IET Wireless Sensor Systems

Special Issue: Smart Cities and Smart Sensory Platforms

IET Journals
The Institution of Engineering and Technology

WiFi-based passive sensing system for human presence and activity event classification

ISSN 2043-6386
Received on 1st November 2017
Revised 11th June 2018
Accepted on 24th July 2018
E-First on 4th October 2018
doi: 10.1049/iet-wss.2018.5113
www.ietdl.org

Wenda Li[1] ✉, *Bo Tan*[2], *Robert J. Piechocki*[1]

[1]*Department of Electronic and Electrical Engineering, University of Bristol, Bristol, UK*
[2]*School of Computing, Electronics and Mathematics, Coventry University, UK*
✉ *E-mail: wenda.li@bristol.ac.uk*

Abstract: Detection of human presence and activity event classification are of importance to a variety of context-awareness applications such as e-Healthcare, security, and low impact building. However, existing radio frequency identification tags, wearables, and passive infrared approaches require the user to carry dedicated electronic devices that suffer from problems of low detection accuracy and false alarms. This study proposes a novel system for non-invasive human sensing by analysing the Doppler information contained in the human reflections of WiFi signal. Doppler information is insensitive to stationary objects, thus there is no need for any scenario-specific calibration which makes it ideal for human sensing. We also introduce the time-frequency domain feature vectors of WiFi Doppler information for the support vector machine classifier towards activity event recognition. The proposed methodology is evaluated on a software defined radio system together with the experiment of five different events. The results indicate that the proposed system is sufficient for indoor context awareness, with 95.3% overall accuracy for event classification and 93.3% accuracy for human presence detection, which outperforms the traditional received signal strength approach where accuracy is 69.3% for event classification and 83.3% for human presence detection.

1 Introduction

Accurate human sensing is essential for context awareness to improve building management system and design of impact building. It also plays an important role in many e-Healthcare applications, such as infant monitoring in rooms, monitoring of the elderly at home, patient monitoring in hospitals, and safety management in offices [1–6]. Traditional approaches for indoor context awareness include vision [7], RFID [8], environment sensors [9], and passive infrared (PIR) sensors [10]. However, PIR and environment sensors suffer from low accuracy problems and require additional spending on infrastructure; RFID sensors need to be carried by the user which can be inconvenient for long-term monitoring, while vision-based techniques have limitations such as dependence on indoor lighting conditions and also privacy concerns. Ultra-wide-band (UWB) radar can provide a high resolution in both range and Doppler information, and has also been used for activity recognition. However, it contains one or more specifically designed transmitters and a commercial frequency allocation is required [11]. Therefore, a passive, cost-effective and low-intrusive technique for accurate human sensing in a residential area is urgently required.

With the widespread deployment of WiFi infrastructure in residential, commercial and industrial buildings, WiFi has become the primary resource for non-invasive human sensing [1, 2, 12, 13]. There are a number of WiFi-based approaches for context-awareness applications such as indoor localisation [1, 14], activity recognition [12], and writing recognition [15]. The basic principle behind WiFi sensing is to measure the variations in WiFi signals caused by human motion. One straightforward approach for human presence detection is measuring the received signal strength (RSS) at the MAC layer that can be easily done by commercial off-the-shelf WiFi devices. Yet, the RSS is limited by its poor sensing accuracy because it cannot measure the multipath effects from human motion [16]. On the other hand, channel state information (CSI), the fine-grained measurement at the physical (PHY) layer, is often used to describe the WiFi signal propagation property which contains impacts from human presence and movements [1, 12, 14–16]. These approaches require specific network cards to exam the phase shifts from multiple radio frequency (RF) channels. For example, [12] quantifies the relationship between human movement speeds and dynamic CSI then pass to a CSI-Speed model to interpret the possible situation, and [1] leverages the signal tendency index to exam the shape similarity in adjacent CSI to detect the possible existing of moving objects. Another approach is to use the near-field angle-of-arrival to recognise arm and hand gesture [17] based on the interpretation of the CSI pattern. Considering those information inherent in WiFi CSI, Doppler shift is the only parameter that represents the dynamic state of motion and is insensitive to stationary background objects. Thus, we envision that the WiFi Doppler has great potential for human detection.

In this study, we focus on context-awareness computing for human presence detection and events classification by extracting the Doppler information from the human-reflected WiFi signal. However, there are two major challenges for the system: firstly, how to collect Doppler information from WiFi signals in an uncontrolled environment without any scenario-specific calibration? Secondly, how to use this Doppler information to detect human presence and classify the activity event? We adopt the cross ambiguity function (CAF) technique to extract the Doppler information which correlates the WiFi signals from surveillance channel (reflected signal from object) and reference channel [signal from WiFi access point (AP)]. The output is a two-dimensional mapping which contains both range (time delay) and Doppler (frequency shift) information. Due to the limits of the bandwidth of the WiFi signal (20 MHz), the range resolution is insufficient (∼7.5 m) for indoor application. Thus, we use only the Doppler information. Using of Doppler information can also avoid scenario-specific calibration. This is because the static background object only caused the zero Doppler in CAF mapping. In addition, we observed that the Doppler information obtained from the human presence channel is much stronger than those from a no-human presence channel because moving personnel generate significant Doppler information from torso and limb motion, while static or inactivity personnel cause periodical Doppler shift due to chest movements during respiration [5]. This Doppler information can be potentially used for presence detection. These observations show the potential of analysing Doppler information for context awareness within a WiFi-covered area.

Afterwards, we treat the non-invasive human presence detection and event awareness as a classification problem. This process includes feature extraction which aims at data dimension reduction and model training to establish the event classifiers. For feature extraction, we propose time–frequency domain feature vectors as the representation for Doppler measurement. This is because of both the time-varying pattern and frequency content of Doppler information vary according to human motion. The reason is that each type of human motion generates a unique time-varying pattern and frequency content [6, 12]. Therefore, robust feature vectors require knowledge on both time and frequency properties. Then, we use the support vector machine (SVM) as the classifier for human presence detection and event classification. The basic principle is to generate an SVM model to separate the scenario between two different classes based on the labelled datasets, then make the decision for the new testing dataset. To evaluate the performance of our proposed concepts, we implement a real-time system on a software defined radio (SDR) platform for a pilot study with five different events. Then, we compute the classification accuracy for each event with the proposed time-frequency feature vectors and SVM classifiers.

