This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

Motion-Fi $^+$: Recognizing and Counting Repetitive Motions with Wireless Backscattering

Ning Xiao, Panlong Yang, Yubo Yan, Hao Zhou, Xiang-Yang Li, Haohua Du

Abstract—Driven by a wide range of real-world applications, several ground-breaking RF-based motion-recognition systems were proposed to detect and/or recognize macro/micro human movements. These systems often suffer from various interferences caused by multiple-users moving simultaneously, resulting in extremely low recognition accuracy. Even if the repetitive motions are fairly well detectable through the wireless signals in theory, in reality they get blended into various other system noises during the motion. Moreover, irregular motion patterns among users will lead to expensive computation cost for motion recognition. To tackle these challenges, we propose a novel wireless sensing system, called *Motion-Fi* $^+$, which marries battery-free wireless backscattering and device-free sensing in one clean sheet. *Motion-Fi* $^+$ is an accurate, interference tolerable motion-recognition system, which counts repetitive motions without using scenario-dependent templates or profiles and enables multi-user performing certain motions simultaneously because of the relatively short transmission range of backscattered signals and dedicated signal separation method. We implement a backscattering wireless platform to validate our design in various scenarios for over 6 months when different persons, distances and orientations are incorporated. In our experiments, the periodicity in motions could be recognized without any learning or training process, and the accuracy of counting such motions can be achieved within 5 % count error. With little efforts in learning the patterns, our method could achieve 95.2 % motion-recognition accuracy for a variety of 7 typical motions. Moreover, by leveraging the periodicity of motions, the recognition accuracy could be further improved to nearly 100 % with only 3 repetitions. Our experiments also show that the motions of multiple persons separating by around 2 meters cause little accuracy reduction in the counting process.

Index Terms—Wireless Sensing, Wireless Backscattering

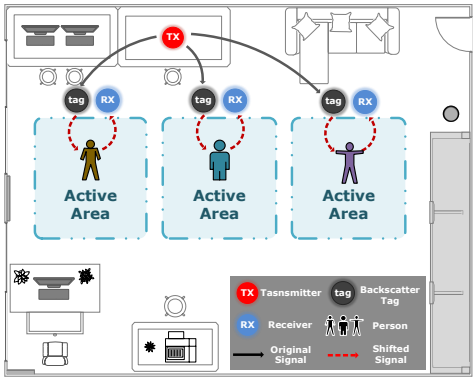


Fig. 1: Working Scenario of *Motion-Fi* $^+$. **TX** node serves as an infrastructure, each **Tag-RX** pair corresponding to one person.

1 Introduction

1.1 Backgrounds and Motivation

Human Activity Recognition (HAR) plays an important role in a wide range of real-world applications, such as smart home, health care and fitness tracking. Traditionally, smart mobile devices, including phones, watches, and other wearables, are widely used to recognize human activities. However, device-based approaches have many limitations due to the extra burden and discomfort brought to those who wear devices. To address this challenge, significant efforts are recently made to explore device-free human activity recognition techniques that utilize the information collected by various wireless infrastructures without the need for the monitored subject to carry a dedicated device. Another inspiring technology advances is battery-free networking, which is built upon wireless backscattering system. In that, the backscattering signals can be leveraged for communication without any active transmitter equipped with radio frequency front-end [1].

In summary, device-free interaction techniques utilize the signals collected from the infrastructures and battery-free networking could effectively use the surrounding RF energy and signals from infrastructures as well. In the following paragraphs, we will introduce this two technologies before present our motivation.

Device-free Interaction: Emerging technologies in the wireless network have brought device-free interactions into reality. Due to the plausible features such as *non-invasive* installment [1]–[4], and *ubiquitous* deployment [5], [6], wireless signals are applied as sensors for human-computer interactions (HCI), behavior identification and movement measurements *et al.*. Although these applications and systems have achieved considerably high success, there are still some intrinsical limitations and deficiencies to conquer. On one hand, most of the existing systems are based on WiFi or RFID signals. These signals would be interfered by nearby wireless devices working in a similar spectrum, and suffer from the multi-path and fading effects as a consequence of surrounding layouts [7]–[9], which could be categorized

- *N. Xiao, P. Yang, Y. Yan, H. Zhou and X.-Y. Li are with the School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, P.R. China. E-mail: xiaoning@mail.ustc.edu.cn, plyang, yuboyan, kitewind, xiangyangli@ustc.edu.cn*
- *H. Du is with the Department of Computer Science, Illinois Institute of Technnology, Chicago, IL 60616 USA. E-mail: hdu4@hawk.iit.edu*

Panlong Yang and Xiang-Yang Li are the corresponding authors.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

TABLE 1: 7 Motion Types.

No.	Motion	Abbr.
1	Squats	SQ
2	Push-ups	PU
3	Sit-ups	SU
4	Leg-raise	LR
5	Step	ST
6	Stoop-down	SD
7	Dumbbell	DB

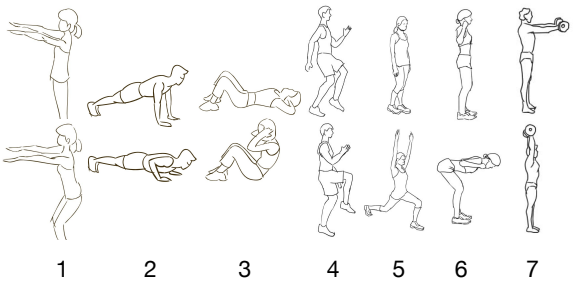


Fig. 2: Seven Regular Motions.

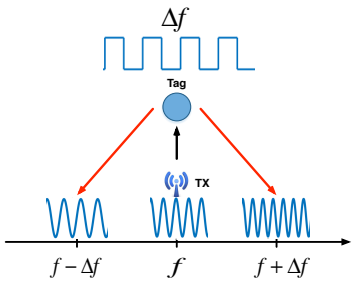


Fig. 3: Frequency Shift.

as *inter-system interference*. On the other hand, with the increased number of Internet of Things (IoT) devices, dense deployment would result in severe interference among wireless devices, which could be categorized as *intra-system interference* [10]. Reducing the negative effect of inter-system and intra-system interferences is one of the grand challenges in these RF-based motion sensing/detection/recognition systems.

Battery-free Networking: Fortunately, battery-free networking has recently become prevalent under the persuasive need for large-scale and long-term IoT applications. With the advances of wireless technology, energy harvesting and backscatter communication have become two prevalent and major schemes for battery-free networking system. For energy harvesting, it relies on the transmission power and decaying law of propagation, which would be subjected to the limited charging distance and low energy transfer efficiency. Inspiringly, Passive-WiFi [1] is proposed, serving as an infrastructure-based powering system, which provides carrier with WiFi-like infrastructure and a high-frequency circuit to backscatter the modulated messages. This innovative design brings two foremost merits. First, the wireless APs can serve as a charging infrastructure to provide pervasive energy sources for devices. This vision is inspiring since the AP deployment is pervasive and available in everyday life. Second, the modulated signal at the backscatter device could be transmitted at a rate up to 20 Mbps due to the high-frequency circuit design. This is impressive, since RF transmission is the major source of energy consumption in such battery-free systems[1].

Motivation - Marrying Device-free Interaction with Battery-free Networking: Our motivation is to design a device-free interaction system with batter-free networking devices, because if this vision could be brought into reality, it would bring in the following irresistible advantages:

- In *Spatio Domain*, the working range of backscattering signals is relatively short, which reduces the *intra-system* interference significantly. This enables our system to support parallel motion recognition for multiple persons.
- In *Spectrum Domain*, our system design introduces a controllable frequency shift (*i.e.*, two mirror shifts in our current design) in the backscattered signal. By carefully selecting the frequency-shift, our system can avoid the *inter-system* spectrum occupancy, thus reducing the interferences.
- In *Deployment Domain*, our tags can be widely deployed just based on few signal sources. Moreover, the tags are cheap and small enough while consume little energy.

In realizing this vision, we propose *Motion-Fi* $^+$, a repetitive motion recognition system leveraging passive wireless backscattering paradigm. Here, we focus on two features of repetitive motions: periodicity (*counting*) and type (*classification*). As depicted in Fig. 1, it uses passive wireless backscattering devices and leverages wireless AP as their powering infrastructure [1]. We define 7 regular motions with abbreviations shown in Table 1 and the sketch of each motion is shown in Fig. 2. Compared with other battery-free systems such as RFID [11], it could be more pervasive due to the prevalent deployment of wireless APs. Moreover, backscatter signals could be emitted by utilizing ambient wireless signals, which differs from previous radar-like system [2], [3]. Compared with our previous work Motion-Fi [12], we have carried out more detailed processing on the feature extraction, added signal separation algorithm to facilitate the multi-user scenes and designed a much smaller backscatter tag to facilitate wider deployment.

1.2 Challenges and Contributions

Challenges: Two challenges need to be formally addressed before realizing the aforementioned inspiring design vision with those favorable merits.

- There exist irregular and user/environment-dependent motion patterns, which naturally lead to incorrect motion recognition among different users/environments. The effect of motion on the received signal varies in environment (*i.e.*, interior furnishings), gender (differences on repetitive motion frequency and amplitude) and body characteristics (differences on height and weight). Even when a person performs the same motion at different time, the motion's speed, direction, and angle may vary a lot. Previous motion recognition solutions base on building profiles as template are invalid due to this ambiguity and diversity.
- Backscattering signals are weak but highly dynamic, especially when motions are incorporated. In multi-user scenes, there are still weakly mutual interferences between/among different user's motion signals, which will weaken the motion signal characteristics of each other, thus increasing the difficulty of recognition.

1. Maintaining frequency oscillator accuracy and ADC installation will cost most of the energy budget. High-speed backscattering scheme borrows the RF carrier from the infrastructure and achieve WiFi standard transmission by changing the resistance in circuit at nearly 20MHz.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

To tackle aforementioned challenges, we demonstrate a prototype system, *Motion-Fi* $^+$, recognizing repetitive human motion frequency with passive wireless backscattering devices. *Motion-Fi* $^+$ supports multi-user motion counting in various scenes and single-user motion recognition in a given environment. Compared with traditional WiFi-based gesture recognition systems, it supports motion recognition/counting in typical indoor scenarios without suffering from surrounding interference. Furthermore, we design a passive backscatter platform to validate our design.

In summary, our contribution is three-folds:

- In our design, we first filter out the noise and observe the motion information through the base-band carrier. Intrinsic properties of human motions are fully explored and we propose a self-adaptive matching scheme, where optimal patterns could be recognized automaticly for motion counting. Different from previous studies, our system works well even when motion patterns are not known in priori (tackling the first challenge).
- Through in-depth observation of the received signals, we found the mixing law of motion signals in multi-user scenes. We formalize the model of mixing signal and use Independent Component Analysis (ICA) based technique to separate the original motion signal, which enables our system to support parallel motion recognition for multiple persons (tackling the second challenge).
- Our extensive long-term experimental evaluation shows that our platform can effectively count the repetitive motions with an error rate less than 5%, across 26 different users, 7 motions, 4 orientations, and 5 distances for more than 6 months. Moreover, with just a little training effort via customized cubic-SVM, our system can accurately recognize different motion types with 95.2% recognition rate for 7 regular motions. Specifically, for periodical motions, the recognition rate could be further improved to nearly 100% after only 3 repetitions of the same motion.

The rest of this paper is organized as follows. Sec. 2 provides preliminary knowledge and presents the intrinsic principle on backscattering signal. Then, we introduce our system design and approaches in Sec. 3, and present the implementation details in Sec. 4. In Sec. 4, we also evaluate and validate our system with extensive experiments and analysis. Finally, we review the related work in Sec. 5 and conclude the paper in Sec. 6.

