Ohara, Kazuya; Maekawa, Takuya ; Sigg, Stephan; Youssef, Moustafa

Preliminary Investigation of Position Independent Gesture Recognition Using Wi-Fi CSI

Published in:
2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)

DOI:
10.1109/PERCOMW.2018.8480253

Published: 01/01/2018

Document Version
Peer reviewed version

Preliminary Investigation of Position Independent Gesture Recognition Using Wi-Fi CSI

Kazuya Ohara
Osaka University, Japan
ohara.kazuya@ist.osaka-u.ac.jp

Takuya Maekawa
Osaka University, Japan
maekawa@ist.osaka-u.ac.jp

Stephan Sigg
Aalto University, Finland
stephan.sigg@aalto.fi

Moustafa Youssef
E-JUST, Egypt
moustafa.youssef@ejust.edu.eg

***Abstract*—This study investigates the feasibility of hand gesture recognition independent of the user position using Wi-Fi channel state information (CSI) obtained from a smartphone carried by a user. In this paper, we investigate the effectiveness of the component corresponding to the velocity of hand movements extracted from CSI for gesture recognition.**

***Index Terms*—Gesture Recognition, Wi-Fi Channel State Information**

I. Introduction

Due to the recent proliferation of IoT, smart home technologies, which enable automatic control and management of networked home appliances, have been attracting attention. To achieve easy and intuitive control of the networked appliances, many researchers have studied gesture recognition methods using various kinds of sensors. Traditional gesture recognition methods employ a depth camera such as Microsoft Kinect [1] or a wearable acceleration sensor such as a smart watch attached to the wrist of a user [2]. However, the depth camera approach has a problem of limited sensing area and the wearable approach requires a user to wear a wristwatch device.

In this study, we investigate a new hand gesture recognition method that employs neither a surveillance camera device nor a hand-worn sensor device. We assume that a user carries a commodity smartphone in, for example, his/her chest pocket and recognize hand gestures based on propagation information of smartphone Wi-Fi signals affected by movements of the hand. We obtain the propagation information from channel state information (CSI) using a computer equipped with a commodity Wi-Fi module that is installed in an environment and communicates with the smartphone. CSI is extracted from the PHY layer of the Wi-Fi communications and describes the changes of the amplitude and the phase of the Wi-Fi signals caused by path loss and multipath effects such as reflection and diffraction. Nowadays, because it is possible to obtain CSI from some advanced Wi-Fi network interface cards (NIC) such as the Intel 5300 Wi-Fi NIC, CSI have been used in some studies on indoor localization and activity recognition. Because collecting CSI from received packets requires modified drivers, we obtain CSI from packets transmitted from the smartphone using the computer with the Intel 5300 NIC.

In this study, we investigate gesture recognition independent of the user position using CSI. The propagation of Wi-Fi signals is greatly affected by the position of the user and the direction of the user body because the transmitter (smartphone) is carried by the user. In this paper, we extract the component corresponding to the velocity of hand movements from CSI based on the Doppler shift, which can be a feature independent of the user position, and investigate the effectiveness of the feature for gesture recognition.

It is difficult to estimate the Doppler shift from the phase components of CSI directly because the transmitter and the receiver are not synchronized precisely and the bandwidth of Wi-Fi signals is much wider than the range of the Doppler shift caused by human movements. To cope with the problem we focus on a fact that, when a target moves, the path length of the signals reflected by the target as well as the phase components also change. Therefore, the velocity component of the moving target can be computed from the difference in the phase components over some packets. However, because the Wi-Fi signals are too noisy to obtain the difference in the phase components, we use the multiple signal classification (MUSIC) algorithm that enables us to estimate the frequency of noisy signals for estimating the difference of the phase components. Moreover, because the phase components of CSI suffer from random offset due to the asynchronization between devices, we remove the offset by conjugate multiplying between CSI elements.

