Exploring Training Options for RF Sensing Using CSI

Simone Di Domenico, Mauro De Sanctis, Ernestina Cianca, Fabrizio Giuliano, and Giuseppe Bianchi

The authors analyze human behavior recognition approaches using WiFi channel state information from the perhaps less usual point of view of training and calibration needs. They classify the diverse forms of training so far employed into three main categories (trained, trained-once, and training-free). They further discuss under which conditions it is possible to move toward lighter forms of calibration or even succeed in devising fully untrained model-based solutions.

Abstract

This work analyzes human behavior recognition approaches using WiFi channel state information from the perhaps less usual point of view of training and calibration needs. With the help of selected literature examples, as well as with more detailed experimental insights on our own Doppler spectrum-based approach for physical motion/presence/cardinality detection, we first classify the diverse forms of training so far employed into three main categories (trained, trained-once, and training-free). We further discuss under which conditions it is possible to move toward lighter forms of calibration or even succeed in devising fully untrained model-based solutions. Our take home messages are mainly two. First, reduced training might not necessarily kill performance (although, of course, trade-offs will emerge). Second, reduced training must come along with a careful customization of the technical detection approach to the specificities of the behavior recognition application targeted, as it seems very hard to find a one-size-fits-all solution without relying on extensive training.

Introduction

In the last few years, we have witnessed significant and growing interest in devising device-free RF sensing for human behavior recognition that does not rely on wearable active sensors placed over a target human body. By itself, this research direction is of course not nearly new; for instance, the radar community has very long experience in engineering specialized waveforms and devising dedicated transmitting devices to detect a variety of physical quantities including human movements through the analysis of the Doppler spectrum [1–4]. However, the challenge that has emerged in the last few years is whether, and to what extent, it is possible to detect human behavior patterns *without any dedicated radar-type system* requiring specialized (and possibly costly) deployment, but rather exploit RF signals of opportunity transmitted by off-the-shelf unmodified RF devices such as legacy WiFi access points (APs). Furthermore, in a classical radar approach, the signal processing involves cross-correlating the reference and surveillance signals, which are not available using commercial devices with omnidirectional antennas. In this work we focus on WiFi-based recognition methods that can be applied to commercial devices without using reference and surveillance signals.

Besides the widespread deployment of WiFi networks (with virtually at least an AP in every home), the research was further boosted by the convenient availability of very flexible/open WiFi receivers that permit the programmer to gather information about the wireless channel characteristics expressed as channel frequency response (CFR) — or, more precisely, its estimation known as channel state information (CSI). The fact that CSI permits us to infer information about the presence/activity/gesture of human beings in the environment is very intuitive, and by itself not a surprise. Indeed, radio signals are attenuated and reflected by human bodies, and, as a consequence, presence and movements of the human body result in a change of the propagation channel between the transmitter and the receiver of the radio signal, which can therefore be perceived by properly analyzing the CSI and its change in time. What is perhaps surprising is the ability of our wireless community to accomplish trailblazing results in many scenarios: from the perhaps more classical gesture recognition applications, to the more challenging crowd counting, up to very creative applications including recognition of vital signs, keystroke detection, and so on.

To Train or Not to Train?

In several cases (but of course not all, as discussed later), detection accuracy and/or ability to discriminate fine-grained behavioral target patterns is accomplished at the expense of what we believe is a severe practical annoyance: the need for the proposed system to rely on a training phase so as to calibrate the system itself. Such a training phase may be very demanding, especially when good classification performance may only be achieved at the expense of dedicated and time-consuming training performed in the very same environment and conditions of the application itself (e.g., in the very same room and/or with the same position of objects including transmitter and receiver). In some scenarios, the need for on-site training may even become a blocker for the deployment of such applications.

In view of the above, a thorough community-wide discussion on the reliance of behavior recognition applications on (different forms and extent of) training would seem very natural and urgent. However, to our surprise, and to the best

Digital Object Identifier: 10.1109/MCOM.2018.1700145

Simone Di Domenico, Ernestina Cianca, and Giuseppe Bianchi are with the University of Rome "Tor Vergata"; Mauro De Sanctis is with the University of Rome "Tor Vergata" and the Peoples' Friendship University of Russia (RUDN University); Fabrizio Giuliano is with the University of Palermo.