2 Preliminaries and System Model

2.1 Preliminaries on Backscattering Signal

In our backscattering system, we have one transmitter (**TX**), one backscatter **Tag**, and multiple receivers (**RX**).

Transmitter Side: It provides a single-frequency tone as the plug-in device in passive WiFi design, and serves as a charging infrastructure. This vision is inspiring and impressive since passive WiFi deployment is considered to be pervasive in IoT industry. The tone signal can be simply represented as $\sin(2\pi f_c t)$.

Backscatter Tag: The backscatter tag is composed of an antenna and a micro-controller, which controls the SPDT RF switch network to generate backscatter signals. By changing the impedance of the antenna, the tag can switch its states between reflecting and non-reflecting. The scattered power of a passive tag is

$$P_{tag} = \frac{P_{tx} G_{tx} \Delta_\Gamma}{4\pi d^2},$$

where P_{tx} and G_{tx} is the transmission power and antenna gain of transmitter respectively, d is the distance between WiFi AP and the tag, and Δ_Γ is the differential Radar Cross Section (RCS) [2], [3] given by

$$\Delta_\Gamma = \frac{\lambda^2}{4\pi} G_{tag}^2 \left|\Gamma_1 - \Gamma_2\right|,$$

where λ is the wavelength of carrier, G_{tag} is the antenna gain of tag and $\Gamma_{1,2}$ are the complex power wave reflection coefficients according to two different load impedances $Z_{c_{1,2}}$ [13]:

$$\Gamma_{1,2} = \frac{Z_{c_{1,2}} - Z_a^*}{Z_{c_{1,2}} + Z_a},$$

where $Z_a = R_a + jX_a$ is the complex antenna impedance, and $Z_{c_{1,2}} = R_{c_{1,2}} + jX_{c_{1,2}}$ is the complex tag circuit impedance.

Frequency Shift (Modulation in the Air): Sideband-backscatter modulation [1], [14] could shift the carrier by a certain frequency. By modulating the RCS of an antenna effectively multiplies the incoming signal by the modulated signal. Thus, modulating the antenna at a frequency Δ_f would create a frequency shift in the incoming signal.

Let a square wave at a frequency of Δ_f, which is generated by the micro-controller and used to control the impedance of the antenna. According to Fourier analysis, the first harmonic of a square wave is a sinusoid signal. Thus, we can simplify the process of square wave switching to sinusoid $\sin(2\pi\Delta_f t)$. Consequently, the process of backscattering can be represented by the product of the aforementioned two sinusoidal signals (we named it modulation in the air), which is given by

$$2\sin(2\pi f_c t)\sin(2\pi\Delta_f t) = \cos(2\pi(f_c - \Delta_f)t) - \cos(2\pi(f_c + \Delta_f)t).$$

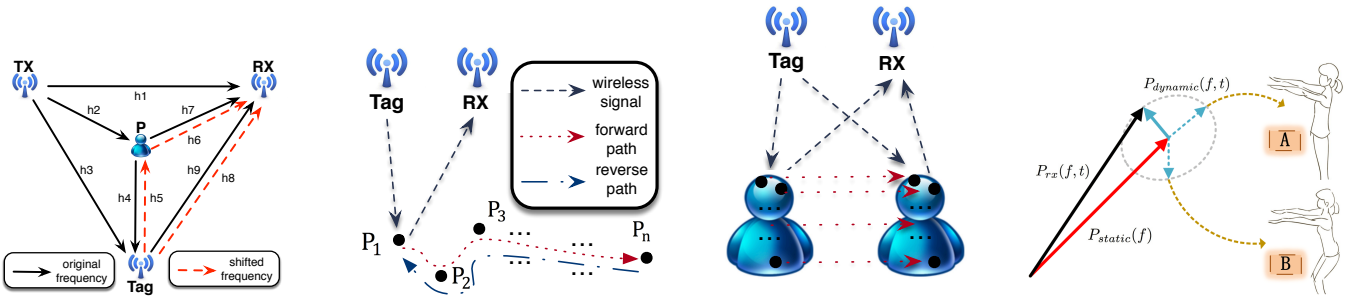
The frequency shifts of generated narrow band signals are $f_c - \Delta_f$ (with a negative shift) and $f_c + \Delta_f$ (with a positive shift). Fig. 3 illustrates the frequency shift. The power consumption of our backscatter tag is near zero (several μW) [1].

Receiver Side: A receiver node tunes its center frequency to one of the shifted signal, which is $f_c - \Delta_f$ in our configuration. As described above, the received signal is generated by modulating the tone signal of transmitter and the sinusoid signal of backscatter tag in the air.

2.2 Signal Reflection Path

To study the signal path of our proposed system, we show the basic reflection model of motion recognition with backscatter signal in Fig. 4a. The **TX** node serves as a powering infrastructure, which sends a sinusoidal tone at a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing



(a) Signal Reflection Path. (b) Model of single point. (c) Model of person. (d) Static and Dyanmic Components.

Fig. 4: Principle on Backscattering based Motion Recognition. (a) depicts the signal reflection path of our backscattering based system. (b) and (c) show the motion model of single point and person respectively. (d) is the sketch of periodical motions.

frequency of f_c [1]. The **Tag** node is a backscattering device, which switches the impedance of its antenna to *reflect* or *absorb* the tone signal. Noted that, in addition to reflect the tone signal transmitted by **TX**, there are two generated signals with frequency $f_c - \Delta_f$ and $f_c + \Delta_f$. A person noted as **P** performs movements next to **Tag**. A receiver noted as **RX** tunes its center frequency to $f_c - \Delta_f$ to capture the generated signal. The sinusoid tone transmitted by **TX** can directly traverse to **RX**, **P** and **Tag**, with path h_1, h_2 and h_3 respectively. Then, **P** reflects the signal to **RX** and **Tag** with path h_7 and h_4 respectively. **Tag** reflects the sinusoidal signal at frequency f_c to **RX** with path h_9.

Meanwhile, the generated signals traverse from **Tag** to **P** and **RX** with path h_5 and h_8. Then, **P** reflects it to **RX** with path h_6. As depicted in Fig. 4a, the traverse of original signal at the frequency f_c is plotted in solid black lines, and the traverse of the shifted signal at the frequency $f_c - \Delta_f$ is plotted in dashed red lines. Since 'RX' receives signals at the frequency $f_c - \Delta_f$, we mainly study the signals plotted in dashed red lines.

2.3 Periodicity in Backscattering Signals

Since the channel is stable over a short period of time [15]–[17], it is believed that the changes of signal are mainly caused by human motion. As shown in Fig. 4a, we classify a path as a dynamic one if the generated signal is affected by the person's motion when traversing across the path, *i.e.*, h_5 and h_6. Similarly, if the generated signal is not affected by the person's motion including both line-of-sight paths as well as the paths reflected from static objects, the path it traverses across is classified as a static one, *i.e.*, h_8. Here we believe that the signals after the human body (*i.e.*, h_2 & h_4.) is very weak compared to the signal that does not pass the human body (*i.e.*, h_3.) .

Let the received energy by **RX** is P_{rx}. Then, we can treat P_{rx} as the sum of P_{static} and $P_{dynamic}$ [4], that is:

$$P_{rx}(f,t) = P_{static}(f) + P_{dynamic}(f,t), \tag{1}$$

where P_{static} is the energy of signals received from static paths, and $P_{dynamic}$ denotes the reflected signal from human motion. According to the previous descriptions, we regard signals from path h_8 as static and signal from path h_5, h_6 as dynamic.

We can find that the dynamic signals is entirely generated by the movement of the human body. Further, we analyze the impact of human motion on the dynamic part of recieved signal.

2.3.1 Model of single point

The simplest reflection model is one point, because there exists only one reflected path. If we consider the influence of both amplitude and phase, we can represent the dynamic singal reflected by the point use $a(f,t)e^{\frac{-j2\pi d(t)}{\lambda}}$, where $a(f,t)$ is the complex representation of amplitude and initial phase offset of the dynamic path (**Tag** $\rightarrow$ object), $d(t)$ denote the distance between the object and receiver at time t, and $e^{\frac{-j2\pi d(t)}{\lambda}}$ is the phase shift along the dynamic path (object $\rightarrow$ **RX**) length $d(t)$. As shown in Fig. 4b, the points named $P_1, P_2, P_3, ..., P_n$ are the positions of the observation point at different sampling times, *i.e.*, $t_1, t_2, t_3, ..., t_n$. The signal from **Tag** to point P_i and then reflected to **RX** is what we concerned as dynamic path.

As previously mentioned, we can get

$$P_{dynamic}(f,t_i) = a(f,t_i)e^{\frac{-j2\pi d(t_i)}{\lambda}}, i = 1,2,3,...,n. \tag{2}$$

When we consider the periodic movement, the point will move from P_1 to P_n and then back to P_1, as the forward path and reverse path depicted in Fig. 4b. Although the round-trip path can not be perfectly coincident, we can see the approximate symmetrical signal changes, which will be validated in our experimental validation.

Moreover, if we denote

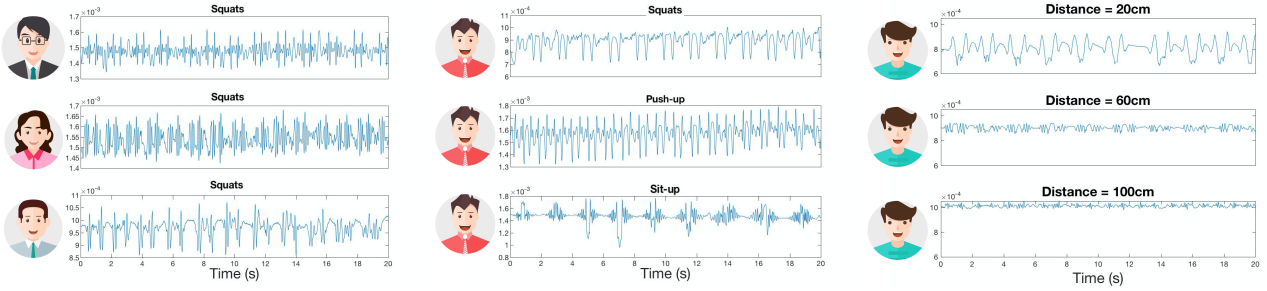
$$P^k{}_{dynamic}(f,t^k{}_i) = a^k(f,t^k{}_i)e^{\frac{-j2\pi d^k(t^k{}_i)}{\lambda}} \tag{3}$$

as the dynamic signal of $k-th$ round-trip for a point periodic movement, we can further see cyclical changes in the signal due to the periodicity. Since the experience time of each cycle may be different, which will be shown on the received signal by stretching or compressing the signal pattern along the time axis, and we will describe how to solve this problem in the algorithm section.

2.3.2 Model of a person's motion

As shown in Fig. 4c, for a person's movement, we can divide a person's body into massive infinitesimal segments, each of these segments can be treat as a point as described above. That is, we can get the dynamic signal by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing



(a) Three persons perform the same action. (b) One person perform three motions. (c) Signal strength under different distances.

Fig. 5: Basic experimental observation. (a) shows the diversities in persons, (b) depicts the differences in motions and (c) shows the received signal strength under different distances.

$$P_{dynamic}(f,t_i) = \sum_{segment_j \in Body} a_j(f,t_i) e^{\frac{-j2\pi d_j(t_i)}{\lambda}}, \quad i = 1,2,3,...,n. \tag{4}$$

Since the motion of each segment can be considered to be periodic, the sum function of the periodic function also has a periodic signal that can be used to obtain periodic motion of the human body.