Compared to previous works [1, 2, 12, 14], the following contributions of the proposed passive WiFi Doppler-based human presence and event recognition system are presented:

- We proved that WiFi signal is a valuable source for human presence and activity event recognition for context awareness in healthcare, security, and low impact building applications. The function is enabled by the innovative passive Doppler capturing technique with the assistance of direct signal cancellation and bed time index removal. Our passive WiFi Doppler-based sensing provides a unique presence and activity event awareness approach in residential and in-building applications without extra infrastructure.
- We discovered that Doppler shift carried in the reflected WiFi signal is a more effective parameter for human presence detection and have the capability to capture human activity details. Comparing with the traditional RSS-based approach, the Doppler-based approach showed significant improvement in accuracy.
- We proposed a novel feature vector which presents the data characteristics in both time and frequency domains. The experiment results indicate that our time-frequency feature vectors outperform the conventional singular value decomposition (SVD) and empirical features.

The rest of the paper is organised as follows: Section 2 briefly reviews previous related works; Section 3 outlines the methodology of the proposed concept including the system overview, measurements setup, preprocessing of WiFi signal, feature extraction, and the classifier; the experiment results are present in Section 4; conclusions and future work are presented in Section 5.

2 Related work

2.1 Human activity detection with WiFi

WiFi-based non-invasive systems analyse the signal properties of reflected signal for context awareness of users. The RSS measurements can be used to detect a moving human by exploiting the variance in its envelope [18]. While, for stationary human and no-human scenario, RSS measurement can hardly provide any sufficient recognition due to it cannot present the changes in the channel. For this reason, CSI at the PHY layer is often used for human sensing recently. Zhou *et al.* [2] developed a density-based spatial clustering of applications with noise to reduce the noise in CSI and used SVM to interpret the status of the user. Lv *et al.* [13] proposed a speed independent device-free entity detection (SIED) for detecting multiple moving speeds. The SIED compares the amplitude variances of each subcarrier and applies to a hidden Markov model for motion detection. Zhou *et al.* [19] used a fingerprinting technique on CSI measurements and used a pre-defined threshold to passively localise objects in an omnidirectional manner. The Deman developed by Wu *et al.* [18] can be used for device-free detection of moving and stationary humans. Deman takes advantage of amplitude and phase information from CSI to detect human breathing and then detects human presence. Zhu *et al.* [20] present a device-free through-the-wall (TTW) system to detect movement of humans by measuring the CSI changes over adjacent subcarriers. They calculate the difference of first-order eigenvector from each subcarrier as the indication of TTW human detection. One common ground for the above-mentioned approaches is that they all use the CSI information which is generated from a specific network card. While in our work, we extract the Doppler information based only on the received signal that avoids the using of CSI measurement. This approach has simpler signal processing compared to the CSI approach which requires a learning process over subcarriers, and also no direct control on WiFi AP is needed.

2.2 Emerging WiFi sensing application

The wireless signals reflection show different characteristics during changing of human gesture [21–23]. Currently, there is number of works which utilise the WiFi signal for many human sensing applications. Han *et al.* [21] propose a fall detection system (WiFall) by using the frequency diversity and time variability of CSI. Hao *et al.* [22] further exploit the phase and amplitude information in CSI and present a more powerful context-awareness system (RT-Fall). It exams the variance of CSI and detects the finishing point of an activity to determine a potential falling event. Khan *et al.* [24] present a passive WiFi system by extracting the 2D phase information for tremor, fall, and breathing detection. Alternatively, WiFinger [23] presents the use of patterns in time series of CSI information to identify the fine-grained finger gestures. They combine 30 subcarriers and concatenate them into a synthetic waveform as the feature vector. Gong *et al.* [25] develop a lightweight and device-free human motion detection with short-term averaged variance ratio and long-term averaged variance ratio which is calculated from the physical layer. Similarly, Wei *et al.* [26] present WiFiU to capture the fine-grained gait patterns for human recognition by converting the CSI waveforms into time-frequency spectrogram with short-time Fourier transform technique. Xie *et al.* [27] illustrate a crow counting system (Electronic Frog Eye), which formulates the monotonic relationship between feature vectors from CSI and the crowd number upon the grey Verhulst model. Those works show the potential and feasibility of using WiFi technology to the near-field activity awareness.

2.3 Other RF-based human activity detection approaches

There are also other approaches for indoor context awareness based on RF techniques. Among them, radar is one of the widely used technologies and has been applied to a number of applications. The basic principle, detecting the frequency shift from the reflected signal, can be used for detection of tiny movements such as chest motion [28]. Ram *et al.* [29] generate micro-Doppler signatures of moving human behind walls and simulate the time-varying radar cross-section with primitive-based prediction methods. Al-Turjman [30] presents the use of embedded sensors (e.g. global positioning system, accelerometers etc.) on the application level for the indoor context awareness. Forouzanfar *et al.* [6] present a phase-modulated continuous wave (CW) radar system for contactless event monitoring. The system analyses the radar returns and extracts both time and frequency feature vectors to receive context-awareness information with Bayesian classifiers. Anderson [31] collects the signatures of humans, as well as birds, dogs, goats, and deer, with an X-band CW Doppler radar. It shows that Doppler shifts from a human are largely different from those of other animals. Another interesting work [32] uses a mini drone that can be used to capture moving objects with on-board sensors and camera for monitoring purposes. Different from the above-mentioned active-radar-based approaches, our system avoids the transmission of any pre-designed waveform and additional hardware. This benefits commercial products as no frequency

