signal feature information from ubiquitous COTS Wi-Fi cards and a common router.
- We introduce algorithms to detect the abnormal CSIs

978-1-4799-9964-4/15/$31.00 ©2015 IEEE

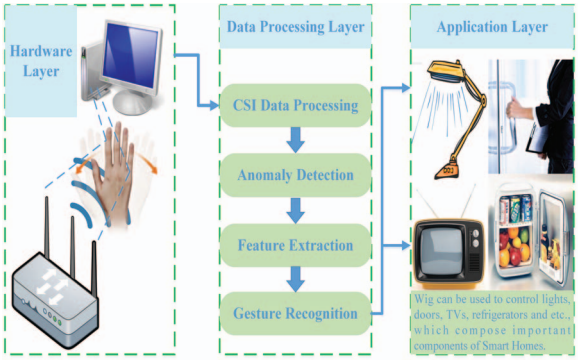


Fig. 1. System overview of WiG

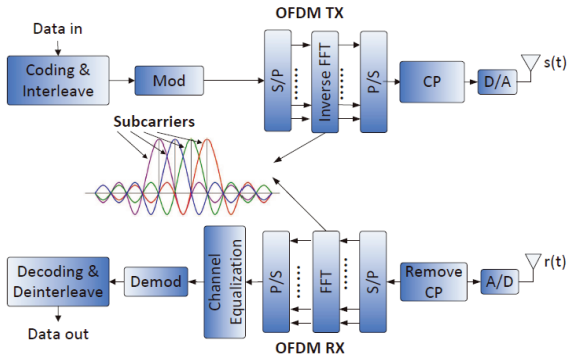


Fig. 2. OFDM Framework

and achieve gesture recognition from COTS devices. And we implement WiG with ubiquitous COTS Wi-Fi cards and a common router. Experimental results demonstrate that WiG can achieve an average recognition accuracy of 92% in line-of-sight scenario and an average accuracy of 88% in the none-line-of-sight scenario.

The rest of this paper is organized as follows. We first summarize the related work in Section 2, followed by section 3, which is an overview of wireless technology that relevant to our work. Then in section 4 and 5, we illustrate the detailed system design and methodology. We show the experimental results and evaluation of the performance in section 5. Finally, we will conclude this work and list our future work in section 6.

II. Related Work

In this section, we introduce the state-of-the-art work of wireless systems and gesture recognition systems, which are related to our work.

(1)**Wireless Systems:** We can classify the related work into two parts: wireless localization and wireless motion detection. When it comes to wireless localization, there are lots of technologies, such as coarse-grained RSSI [21], [38], [39], fine-grained CSI [36], [34], fingerprint method [35], sensor-based method and MIMO technology [34]. FILA [34] is a good case in point, which leverages the CSI to alleviate multipath effect on COTS 802.11 NIC. WiG builds on this prior work but aims to achieve gesture recognition. What's more, WiG is relevant to wireless motion detection. With wireless technology developing so fast, recent years have witnessed a trend in motion detection systems, owing to existing infrastructure without additional cost. Wi-Vi [9], Wi-See [24], All-See [16] and WiTrack [8] can be the good examples for it, which are based on feature information extracted form wireless signals. They have showed the feasibility of motion detection even gesture recognition by leveraging the wireless signals.

(2)**Gesture Recognition Systems:** To the best of our knowledge, the existing gesture recognition systems can be classified into vision-based, sensor-based, RF-based. The research of vision-based gesture recognition can date back in the 1990's, Rehg et al. [25] used cameras to gain hand gesture. In recent years, we can enable in-air 3D gesture-based interaction by using depth sensing and computer vision [27]. The Xbox Kinect [4] and Leap Motion [2] can be the typical examples of applications, which are examples of the successful commercial applications. However, vision-based approaches cannot work in dark scenarios due to the requirement of light and may raise privacy issues, which is a very sensitive topic among the public. With the prevalence of wearable sensing device, such as Microsoft Band [3], Moto 360 [5] and Apple Watch [1], we can get a rich series of signal features from them, which help us to achieve gesture recognition. Some researchers even use accelerometer sensors, surface electromyography(SEMG) sensors [40], [37] and magnetic sensors [17] to recognize different gestures. Though they can work in the dark scenarios, they usually need people to wear additional devices, which is inconvenient, sometimes.