We demonstrate this principle in Fig. 4d. So, if we perform periodical motions, the signal of $P_{dynamic}(f,t)$ also changes periodically, thus the energy of $P_{rx}(f,t)$ is periodical.

2.3.3 Experimental Validation

To examine the effect of person's motion in received signal, we first observe the received signal when one person performs different motions, different persons perform the same motions and one person perform the same action at different distances, the result of which are shown in Fig. 5.

Form Fig. 5a, we can see the basic periodicity in the time series and for each fragment we can also see the symmetry when different persons perform the same motion (SQ), which is consistent with what we described in Sec. 2.3.1 & 2.3.2. Besides, we can find the difference (*i.e.*, different waveforms, different cycle times and different amplitude changes) among these signals, because there exist different habits when different people perform the same action, such as the speed of motion, the degree of leg bending and body differences (*i.e.*, height and weight).

Fig. 5b shows the result when one person performs 3 kinds of motions (SQ, PU and SU), from which we can also see the periodicity and symmetry. However, the waveforms of them vary a lot since different motions have distinctly different patterns, which can be used to distinguish different motions.

Fig. 5c depicts the distance[2] effect when one person perform squats (SQ) at 3 distances (20cm, 60cm and 100cm). Since the signal is attenuated square inversely with the increase in distance, we can see the dynamic part of received signal decay quickly. Therefore, the motion signals of multiple people have little influence on each other as long as they are separated by a certain distance. This enables our system to support multiple persons perform simultaneously.

3 System Design

3.1 Design Overview

Our *Motion-Fi* $^+$ design can recognize/count repetitive motions regardless of the motion, even when multiple users are performing movements at the same time. The system overview of *Motion-Fi* $^+$ is illustrated in Fig. 6. There are mainly four modules in *Motion-Fi* $^+$, which are *Smoothing Filter*, *Adaptive Counter*, *SVM Classifiers* and *Signal Separation Module*. When the frequency shifted backscattering signal is received, the energy of signal is computed according to Sec. 3.2. Then, an LPF filter is applied to the received signal to smooth the amplitude. The adaptive counting module is used to count the repetitive motions without using scenario-dependent templates or profiles. After that, an enhanced SVM classifier is used to recognize the motions. Finally, *Motion-Fi* $^+$ outputs the counting number and type of motions. For multi-user scenes, we use signal separation method to separate the mixed signals.

3.2 Preprocessing of the Signal

As is known, $I(t)$ and $Q(t)$ are orthogonal signals in wireless communications. They can be represented as

$$\begin{aligned} I(t) &= A(t)A_0\cos(2\pi ft) + N_I(t), \\ Q(t) &= A(t)A_0\sin(2\pi ft) + N_Q(t), \end{aligned} \tag{5}$$

where A_0 and f are the amplitude and frequency of the sinusoidal signal. $N(t)$ is the white noise. $A(t)$ denotes the influence coefficient[3] of motions, which is equal to 1 when there are no motions.

Then the energy of received signal is $E(t) = \sqrt{(I(t)^2 + Q(t)^2)}$, and replace $I(t)$ and $Q(t)$ with Eq. (5). After that, $E(t)$ could be represented as $E(t) = A(t)A_0 + N'(t)$, where $N'(t)$ is a term relevant to the white noise. Compared with human movements, $N'(t)$ can be treated as high frequency noise, which could be filtered out with LPF (Low Pass Filter). After applying LPF to $E(t)$, we have:

$$E'(t) = A(t)A_0. \tag{6}$$

2. As shown in Fig. 1, 'distance' here indicates the distance from the person to the **Tag-RX** pair.

3. In our system, the effect of human motions on signals is reflected in the change of signal amplitude, which is equivalent to multiplying a coefficient in the original amplitude.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

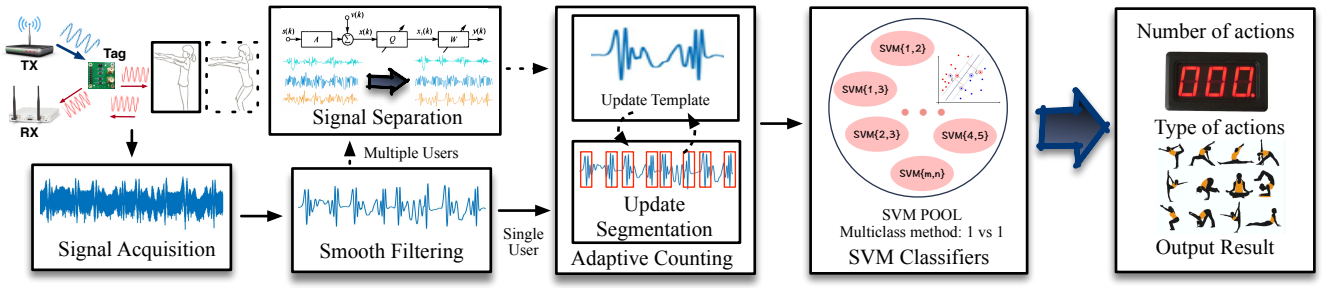


Fig. 6: System Overview. The received signal is first processed in a *smooth filtering* module, then it will be sent to our *adaptive counting* module and *SVM classifiers* module, finally output the motion type and counting result.

Finally, the filtered signal could reflect human movements approximately, since A_0 is constant in this scenario.

3.3 Adaptive Counting

3.3.1 Motion Segmentation

We formulate an optimization problem that jointly recovers the morphology of the motions and the segmentation [5]. We adjust the algorithm to fit different motions, whose cyclical features are weaker than those of heart beats. The intuition underlying this optimization is that successive human motions should have similar morphology. Since each motion lasts different time duration, we can define the similarity of the two motions by stretching or compressing the motion signal to an equal length. So the goal of the algorithm is to find an optimized segmentation which makes the difference between each segmentation as small as possible, while accounting for the fact that we do not know a priori the shape of a motion.

Let $\vec{x} = (x_1, x_2, ..., x_n)$ denote a sequence of length n. A segmentation $\mathbf{S} = \{\vec{s_1}, \vec{s_2}, ...\}$ of $\vec{x}$ is a partition that divides $\vec{x}$ into non-overlapping continuous subsequences (segments), where each segment $\vec{s_i}$ consists of $|\vec{s_i}|$ points and $|\vec{x}| = \sum_{\vec{s_i} \in \mathbf{S}} |\vec{s_i}|$.

In order to identify each motion cycle, our idea is to find a segmentation with segments most similar to each other, *i.e.*, to minimize the variation across segments. Since statistical distance is only defined for scalars or vectors with the same dimension, we extend the definition of distance for vectors with different lengths as follows:

Definition 1. The distance between vectors $\vec{a_1}$ and $\vec{a_2}$ is:

$$Dist(\vec{a_1}, \vec{a_2}) = dtw(\vec{a_1}, \vec{a_2}),$$

where $dtw(\vec{a_1}, \vec{a_2})$ is dynamic time warping function to measure the similarity between two temporal sequences. It stretches two vectors, $\vec{a_1}$ and $\vec{a_2}$, onto a common length L $(\max(|\vec{a_1}|, |\vec{a_2}|) \leq L \leq |\vec{a_1}| + |\vec{a_2}|)$, making the sum of the Euclidean distances between corresponding points the smallest. To stretch the inputs, dtw will repeat each element of $\vec{a_1}$ and $\vec{a_2}$ as many times as necessary.

Following that, we define the distance between segmentation $\mathbf{S}$ and motion template $\boldsymbol{\xi}$[4]:

Definition 2. The distance between segments $\mathbf{S} = \{\vec{s_1}, \vec{s_2}, ...\}$ and a template $\boldsymbol{\xi}$ is:

$$Dist(\mathbf{S}, \boldsymbol{\xi}) = \sum_{\vec{s_i} \in \mathbf{S}} Dist(\vec{s_i}, \boldsymbol{\xi}) = \sum_{\vec{s_i} \in \mathbf{S}} dtw(\vec{s_i}, \boldsymbol{\xi}).$$

The goal of our algorithm is to find the optimal segmentation $\mathbf{S}^*$ and template $\boldsymbol{\xi}^*$ that minimizes $Dist(\mathbf{S}, \boldsymbol{\xi})$. Since both segmentation $\mathbf{S}$ and template $\boldsymbol{\xi}$ are unknown for us, we can rewrite it as the following joint optimization problem:

$$\begin{aligned}(\mathbf{S}^*, \boldsymbol{\xi}^*) &= \arg\min_{\mathbf{S}, \boldsymbol{\xi}} Dist(\mathbf{S}, \boldsymbol{\xi}) \\ &= \arg\min_{\mathbf{S}, \boldsymbol{\xi}} \sum_{\vec{s_i} \in \mathbf{S}} dtw(\vec{s_i}, \boldsymbol{\xi}), \\ \text{s.t.}: \ t_{min} &\leq \frac{|\vec{s_i}|}{C} \leq t_{max}, \vec{s_i} \in \mathbf{S},\end{aligned}$$

where t_{min} and t_{max} are constraints on the length of each motion cycle[5], and C denotes the sample rate. It tries to find the optimal segmentation $\mathbf{S}$ and template (*i.e.*, morphology) $\boldsymbol{\xi}$ that minimize the sum of the distance between segments and template. This optimization problem is difficult as it involves both combinatorial optimization over $\mathbf{S}$ and numerical optimization over $\boldsymbol{\xi}$. Exhaustive search of all possible segmentations will lead to exponential complexity.

3.3.2 Iterative Segmentation

Instead of estimating the segmentation $\mathbf{S}$ and the template $\boldsymbol{\xi}$ simultaneously, our algorithm alternates between updating the segmentation and template. During each iteration, our algorithm updates the segmentation given the current template, then updates the template given the new segmentation. For each of these two sub-problems, our algorithm can obtain the global optimal with linear time complexity.

Update Segmentation: In the i-th iteration, segmentation $\mathbf{S}^{i+1}$ is updated given template $\boldsymbol{\xi}^i$ as

$$\mathbf{S}^{i+1} = \arg\min_{\mathbf{S}} Dist(\mathbf{S}, \boldsymbol{\xi}^i). \tag{7}$$

Though the number of possible segmentations grows exponentially with the length of x, the above optimization problem can be solved efficiently using dynamic programming [18]. The recursive relationship for the dynamic program is as follows: if D_l denotes the minimal cost[6] of

4. $\boldsymbol{\xi}$ is a vector represents the central tendency of all the segments, i.e., a template for the motion shape (or morphology). We denote m=$|\boldsymbol{\xi}|$ in the following content.

5. t_{min} and t_{max} capture the fact that human motion cannot be too short or too long.