The 2.4 GHz band and 5 GHz band are commonly used in Wi-Fi communication. In the 2.4 GHz band, the phase components obtained from CSI using some NICs such as the Intel 5300 NIC are shifted $\pi/2$ randomly for each packet, making it impossible to compute the difference in the phase components, i.e., the velocity component. Therefore, we use the 5 GHz band of Wi-Fi signals that do not suffer from the random shift [3] to compute the velocity component. However, using Wi-Fi modules as 5 GHz access points is restricted by regulations regarding

radio communication. Instead of using the Wi-Fi module as 5 GHz access point, we employ a tethering function of the smartphone to use the smartphone as 5 GHz access point in order to communicate with the computer for CSI acquisition and design a method to derive the velocity component under this setting.

II. Related Work

A. Channel State Information

CSI is information obtained from the PHY layer of the Wi-Fi communication that describes states of propagation path for each pair of transmit and receive antennas in a multiple input/multiple output (MIMO) and for each subcarrier in orthogonal frequency division multiplexing (OFDM). CSI represents the change of amplitude and phase cased by path loss and multipath effects such as reflection and diffraction. The dimension of CSI is $N_T \times N_R \times N_S$ when N_T, N_R and N_S are the number of transmit antennas, the number of receive antennas and the number of subcarriers, respectively, and each element of CSI h is described as

$$h = Ae^{-j\theta}, \tag{1}$$

where A is attenuation of the signals and θ denotes the change of the phase.

B. Gesture Recognition using Wi-Fi Signals

Abdelnasser et al. [4] propose a gesture recognition method using the received signal strength indicator (RSSI) of the Wi-Fi signals obtained from a laptop computer. They remove noises in RSSI by the Discrete Wavelet Transform (DWT) and then recognize gestures within proximity of the receiver by associating between hand motion and three unique signal states of RSSI: a rising edge, a falling edge, and a pause. They use a laptop computer carried by a user. In contrast, we attempt to use a smartphone for gesture recognition.

Pu et al. [5] achieve whole-home gesture recognition by calculating the Doppler shift caused by gestures from Wi-Fi signals obtained by the USRP-N210 hardware. Because the bandwidth of the subcarriers, which is 312.5 KHz, is wider than the range of the Doppler shift caused by the gestures which is smaller than 20 Hz, the Doppler shift cannot be calculated from the amplitude for each subcarrier obtained by the USRP-N210 hardware. Therefore, they simulate the signals with a few Hertz bandwidth from multiple packets of the Wi-Fi communications in order to calculate the Doppler shift. They use USRP-N210, which is an expensive device for experts. In contrast, we use commodity devices, i.e., a desktop computer and a smartphone.

Li et al. [6] assume 9 digits finger-grained gestures performed between a transmitter and a receiver that are located at few meters away from each other, and identify the digits using CSI. They detect the duration of the gestures based on the fact that, when the gestures are performed, the subcarriers become correlated. They then classify the CSI time-series data corresponding to gestures by the kNN based on the Dynamic Time Warping (DTW). Al-qaness et al. [7] propose a hand motion gesture recognition method using CSI. They extract the duration of the gestures using the short-time energy algorithm, and then classify the gestures based on DTW. These methods require the training data obtained at the positions where gestures are performed. Sun et al. [8] try to track the user's hand using CSI. They estimate the angle of arrival of the Wi-Fi signals obtained by a receiver computed from CSI, and track the user's hand based on the fact that, when the user's hand occludes a signal coming from a certain direction, the signal strength of the signal decreases. In this method, the user is required to perform gestures close to the computer installed in an environment. We try to achieve gesture recognition independent of the user position.

Li et al. [9] propose a method that computes the Doppler shift caused by the user's movement from CSI using the multiple signal classification (MUSIC) algorithm, and use the method for the device-free passive indoor localization.

The above CSI-based methods assumes multiple MIMO enabled Wi-Fi devices installed in an environment of interest. In contrast, we assume a smartphone carried by a user and a Wi-Fi device communicates with the smartphone. Because many smartphone products employ single input single output (SISO) when communicating with a CSI-enabled Wi-Fi device, we design a method for hand movement detection for SISO based on the method for MIMO [9].