Specifically, there are some approaches to achieve gesture recognition by using radio frequency(RF) for the reason that it is not only independent of light conditions, but also can work in dark scenarios. Wi-Vi [9] uses ISAR technology to detect motions. Wi-See [24] extracts Doppler effect to recognize the gestures. All-See [16] and WiTrack [8] achieve gesture recognition based on home-made hardware or USRP [6]. Can we achieve a gesture recognition only by leveraging wireless signal feature information from ubiquitous COTS Wi-Fi cards and a common router instead of USRP or expensive WARP [7], [20] or some home-made hardware [16], [8]? In this paper, we present algorithms that allow us to extract feature information of gesture without additional devices based on COTS Wi-Fi cards.

III. Overview of WiG System

In this section, we first briefly introduce the background knowledge of OFDM system and CSI value from it which is the foundation of WiG system, then give an overview of the system.

A. Background

Orthogonal Frequency Division Multiplexing(OFDM) [31], [35] is a multi-carrier modulation scheme that is used for the wireless and telecommunications standards, such as IEEE 802.11a/g/n. In OFDM system, the incoming data stream is split into multiple narrow and orthogonally overlapped sub-carriers, as depicted in Fig.2. Then the data on each subcarrier is modulated and converted to time domain via an inverse Fast Fourier Transform(IFFT), followed by paralled to serial(P/S) and digital to analog(DAC) conversion process. The analog signals are summed to give the transmission signal. On the receiver, the signals are sampled, passed on to a demodulation

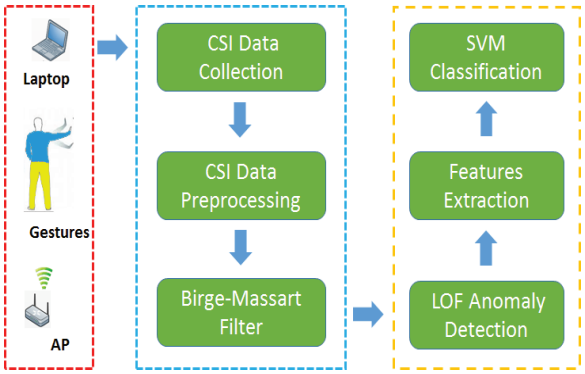


Fig. 3. Framework of WiG

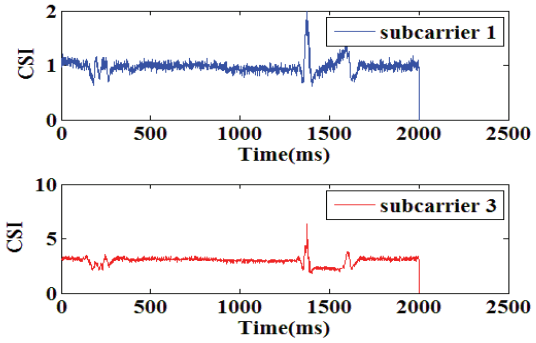


Fig. 4. CSI variance of different subcarriers in the same stream

process chain and digitized by analogue-to-digital converters(ADCs). Next, the FFT procedure processes the data sample blocks to convert back into the frequency domain.

In OFDM system, channel response can be extracted in the format of Channel State Information(CSI) [15], which is a fine-grained PHY layer information that estimates the channel property of a communication link at the subcarrier level. Further, CSI reveals a set of channel measurements describing the amplitudes and phases of every subcarrier and the combined effect, for instance, fading, scattering, and power decay with distance, which helps us to analyze the signal propagation while performing the gestures.