6. The minimal cost of sequence $\vec{x}_{1:l}$ is the minimum of $Dist(\mathbf{S}, \boldsymbol{\xi})$ in this sequence.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

algorithm 1 Motion Segmentation

Input: sequence $\vec{x}$, time constraint ϱ= $\{t_{min}$,$t_{max}\}$
Output: segments $\mathbf{S}$, template $\boldsymbol{\xi}$

initialize $\boldsymbol{\xi}^0$ as zero vector ;
$m \leftarrow \frac{1}{2} * (t_{max} + t_{min}) * C$;
$i \leftarrow 0$; //number of iterations
$Similarity \leftarrow Inf$;
while $(Similarity \geq Threshold)$ **do**
  $\mathbf{S}^{i+1} \leftarrow UpdateSegmentation(\vec{x}, \boldsymbol{\xi}^i)$;
  $\boldsymbol{\xi}^{i+1} \leftarrow UpdateTemplate(\vec{x}, \mathbf{S}^i)$;
  $i \leftarrow i + 1$;
  $Similarity \leftarrow Dist(\boldsymbol{\xi}^{i+1}, \boldsymbol{\xi}^i)$;
end while
return $\mathbf{S}^i$ *and* $\boldsymbol{\xi}^i$

function UPDATESEGMENTATION($\vec{x}, \boldsymbol{\xi}$)
  $\mathbf{S}_0 \leftarrow \emptyset$;
  $D_0 \leftarrow 0$;
  for $t = 1 \rightarrow length(\vec{x})$ **do**
    $\tau^* \leftarrow arg\min_{\tau \in \tau_{t,\varrho}} \{D_\tau + Dist(x_{\tau+1:t}, \boldsymbol{\xi})\}$;
    $D_t \leftarrow D_{\tau^*} + Dist(\vec{x}_{\tau+1:t}, \boldsymbol{\xi})$;
    $\mathbf{S}_t \leftarrow \mathbf{S}_{\tau^*} \cup \{\vec{x}_{\tau*+1:t}\}$;
  end for
  return $\mathbf{S}_n$
end function

function UPDATETEMPLATE($\vec{x}, \mathbf{S}$)
  $n \leftarrow length(\vec{x})$;
  $\boldsymbol{\xi} \leftarrow \frac{1}{n} \sum_{\vec{s_i} \in \mathbf{S}} |\vec{s_i}| \omega(\vec{s_i}, m)$;
end function

segmenting sequence $\vec{x}_{1:l}$, then we have

$$D_l = \min_{\tau \in \tau_{l,\varrho}} \{D_\tau + Dist(\vec{x}_{\tau+1:l}, \boldsymbol{\xi})\}, \tag{8}$$

where $\varrho = \{t_{min}, t_{max}\}$ and $l - C * t_{max} \leq \tau_{l,\varrho} \leq l - C * t_{min}$ specifies possible choices of τ considering the segment length. According to Eq. (8), the time complexity of the dynamic programing is $O(C(t_{max} - t_{min})n)$ and the global optimum could be guaranteed.

Update template: In the i-th iteration, template $\boldsymbol{\xi}^{i+1}$ is updated given segmentation $\mathbf{S}^i$ as

$$\boldsymbol{\xi}^{i+1} = \arg\min_{\boldsymbol{\xi}} \sum_{\vec{s_j} \in \mathbf{S^i}} Dist(\vec{s_j}, \boldsymbol{\xi}). \tag{9}$$

Obviously, the above optimization is a weighted least squares problem and it is impossible to meet the above optimization goals $\boldsymbol{\xi}^{i+1}$ by using brute-force search. Here we present an analogy method to get a feasible solution.

Consider another similar problem: given an array $\vec{a}$, find a number x such that the sum of the distance between x and each element in array $\vec{a}$ is minimized. If y denotes the sum of distance, we can get the expression as follows easily: $y = \sum_{i=1}^{|\vec{a}|} (a[i] - x)^2 = |\vec{a}|x^2 - 2C_1 x + C_2$, where C_1 and C_2 are two constants, $C_1 = \sum_{i=1}^{|\vec{a}|} a[i]$, $C_2 = \sum_{i=1}^{|\vec{a}|} a[i]^2$. Obviously, this is a quadratic function optimization problem, and it is easy to get the solution that makes y minimal: $x^* = \frac{(-2*C_1)}{-2*n} = \frac{1}{|\vec{a}|} \sum_{i=1}^{|\vec{a}|} a[i] = mean(\vec{a})$.

From the above equation we can see that the optimal solution x^* is the average of the array $\vec{a}$. Similarly, we map the problem to Eq. (9), the array element $a[i]$ becomes the subsequence $\vec{s_i}$, x becomes the template vector $\boldsymbol{\xi}$, the distance between two values becomes the distance of two vectors, in the end, the solution x^* is mapped to the weighted average of the subsequences, and a feasible solution is given as follows:

$$\boldsymbol{\xi}^{i+1} = \frac{\sum_{\vec{s_j} \in \mathbf{S}^i} |\vec{s_j}| \omega(\vec{s_j}, m)}{\sum_{\vec{s_j} \in \mathbf{S}^i} |\vec{s_j}|} = \frac{1}{|\vec{x}|} \sum_{\vec{s_j} \in \mathbf{S}^i} |\vec{s_j}| \omega(\vec{s_j}, m), \tag{10}$$

where the weight of each subsequence is the proportion of the length of the sequence to the entire sequence, $\omega(\vec{s_j}, m)$ is linear warping of $\vec{s_j}$ into length m, which is realized through a cubic spline interpolation [19].

The algorithm pseudo-code is shown in Algorithm 1. The stopping condition represents the convergence state. The difference between two conjunctive iterations are inspected as the criterion of convergence. For practical consideration, we set a feasible threshold, where system efficiency and accuracy can be achieved.

3.4 Motion Signal Separation (Multi-User Scenes)

In this part, we discuss how we count the repetitive motions for the target person when multi-user performing certain motions simultaneously. Due to the nature of the relatively short transmission range and frequency shift, backscatter signal can significantly reduce inter-signal interference. Even though earlier work *MotionFi* [20] has shown that backscatter signal are suitable to monitor motion rate when multiple persons involved, the counting error is a bit high (up to 15%). So, we attempt to find new approach for improving the accuracy.

Through in-depth observation of the received signal, we find that the periodic pattern of a multi-user scene is less noticeable than that of a single person on account of the interaction of multiple motions. Hence, in order to improve the counting accuracy, we must separate each motion signal from these mixed signals. In this study, we aim to extract every motion signal from the received backscatter signals, as shown in Fig. 7. That way we can count the motions when multi-user move together.

3.4.1 Modeling Mixture Signals

We first formulate the signal separation problem by formally defining the mixed signal. Suppose we have N signal sources, where the n^{th} signal, *i.e.*, $s_n t$ is corresponding to the motion of the n^{th} user. In our deployment scenario, we also have N receivers, where $x_n(t)$ denotes the mixed signal at the n^{th} receiver, such that

$$x_n(t) = \sum_{k=1}^{N} a_{k,n} s_k(t), \forall 1 \leq n \leq N. \tag{11}$$

where $a_{k,n}$ represents the mixing coefficients associated with the path from s_k to x_n. Here we ignore the negligible signal delay due to the high speed of wireless signals ($\sim 3 * 10^8 m/s$) and relative short distances among

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

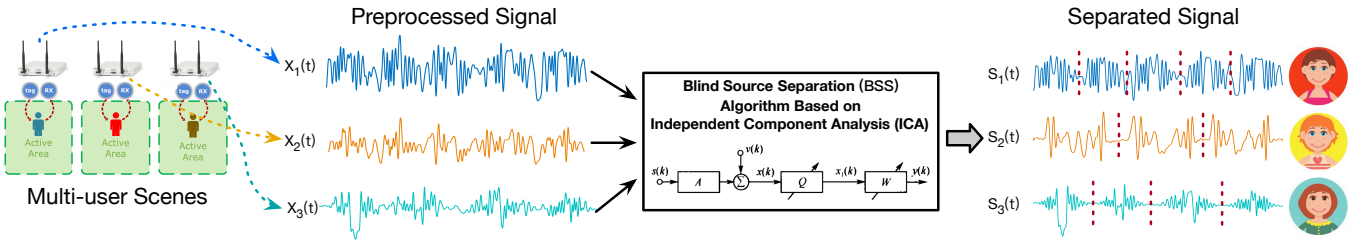


Fig. 7: Signal Separation Overview. In multi-user scenes, the received signals will interact with each other, we separate each motion signal based on Blind Source Separation (BSS) technique.

users. If we denote $X(t) = [x_1(t), x_2(t), ..., x_n(t)]^T$, $S(t) = [s_1(t), s_2(t), ..., s_n(t)]^T$, then we can rewrite Eq. (11) as

$$X(t) = AS(t),$$
$$A = \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{bmatrix}_{n \times n}, \quad (12)$$

where A is the mixing coefficients matrix, hence we can represent $S(t) = A^{-1}X(t)$. Therefore, our objective is to find a matrix $W = A^{-1}$ that satisfies the constraint to separate the original signal $S(t)$.

3.4.2 Background on Blind Source Separation (BSS)

Our signal separation problem is similar to the cocktail party problem [21] where an arbitrary number of people are talking simultaneously at a cocktail party and a listener is trying to identify and follow one particular discussion. In many cases, the measurements are given as a set of parallel signals or time series. Typical examples are mixtures of simultaneous sounds or human voices that have been picked up by several microphones, brain signal measurements from multiple EEG sensors, several radio signals arriving at a portable phone, or multiple parallel time series obtained from some industrial process. The Blind Source Separation (BSS) is used to characterize this problem. In our scenario, we have no control of, nor a priori knowledge of, the frequencies of motions, but we can assume that the motion signal from each user is independent because the physical processes of each user's movements are not related to each other. Naturally, we treat our problem as a BSS problem.

To tackle BSS problem, many population solutions have been proposed, such as Independent Component Analysis (ICA) [22], Principal Component Analysis (PCA) [22] and Degenerate Unmixing Estimation Technique (DUET) [23]. Among these solutions, ICA is a much richer technique, which can be seen as an extension to PCA and Factor Analysis, capable of finding the sources when these classical methods fail completely. Thus we choose ICA, which is well-suited to address our needs and easy to implement.

3.4.3 Signal Separation Based on ICA

Next, we present our algorithm that separates individual motion signals. Independent Component Analysis (ICA) is a computational technique for revealing hidden factors that underlie sets of measurements or signals. In our deployment, there are 3 signal sources and 3 signal receivers ($|X| = |S| = N = 3$). After preprocess the 3 signals as described in Sec. 3.2, we apply the following signal processing steps on each signal within the same window.