 0163-6804/18/$25.00 © 2018 IEEE

of our knowledge, so far very little *systematic* attention has been given not only to techniques devised to reduce the training requirements, but even to more basic aspects such as the classification of applications on the basis of their training requirements.

The goal of this article is to provide some insights on human behavior recognition approaches using WiFi CSI from the less usual point of view of training and calibration needs. We believe that a good starting point is to introduce a simple categorization, in just three families, of the different types of training so far considered by literature works:

1. Trained: The recognition system uses a classifier that is trained and tested *under the same environmental conditions in which the system will be deployed* (very same room, position of the transmitter/receivers, etc). As a consequence, the training phase is specific for each environment and for each system setup. In particular, the set of the best features for one environment may be different from the best set of features for another environment. On one hand, with a dedicate training and a large set of features, very good accuracies may be achieved. On the other hand, this leads to a complex recognition system that is very sensitive to environmental changes (over-fitting problem).

2. Trained-once: The recognition system uses a classifier that is trained using data collected under certain conditions (room, position of Tx and Rx, position of objects, etc.), but it is then tested using data collected under different conditions.

3. Training-free: To get rid of training, it is necessary to introduce an explicit model that correlates the activity (e.g., human presence, gestures, or number of people) to one or more descriptors that are independent of the environment. The recognition process in this case consists of comparing these descriptors with some thresholds. When the thresholds are *theoretically* defined, making some usually strict assumptions (e.g., the heart movement is sinusoidal with a given frequency), we talk about a *fully modeled* approach. When these thresholds are defined under relaxed assumptions but using some (even slight) calibration, we talk of an *empirically modeled* approach. To the best of our knowledge, only part of the approach in [5], in particular the method for stationary human detection, can be defined as *fully modeled*.

Our Contribution

With this article, our goal is to focus the reader's attention on the role of (and possible alternatives to) training in human behavior recognition systems that rely on WiFi CSI. Besides contributing with the above introduced categorization into three categories, the main goal of this work is to provide insights on when (i.e., for which applications and under which conditions) it may be possible to reduce the amount of training and, for instance, re-design an application so that it can be moved from the "trained" category to the "trained-once" or even the "training-free" category, thus gaining in practicality and deployability.

First, we discuss *selected* literature examples so as to provide more insights on how, and through which solutions or optimizations, specific literature approaches have reduced their need for training and have succeeded in becoming either "trained-once" approaches or "training-free" ones. Of course, we do not claim completeness for such an analysis of the literature, but we hope the selected examples will be sufficiently representative of different applications and of the relevant diverse technical issues emerging.

Second, and the original contribution of this work, we present two different technical solutions and relevant experimental results: one falling into the "trained-once" category and a second approach trying to get as close as possible to a "training-free" approach, at the expense of accuracy trade-offs. In particular, it is shown that a *fully modeled* approach is feasible for the simpler recognition task of presence detection of moving persons. For more complex tasks (e.g., recognizing whether in a room there is one person or more than one person, or there is someone sitting or standing rather than walking), the article shows that it is still possible to effectively reduce the need for training. More specifically, the article shows that in these cases it is possible to define thresholds, on some descriptors that must be carefully selected, which are independent of the environment, even if they must be empirically derived.

Toward a Training-Free Recognition System

This section reviews and discusses the current effort toward reduced need for traning

Trained-Once Approach

In this class of approaches, recognition is still based on a classification process. This means that a set of features must be extracted from the collected data and then selected. To make a recognition system less sensitive to the environment, features should be more "sensitive" to the changes of the propagation environment caused by any human motion and less "sensitive" to characteristics of the propagation environment in completely static conditions.

One of the first papers explicitly tackling the issue of reducing the training is [6], where robustness with respect to changes in the environment is achieved by using features related to the rate of change of the CSI rather than to its absolute value. In [7], the authors focus on the Doppler spectrum as a measure of the channel dynamics more correlated to the human presence and motion and less sensitive to the background environment. The proposed recognition system is used for crowd density estimation, and the intuitive idea is that the width, the height, and the number of peaks of the Doppler spectrum increase as the crowd density increases. More specifically, in [6], features are extracted from a mean Doppler spectrum computed using the CSI, where the mean is performed over the subcarriers.