From prior work, we can know that in a narrowband flat-fading channel, the OFDM system in frequency domain can be modeled as

$$Y = HX + N \tag{1}$$

where Y and X are the received and transmitted vectors, respectively, and H and N are the channel matrix and the noise vector which is named as additive white Gaussian noise, respectively. According to (1), the value of H can be estimated as

$$\hat{H} = \frac{Y}{X} \tag{2}$$

CSI of a single subcarrier is mathematically defined as

$$h = |h|e^{j \sin \angle h} \tag{3}$$

where $|h|$ is the amplitude while the $\angle h$ is the phrase of each carrier.

B. Overview of WiG System

WiG is a wireless system that enable commercial Wi-Fi devices to recognize people's gestures based on the WLAN infrastructure without additional deployment or home-made hardware or other special devices. WiG leverages CSI information to achieve gesture recognition in indoor scenarios. On the whole, WiG includes hardware performing as gesture sensing and software acting data analysis. Fig.3 gives an overview of the system. In our design, WiG consists of two hardware elements: a common router acted as access point(AP), a desktop with off-the-shelf wireless cards(e.g., Intel 5300 NIC) as detecting point(DP). The AP propagates wireless signals while the person performs gestures between the AP and DP. In the meanwhile, the DP collects the CSI information and sends to further analysis.

As depicted in the Fig.3, There are six functional blocks in the software area, i.e., CSI data collection, CSI data preprocessing, Birge-Massart filter, LOF anomaly detection, features extraction and SVM classification. Since the wireless signals can be affected by the change of environment, the noise should be filtered out. After CSI data preprocessing block, we choose birge-massart method to denoise the wireless signal. Then WiG performs LOF anomaly detection algorithm to detect outliers in time series, followed by feature extraction, which helps us to obtain characteristics of gesture. At last, a SVM classifier is used to classify the gestures which have been performed.

IV. Methodology

In this section, we elaborate the methodology of WiG relied on six functional blocks as mentioned before, i.e., CSI data collection and preprocessing, Birge-Massart filter, LOF anomaly detection, features extraction and SVM classification. Based on Fig.5, the methodology of WiG system can be broken down into four following steps.

A. CSI Data Collection and Preprocessing

We start by using a desktop(DP) equipped with Intel 5300 NIC to receive the beacon message from the AP in the indoor scenario. From the DP, we obtain the raw CSI data based on OFDM system. Since multiple antennas can bring spatial diversity in wireless communication system, MIMO technology has been employed in IEEE 802.11n/ac devices. In terms of communication theory, the capacity of a MIMO channel is $min\{m, n\}$ times of a corresponding channel with a single antenna, where m and n are the number of antennas of receiver and transmitter, respectively. In order to get more information about CSI, we use MIMO technology for multiplying the capacity of a radio link using three antennas for transmitting and two antennas for receiving to form a 3×2 MIMO system to exploit multipath propagation. As a result, the CSI data can be divided into 6 streams and has 30 subcarriers in each stream while we can expect better accuracy with additional CSI information. Specifically, there are 180 groups of CSI data from each packet, which can be described in the following format:

$$\mathbf{CSI} = \begin{bmatrix} H_{1,1} & H_{1,2} & \dots & H_{1,30} \\ H_{2,1} & H_{2,2} & \dots & H_{2,30} \\ \vdots & \vdots & \ddots & \vdots \\ H_{6,1} & H_{6,2} & \dots & H_{6,30} \end{bmatrix} \quad (4)$$

where $H_{i,j}$ is the CSI value of each subcarrier and in $H_{i,j}$, i is the indicator of stream and j is the indicator of subcarrier number.

In Fig.4, the blue line refers to the CSI in the $1th$ subcarrier of a stream, while the red one represents for the CSI of the $3th$ subcarrier in the same stream. Maybe the absolute value of the two CSIs are different, they have the similar pattern under the influence of gesture motions, which help us to analyze and recognize different gestures.