III. Proposed method

Our method computes the velocity components of the hand movement from CSI, and recognize hand motion gestures by hidden Markov models (HMMs) [10]. We use the tethering function of a smartphone to obtain CSI transmitted from the smartphone as mentioned in Section I. The tethering function of many smartphone employ SISO, which use only one antenna for transmitting and receiving signals. Therefore, we extract the velocity components from CSI obtained from the single antenna.

A. Computing Doppler Shift

The path length of the Wi-Fi signals reflected by a target changes between time 0 and time t when the target moves. Because the change in the path length affects the phase of the signals, the velocity of the target toward the Wi-Fi devices is computed from the difference in the phase over some Wi-Fi packets. However, because the Wi-Fi signals are too noisy to obtain the difference in the phase, we employ the MUSIC algorithm, witch enables us to estimate the frequency of noisy signals.

Each CSI element $h(t)$, which includes the signals reflected by moving targets, at time t is described as,

$$h(t) = e^{-j\theta_{offset}}(h_s + \sum_{i=1}^{M} A_i e^{-j2\pi f(\tau_i + \frac{v_i t}{c})}), \quad (2)$$

where h_s is the static component of CSI derived from, for example, the direct path and the signals reflected by walls, which is not affected by the moving targets. M is the number of targets, and v_i denotes the velocity of the ith target toward the Wi-Fi device. A_i is attenuation of the signals reflected by each target, and τ_i denotes the time delay of the signals reflected by each target. The phase component of CSI suffers from the random value offset θ_{offset} for each packet because commodity Wi-Fi receivers are not tightly synchronized with the transmitter. f is the frequency of the subcarrier, and c denotes the speed of the light.

To compute the velocities of the targets from CSI, we use the following 3 steps.

1) **Conjugate Multiplication:** To remove θ_{offset}, we apply conjugate multiplication between the elements of CSI for each packet. Li et al. multiply between CSI of the two receive antennas. In contrast, because the tethering function of the smartphone uses only one receive antenna, we multiply between CSI of adjacent subcarriers. The CSI conjugate multiplied between subcarriers whose frequencies are f_1 and f_2 is described as,

$$\begin{aligned} &h(f_1,t)\overline{h}(f_2,t) = h_s(f_1)\overline{h}_s(f_2) \\ &+ \sum_{i=1}^{M}\sum_{k=1}^{M} A_i(f_1)A_k(f_2)e^{-j2\pi(f_1(\tau_i+\frac{v_i t}{c}) - f_2(\tau_k+\frac{v_k t}{c}))} \\ &+ \overline{h}_s(f_2)\sum_{i=1}^{M} A_i(f_1)e^{-j2\pi f_1(\tau_i+\frac{v_i t}{c})} \\ &+ h_s(f_1)\sum_{k=1}^{M} A_i(f_2)e^{j2\pi f_2(\tau_k+\frac{v_k t}{c})}. \end{aligned} \quad (3)$$

2) **Removing Static Component:** The first term of Equation 3 contains only the static components. Because the power of the the static components is very higher than the power of the dynamic components, we remove this term by subtracting the averaged CSI within a certain time window, which is regarded as the static components. The second term is ignored because the power of the second term is very low due to the product of dynamic components. v_i in the third term and v_k in the fourth term are the velocities of the targets.
3) **MUSIC Algorithm:** In the MUSIC algorithm, we obtain CSI $\hat{\boldsymbol{h}} \in \mathbb{C}^{N_S}$, which is conjugate multiplied and removed static component, over N_p packets and create data matrix $\boldsymbol{X} = [\hat{\boldsymbol{h}}_1, \hat{\boldsymbol{h}}_2, \cdots, \hat{\boldsymbol{h}}_{N_p}]$. And then we calculate the correlation matrix $\boldsymbol{R}_x \in \mathbb{C}^{N_p \times N_p}$ of $\boldsymbol{X}$. The eigenvectors of $\boldsymbol{R}_x$ are composed of two components; the signal component with large eigenvalues and the noise component with small eigenvalues. We compute $P(v)$ using the noise component $\boldsymbol{E}_n$ as follows.