It is worth noting that in the mentioned "trained-once" approaches, in the effort to find features that are more robust to the background environment, features measuring channel variations represent good candidates. In particular, the Doppler spectrum [7] and the CSI vectors distance [6] implicitly characterize the channel

To make a recognition system less sensitive to the environment, features should be more "sensitive" to the changes of the propagation environment caused by any human motion and less "sensitive" to characteristics of the propagation environment in completely static conditions.

In general, radar applications offer many hints in the effort of building physical-mathematical models correlating the human motion to the characteristics of the backscattered signals. Nevertheless, there are differences that must be taken into account.

variation and hence the motion of the scatterers, including the human scatterers. Therefore, any proper measure applied on the Doppler spectrum or on the CSI vectors distance is also an indirect measure of the human motion.

It is also worth outlining that in [7], classification is performed using only one feature: the spectral kurtosis of the mean Doppler spectrum. This is not in line with what is usually done when dedicated training is used. In that case, assuming good and uncorrelated features, increasing the number of selected features helps in improving the accuracy. This could be intuitively explained by the fact that while one single feature could be less dependent on the specific environment, a combined set of features is more likely dependent on it.

Training-Free

To go toward a completely training-free system, we need:

- A model that correlates the activity/gesture or the number of people with one descriptor that is uncorrelated to the environment;
- To estimate this descriptor from noisy measurements that are mostly correlated to the environment

When the parameters of the model do not need any calibration phase, we talk of a *fully modeled* approach. When the model needs some, even slight, calibration, we talk about an *empirically modeled* approach.

In [8] authors present a human detection system that is claimed to be "calibration-free." As a matter of fact, results in [8] prove the effectiveness of their proposed system based on the use of rate of change of the phase of the CSI, and also the robustness to changes in the testing environment. However, strictly speaking, the system cannot be defined as "calibration-free" as thresholds, even if by nature they are less sensitive to the background environment, are calculated by the data collected in many experiments performed in different rooms.

The fully modeled approach is a rather challenging task for activity recognition or crowd counting purposes. Different types of movements (slower vs. faster, periodic vs. random) might require different models and a multi-step recognition system.

Also, in the simpler case of human presence detection, we must distinguish between the detection of moving targets or stationary targets and use different models in the two cases, as done in [5]. In the case of moving targets, in [5] the idea is that in the case of no human presence or merely a stationary human, successive measurements would exhibit a high correlation factor, that is, large eigenvalues (close to 1). In the case of a moving human, eigenvalues tend to be small. Therefore, they calculate the normalized maximum eigenvalues λ_A and λ_C of the correlation matrices for both the amplitude and the phase of a set of consecutive CSI vectors. Therefore, their moving target detection system is based on the following model: if $\lambda_A \geq C_{thr}$, then $p = 0$ (empty), while if $\lambda_A < C_{thr}$, then $p = 1$ (presence), where the threshold C_{thr} is scenario-independent as they use eigenvalue-based features that are independent of absolute signal powers, which vary over different scenarios and time. However, as the authors of [5] also mention, the threshold C_{thr} also needs a calibration, even if just a slight one. Therefore, this approach is not strictly speaking a fully modeled approach.

Closer to a fully modeled approach is what the authors of [5] propose for stationary human detection. The basic observation is that humans, even if stationary, continuously breathe. Therefore, a system able to detect the subtle chest motions due to breathing could detect the presence of a stationary human. Human breathing changes the signal propagation, and, whatever is the relative position of the human in the room, these signal changes have a periodicity that should be equal to the breathing periodicity. Basically, they assume that in the measured signal there should be a periodical signal component with the same frequency as human breathing, which ranges between 0.167 Hz and 0.667 Hz. Therefore, they detect the human breathing using a parameter estimation technique for estimating frequency, amplitude, and phase of this sinusoidal component. It is worth noting that similar model-based approaches are used in several radar-based systems for vital signs monitoring.