B. Birge-Massart Filter

The CSI can be the raw data that with some random noise for the reason that the wireless signal may be affected by the change in the indoor environment such as air pressure and temperature. As a result, we apply a wavelet-based denoising scheme to remove any random noise and smooth the CSI data. Compared with fourier-based denoising method, it captures both frequency and time domain information. The denoising key idea is to conserve only the greatest wavelet coefficients and put the noise at zero in thresholding step before reconstruction of the signal. The selection of the threshold is so import in denoising that we choose the non-parameter adaptive density estimate theory of Birge-Massart [10], which is provided by Birge and Massart in 1997. We can get the threshold from the following rules.

1)Let m be the decomposition level, to decompose the signal in p layers and reserve all coefficients in the $p+1$ layer;

2)To the coefficient of the q layer($1 \leq q \leq p$), reserving the maximal absolute value of n_q coefficients by equation(5).

$$n_q = M(p+2-q)^b \quad (5)$$

where M and b are the length of the first layer decompose coefficients and 3 in the denoising situations, respectively.

C. LOF Anomaly Detection

After CSI data processing phase, we obtain the cleaned CSI data. Then anomaly detection should be performed, which plays an important role in the overall performance of the WiG system while is full of challenge. On one hand, we should detect the anomaly segment corresponding to gestures exactly. On the other hand, the pattern of gestures are not always the same for the reason that we may perform at different speed or at different range at each time. For example, when we keep on performing the same gesture for about 50 times, we may feel tired and slow down the speed.

In order to solve the problems, we choose Local Outlier Factor(LOF) [11] based anomaly detection algorithm to detect the anomaly segment. LOF was first put forward by Markus et al. to detect the anomalous data points based on the method of

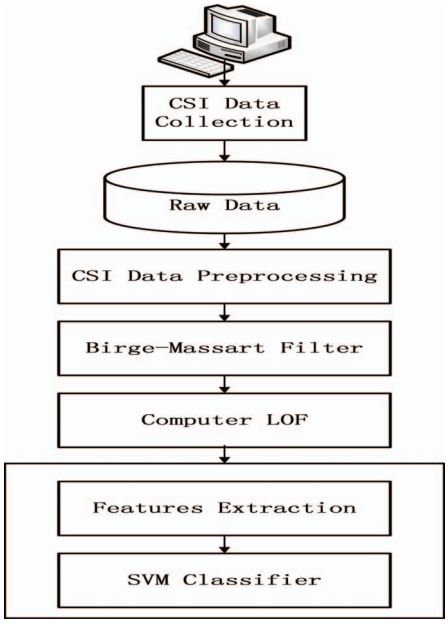


Fig. 5. WiG Flow Chart

comparing the local density of an object to the local densities of its neighbors. Previously, We introduce some basic concepts such as the $k\text{-}distance$ of an object p and the $reachability\ distance$ of p w.r.t. object t to pave the way for LOF.

Let k and S be a positive integer and a set of data points, respectively. The $k\text{-}distance$ of an object p is called $k\text{-}D(p)$ and $D(p,x)$ means the distance between p and x. They can be defined as:

1)There are at least k objects in S when $x \in S\backslash\{p\}$, $D(p,x) \leq k\text{-}D(p)$.

2)There are at most $k-1$ objects in S when $x \in S\backslash\{p\}$, $D(p,x) < k\text{-}D(p)$.

Then comes the $reachability\ distance$ of p w.r.t. t, named $R\text{-}D_k(p,t)$. The reachability distance between p and t can be modeled as:

$$R\text{-}D_k(p,t) = max(k\text{-}D(p), D(p,t)) \quad (6)$$

where $D(p,t)$ describes the distance between p and t.

Based on the above concept, the local reachability density of an object p is defined as

$$Lrd(p) = \frac{k}{\sum_{x \in k(p)} R\text{-}D_k(p,x)} \quad (7)$$

where $k(q)$ is the set of $k-nearest$ neighbors of p. From equation(7), we know that $Lrd(p)$ is the inverse of the average reachability distance of object p from its neighbors. More specifically, it reveals the density of these points around p. The more dense, the closer they are to p. In other words, p has less possibility to be a outlier.