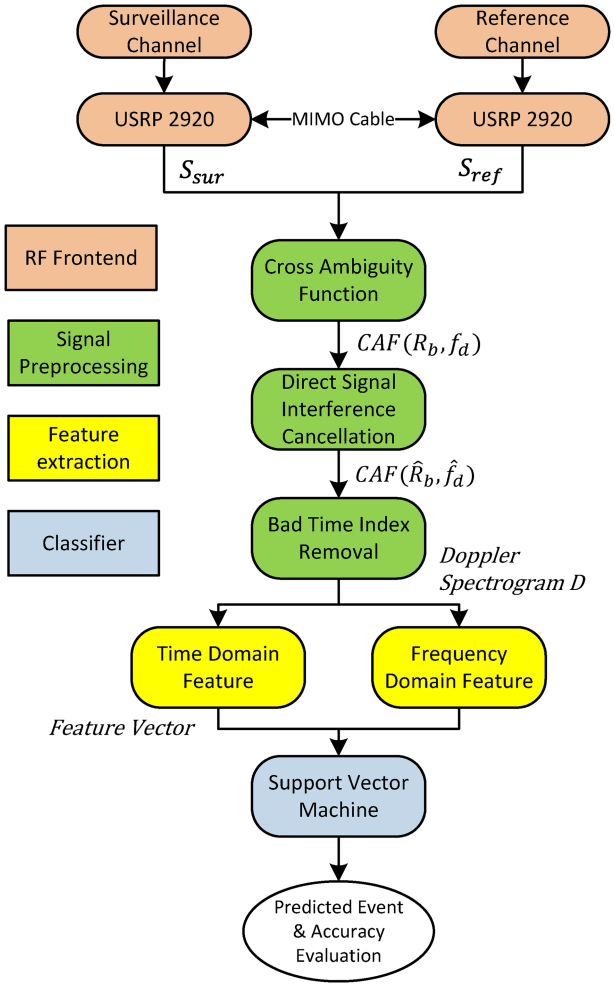

Fig. 1 *Block diagram of the proposed passive sensing system*

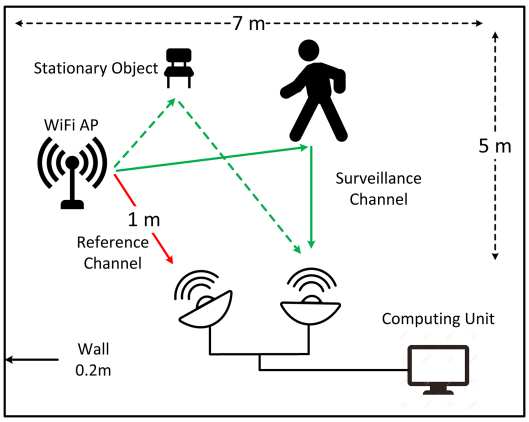

Fig. 2 *Experiment layout*

allocation is required and allows for very flexible deployment of the products within any WiFi-enabled area.

3 Methodology

3.1 Passive sensing system

In this work, we design a passive WiFi Doppler sensing system and implement on an SDR platform to prove our concept. The system block diagram of the system is shown in Fig. 1. There are four main components in the system as RF front-end, signal preprocessing, feature extraction, and classifier. The data formats of each function's input/output port are also illustrated. The details of the system RF front-end are listed in Table 1, where two USRP 2920 are used to collect the wireless signal from a WiFi AP. The signal preprocessing includes three parts: CAF, direct signal cancellation, and bad time index removal. It first computes a CAF mapping based on the wireless signal from the surveillance channel and reference channel, and then performs a CLEAN process [3] to reduce the direct signal and clutter effect. Afterwards, in order to characterise the different scenario of interest and to distinguish the occupied and non-occupied channel, a comprehensive feature vector is extracted from the Doppler spectrogram in both time and frequency domains, respectively. Finally, we use SVM as the classifier to train those feature vectors and decide the presence of a human. For convenient, the feature extraction and SVM are implemented in MATLAB for the off-line process while the preprocessing on wireless signal is implemented in LabVIEW with real-time processing ability. The system is processed at 5 Hz which outputs a Doppler measurement every 0.2 s.

Table 1 System RF front-end

WiFi AP	Netgear R6300 working at 13th channel
WiFi protocol	802.11 g
bandwidth	20 MHz
RF receiver	USRP 2920 x2 with SBX-12 daughter board
antenna type	SCR12-2400 Corner Antenna x2 (22 dBi)
computing unit	I7 Laptop running LabVIEW 2016

Table 2 Activity event description

Event	Description	Activity level
(a)	no-human presence for 30 s	no-human
(b)	standing still at one location and breathing normally for 30 s	no-activity
(c)	standing at one location and continuously rotating the head gently for 30 s	low-level
(d)	standing at one location and continuously waving the hand slowly for 30 s	middle-level
(e)	walking continuously at normal speed between two random locations, forward and backward	high-level

3.2 Measurement setups

The measurements were performed in a lab environment at the University of Bristol, CSN group. The surveillance area is 7m × 5 m with chairs, desks, and equipment. The WiFi AP was placed in one corner of the room, and the passive sensing system was placed on a desk at ~80 cm in height. One antenna was pointed to the surveillance area as the surveillance channel which received the signal from both human objects and stationary objects, and the other antenna was pointed to the WiFi AP as the reference channel to recreate the transmitted WiFi signal. The experiment layout is shown in Fig. 2. To present the proposed system in context awareness, we carry out five different events that relate to daily life. Three volunteers (three males, age 27, 27, and 38) were asked to perform the following activities with respect to each event as described in Table 2.

Each event is repeated 10 times by each volunteer. Total 150 (3 × 5 × 10) measurements are recorded. Four different locations have been tested in (b), (c) and (d) events with 1, 2, 3, and 4 m towards the surveillance antenna.

3.3 Preprocessing

3.3.1 Cross ambiguity function: CAF is an effective tool to extract the range R_b and Doppler f_d information from surveillance channel $S_{sur}(t)$ and reference channel $S_{ref}(t)$. The general CAF mapping CAF(R_b, f_d) can be expressed as

$$\mathrm{CAF}(R_b, f_d) = \sum_{0}^{T} S_{sur}(t) S_{ref}^{*}\left(t - \frac{R_b}{c}\right) e^{j2\pi f_c f_d t}\, dt, \quad (1)$$

where $\lfloor * \rfloor$ denotes the complex conjugate, c represents the light speed, and f_c is the carrier frequency. However, this equation has

one major limitation: the wide bandwidth signal and long integration time will contain such a significant amount of data that the fast Fourier transform (FFT) process becomes impractical due to the sequence transformation being too long. In contrast, the decimation technology can solve the computational issue by dividing a long sequence into groups of short sequences before the FFT transformation. One such technology is known as the batching process [33], which when applied to the CAF can be expressed as

$$\mathrm{CAF}(R_\mathrm{b}, f_\mathrm{d}) = \sum_{k=0}^{n_\mathrm{b}-1} \int_{T_{i-1}}^{T_i} S_\mathrm{sur}^i(t) \, S_\mathrm{ref}^{i\,*}\left(t - kT_\mathrm{B} - \frac{R_\mathrm{b}}{c}\right) \mathrm{e}^{\mathrm{j}2\pi f_\mathrm{c} f_\mathrm{d} t} \, \mathrm{d}t, \tag{2}$$

where $S_\mathrm{sur}^i(t)$ and $S_\mathrm{ref}^i(t)$ are down-converted baseband signals from surveillance and reference channels with batching length T_B and n_b is the number of batches. The range resolution is defined as $\Delta R_\mathrm{b} = c/2B$, where B is the bandwidth of the WiFi signal, which is obviously too coarse for indoor application. In addition, the range information requires a scenario-specific calibration under human-free environment as primitive background to compare with the human occupied situation. On the other hand, Doppler resolution is adaptable by adjusting the integration time T_i as $\Delta f_\mathrm{d} = (1/T_\mathrm{i})$. Since the Doppler information is insensitive to the stationary object which means the scenario-specific calibration is not required for our system. This presents a more robust approach for WiFi-based human sensing system.