1) *Remove the Direct Current (DC) Component.* ICA requires that the source signals be real random variables of zero mean, so we first approximate a zero-mean signal by subtracting the average of the signals at each receiver[7]. This process can be expressed as:
$$x'_n(t) = x_n(t) - E\{x_n\}, n = 1, 2, 3, \quad (13)$$
where $E\{x_n\}$ denotes the mean value of received signal x_n.
2) *Signal Whiting.* ICA assumes that the individual source signals are independent of each other. We say a zero-mean random vector v is a white signal, meaning that each component of v has a unit variance and is uncorrelated with each other, that is, the covariance matrix of v is an identity matrix. Thereby, the process of applying a linear transformation to a multi-dimensional signal to turn it into a white signal is called whitening. Denote Q as the whiting matrix, we can represent the process as follow:
$$Z = QX', \quad \text{s.t. } Cov(Z) = I,$$
where $X' = [x'_1(t), x'_2(t), ..., x'_N(t)]^T$ denotes the signal after removing the DC component in step 1), Cov(*) represents the calculation of the covariance matrix and I is the identity matrix. The covariance matrix of X' is $C_{X'} = Cov(X')$. $C_{X'}$ can be replaced by PDP^T using singular value decomposition (SVD), where P is the unit eigenvector of $C_{x'}$, and D is a diagonal matrix composed of the eigenvalues of $C_{x'}$. We can finally get the solution of Q [8] as $Q = D^{-\frac{1}{2}}P^T$.
3) *Iterative Update Based on FastICA.* By now, our goal is to recover the original signal S by using the whiting signal Z, that is, find a transformation W so that $Y = WZ(Y \approx S)$. Technically, ICA separates the original signals based on the independence between signals. The Central Limit Theorem [24], a classical result in probability theory, tells that the distribution of a sum of independent random variables tends toward a gaussian distribution, under certain conditions. Thus, a sum of $N(N \geq 2)$ independent random variables usually has a distribution that is closer to gaussian than any of the original random

7. It is worth noting that ICA will change the amplitude of the original signals, but this has no effect on our counting, since we only care about the periodicity of signals

8. The whitening transformation is always possible. It can be proved by the method of linear algebra and we have omitted the proof process due to space limitation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

algorithm 2 Signal Separation Based on FastICA

Input: whiting matrix $Z = [z_1, z_2, z_3]$
Output: separated original signal $Y = [y_1, y_2, y_3]$

initialize a normalized transformation matrix $W(0)$ randomly;
$\eta \leftarrow 0$;
while not converged **do**
  for $i = 1 \rightarrow 3$ **do**
    $w_i(\eta+1) \leftarrow E\{ZG(w_i(\eta)Z)\} - E\{G'(w_i(\eta)Z)\}w_i(\eta)$;
    for $j = 1 \rightarrow i-1$ **do**
      $w_i(\eta+1) \leftarrow w_i(\eta+1) - w_i(\eta+1)w_j^T(\eta+1)w_j(\eta+1)$;
    end for
    $w_i(\eta+1) \leftarrow \frac{w_i(\eta+1)}{\sqrt{w_i(\eta+1)w_i^T(\eta+1)}}$;
  end for
  $\eta \leftarrow \eta + 1$;
end while
return Y

variables. Intuitively speaking, the key to estimating the ICA model is nongaussianity. In our practice, we use negentropy to estimate the nongaussianity of signals, which is based on the information-theoretic quantity of entropy [25]. Our optimization objective function is as follows:

$$max\ J(y), \text{s.t. J(y)} \propto [E\{G(y)\} - E\{G(v)\}]^2, \tag{14}$$

where y denote a recovered signal, v denote a zero-mean Gaussian random variable, $G(*)$ is a non-quadratic function. It should be noted that by choosing G wisely, we can obtain more robust estimators. In our scenario, the motion signal has symmetry and periodicity, and we assume it is a super Gaussian distribution, thus we choose

$$G(u) = tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}, G'(u) = 1 - G(u)^2$$

for our FastICA Algorithm [25]. Therefore, for weight vector w_i (the i-th row vector in transformation matrix W), the update of w_i can be represented as

$$w_i^+ = E\{ZG(w_iZ)\} - E\{G'(w_iZ)\}w_i, i = 1, 2, 3, \tag{15}$$

where we calculate the expectation by treating Z as 3 random variables. It is obvious that we should execute Eq. (15) 3 times (for w_1, w_2, w_3 respectively) during each iteration. To prevent different vectors from converging to the same maxima, we must decorrelate the output after every iteration. A simple way of achieving decorrelation is a deflation scheme based on a Gram-Schmidt-like decorrelation [26]. This means that we estimate the independent components one by one. When we have estimated p vectors $w_1, ..., w_p$, we run the algorithm for w_{p+1}, and after every iteration step subtract from w_{p+1} the *"projections"* $w_{p+1}w_j^T w_j, j = 1, ..., p$ of the previously estimated p vectors, and finally we normalize w_{p+1} as $w_{p+1} = \frac{w_{p+1}}{\sqrt{w_{p+1}w_{p+1}^T}}$. In the end, we repeat the previous iterative process until it converges, which means the change of W is less than a given threshold between two successive iterations. A schematic process is depicted in Fig. 7. The algorithm pseudo-code is shown in Algorithm 2.

TABLE 2: Recogntion Accuracy for Candidate Methods.

Classification Method	Accuracy
MLP	87.8%
Ensemble	92.9%
Decision Tree	89.4%
Discriminant Analysis	77.3%
Nearest Neighbor Classifier	84.8%
Cubic Kernel SVM	**95.2%**
Linear Kernel SVM	88.7%
Gaussian Kernel SVM	91.2%
Quadratic Kernel SVM	93.1%

3.5 Motion Classification

Now we briefly introduce our classification method, which mainly consists of three parts, signal normalization, feature selection and classifier algorithm.

3.5.1 *Signal Normalization:*

When performing a motion, the received signal strength (RSS) fluctuates as the user's body moves. However, the baseline of RSS varies due to different body characteristics (*i.e.*, different height or weight) or environment. In order to retrieve the relative fluctuation pattern, after segmenting the signal, we carry out the signal normalization processing. For a sequence $\vec{x} = (x_1, x_2, ..., x_n)$, the processing method is:

$$x_i' = \frac{x_i - \min(\vec{x})}{\max(\vec{x}) - \min(\vec{x})}, i = 1, 2, 3, ..., n, \tag{16}$$

where $\max(\vec{x})$ and $\min(\vec{x})$ denote the maximum and minimum of $\vec{x}$.

3.5.2 *Feature Selection:*

From Fig.5 we can find that the user moves his body in different patterns when performing different motions. As a result, the fluctuations of the RSS also change in different patterns. In order to differentiate these patterns, we need to find features that accurately depict them. An intuitive idea is to find some statistical features to distinguish them.

After normalization, we extract the following statistical features [6] of new sequence $\vec{x'}$. First of all, the mean and standard deviation values are incorporated, which are given by $\mu = \frac{1}{|\vec{x'}|}\sum_{i=1}^{|\vec{x'}|} x'_i$, and $\sigma = \sqrt{\frac{\sum_{i=1}^{|\vec{x'}|}(x'_i - \mu)^2}{|\vec{x'}|}}$. Second, the maximum and minimum values of $\vec{x'}$ are considered. Typically, the 3 quantiles ($p = 0.25, 0.5, 0.75$), skewness ($Skew(\vec{x'}) = \frac{E(x'-\mu)^3}{\sigma^3}$), kurtosis ($Kurtosis(\vec{x'}) = \frac{E(x'-\mu)^4}{\sigma^4}$) and $\theta = \sqrt{\sum_{n=1}^{|\vec{x'}|} {x'_i}^2}$ are incorporated as features of our classification model.

Although the above 10 features have been able to achieve decent classification accuracy [20], we still want to further improve it. In order to achieve higher accuracy, we need to extract more effective features. A feasible approach is to divide a signal sequence of a motion into slices, and calculate each slice's average RSS as the features [27]. In practice, we divide a motion signal sequence into 20 slices [28]. Up

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

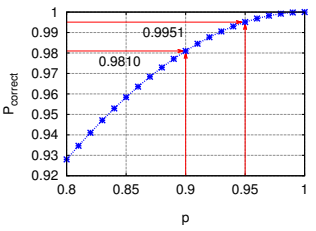


Fig. 8: Voting Strategy.

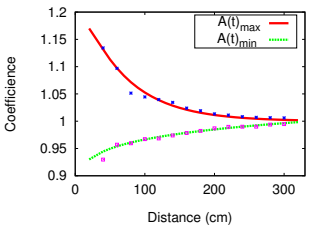


Fig. 9: Impact of Distance.

to now, we have only used the features of the signal in the time domain. Further, we perform Fourier transform (FFT) on the signal to extract the features in the frequency domain, such as spectral centroid, spectral entropy, spectral spread, spectral skewness, spectral kurtosis and apectral flatness [29].

Now, we can represent a motion's overall signal sequence with the 36 features mentioned above. In addition, it is worth noting that we have only used the RSS of our received signals. To make full use of the signals, we calculate their phase sequence and not surprisingly, we find that the fluctuation patterns that exit in the RSS also exit in the phase sequence. We assume that adding features from the phase domain will elevate the recognition accuracy, which is confirmed and validated in our experiments. Similarly, we obtain another 36 features from the phase domain and together we get a feature set of 72 features.

3.5.3 Classifier Algorithm

We collect more than 5000 samples and use the 10-fold cross validation to test different classification methods such as SVM (with various kernel functions), decision trees, discriminant analysis and nearest neighbor classifiers (KNN). Table 2 shows the average classification accuracy. We find that SVM with cubic kernel is the most stable and effective one for our task. Consequently, the kernel function we use is

$$K(\vec{\nu_{\mathbf{i}}}, \vec{\nu_{\mathbf{j}}}) = (1 + \vec{\nu_{\mathbf{i}}}^T \vec{\nu_{\mathbf{j}}})^3, \tag{17}$$

where $\vec{\nu_{\mathbf{i}}}$ and $\vec{\nu_{\mathbf{j}}}$ represent the feature vectors. Besides, we use a one-vs-one multi-classification approach for our model.

Moreover, since our counting algorithm has segmented the experimental data, given a series of k motions, we can provide k segmentations to the classifier and will get k results in return. Thus we can use voting method to get a more accurate result. Owing to the judgment for each segmentation is independent, we can denote the probability of judging correctly once as p, and for three (here we denote $k = 3$) consecutive segments (they are the same motion) the correct probability is $P_{correct} = p^3 + \binom{3}{2} \cdot p^2(1-p) + \binom{3}{1} \cdot p(1-p)^2 \cdot \frac{1}{3} = p + p^2 - p^3$.

Fig. 8 shows the relationship between $P_{correct}$ and p. As depicted in Fig. 8, when p = 0.9, our accuracy can further be improved to 98.10 %, when p increased to 0.95, the accuracy will be raised to 99.51 % using three repetitions!

4 Implementation and Evaluation

4.1 Implementation

We implement a prototype of *Motion-Fi* $^+$ using COTS components for signal backscatter and an USRP-RIO for signal receiving and processing. The prototype of *Motion-Fi* $^+$ is built with our customized design of passive tag, which backscatters the signal of plug-in device to the receiver. Fig. 10a shows our prototype of a passive tag, whose main components are SPDTs, antenna interfaces, pins and resistances, the size of which is about 3.4×3.4 cm^2. In order to enable our tag adapt to high-frequency signal, we choose the HMC190BMS8 SPDT. An additional FPGA is used to control the backscatter state of the passive tag. The reflected signal is received by a NI USRP-2953R, and then fed to our proposed scheme to recognize repetitive motions. Another USRP serves as a plug-in device to transmit signal in the environment. Our equipments are shown in Fig. 10b. It is worth noting that the power consumption of backscatter process can be as low as several μW [1] in theory, but we actually use FPGA to implement this process with no energy harvesting module due to the cost and professionalism of circuit design.

Besides, comparing with our previous work *MotionFi* [20], we continuously improve the reflection effect of the tag and reduce the composition and area of the Printed Circuit Board (PCB). In this work, we designed a brand new backscatter tag (shown in Fig.10c), whose size is about 2.2×1.2 cm^2 , smaller than a one US dollar coin (diameter of 2.64 cm), $\sim$ 80 % reduction in area compared with the previous one (shown in Fig.10a). Its main components are reflective switch (ADG902), SMA (SubMiniature version A) connector (where an extra antenna is connected), control pins and capacitance. We believe that smaller size can facilitate wider deployment, and we use this new tag in our multi-user experiments.