Local Outlier Factor of an object p is defined as

$$LOF(p) = \frac{\sum_{o \in k(p)} \frac{Lrd(o)}{Lrd(p)}}{k} \quad (8)$$

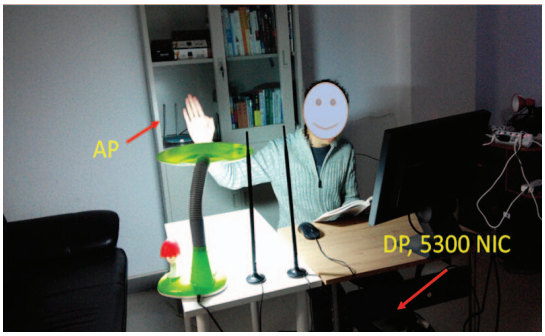


Fig. 6. Line-of-sight Scenario

Fig. 7. None-line-of-sight Scenario

From the equation(8), we know that LOF is the ratio of average of local densities of $p's$ neighbors to the local reachability density of p. It is not hard to find that LOF demotes the degree of outlier-ness.The higher the local reachability densities of $p's$ neighbors are and the lower $p's$ local reachability density is, the highter is the LOF value of p.

D. Features Extraction and SVM Classification

After the above steps, we obtain the anomalous data series, which are caused by different gestures. Now it's of importance for us to extract the features that can stand for different gestures from the anomaly data series. Based on lots of experiment, we choose the follow features: (1)the mean value(MEAN) of the anomaly patterns, (2)the standard deviation(STD) of the anomaly patterns, (3) the median absolute deviation(MAD) of the anomaly patterns, (4)the max value(MAX) of the anomaly patterns. Fig.8 shows the examples of the four features of some gestures.

As we have extracted the most representative features of different gestures, we should choose a correct classifier to recognize gesture signatures. In this paper, we apply the Support Vector Machine(SVM) [29] to classify different gestures. In machine learning, SVM is a supervised learning models that usually involves separating data into training and testing sets. In training sets, each instance contains several features called attributes and one target value named label. The main idea of SVM is to train a model that can predict the label of the test data given by its attributes. As a result, We select a subset of representative gesture data sets to train a model.

To put it simply, a SVM constructs a hyperplane, by which a good separation is achieved that has the largest distance to the nearest data point of any class. However, the data sets are not always linearly separable in original space. Therefore, the original space is required to be mapped into a higher-dimensional space, which can discriminate the data sets. The non-linear SVM model can formally described as:

$$\min_{w} \frac{\|\mathbf{w}\|^2}{2} \tag{9}$$

$$\text{subject to } y_i(\mathbf{w}\cdot(\Phi(\mathbf{x}_i)+b)\geq 1, \quad i=1,2,\cdots,N$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id})^T$ is corresponding to the attributes of the ith sample and $\mathbf{w}$ is the parameter of SVM model.

Since the objective function is a quadratic function, which can be solved by using Lagrange multipliers, we can get the the Lagrangian function of it as

$$L_D = \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\sum_{i,j}\lambda_i\lambda_j y_i y_j \Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j) \tag{10}$$

Based on the quadratic programming, we can obtain the λ_i, w and b. According to SVM, a discriminant function can be defined as:

$$f(\mathbf{z}) = sign(\mathbf{w}\cdot\Phi(\mathbf{z})+b) \tag{11}$$

$$= sign(\sum_{i=1}^{n}\lambda_j y_j \Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{z})+b) \tag{12}$$

On account of keeping equation(12)'s computational load reasonable, kernel function is introduced to solve this problem. In this paper, we choose the radial basis function(RBF) kernel that can nonlinearly map samples into a higher dimensional space, so it can deal with the case that the class attributes and labels are not linear.