3.3.2 Direct signal cancellation: One of the major limitations of the CAF method is that the direct signal between surveillance antenna and WiFi AP generates a significant Doppler peak in the zero range bin, which contains much higher energy than the target's return. This pulse could bury the desired target and reduce the detection capability of a moving object. For this reason, we apply the CLEAN algorithm [3] to remove the direct signal interference. The basic idea is to suppress the ith CAF mapping $\mathrm{CAF}^i(R_\mathrm{b}, f_\mathrm{d})$ with the self-ambiguity function $\mathrm{CAF}_\mathrm{self}^i(R_\mathrm{b}, f_\mathrm{d})$. The CLEANed CAF mapping $\mathrm{CAF}^i(\hat{R}_\mathrm{b}, \hat{f}_\mathrm{d})$ can be then calculated as

$$\mathrm{CAF}^i(\hat{R}_\mathrm{b}, \hat{f}_\mathrm{d}) = \mathrm{CAF}^{i-1}(R_\mathrm{b}, f_\mathrm{d}) - \alpha^i \mathrm{CAF}_\mathrm{self}(R_\mathrm{b} - f_i, f_\mathrm{d}), \tag{3}$$

where α^i is the scaling factor related to the location of α^i and f_i is the corresponding phase shift factor. Afterwards, the desired Doppler information can be revealed. Note that direct signal cancellation has also been applied in the batching process because the processing on self-ambiguity $\mathrm{CAF}_\mathrm{self}^i$ is similar to the CAF process (the only difference is in the inputs). Then, we pick a column which contains the highest Doppler peak and generates a Doppler spectrogram $\boldsymbol{D} = \left[D_{t,f}\right]_{t=1, f=-(n_\mathrm{b}/2)}^{T, (n_\mathrm{b}/2)}$ by combining T seconds of CAF mappings.

3.3.3 Bad time index removal: Another interference of our system is the 'bad time index' during the processing on Doppler spectrogram. Previous works [12, 15, 17] manually set WiFi AP with high package transmission rate (more than 4000 frames per second) only for sensing purpose. However, this setting occupies significant bandwidth that obstructs the original communication purpose of WiFi AP. Thus, it is not an ideal solution for a commercial system. In comparison, our system 'passively' receives RF signals with no direct control over the broadcasting of WiFi AP. This means the WiFi signal can be very much irregular, with little dependence on Internet usage [3]. As a result, we found the irregular and non-continuous WiFi signal would harm the correlation process in CAF and generate bad time index in the Doppler spectrogram. Note that the signal from other WiFi AP but in the same channel could also affect the Doppler spectrogram. Those errors would potentially affect the classification on human presence. Our solution is to detect and remove this bad time index from the spectrogram and leave it as blank. It is observed that the bad time index normally contains multiple Doppler peaks spanning entire frequencies that have much higher Doppler power than the normal measurement (five times higher than the Doppler information on walking). For this reason, we calculate the Doppler power P_t at each time index as the indicator with a pre-defined threshold

$$P_t = \sum_{i=-(n_\mathrm{b}/2)}^{(n_\mathrm{b}/2)} D[t, f] \quad \begin{cases} P_t < \text{threshold} & \text{good time index} \\ P_t > \text{threshold} & \text{bad time index} \end{cases} \tag{4}$$

For the Doppler power lower than the threshold, the corresponding time index is considered as good and will be kept. For the Doppler power higher than the threshold, the corresponding time index is considered as an error and will be removed. The threshold is defined through the observation over a number of bad time indices. This could be optimised by applying Hansen's regression [34] to automatically estimate the threshold in our future work.

The effectiveness of the direct signal cancellation and bad time index removal will be shown in Section 4.1.

3.4 Feature selection and extraction

After the preprocessing, we need to extract feature vectors from the Doppler spectrogram in order to reduce the dimension of a dataset while maintaining the data diversity. The time-varying pattern and frequency content of the Doppler spectrogram vary according to the presence of a human. Therefore, robust classification will require the use of appropriate features with both time and frequency properties.

3.4.1 Time feature: Time domain feature shows the dynamic change of human motion varies across the time which is an important information for our event classification. Let the time feature vector as $\boldsymbol{G}$, it is formulated as $\boldsymbol{G} = [G_t]_{t=1}^{\mathrm{T}}$, where G_t is calculated as

$$G_t = \arg\max_f \left\{ \left[D_{t,f}\right]_{f=-(n_\mathrm{b}/2)}^{n_\mathrm{b}/2} \right\}, \tag{5}$$

where $\max_f$ denotes the Doppler shift index with the largest value, n_b is the batching number which defines the border of Doppler shift, the division by 2 is because of both positive and negative region.

3.4.2 Frequency feature: Frequency component in Doppler spectrogram varies among different motions which are also considered as an ideal feature for event classification. With some prior knowledge about the Doppler frequency of different activities, it is possible to discriminate the peaks caused by different movements. For example, the breathing motion constraints in +1 to −1 Hz due to the slow chest movement speed, while the walking motion is normally >+4 or −4 Hz as relatively high speed. The spectral power along Doppler frequency is used as the frequency feature vector $\boldsymbol{E}$, it is formulated as $\boldsymbol{E} = \left[E_f\right]_{f=-(n_\mathrm{b}/2)}^{n_\mathrm{b}/2}$, where $\left[E_f\right]$ is calculated as

$$E_f = \left[\frac{1}{T}\sum_{t=1}^{T} D_{t,f}\right]_{f=-(n_\mathrm{b}/2)}^{n_\mathrm{b}/2}. \tag{6}$$

In this work, we use both time feature vector $\boldsymbol{G}$ and frequency feature vector $\boldsymbol{E}$ as the representation of Doppler spectrogram.

3.5 Classification by SVM classifier

The classification includes both training and testing process. Training samples contain several feature vectors and corresponding labels. The purpose of classification is to judge the labels of testing samples according to their feature vectors, on the models estimated based on the training samples. For example, in the case of human

presence detection, two kinds of Doppler information are needed: (i) +1 (positive) samples for human presence scenario and (ii) −1 (negative) samples for no-human presence. Let n be the number of total training samples, l is the dimension of feature vectors. For each training sample, it can be presented as a pair (r_i, c_i), where $r_i = (r_{i1}, r_{i2}, \ldots r_{il})$ is the feature vector, e.g. the time and frequency feature from Doppler information, and $c_i \in \{-1, +1\}$ is the label of samples. Those labelled training samples are used by SVM classifiers to train the model. Considering (r, c) as a testing sample, with $r \in R^l$ as the feature vector, SVM is used to determine the sign of c, irrespective of human presence in the monitoring area.