Our software platform is built upon LabVIEW, where coding program could be put into our aforementioned hardware platform. Our FPGA program built with LabView is applied to produce a square wave at a frequency shift. It is worth noting that, in practical design, a 'perfect' square wave is not available[9]. Thus we use the sine wave to approximate. From Fourier analysis, a square wave can be written as:

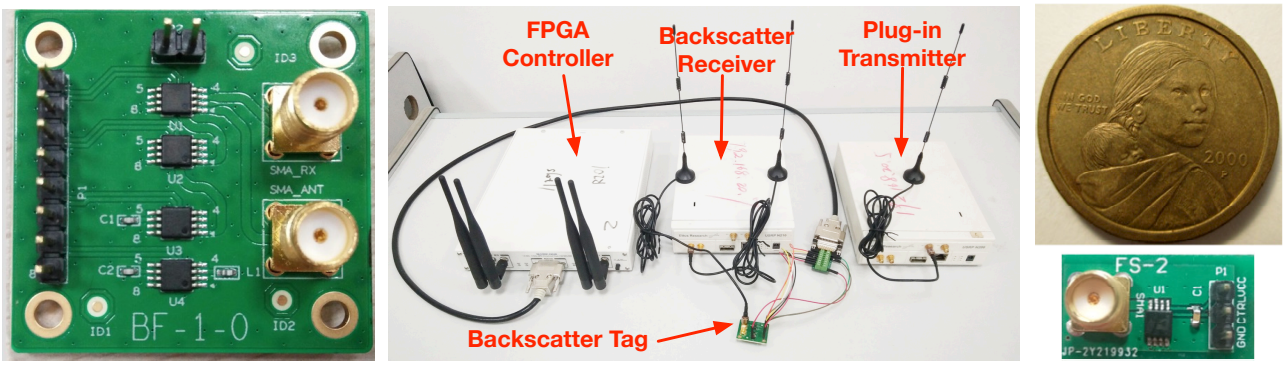
$$Square(\Delta f * t) = \frac{4}{\pi} * \sum_{n=1,3,5,\ldots}^{\infty} \frac{1}{n} \sin(2\pi n \Delta f * t) \tag{18}$$

Here the first harmonic is a sinusoidal signal at the desired frequency Δf. Note that the power in each of these harmonic scales as $\frac{1}{n^2}$. So the third and the fifth harmonic are around 9.5 dB and 14 dB lower than the first harmonic. Thus, we can approximate a square wave as just the sinusoidal signal, $\frac{4}{\pi}\sin(2\pi\Delta f * t)$. In practice, we use a carrier centering at 2 GHz, and denote Δf equals 20 MHz.

4.2 Experiment Settings

Deployment and Layout: We analyze the performance of our system in a typical office environment, which covers a 6.3×5.7 m^2 area, consisting of some bookcases and office furniture, including desks, chairs and computers. One USRP serves as a plug-in device for single-frequency tone signal transmission and another one serves as a receiver. The

9. The transition between minimum to maximum is instantaneous for an ideal square wave, but this is not realizable in physical systems.



(a) Prototype of Tag. (b) Equipments. (c) New Tag.

Fig. 10: Demonstration Equipments. (a) and (c) are our two versions of backscatter tag. (b) shows the USRP platform.

transmitter is placed in a desktop as shown in Fig. 1, and the receiver and tag pair are placed in the front of the gym area. When multiple persons performing motions, each person does it in an area $\sim 2m^2$ in front of one tag (as shown in Fig. 1).

Fig. 11 shows two typical deployment of **Tag**, **TX** and **RX**. As depicted in Fig. 11a, the **Tag** is away from the very near **TX** and **RX** pair. It works as if **Tag** is a client, which is affected by the server side **TX** and **RX** pair. The near deployment of **TX** and **RX** is similar to RFID reader. While for Fig. 11b, the difference is **TX** and **RX** are deployed seperately with each other. This mode could be utilized when WiFi infrastructure and receiving devices are working independently and distributedly.

Moreover, we find that the strength of the received signal is only related to the path **Tx** $\rightarrow$ **Tag** $\rightarrow$ **Rx**, and has nothing to do with the path between **Tx** and **RX**. Consequently, we fix the positions of **Tx** and **Rx**, observe the change of received signal when **Tag** moved between them, as shown in Fig. 11c. We can see that the strength of the received signal increases when the tag get close to either **Tx** or **Rx** and it achieves the minimum at the midpoint of the connection between **Tx** and **Rx**. This is well understood since the signal strength is proportional to $\frac{1}{d_1 \times d_2}$, and the experimental result is consistent with this mathematical meaning. Finally, we treat **Tx** as a infrastructure that can not moved randomly. So, in order to obtain a stronger received signal and enable **Tx** to cover further distance, we place the **Tag** and **Rx** together as depicted in Fig. 1.

Fig. 12 depicts the orientations and distances of the persons involved in evaluations. Basically, there are four typical directions (Front, Rear, Left and Right) to evaluate the impact of orientations. Each orientation denotes the relative position between a person and tag. *i.e.*, Front means the tag is in front of the person. Meanwhile, we test five distances from 0.6 m to 3 m. The maximum distance is set to 3 m, because the backscattering signal is too weak to be used for recognition at this distance according to our evaluations (Fig. 9).

Volunteers and Concurrent Motions: We recruited 26 volunteers, including 16 males and 10 females for testing over 6 months. They vary in age (18 - 45 years old), stature (155 - 188 cm) and weight (45 - 90 kg). During the experiments, they wore their daily attire with different fabrics and they performed motions at the specified location within the scope of the 2 m^2 square area. The variance in experimental environment and the presence of other users had a negligible impact on the results, because the backscatter signal is relatively weak, where only nearby motions can cause discernible volatility on the receiver signal strength. In our scenario, up to 3 persons could perform motions concurrently. We then show all experimental results in the following benchmark tests.

4.3 Long-term Experimental Results

We set up a gym area to evaluate the performance of *MotionFi* $^+$ for more than half a year. Volunteers perform daily exercise in the gym, and our equipments record the data at the same time. We record the meta-data for each exercise, including name, time, motion, number, *etc.*. We totally record thousands test data. Each test contains one of the 7 motions as listed in Table 1. One motion is repeated from 20 to 80 times during each test. We evaluate our proposed scheme on each test data, and examine the error ratio of counting for motions. The error ratio of counting is defined as

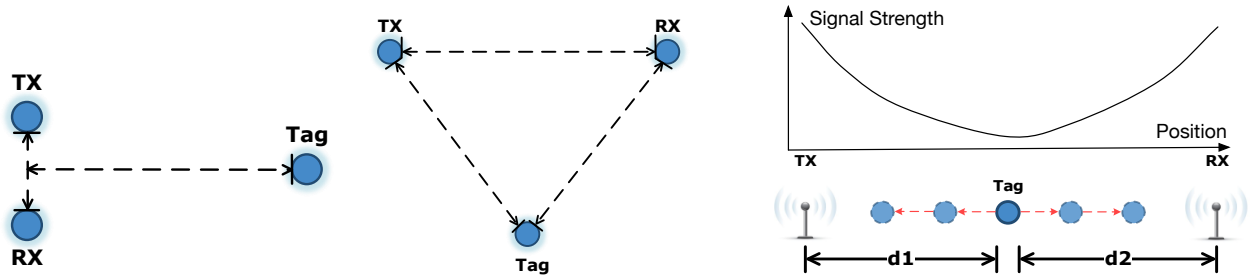
$$error\ ratio = \frac{N_{est} - N_{truth}}{N_{truth}} \times 100\%,$$

where N_{est} is the number of estimated motion counting, and N_{truth} is the number of recorded motion counting. Note error ratio in one test could be negative. The evaluation results are averaged on the test records.

Performance of single person: We evaluate the error ratio of motion counting for various motions, persons, orientations and distances when only one person performs motions. To do this, we place a transmitter, a receiver and a tag in the room as shown in Fig. 12, the transmission power is set to 10 dBm.

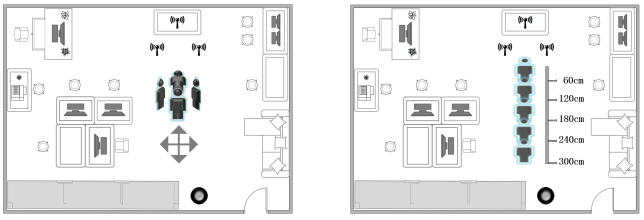
We firstly study the impact of distance on the strength of received signal. The distance from tag to person varies from 40 cm to 300 cm with step of 20 cm. Specifically, a volunteer performs 100 motions at each location apart from tag 20, 40, $\cdots$, 300 cm respectively. Fig. 9 shows the changes of $A(t)$ with the increase in distance, where $A(t)$ denotes the coefficient on received signal with human motions, A_0 denotes the amplitude without human motions (Eq. (6)), $A(t)_{max} = \max(E'(t))/A_0$ is the maximum value of $A(t)$, and $A(t)_{min} = \min(E'(t))/A_0$ is the minimum value of $A(t)$. By showing the difference of $A(t)_{max}$, $A(t)_{min}$ respectively, we conclude that the effect of human motion on

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing



(a) 'Client-Server' mode (b) 'Peer-to-Peer' mode (c) Testing Scenarios

Fig. 11: Deployment of Tags, Transmitter and Receiver.



(a) Orientations. (b) Distances.

Fig. 12: Configurations of Persons on Orientation and Distance.

received signal intensity is less than 2% when the distance is beyond 2 meters. This is a favorable property for the application such as exercising in a gym as this observation enables motion counting when a lot of people doing exercise simultaneously with a small seperation distance among them.

We choose one volunteer to perform all the seven motions, and examine the error ratio of counting for each of the motions. The results are shown in Fig. 13a. It can be seen that the error ratio of counting are different among motions, from 0 to 2 %. Noted that the error ratio of Push-ups (PU), Stoop-down (SD) and Dumbbell (DB) are relative small, while the error ratio of Squats (SQ), Sit-ups (SU), Leg-raise (LR) and Step (ST) are larger. The reason is that when a person performs these type of motions, part of the limb's range of activities is relative large, which lead to larger variances of received signal.

To study the changes of error ratio when different persons perform the same motion, we arrange for 6 volunteers labeled with P1 to P6 to perform squats one by one. The results are shown in Fig. 13b. In general, the average of the error ratio is no more than 2 %. Though there do exist difference among different people, *i.e.*, gender, height, weight and habit, there is no significant difference in the counting accuracy of different people. Since our counting algorithm is adaptive, independent of the template, focusing only on the periodicity of the signal.

To study the impact of orientations when people do motions, we let a volunteer do squats in 4 orientations as shown in Fig. 12a. The evaluation results are shown in Fig. 13c. It can be seen that the error ratio of Front and Rear are lager than that of Left and Right. The reason is that when people do motions with orientations of Front and Rear, the cross section of their body is larger than that of Left and Right. Thus, inconsistency of motions with orientations of Front and Rear will cause large fluctuations in backscattered signals.

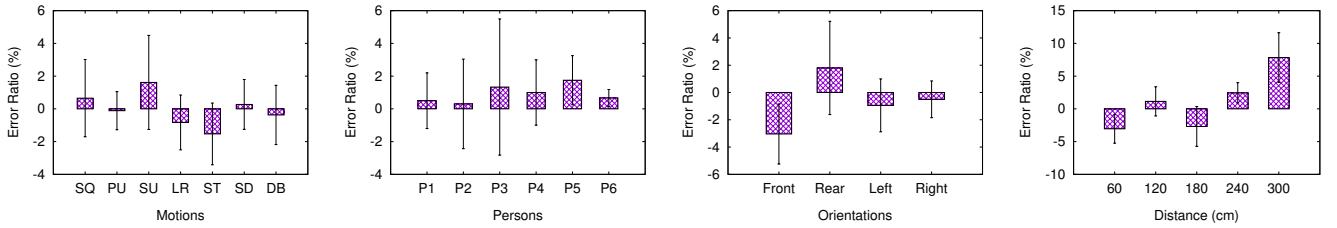
Finally, we study the impact of distance between person and tag on the counting error ratio. A volunteer performs squats at 5 distances ranging from 60 cm to 300 cm as shown in Fig. 12b. The error ratio at different distances are shown in Fig. 13d. We can see that when the distance is $\geq$ 100 cm, the error ratio increases because the strength of backscattering signal decreases with distance. Noted that, when the distance is $\leq$ 100 cm, *e.g.* 60 cm in our tests, the error ratio increases as distance decreases. Although the strength of backscattering signal increases when distance decreases, the interference caused by the irregular movement of limbs also increases. Thus, the distance $\sim$ 1 m is ideal for our applications.