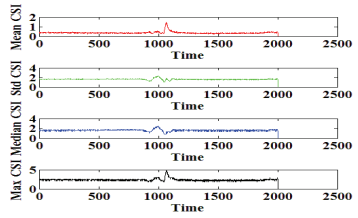
Generally speaking, SVM is designed for binary classification. But in our system, there are different gestures to be classified. Here we choose LIBSVM, a open source machine learning libraries written by Chang and Lin for classify. LIBSVM [12] applys a method named one-against-one method to construct $k(k-1)$ classifiers where each one is trained on data from two classes to classify these two classes and k is the number of classes. Then a voting strategy is used to help us predict the test data.

V. Experimentation And Evaluation

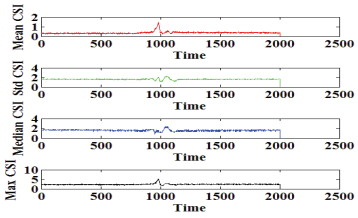
In this section, we first introduce the prototype implementation of WiG and describe the details of experimental settings. Then we evaluate the performance of WiG.

A. implementation

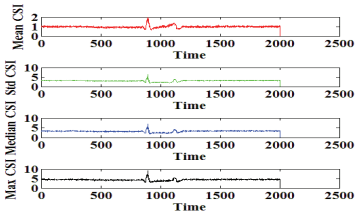
1) Hardware and sofware: In our experiment, the proposed methodology is implemented on TL-WR882N router manufactured by TP-LINK technologies CO.,Ltd. as the transmitted AP and a LENOVO desktop with 3.2GHz Intel(R) Pentium 4 CPU and 4G RAM. The desktop is equipped with Intel 5300 NIC as the receiver. The AP runs in the 2.4GHz frequency and has three antennas while the desktop has two antennas. As a result, they can form 3×2 MIMO system to achieve spacial diversity. Based on the method proposed by Dan [15],



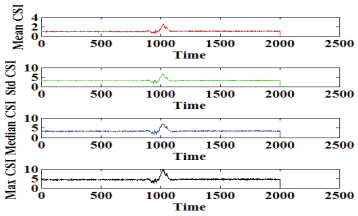
(a) 'Right'



(b) 'Left'

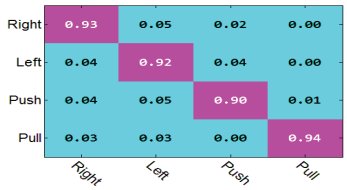


(c) 'Push'

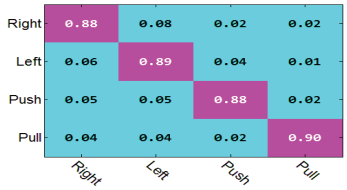


(d) 'Pull'

Fig. 8. Features of different gestures.



(a) Confusion Matrix for LOS Scenario



(b) Confusion Matrix for NLOS Scenario.

Fig. 9. Results of two different Scenario.

we modify the driver to collect CSI values from Intel 5300 NIC.

As for software in our system, to accomplish the task of data collection and analysis, our testbed runs 32-bit Ubuntu Linux, version 10.04LTS of the Desktop Edition to collect CSI data, which is a fine-grained PHY layer information that describes the channel property of a radio frequency link at the subcarrier level and leverages MATLAB 7.11.0 (R2010b) to analyze CSI data.

2) Experiment scenario: We conduct experiments to show the performance of WiG system in two typical indoor scenarios in the research laboratory of Shenzhen University as follows:

1)**Line of sight**: Firstly, we set up a testbed in a $5m\times 6m$ single room. The AP is placed in the bookshelf. The DP continuously pings packets from AP at the rate of 100 packets per second. The volunteers perform predefined gestures on the line of sight range between AP and DP(Fig.6). Specially, we make gestures in front of the receiving antennas for very close distance to reduce the irrelevant multipath effects.

1)**None line of sight**: Secondly, the experiment is conducted in a a place that covers two rooms. The AP and DP are separated by a wall(roughly $10cm$). The volunteers are on the same side as the DP(Fig.7). They perform the same gestures as before.