We adopt the C-support vector classification (C-SVC) for this work as it is an effective machine learning tool with the low requirement on training set [35]. The C-SVC solves the following problem:

$$\min \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \varepsilon_i, \tag{7}$$

where $w \in R^l$ is the vector which represents the direction of the hyperplane, constant C defines the trade-off between the misclassification errors and separation region, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i)$ are the slack variables which represent samples on the wrong side in a hyperplane. They obey that $C > 0$ and $\varepsilon_i \geq 0$. Solving (7) gives the classification function as

$$f(r) = \text{sign}\left(\sum_{i=1}^{n} c_i \alpha_i K(r_i, r) + b \right), \tag{8}$$

where α_i is the Lagrange multiplier, $K(r_i, r)$ is the mapping of the kernel function for Doppler information into a higher-dimensional representative, and $b \in R$ is the constant of the location of the hyperplane. When $f(r) > 0$, it indicates the presence of a human in the monitoring area, while $f(r) < 0$ indicates there is no human detected. Two parameters are essential for C-SVC: the kernel parameter ε and trade-off C which are calculated by cross-validation and grid-searching, respectively.

4 Experiment results

In this section, we present the performance of our WiFi-based system for context awareness including human presence detection and event classification. We first present the effectiveness of the signal preprocessing in generating the Doppler information with respect to Section 3.3. Then we compare the performance by using Doppler and RSS information based on the dataset of 150 measurements from five different activity events as introduced in Section 3.2. Afterwards, we present the robust of using the proposed time-frequency feature vector that outperforms the conventional SVD and empirical features.

4.1 Effectiveness of proposed concept

For this demonstration, we provide a 30 s measurement from a standing still human (event (b)). The measured Doppler spectrogram is shown in Fig. 3: (a) the original Doppler spectrogram ((2), Section 3.3.1), (b) with direct signal cancellation ((3), Section 3.3.2) and (c) with bad time index removal ((4), Section 3.3.3). As can be seen, the Doppler spectrogram in Fig. 3*a* shows insufficient information due to the strong peaks along the zero Doppler shift (the red line). This is because of the direct signal interference outputs a Doppler peak with a significant amplitude that buries other peaks. For this reason, we apply the self-ambiguity cancellation technique to remove this dominant peak. The result is shown in Fig. 3*b*, where the peaks due to the direct signal interference have been effectively removed. However, there are four bad time indices appearing in the spectrogram and contain peaks that span all the frequencies. They are weaker when compared to the direct signal interference but still stronger than the peak of breathing motion. The reason for those peaks is because of the irregular and non-continuous WiFi signal that causes an error in

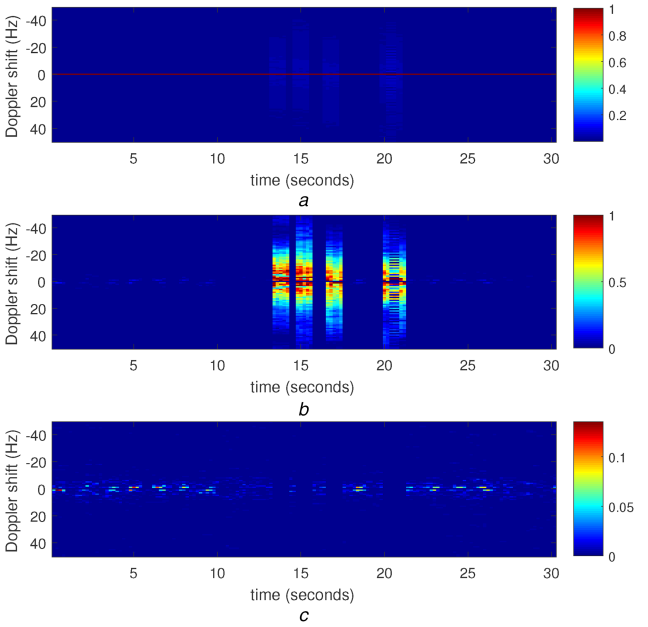

Fig. 3 *Effectiveness of the preprocessing*
(a) Doppler spectrogram from CAF, ***(b)*** with direct signal cancellation, ***(c)*** With bad time index removal

the CAF process. For this type of interference, we calculate the Doppler power at each time index and clean the Doppler spectrogram with the pre-defined threshold. The final output is shown in Fig. 3*c*; it is observed that all bad time indices have been successfully removed, while several clear and weak peaks are now visible. Those peaks represent the chest motion due to the inhalation and exhalation within each respiration cycle.

Here, we also provide the Doppler spectrogram for all five events as shown in Fig. 4. As can be seen, the Doppler spectrogram is almost clean if there is no human presence, event (a), as expected. This illustrates that the proposed concept is insensitive to background clutters even without scenario-specific calibration. For events (b), (c), (d), and (e), where there is a human presence, there are clear Doppler peaks in the spectrogram. Especially, for the walking event (e), it contains a noticeable Doppler trace compared to others. For activities such as waving hand (d), there is a CW at low Doppler frequency. Both standing still (b) and rotating head (c) activities have non-continuous and short peaks. Since these activities are relatively short and slow, they result in weak reflections. Yet, the Doppler spectrogram can visualise the difference between chest motion and rotating hand. Understandably, the Doppler information can be further used for event interpretation with proper feature extraction methods and classifiers.

4.2 Detection performance

It is important to evaluate the detection accuracy of our proposed human sensing system with an unbiased estimation of classification errors. For this purpose, cross-validation has been applied. That is, we divide the dataset into 20 training sets and 10 testing sets. The SVM classifier then evaluates the testing sets upon the trained model. This procedure is repeated three times to allow for all samples to be tested. In addition, we conduct the testing of both Doppler and RSS information to demonstrate the robust of the proposed concept. The RSS measurement is collected within the same testbeds in parallel.