Performance of multiple persons: We then evaluate the performance of *Motion-Fi* $^+$ when multiple persons perform motions simultaneously. We deploy three sets of equipments in our laboratory as shown in Fig. 1. Specifically, we deploy one transmitter to transmit the tone signal. For each person, a pair of tag and receiver (all receivers are synchronous) is used to recognize his/her motion. During our evaluation, three persons perform motions in the test area simultaneously. Each person stands before the tag at a distance about 1 m. Three persons are labeled with 'Left', 'Middle' and 'Right' respectively. We examine the error ratio of motion counting under various conditions.

We firstly study the impact of transmission power of the tone signal. To do this, we configure the transmission power to 0, 10 and 20 dBm respectively. Three persons perform squats in front of the tag at a distance of 1 m simultaneously and the separation between them is 2 m. Fig. 14a shows the error ratio of motion counting. It can be seen that there is a trade-off between transmission power and error ratio. In our experiment configuration, a transmission power of 10 dBm can achieve the lowest error ratio. Noted that when transmission power is too weak, for example 0 dBm in our case, the backscattered signal becomes too weak to discern. However, when transmission power is too strong, the interference from adjacent person becomes more serious. So, we configure the power to 10 dBm in the following tests.

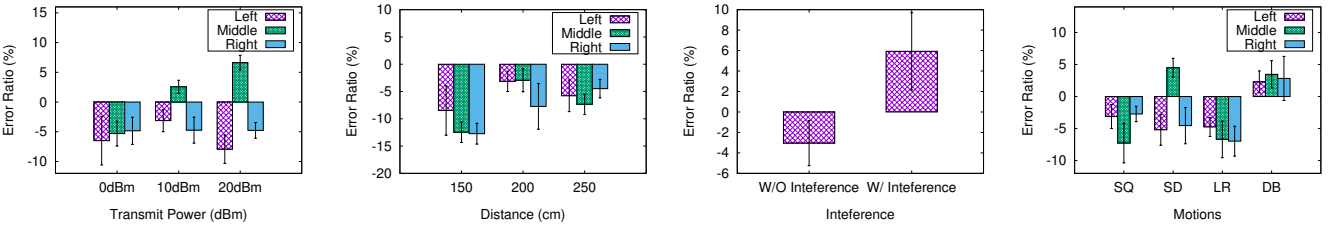
We then examine the impact of separation distance between users. Again, three volunteers perform squats simultaneously. The distance between tag and person is about 1 m. We change the separation between persons from 150 cm to 250 cm at a interval of 50 cm. The results are shown in Fig. 14b. When the separation is small, the interference

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing



(a) Impact of Motions. (b) Impact of Persons. (c) Impact of Orientations. (d) Impact of Distances.

Fig. 13: Impact on Different Motions, Persons, Orientations and Distances.



(a) Impact of Signal Strength. (b) Impact of Separation Distcance. (c) Impact of Interference. (d) Impact of Same Motion.

Fig. 14: Impact of Multiple Persons.

from a neighbor person is serious, leading to large error ratio of motion counting. However, there are no significant difference between the results of 200 cm and 250 cm, which indicates larger separation incurs less interference. It is recommended that a separation distance of 2 m is appropriate for our configuration.

We study the impact of interference from other persons' motions (even undefined motions). In the evaluation, one volunteer performs squats in the front of a tag while another volunteer perform some irregular movement as interference at a distance of more than 2.5 m away. We compare the error ratio of counting with and without interference. Fig. 14c shows the results. We can see that even with interference from other person's irregular movement, the error ratio of counting is still less than 6 %, which shows that our system works acceptably well even with some interfering activities.

We then examine the error ratio of counting when multiple persons perform same motions simultaneously. In our test, three volunteers perform the same motion. The separation between them is 2 m. The experimental results are shown in Fig. 14d. The error ratio of counting for different motions are relatively small, about 7 % in our test results. Noted that, the error ratio of 'Middle' persons are larger than that of 'Left' and 'Right' person. The reason is that the signal that monitor the middle person contains more interference than that of the boundary ones.

At last, we evaluate the performance when multiple persons perform different motions simultaneously. Three volunteers perform different combination of motions. The separation between them is also set as 2 m. As shown in Fig. 15, the tick name of x-axis denotes the combination of motions. For example, 'LR/SD/DB' denotes the 'Left' person performs leg-rise (LR), the 'Middle' person performs stoop-down (SD) and the 'Right' person performs dumbbell (DB). According to the result, it can be seen that the error ratio is relatively small regardless of the combination of motions. Noted that, the error ratio of Dumbbell (DB) is about -4 %. The main reason for this is that when people do

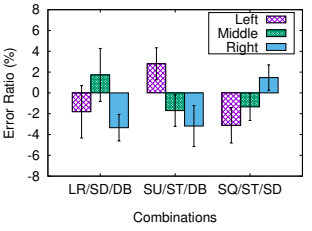


Fig. 15: Impact of Different Motions.

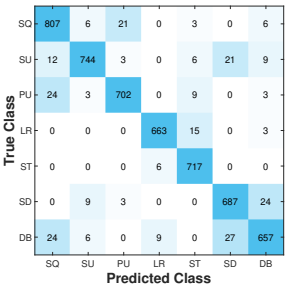


Fig. 16: Confusion Matrix.

TABLE 3: Accuracy of Unknown Subjects.

Motion	SQ	PU	SU	LR
Accuracy	143/150	133/150	144/150	136/150
Motion	ST	SD	DB	**Total**
Accuracy	139/150	140/150	137/150	**972/1050**

dumbbell, the cross section of moving upper limbs is small, leading to small variation in received signal, thus making it susceptible to interference from others. It's recommended that people can do motions with obvious limbs movement to achieve better counting accuracy.

In general, thanks to the design of signal separation algorithm (ICA), the counting results of multi-person scenes have been improved compared to the previous work MotionFi [20]. By averaging all the testing items, we found that the calculation error was reduced a lot and the overall error has controlled within 7 % (originally 15 % without ICA).

Classification: We also examine the performance of single-user motion recognition in a given environment. To do this, we collect lots of environmental data and manually divide more than 5000 samples of 7 motions. After that, we extract 10 features from each sample compositing characteristic matrix for cubic SVM model training. The training process only takes a few minutes. We use the 10-fold cross validation to test our model, which reaches 95.2 % recognition accuracy. The classification speed of our SVM model is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2971996, IEEE Transactions on Mobile Computing

TABLE 4: Distribution of Misjudgement Times.

Misjudgement Times	0	1	2	3	Total
Test Set	88	10	2	0	100

$\sim 3000/s$. The confusion matrix for classification is shown in Fig. 16, which sums up the result of 10-CV. Besides, in order to test the generalization of our classification model, we recruit 5 additional volunteers, each volunteer performing each motion 30 times and we obtain 1050 (5*7*30) samples in total. Then, we used the trained model to classify them. The classification accuracy is 92.6 % (shown in Table 3), which shows that our model has good generalization ability.

Moreover, in order to further improve the prediction accuracy, we choose $k = 3$ to make multiple judgments (as described in Sec. 3.5). To verify this result, we select 100 sets of motion data, and only take the first three segments of each group to judge. Our experimental results reach a correct rate of 100 %, Table 4 shows the distribution of the number of misjudgments among 100 sets of data. The results show that for continuous repetitive motion, after three or more motions, the classifier can achieve almost error-free recognition.

5 Related Work

This section will be divided into two parts, the first part introduces the related work on the passive platform, and the second part is about the relevant work about perception.

5.1 Battery-free Backscattering Network

Beginning in 2013, many sessions began to emerge articles about the results of passive platforms. There are many sources of energy in the environment, such as light energy, wind energy, electromagnetic energy, radio (RF) energy [14], [30]. Backscatter based communication received considerable research interests recently, including, ubiquitous energy acquisition, passive protocol optimization and passive component separation. Liu *et al.* [31] use the signal from TV towers to obtain energy for the tag-to-tag communication. Inspiringly, Kelloggvasisht2018body *et al.* [32] get the energy from the WiFi APs for backscatter communication. Based on this innovative design, Bharadia *et al.* [9] realize the optimization on conventional WiFi protocol. Kellogg *et al.* [1] propose a revolutionary design, which uses the backscattering signals from the air and keep the ADC component processing. Zhang *et al.* [33] present a low power backscatter system, which allows a tag to embed its information on standard 802.11b codeword to another valid 802.11b codeword. Many solutions have explored the passive projects, One of the important theories they use is frequency shift [34]. Recently, backscatter-based systems are gradually proposed for motion sensing [35], localization [36] and tracking [37] In this work, we study how to use backscattering signals to realize repetitive motion recognition and counting.

5.2 Device-free Sensing

In recent years, exponential explosive growth of mobile devices once again ignited the people of the new form of human-computer interaction exploration, the use of human-computer interaction to control a wide range of applications [12]. Gesture recognition system, as the basic solution of human-computer interaction, has become more and more popular. The current system allows users to use not only dedicated equipment, but also natural body movements and context-related information. In fact, implantation of gesture recognition systems in electronic products and mobile devices has become commonplace and is on the rise, such as smartphones [38], laptop [39], navigation facility [40] and some game control system [41]. These systems typically implement gesture recognition by utilizing a variety of sensors available on the device, such as computer vision (camera, vidicon, etc.) [41], inertial sensors [42]–[44], vibration sensor [45], acoustics [38], [46], light sensors [47]. However, these technologies still encounter a lot of limitations when they are implemented, such as being customized for specific applications, sensitive to light, high installation costs or high equipment costs, requiring handheld devices, or the need to install additional sensors.

To address these challenges, researchers start to leverage wireless signals (e.g., ultrasound [48], RFID [49], 60GHz mmWave [50], etc.) to achieve device-free human activity recognition. With the proliferation of ubiquitous wireless devices and the establishment of wireless network infrastructure, WiFi-based recognition systems [7], [8], [51], [52] gradually being put forward. WiFi signal can be used not only in gesture recognition, but also localization [8], human identification [6], vibration detection [53] and so on. These WiFi-based systems operate by analyzing the characteristic changes of the wireless signal, such as analyzing changes in CSI (channel state information) or RSSI (received signal strength indication) caused by human motion [54]. Our work leverages the backscattering signals to recognize motions, in order to overcome the limitations in directly applying WiFi signals when multiple users and complicated scenarios are incorporated.

6 Conclusion

We present an accurate wireless sensing system building upon a passive backscattering platform. In wireless backscattering systems, human motions could be effectively explored using our customized noise taming and self-adaptive pattern matching algorithms. We evaluate our design with extensive experiments, which shows a satisfying recognition accuracy. Different from previous studies leveraging backscattering technology, our system leverages pervasive wireless APs as RF source instead of customized readers in RFID systems, and enables multiple users peform simultaneously.

Acknowledgement

The work is partially supported by the National Key R&D Program of China 2018YFB0803400, China National Funds for Distinguished Young Scientists with No. 61625205, NSFC with No. 61751211, No. 61520106007, Key Research Program of Frontier Sciences, CAS, No. QYZDY-SSW-JSC002, and NSF CNS 1526638.