B. Evaluation

We use confusion matrix, which each column represents the instances in a predicted class and each row represents the instances in an actual class, to evaluate the performance of WiG system.

1)WiG's Performance in Line-of-sight Scenario: We investigate the performance of WiG in the Line-of-sight scenario. Fig.9(a) shows the gesture recognition results in the form of a confusion matrix. The average recognition accuracy(i.e, the true positive rate) of four gestures is about 92%, which indicates that all the predefined gestures can be recognized with high accuracy. What's more, it reveals that we can achieve a fine-grained gesture recognition by leveraging wireless signal feature information from ubiquitous COTS Wi-Fi cards and a common router.

2)WiG's Performance in None-line-of-sight Scenario: We also evaluate the performance in the none-line-of-sight scenario. Fig.9(b) shows the confusion matrix of it, which can be clearly shown that the average accuracy of gesture recognition even in such a none-line-of-sight scenario can remain relatively high, with a average value of 88%. In addition, compared to the line-of-sight case, there is only slight performance degradation of 4%. Though the thick wall and multipath effects have some negative influence on the accuracy of gesture recognition, applying the wireless signal feature information and matching algorithm, we still get a acceptable accuracy of gesture recognition, which ensures the robustness of this system.

VI. Conclusions and Future Work

In this paper, we propose to accomplish gesture recognition by leveraging a more fine-grained indicator Channel State Information (CSI), which can be output by COTS WiFi devices without hardware modification. WiG precedes other relevant systems with a key property that WiG achieves a relatively high accuracy in gesture recognition without any specialized devices and hardware modification. We implemented WiG by carefully addressing feature extraction and gestures classification problems. Experimental results show that WiG can achieve an average accuracy of 92% in LOS scenario and 88% in NLOS scenario. Moreover, Our experimental results show that WiG can achieve sufficient accuracy with the proposed CSI-based features.

In the perspective of future research, we intend to recognize more different gestures. Furthermore, we may try to recognize gestures of different objects. Since 802.11ad operates in the 60 GHz band, it has more bandwidth than what's available in the 2.4 GHz and 5 GHz bands combined. Additionally, 802.11ad

can use directional antennas to focus the radio beam into a six-degree angle, which can not only improve the accuracy of gesture recognition, but also be used for in-room applications. We leave this more interesting and challenging topic as our future work.

ACKNOWLEDGMENT

This research is supported in part by Program for New Century Excellent Talents in University (NCET-13-0908), Guangdong Natural Science Funds for Distinguished Young Scholar (No.S20120011468), the Shenzhen Science and Technology Foundation (Grant No.JCYJ20140509172719309), China NSFC Grant 61472259, 61170077, NSF , S&T project of GDA:2012B091100198, S&T project of SZ JCYJ20130326110956468.

REFERENCES

[1] Apple watch. https://www.apple.com/watch.

[2] Leap motion. https://www.leapmotion.com.

[3] Microsoft band. http://www.microsoft.com/Microsoft-Band/en-us.

[4] Microsoft kinect. http://www.microsoft.com/en-us/kinectforwindows.

[5] Moto 360. https://moto360.motorola.com.

[6] Usrp. http://www.ni.com/sdr/usrp.

[7] Warp v3. http://mangocomm.com/products/kits/warp-v3-kit.

[8] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller. 3d tracking via body radio reflections. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 317–329, 2014.

[9] F. Adib and D. Katabi. See through walls with wifi! In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pages 75–86. ACM, 2013.

[10] A. Bhandari, V. Khare, M. Trikha, and S. Anand. Wavelet based novel technique for signal conditioning of electro-oculogram signals. In *India Conference, 2006 Annual IEEE*, pages 1–6. IEEE, 2006.

[11] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. Lof: identifying density-based local outliers. In *ACM sigmod record*, volume 29, pages 93–104. ACM, 2000.

[12] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[13] P. Eye. http://asia.playstation.com/hk/en/ps4.

[14] P. Garg, N. Aggarwal, and S. Sofat. Vision based hand gesture recognition. *World Academy of Science, Engineering and Technology*, 49(1):972–977, 2009.