The classification confusion matrices of both Doppler and RSS approaches are presented in Figs. 5 and 6, respectively. The rows in the matrix represent the predicted classes and the columns represent the actual classes. Elements at the diagonal show the number and percentage of the correct predicted class, whereas the wrong prediction assigns to wrong classes. All corresponding predictions are indicated in the matrix. There are four matrices that need to be defined: true positive (TP) is the probability of

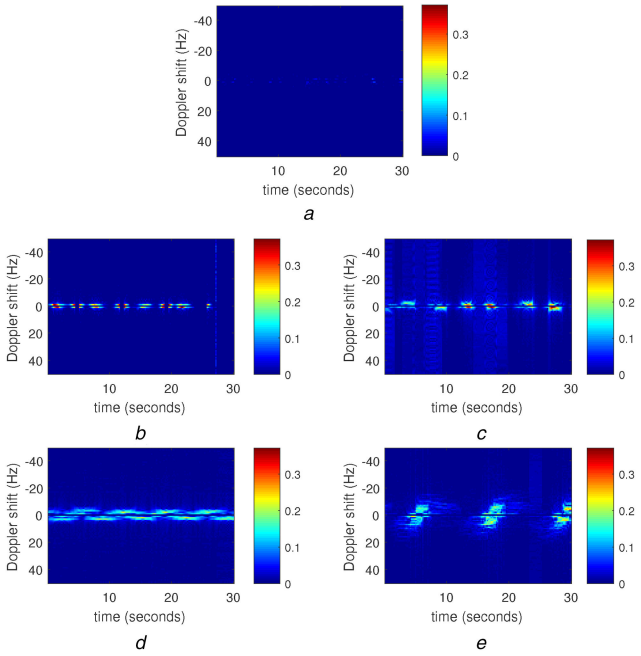

Fig. 4 *Examples of Doppler spectrogram belonging to five different events*
(a) No human, ***(b)*** Standing still and no activity, ***(c)*** Standing still and moving head, ***(d)*** Standing still and waving arm, ***(e)*** Walking forward and backward

successful prediction of an event i over total predictions of this event; false positive (FP) is the probability of failure prediction of an event i over total predictions of this event or $1 - \mathrm{TP}$; true negative (TN) is the probability of unsuccessful prediction of an event i over total predictions of this event; false negative (FN) is the probability of prediction of an event i over total samples of this event or $1 - \mathrm{TN}$. The right column of the confusion matrix shows the positive predicted value (PPV) and the false discovery rate (FDR) in green and red, respectively, where PPV illustrates the fraction of positive predictions that are true and FDR illustrates the fractions that are false. The last row of the confusion matrix shows the true positive rate (TPR) and false negative rate (FNR) in green and red, respectively, where TPR illustrates the fraction of events that are correctly predicted and FNR illustrates the fraction of classes that are wrongly predicted. PPV, FDR, TPR, and FNR are calculated as [6]

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{9}$$

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}} = 1 - \mathrm{PPV}, \tag{10}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{11}$$

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{TP} + \mathrm{FN}} = 1 - \mathrm{TPR}\,. \tag{12}$$

As can be seen, the proposed Doppler approach achieves higher PPV and lower FDR than those achieved when using the RSS approach. The system receives 100% PPV and 0% FDR for events (b), (d) and (e). The lowest is event (a) with 84.8% PPV and 15.2% FDR due to the misclassification of event (b). This can be explained as follows: event (b), standing still, is similar to event (a), no human presence, with only Doppler information from chest motion. However, this Doppler information is so very weak and non-continuous (shown in Fig. 4) that it may not get captured. It is observed that event (b) becomes closer to event (a) when the distance increases. It is also observed that event (b) has the lowest TPR at 83.3% and highest FNR at 16.7% due to the error assignments to event (a), while events (c), (d), and (e), all have 100% TPR and 0% FDR, which means that the proposed Doppler approach provides sufficient diversity to classify between different

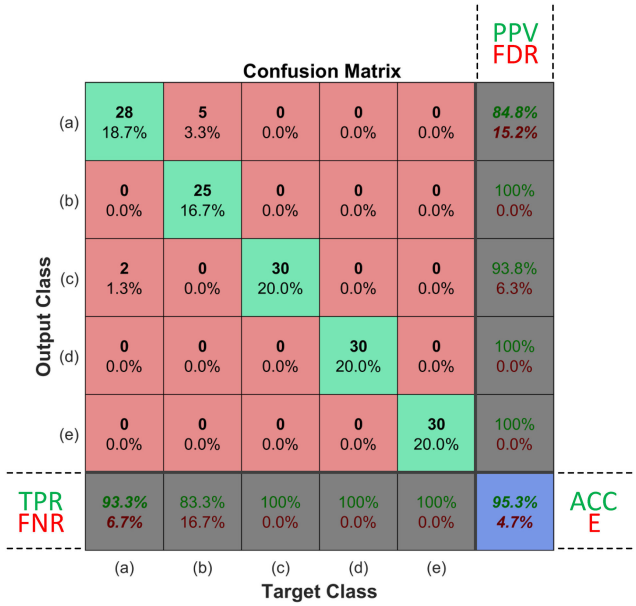

Fig. 5 *Classification confusion matrix for different activity event with our proposed Doppler approach*

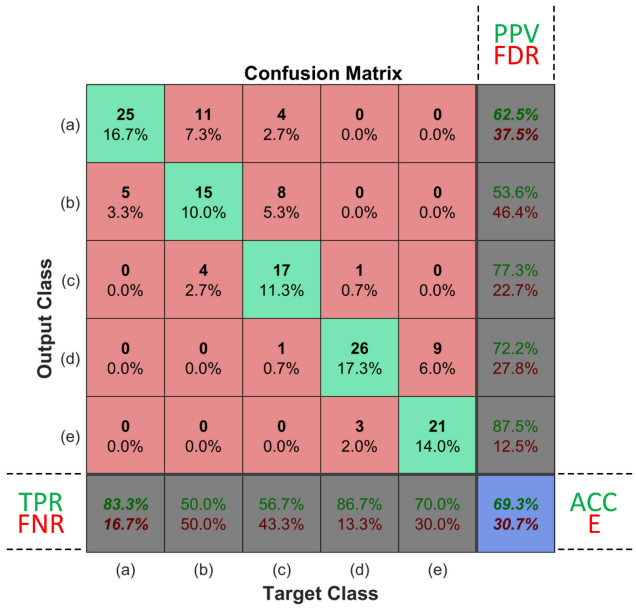

Fig. 6 *Classification confusion matrix for different activity event with the classical RSS approach*

activity events. In comparison, the RSS approach has significant downgraded performance, with only 62.5% PPV and 37.5% FDR due to the misclassification from events (b) and (c). As explained before, RSS information cannot sufficiently present changes in channels.

The performance for events (b), (c), and (d) are all below 80% PPV, which indicates that there are more misclassified samples from other events. The best classification is observed in event (e) with 87.5% PPV and 12.5% FDR. This is because the high-level activity generates more fluctuations in RSS information which makes it easier to be separated. Event (a) has 83.3% TPR and 16.7% FNR which slightly drops compare to the Doppler approach. However, event (b) has the worst 50% TPR and 50% FNR that half of its samples are assigned wrongly. Also, all events are below 90% TPR and 10% FNR, which indicated that there is no perfect prediction with RSS information. Moreover, PPV and TPR from event (a) can be treated as the classification accuracy for human presence detection. They represent the predictions between no-human presence event (a) and human presence events (b)-(e). As expected, Doppler information gives 94.8% in PPV and 93.3% in

Table 3 Accuracy, sensitivity, and specificity of each event

Event	Accuracy, %	Sensitivity, %	Specificity, %
(a)	96.33	84.51	100
(b)	95.33	96	95.2
(c)	98.67	98.28	98.76
(d)	99.67	98.36	100
(e)	100	100	100

TPR which are much higher than that in RSS information as 62.5% PPV and 83.3% TPR.