REFERENCES

[1] B. Kellogg, V. Talla, S. Gollakota, and J. R. Smith, "Passive Wi-Fi: Bringing low power to Wi-Fi transmissions," in *NSDI'16*, vol. 16. USENIX Association, 2016, pp. 151–164.

[2] C. C. Yen, A. E. Gutierrez, D. Veeramani, and D. van der Weide, "Radar cross-section analysis of backscattering rfid tags," *IEEE Antennas and Wireless Propagation Letters*, vol. 6, pp. 279–281, 2007.

[3] P. V. Nikitin, K. V. S. Rao, and R. D. Martinez, "Differential RCS of RFID tag," *Electronics Letters*, vol. 43, no. 8, pp. 431–432, April 2007.

[4] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with Wi-Fi signals," *Computer*, vol. 50, no. 1, pp. 48–57, 2017.

[5] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *MobiCom'16*. ACM, 2016, pp. 95–108.

[6] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: Wifi-based person identification in smart spaces," in *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*. IEEE Press, 2016, p. 4.

[7] 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*. ACM, 2013, pp. 27–38.

[8] F. Adib and D. Katabi, "See through walls with WiFi!" in *SIGCOMM*, vol. 43, no. 4. New York, NY, USA: ACM, aug 2013, pp. 75–86.

[9] D. Bharadia, K. R. Joshi, M. Kotaru, and S. Katti, "Backfi: High throughput wifi backscatter," *ACM SIGCOMM Computer Comm. Review*, vol. 45, no. 4, pp. 283–296, 2015.

[10] Y. Yan, P. Yang, X.-Y. Li, Y. Zhang, J. Lu, L. You, J. Wang, J. Han, and Y. Xiong, "Wizbee: Wise zigbee coexistence via interference cancellation with single antenna," *IEEE Transactions on Mobile Computing*, vol. 14, no. 12, pp. 2590–2603, 2015.

[11] J. Han, C. Qian, P. Yang, D. Ma, Z. Jiang, W. Xi, and J. Zhao, "Geneprint: Generic and accurate physical-layer identification for uhf rfid tags," *IEEE/ACM Transactions on Networking (TON)*, vol. 24, no. 2, pp. 846–858, 2016.

[12] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *INFOCOM*. IEEE, 2015, pp. 1472–1480.

[13] K. Kurokawa, "Power waves and the scattering matrix," *IEEE transactions on microwave theory and techniques*, vol. 13, no. 2, pp. 194–202, 1965.

[14] J. F. Ensworth and M. S. Reynolds, "Every smart phone is a backscatter reader: Modulated backscatter compatibility with bluetooth 4.0 low energy (ble) devices," in *RFID (RFID), 2015 IEEE International Conference on*. IEEE, 2015, pp. 78–85.

[15] B. Li, P. Yang, J. Wang, Q. Wu, S. Tang, X.-Y. Li, and Y. Liu, "Almost optimal dynamically-ordered channel sensing and accessing for cognitive networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2215–2228, 2014.

[16] P. Yang, B. Li, J. Wang, X.-Y. Li, Z. Du, Y. Yan, and Y. Xiong, "Online sequential channel accessing control: A double exploration vs. exploitation problem," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4654–4666, 2015.

[17] X.-Y. Li, P. Yang, and Y. Yan, "Spa: Almost optimal accessing of nonstochastic channels in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1540–1553, 2016.

[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. The MIT Press, 2009.

[19] S. McKinley and M. Levine, "Cubic spline interpolation," *College of the Redwoods*, vol. 45, no. 1, pp. 1049–1060, 1998.

[20] N. Xiao, P. Yang, Y. Yan, H. Zhou, and X.-Y. Li, "Motion-fi: Recognizing and counting repetitive motions with passive wireless backscattering," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2024–2032.

[21] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the acoustical society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[22] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. John Wiley & Sons, 2004, vol. 46.

[23] Z. Jia, A. Bonde, S. Li, C. Xu, J. Wang, Y. Zhang, R. E. Howard, and P. Zhang, "Monitoring a person's heart rate and respiratory rate on a shared bed using geophones," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. ACM, 2017, p. 6.

[24] Central Limit Theorem (CLT), "https://en.wikipedia.org/wiki/Central_limit_theorem/."

[25] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[26] Gram–Schmidt Process, "https://en.wikipedia.org/wiki/Gram–Schmidt_process/."

[27] 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*. ACM, 2014, pp. 541–551.

[28] D. Ren, Y. Zhang, N. Xiao, H. Zhou, X. Li, J. Qian, and P. Yang, "Word-fi: Accurate handwrite system empowered by wireless backscattering and machine learning," *IEEE Network*, vol. 32, no. 4, pp. 47–53, 2018.

[29] Spectral Features, "http://docs.twoears.eu/en/latest/afe/available-processors/spectral-features/."

[30] P. Zhang and D. Ganesan, "Enabling bit-by-bit backscatter communication in severe energy harvesting environments." in *NSDI*, 2014, pp. 345–357.

[31] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: wireless communication out of thin air," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 39–50, 2013.

[32] B. Kellogg, A. Parks, S. Gollakota, J. R. Smith, and D. Wetherall, "Wi-Fi backscatter: Internet connectivity for RF-powered devices," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 607–618, 2015.

[33] P. Zhang, D. Bharadia, K. R. Joshi, and S. Katti, "Hitchhike: Practical backscatter using commodity wifi." in *SenSys*, 2016, pp. 259–271.

[34] V. Iyer, V. Talla, B. Kellogg, S. Gollakota, and J. Smith, "Inter-technology backscatter: Towards internet connectivity for implanted devices," in *SIGCOMM*. ACM, 2016, pp. 356–369.

[35] N. Xiao, P. Yang, Y. Yan, H. Zhou, J. Hou, and X.-Y. Li, "Dfmose: Device-free motion sensing with wireless backscattering," in *The 25th Annual International Conference on Mobile Computing and Networking*. ACM, 2019, p. 72.

[36] D. Vasisht, G. Zhang, O. Abari, H.-M. Lu, J. Flanz, and D. Katabi, "In-body backscatter communication and localization," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. ACM, 2018, pp. 132–146.

[37] N. Xiao, P. Yang, X.-Y. Li, Y. Zhang, Y. Yan, and H. Zhou, "Milliback: Real-time plug-n-play millimeter level tracking using wireless backscattering," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, p. 112, 2019.

[38] S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: using the doppler effect to sense gestures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 1911–1914.

[39] R. Block, "Toshiba qosmio g55 features spursengine," *Visual Gesture Controls*.

[40] A. Santos, "Pioneer's latest raku navi gps units take commands from hand gestures," 2014.

[41] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Comm. of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[42] G. Cohn, D. Morris, S. Patel, and D. Tan, "Humantenna: using the body as an antenna for real-time whole-body interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 1901–1910.

[43] C. Harrison, D. Tan, and D. Morris, "Skinput: appropriating the body as an input surface," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 453–462.

[44] D. Kim, O. Hilliges, S. Izadi, A. D. Butler, J. Chen, I. Oikonomidis, and P. Olivier, "Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor," in *Proceedings of the 25th annual ACM symposium on User interface software and technology*. ACM, 2012, pp. 167–176.

[45] Z. Jia, M. Alaziz, X. Chi, R. E. Howard, Y. Zhang, P. Zhang, W. Trappe, A. Sivasubramaniam, and N. An, "Hb-phone: a bed-mounted geophone-based heartbeat monitoring system," in *Information Processing in Sensor Networks (IPSN), 2016 15th ACM/IEEE International Conference on*. IEEE, 2016, pp. 1–12.

[46] L. Zhang, K. Liu, Y. Jiang, X.-Y. Li, Y. Liu, P. Yang, and Z. Li, "Montage: Combine frames with movement continuity for real-time multi-user tracking," *IEEE Transactions on Mobile Computing*, vol. 16, no. 4, pp. 1019–1031, 2017.

[47] T. Li, C. An, Z. Tian, A. T. Campbell, and X. Zhou, "Human sensing using visible light communication," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 331–344.

[48] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 289–304.

[49] L. Yang, Q. Lin, X. Li, T. Liu, and Y. Liu, "See through walls with cots RFID system!" in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 487–499.

[50] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, p. 142, 2016.

[51] M. Scholz, S. Sigg, H. R. Schmidtke, and M. Beigl, "Challenges for device-free radio-based activity recognition," in *Workshop on Context Systems, Design, Evaluation and Optimisation*, 2011.

[52] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using wifi," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2017, pp. 252–264.

[53] T. Wei, S. Wang, A. Zhou, and X. Zhang, "Acoustic eavesdropping through wireless vibrometry," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 130–141.

[54] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and Modeling of WiFi Signal Based Human Activity Recognition," in *Proceedings of the 21st annual international conference on mobile computing and networking*. ACM, 2015, pp. 65–76.

Ning Xiao received the B.S. degree in computer science and technology from the College of Computer Science and Technology, University of Science and Technology of China (USTC), China, in 2017. He is now a postgraduate student in USTC. His current research interests include wireless sensing systems, wireless networks and IoT Technology.

Panlong Yang (M'02) received his B.S. degree, M.S. degree, and Ph.D. degree in communication and information system from Nanjing Institute of Communication Engineering, China, in 1999, 2002, and 2005 respectively. Dr. Yang is now a professor in the College of Computer Science and Technology, University of Science and Technology of China. His research interests include wireless sensor networks, battery-free communication, wireless sensing and software defined radio networks. He is a member of the IEEE Computer Society and ACM SIGMOBILE Society. He has published over 150 papers including 36 papers on CCF Rank A list, such as ACM MobiCom, ACM Ubicomp, IEEE JSAC, IEEE TMC etc. Professor Yang has won best paper awards on IEEE MSN 2013, and best paper runner up awards on ACM MobiHoc 2014.

Yubo Yan (S'10, M'19) received the B.S. degree, M.S. degree and Ph. D degree in communication and information system from PLA University of Science and Technology, China, in 2006, 2011 and 2017 respectively. He is now an associate professor in University of Science and Technology of China. His current research interests include Internet of Things, wireless networks, smart sensing, and mobile computing. He is a member of the IEEE and the IEEE Communications Society.

Hao Zhou (M'15) received his B.S. and Ph.D. degrees in Computer Science from the University of Science and Technology of China, Hefei, China in 1997 and 2002. He is now an associate professor at the University of Science and Technology of China, and he worked as a project lecturer in the National Institute of Informatics (NII), Japan, from 2014 to 2016. His research interests are in the area of Internet of Things,wireless communication and software engineering.

Xiang-Yang Li (SM'08, F'15) is a professor and executive dean at School of Computer Science and Technology at University of Science and Technology of China. He is an IEEE Fellow and an ACM Distinguished Scientist. He was a professor at the Illinois Institute of Technology. Dr. Li received MS (2000) and PhD (2001) degree at Department of Computer Science from University of Illinois at Urbana-Champaign, a Bachelor degree at Department of Computer Science and a Bachelor degree at Department of Business Management from Tsinghua University, P.R. China, both in 1995. His research interests include wireless networking, mobile computing, security and privacy, cyber physical systems, and data sharing. He and his students won several best paper awards and best demo award.

Haohua Du is a Ph.D. student in Computer Science at Illinois Institute of Technology. She received the B.E. degree from Beihang University, Beijing, P.R. China. Her research interests include mobile computing, wireless sensing and data privacy.