$$P(v) = \frac{1}{\boldsymbol{a}^H(v)\boldsymbol{E}_n\boldsymbol{E}_n^H\boldsymbol{a}(v)}, \quad (4)$$

where

$$\boldsymbol{a}(v) = [1, e^{-j2\pi f\frac{vt_2}{c}}, \cdots, e^{-j2\pi f\frac{vt_{N_p}}{c}}]^\top. \quad (5)$$

$P(v)$ has a sharp peak when v is identical to v_i, i.e., actual speed of the target.

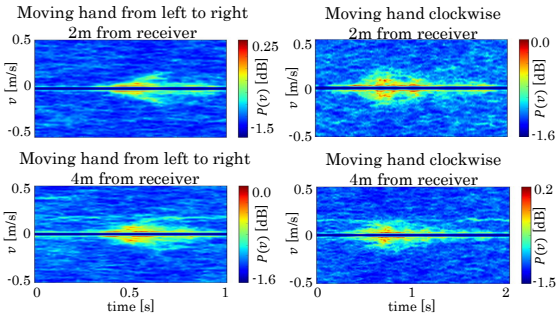

Fig. 1. Pseudo spectrogram during gestures

B. Classification by HMM

Because the whole hand moves in the hand motion gestures, $P(v)$ has many peaks corresponding the velocities of the various parts of the hand. Therefore, we make a pseudo spectrogram by computing $P(v)$ while changing v for each time sliding window. Figure 1 shows the pseudo spectrogram observed while gestures are performed. We compute the variance value, the maximum value and the kurtosis value as classification features from this pseudo spectrogram for each time slice, and then classify gestures by a 10-state left-to-right HMM prepared for each gesture.

IV. Evaluation

A. Dataset

Figure 2 shows our experimental environment. A computer equipped with the Intel 5300 NIC was installed in the environment as a receiver, and a modified NIC driver developed by Halperin et al. [11] was installed on the computer. As shown in Figure 3, a smartphone (FREETEL Raijin) was carried by a participant in front of the chest. The smartphone was connected to the computer using the tethering function with 5.2 GHz center frequency, and sent Wi-Fi packets at a rate of approximately 200 Hz.

The participant stood facing toward the computer 2 m, 4 m and 6 m away from the computer (these setups are called 2m-front, 4m-front and 6m-front, respectively), and performed 10 sessions where each session consists of 6 kinds of hand gestures 10 times: moving hand "up", "down", "left", "right", "clockwise" and "anticlockwise".

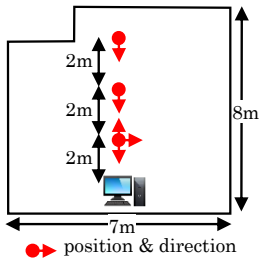

Fig. 2. Experimental environment

Fig. 3. Smartphone carried by the participant

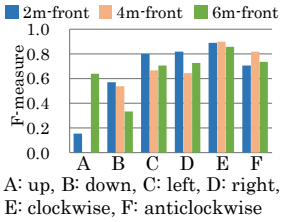

Fig. 4. Accuracy for each gesture

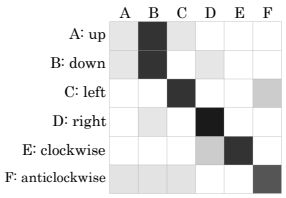

Fig. 5. Confusion matrix for 2m-front

TABLE I
ACCURACY FOR EACH POSITION AND DIRECTION OF THE BODY

	precision	recall	fmeasure	accuracy
2m-front	0.68	0.678	0.656	0.678
4m-front	0.615	0.636	0.594	0.639
6m-front	0.678	0.667	0.667	0.656
2m-left	0.365	0.433	0.379	0.433
2m-back	0.468	0.417	0.354	0.417

Moreover, the participant performed the same gestures at 2 m away from the computer when the participant stood while turning away from the computer to his left and while turning back to the computer as shown in Figure 2 (these setups are called 2m-left and 2m-back, respectively).