The last element (right-bottom) of the confusion matrix presents the total accuracy (ACC) and total error (E) about the classification in green and red. The ACC and E are defined as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}, \tag{13}$$

$$E = \frac{\mathrm{FP} + \mathrm{FN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} = 1 - \mathrm{ACC}, \tag{14}$$

where $\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN} = 1$ is the size of our dataset. It is observed that the Doppler approach has 95.3% ACC and 4.7% E that outperform the RSS approach which is 69.3% ACC and 30.7% E. Those results demonstrate that Doppler information is a more robust approach for human sensing applications.

It is also important to evaluate the accuracy $((\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}))$, sensitivity $(\mathrm{TP}/(\mathrm{TP} + \mathrm{FN}))$, and specificity $(\mathrm{TN}/(\mathrm{TN} + \mathrm{FP}))$ for each event to compare the detection performance. Sensitivity determines the probability of detection. It is usually necessary for a system to maximise its precision on a different event, therefore, sensitivity is important. Specificity measures the probability of negative samples that are correctly labelled as such. The classification results are shown in Table 3. It is observed that almost all elements are more than 95%. Yet, the lowest element exhibits the sensitive of event (a) at 84.51%, and event (b) has both lowest accuracy at 95.33% and specificity at 95.2%. This means the system sometimes makes errors in separating no-human events and no-activity events. As discussed before, the reason is that Doppler information from chest motion is comparably short and weak which may not be captured by our system. In comparison, event (e) shows the best performance: 100% accuracy, sensitivity, and specificity. This indicates that our proposed system is more sensitive to high-level activity events compared to low- and middle-level events, as expected.

Furthermore, we compare the performance of the proposed time–frequency feature vector (described in Section 3.4) with other features including SVD and empirical features [4, 36] and test using four different SVM kernels. The results are shown in Fig. 7. It is observed that both proposed feature vectors are able to achieve recognition accuracy of more than 85%. The best accuracy of 96% is generated with a polynomial kernel with our feature vector, and the worst is at empirical with a radial basis function kernel. Especially, empirical features have the worst performance that are below 85%. The reason is that empirical features represent the physical characters of a Doppler spectrogram and are insufficient compared to eigenvalue-based features, and for long duration datasets (in this work samples are 20 s), empirical features show even worse performance. On the other hand, the proposed feature vector outperforms the classical SVD methods. This is because we use the features from both time and frequency domains which provide more diversity in the kernel.

5 Conclusion

This study presents a novel WiFi-based system for context awareness including human presence detection and event classification. The human-reflected WiFi signal contains wide-ranging information from torso movement, limb movement, chest movement during breathing, surrounding stationary objects, and interference. Our system, therefore, performed the following processing: CAF to generate Doppler information, direct signal interference cancellation, and bad time index removal for irregular WiFi signal. The proposed concept is evaluated with a pilot study on a dataset with measurements from five different activity events. It is shown that the overall accuracy of our proposed Doppler approach reaches up to 95% and significantly outperformed the RSS approach whose accuracy is 69.3%. Moreover, Doppler information also shows better performance in human presence detection with 93.3% TPR and 84.8% PPV whereas RSS information is at 83.3% TPR and 62.5% PPV. This gives similar detection performance compared to the UWB approach [11] with 90% accuracy among eight activities. Moreover, we have demonstrated the robustness of using time-frequency feature vectors from both time and frequency domains rather than the traditional SVD and empirical features. This method further benefits the classification process.

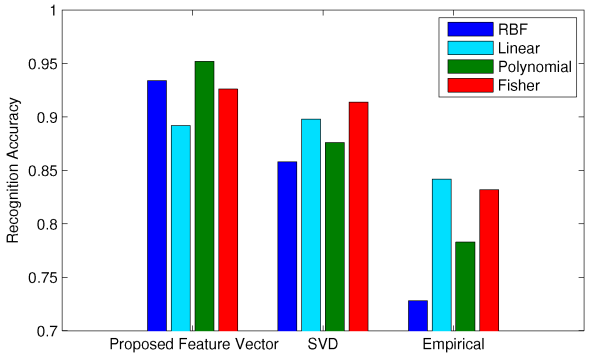

Fig. 7 *Recognition accuracy with different feature by four SVM kernels*

The proposed system can be seamlessly integrated into a WiFi-enabled area and provide a practical solution for indoor context-awareness sensing. For instance, the system can first classify the activity event as discussed in this paper. Then it can detect breathing in a stationary human by estimating the frequency of occurrence of the Doppler peak or perform activity recognition for a dynamic human by examining the difference in the Doppler pattern. Further work will include the study on the feasibility of detecting multiple objects and emergency contexts such as falls.

6 Acknowledgment

This work was partially funded under the SPHERE IRC, the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1.

7 References

[1] Zou, H., Zhou, Y., Yang, J., *et al.*: 'Freedetector: device-free occupancy detection with commodity WiFi', 2017 IEEE Int. Conf. on Sensing, Communication and Networking (SECON Workshops), San Diego, CA, 2017, pp. 1–5

[2] Zhou, R., Lu, X., Zhao, P., *et al.*: 'Device-free presence detection and localization with SVM and CSI fingerprinting', *IEEE Sens. J.*, 2017, **PP**, (99), pp. 1–1

[3] Li, W., Tan, B., Piechocki, R.J., *et al.*: 'Opportunistic physical activity monitoring via passive WiFi radar', 2016 IEEE 18th Int. Conf. on e-Health Networking, Applications and Services (Healthcom), Munich, 2016, pp. 1–6

[4] Li, W., Xu, Y., Tan, B., *et al.*: 'Passive wireless sensing for unsupervised human activity recognition in healthcare', 2017 13th Int. Wireless Communications and Mobile Computing Conf. (IWCMC), Valencia, 2017, pp. 1528–1533

[5] Li, W., Tan, B., Piechocki, R.J.: 'Non-contact breathing detection using passive radar', 2016 IEEE Int. Conf. on Communications (ICC), Kuala Lumpur, 2016, pp. 1–6

[6] Forouzanfar, M., Mabrouk, M., Rajan, S., *et al.*: 'Event recognition for contactless activity monitoring using phase-modulated continuous wave radar', *IEEE Trans. Biomed. Eng.*, 2017, **64**, (2), pp. 479–491