[15] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM CCR*, 41(1):53, Jan. 2011.

[16] B. Kellogg, V. Talla, and S. Gollakota. Bringing gesture recognition to all devices. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 303–316, 2014.

[17] H. Ketabdar, P. Moghadam, B. Naderi, and M. Roshandel. Magnetic signatures in air for mobile devices. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion*, pages 185–188. ACM, 2012.

[18] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6):657–675, 2009.

[19] V.-M. Mantyla, J. Mantyjarvi, T. Seppanen, and E. Tuulari. Hand gesture recognition of a mobile device user. In *Multimedia and Expo, 2000. IEEE International Conference on, (ICME)*, volume 1, pages 281–284.

[20] P. Melgarejo, X. Zhang, P. Ramanathan, and D. Chu. Leveraging directional antenna capabilities for fine-grained gesture recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 541–551. ACM, 2014.

[21] D. Moore, J. Leonard, D. Rus, and S. Teller. Robust distributed network localization with noisy range measurements. In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 50–61. ACM, 2004.

[22] Myo. https://www.thalmic.com/en/myo/.

[23] PointGrab. http://www.pointgrab.com/.

[24] Q. Pu, S. Gupta, S. Gollakota, and S. Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 27–38. ACM, 2013.

[25] J. M. Rehg and T. Kanade. Visual tracking of high dof articulated structures: an application to human hand tracking. In *Computer VisionECCV'94*, pages 35–46. Springer, 1994.

[26] J. Rekimoto. Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120. ACM, 2002.

[27] E. N. Saba, E. C. Larson, and S. N. Patel. Dante vision: In-air and touch gesture sensing for natural surface interaction with combined depth and thermal cameras. In *Emerging Signal Processing Applications (ESPA), 2012 IEEE International Conference on*, pages 167–170. IEEE, 2012.

[28] T. Schlömer, B. Poppinga, N. Henze, and S. Boll. Gesture recognition with a wii controller. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 11–14. ACM, 2008.

[29] J. A. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[30] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[31] J.-J. Van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. Ola Borjesson. On channel estimation in ofdm systems. In *Vehicular Technology Conference, 1995 IEEE 45th*, volume 2, pages 815–819. IEEE, 1995.

[32] M. Van den Bergh and L. Van Gool. Combining rgb and tof cameras for real-time 3d hand gesture interaction. In *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, pages 66–72. IEEE, 2011.

[33] J. Wu, G. Pan, D. Zhang, G. Qi, and S. Li. Gesture recognition with a 3-d accelerometer. In *Ubiquitous intelligence and computing*, pages 25–38. Springer, 2009.

[34] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni. Fila: Fine-grained indoor localization. In *INFOCOM, 2012 Proceedings IEEE*, pages 2210–2218. IEEE, 2012.

[35] J. Xiao, K. Wu, Y. Yi, and L. M. Ni. Fifs: Fine-grained indoor fingerprinting system. In *Computer Communications and Networks (ICCCN), 2012 21st International Conference on*, pages 1–7. IEEE, 2012.

[36] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni. Fimd: Fine-grained device-free motion detection. In *ICPADS*, pages 229–235, 2012.

[37] R. Xu, S. Zhou, and W. J. Li. Mems accelerometer based nonspecific-user hand gesture recognition. *Sensors Journal, IEEE*, 12(5):1166–1173, 2012.

[38] D. Zhang, J. Ma, Q. Chen, and L. M. Ni. An rf-based system for tracking transceiver-free objects. In *Pervasive Computing and Communications, 2007. PerCom'07. Fifth Annual IEEE International Conference on*, pages 135–144. IEEE, 2007.

[39] D. Zhang and L. M. Ni. Dynamic clustering for tracking multiple transceiver-free objects. In *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, pages 1–8. IEEE, 2009.

[40] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang. A framework for hand gesture recognition based on accelerometer and emg sensors. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(6):1064–1076, 2011.