B. Result of Position Dependent Estimation

Table I shows the result of the leave-one-out cross-validation, where one instance of a gesture is used as test data and the remaining gestures collected at the same position and direction as the test gesture. As shown in the results, although the accuracies when the participant faced toward the computer are higher than 60 % at any distances, the accuracies under the left and back conditions are poor. This is because, the smartphone was attached to the front part of the chest in this experiment, the body greatly affected the propagation of the Wi-Fi signals under the left and back conditions, increasing the signal noise ratio.

Figure 4 shows the result for each gesture class at each position when the participant faced toward the computer and Figure 5 shows a confusion matrix of 2m-front. As shown in the results, the accuracies for the ”up” and ”down” gestures are poor. It is difficult to distinguish between the ”up” and ”down” gestures from the velocity components because the velocity component parallel to the direction of the radio transmission, i.e., the direction away from the front side of the chest, does not change significantly when these gestures are performed. (The ”up” and ”down” gestures consist of vertical hand movements.) In contrast, our method could classify the other gestures with high accuracies about 70 % to 80 % F-measure.

C. Result of Position Independent Estimation

We investigated the effects of the velocity components on position independent recognition using the leave-one-position-out cross-validation where data collected at two different are used as training data and data collected at the remaining position are used as test data. The average accuracy and F-measure are poor and 42.4 % and 37.4 %, respectively while the spectrograms of the same gesture performed at different positions seems to be similar to each other. This may be because the amplitudes of the spectrograms of the different positions are different from each other. Therefore, we believe that adjusting the amplitude components for each position improves the recognition accuracy under the position independent setting.

V. Discussion

A. Limitation

Because the proposed method captures the velocity component parallel to the direction of the radio transmission, it is difficult to recognize gestures that do not contain the component such as the ”up” and ”down” gestures.

B. Feature work

As mentioned in Seciton IV-B, the noises of Wi-Fi signals caused by the body existing between the smartphone and the computer deteriorated the classification accuracy poor. Removing the noises from CSI without influencing on the velocity component estimation is one of our future tasks.

Furthermore, as mentioned in Seciton IV-C, the recognition accuracy was poor when training data collected in different positions were used. However, the pseudo spectrograms of the same gesture performed at different positions seems to be similar to each other as shown in Figure 1. Because the accuracy decline may be caused by the difference in the amplitudes of the spectrograms for each position, adjusting the amplitude components for each position is also one of our future tasks.

Acknowledgment

This work is partially supported by JST CREST JPMJCR15E2, JSPS KAKENHI Grant Number JP16H06539, JP17J06602, and JP17H04679.

REFERENCES

[1] Z. Ren, J. Meng, J. Yuan, and Z. Zhang, "Robust hand gesture recognition with kinect sensor," in *Proceedings of the 19th ACM International Conference on Multimedia*, 2011, pp. 759–760.

[2] J. Korpela, K. Takase, T. Hirashima, T. Maekawa, J. Eberle, D. Chakraborty, and K. Aberer, "An energy-aware method for the joint recognition of activities and gestures using wearable sensors," in *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, 2015, pp. 101–108.

[3] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: Enabling phased array signal processing on commodity wifi access points," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, 2014, pp. 153–164.

[4] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *Proceedings of 2015 IEEE Conference on Computer Communications*, 2015, pp. 1472–1480.

[5] 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*, 2013, pp. 27–38.

[6] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "Wifinger: talk to your smart devices with finger-grained gesture," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 250–261.

[7] M. A. A. Al-qaness and F. Li, "Wiger: Wifi-based gesture recognition system," *ISPRS International Journal of Geo-Information*, vol. 5, no. 6, p. 92, 2016.

[8] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 2015, pp. 77–89.

[9] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, and H. Mei, "Indotrack: Device-free indoor human tracking with commodity Wi-Fi," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, 2017, p. 72.

[10] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[11] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: gathering 802.11 n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.