[7] Lu, X., Aggarwal, J.K.: 'Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera', Proc. IEEE Conf. on Computer Vision and Pattern Recognition 2013, Portland, USA, June 2013, pp. 2834–2841

[8] Riekki, J., Salminen, T., Alakarppa, I.: 'Requesting pervasive services by touching RFID tags', *IEEE Pervasive Comput.*, 2006, **5**, (1), pp. 40–46

[9] Maenaka, K., Masaki, K., Fujita, T.: 'Application of multi-environmental sensing system in MEMS technology – monitoring of human activity', 2007 Fourth Int. Conf. on Networked Sensing Systems, Braunschweig, 2007, pp. 47–52

[10] Mashiyama, S., Hong, J., Ohtsuki, T.: 'Activity recognition using low resolution infrared array sensor', 2015 IEEE Int. Conf. on Communications (ICC), London, 2015, pp. 495–500
[11] Bryan, J.D., Kwon, J., Lee, N., *et al.*: 'Application of ultra-wide band radar for classification of human activities', *IET Radar Sonar Navig.*, 2012, **6**, (3), pp. 172–179
[12] Wang, W., Liu, A.X., Shahzad, M., *et al.*: 'Device-free human activity recognition using commercial WiFi devices', *IEEE J. Sel. Areas Commun.*, 2017, **35**, (5), pp. 1118–1131
[13] Lv, J., Yang, W., Gong, L., *et al.*: 'Robust WLAN-based indoor fine-grained intrusion detection', 2016 IEEE Global Communications Conf. (GLOBECOM), Washington, DC, 2016, pp. 1–6
[14] Lymberopoulos, D., Liu, J., Yang, X., *et al.*: 'A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned', Proc. 14th Int. Conf. on Information Processing in Sensor Networks, Seattle, USA, April 2015, pp. 178–189
[15] Cao, X., Chen, B., Zhao, Y.: 'Wi-Wri: fine-grained writing recognition using Wi-Fi signals', 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, 2016, pp. 1366–1373
[16] Zheng, Y., Zhou, Z., Liu, Y.: 'From RSSI to CSI: indoor localization via channel response', *ACM Comput. Surv.*, 2013, **46**, (2), p. 25
[17] Abdelnasser, H., Youssef, M., Harras, K.A.: 'Wigest: a ubiquitous WiFi-based gesture recognition system', 2015 IEEE Conf. on Computer Communications (INFOCOM), Kowloon, 2015, pp. 1472–1480
[18] Wu, C., Yang, Z., Zhou, Z., *et al.*: 'Non-invasive detection of moving and stationary human with WiFi', *IEEE J. Sel. Areas Commun.*, 2015, **33**, (11), pp. 2329–2342
[19] Zhou, Z., Yang, Z., Wu, C., *et al.*: 'Omnidirectional coverage for device-free passive human detection', *IEEE Trans. Parallel Distrib. Syst.*, 2014, **25**, (7), pp. 1819–1829
[20] Zhu, H., Xiao, F., Sun, L., *et al.*: 'R-TTWD: robust device-free through-the-wall detection of moving human with WiFi', *IEEE J. Sel. Areas Commun.*, 2017, **35**, (5), pp. 1090–1103
[21] Wang, Y., Wu, K., Ni, L.M.: 'Wifall: device-free fall detection by wireless networks', *IEEE Trans. Mob. Comput.*, 2017, **16**, (2), pp. 581–594
[22] Wang, H., Zhang, D., Wang, Y., *et al.*: 'RT-Fall: a real-time and contactless fall detection system with commodity WiFi devices', *IEEE Trans. Mob. Comput.*, 2017, **16**, (2), pp. 511–526
[23] Li, H., Yang, W., Wang, J., *et al.*: 'Wifinger: talk to your smart devices with finger-grained gesture', Proc. 2016 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing, Heidelberg, Germany, September 2016, pp. 250–261
[24] Khan, U.M., Kabir, Z., Hassan, S.A.: 'Wireless health monitoring using passive WiFi sensing', 2017 13th Int. Wireless Communications and Mobile Computing Conf. (IWCMC), Valencia, Spain, June 2017, pp. 1771–1776
[25] Gong, L., Yang, W., Man, D., *et al.*: 'WiFi-based real-time calibration-free passive human motion detection', *Sensors*, 2015, **15**, (12), pp. 32213–32229
[26] Wang, W., Liu, X., Shahzad, M.: 'Gait recognition using WiFi signals', Proc. 2016 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing, Heidelberg, Germany, September 2016, pp. 363–373
[27] Xi, W., Zhao, J., Li, X-Y., *et al.*: 'Electronic frog eye: counting crowd using WiFi', IEEE INFOCOM 2014 – IEEE Conf. on Computer Communications, Toronto, ON, 2014, pp. 361–369
[28] Li, C., Lubecke, V.M., Boric-Lubecke, O., *et al.*: 'A review on recent advances in Doppler radar sensors for noncontact healthcare monitoring', *IEEE Trans. Microw. Theory Tech.*, 2013, **61**, (5), pp. 2046–2060
[29] Ram, S.S., Christianson, C., Kim, Y., *et al.*: 'Simulation and analysis of human micro-Dopplers in through-wall environments', *IEEE Trans. Geosci. Remote Sens.*, 2010, **48**, (4), pp. 2015–2023
[30] Fadi, A.T.: '5G-enabled devices and smart-spaces in social-IoT: an overview', *Future Gener. Comput. Syst.*, 2017
[31] Anderson, M.G.: 'Design of multiple frequency continuous wave radar hardware and micro-Doppler based detection and classification algorithms', The University of Texas, Austin, 2008
[32] Kaleem, Z., Rehmani, M.H.: 'Amateur drone monitoring: state-of-the-art architectures, key enabling technologies, and future research directions', *IEEE Wirel. Commun.*, 2018, **25**, (2), pp. 150–159
[33] Ansari, F., Taban, M.R.: 'Implementation of sequential algorithm in batch processing for clutter and direct signal cancellation in passive bistatic radars', 2013 21st Iranian Conf. on Electrical Engineering (ICEE), Mashhad, 2013, pp. 1–6
[34] Hansen, B.E.: 'Sample splitting and threshold estimation', *Econometrica*, 2000, **68**, (3), pp. 575–603
[35] Hastie, T., Tibsbirani, R., Friedman, J.: '*The elements of statistical learning*' (Springer, Stanford, USA, 2008, 2nd edn.)
[36] Li, W., Tan, B., Piechocki, R.J.: 'Passive radar for opportunistic monitoring in E-health applications', *IEEE. J. Transl. Eng. Health. Med.*, 2018, **6**, pp. 1–10