In general, radar applications offer many hints in the effort to build physical-mathematical models correlating the human motion to the characteristics of the backscattered signals. Nevertheless, there are differences that must be taken into account. First of all, in a classical radar-based approach, the backscattered signal is usually transmitted by a dedicated transmitter and has some favorable characteristics, while in this case there is no control on the transmitter at all. Moreover, the parameter estimation should not be done by trying to remove the multipath but rather by trying to take advantage of the multipath itself (collecting as much as possible the energy of the signal paths reflected by the target and removing the ones not reflected by the target). This approach requires methods that are rather different from the ones used in typical radar systems.

Doppler Spectrum Using CSI

Correlation between Doppler Spectrum and Human Motion

In a propagation environment including a fixed transmitter, a fixed receiver, and a moving scatterer, different multipath components arrive at the receiver from different directions and with different Doppler frequency shifts. The Doppler frequency shift of a single multipath component reflected by a moving scatterer is given by

$$f_D = 2 f_c \frac{v}{c} \cos\left(\frac{\beta}{2}\right) \cos(\delta) \qquad (1)$$

where f_c is the carrier frequency of the transmitted signal, c is the speed of light, v is the magnitude of the velocity vector of the scatterer, δ is the angle between the bistatic bisector and the velocity vector of the scatterer, and β is the angle subtended between the transmitter, the scatterer, and the receiver. This equation is derived from a bistatic radar system [9].

The mean Doppler spectrum describes the distribution of the Doppler shift considering all the scatterers and all the multipath components. It

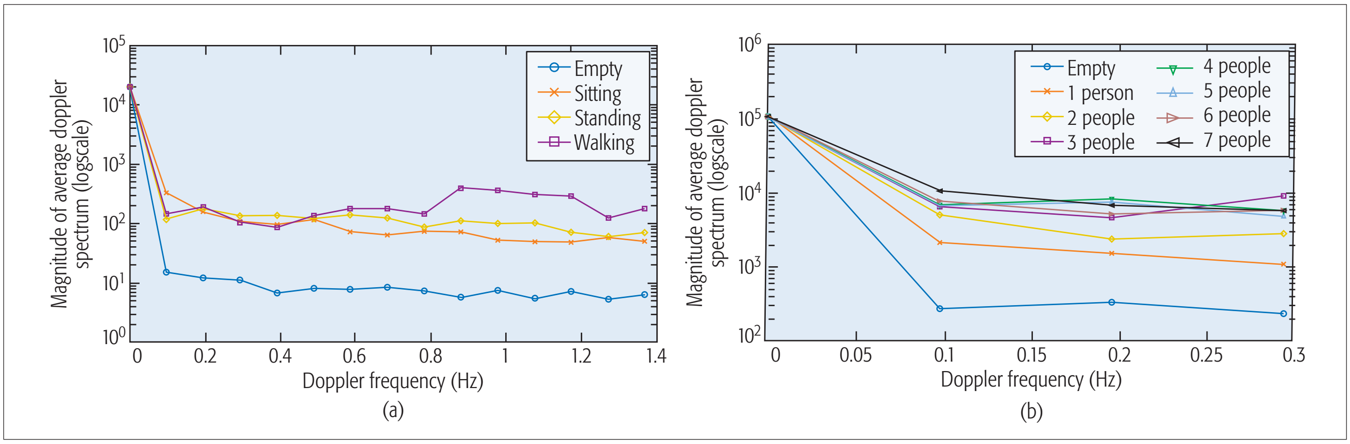

Figure 1. Doppler spectrum obtained for: a) different human activities; b) different number of people in the room.

is worth noting that even in a crowded environment, most of the scatterers are fixed, and hence, the dominant component of the mean Doppler spectrum is the static component with $f_D = 0$ Hz. Furthermore, for a scatterer moving with speed v, the Doppler shift of the received signal depends on the pair of angles (β, δ), achieving a minimum of 0 Hz and a maximum of

$$f_D = 2 f_c \frac{v}{c} \text{Hz}.$$

Each part of the human body may be modeled as a single scatterer. The range of walking speeds that humans may have in a confined indoor space is fairly narrow. Concerning the human motion in static conditions (sitting or standing still), the maximum speed of gesture/movement speed is usually low and, in any case, lower with respect to walking movements. Furthermore, the set of possible movements is larger in a standing position than in a sitting position.

The concept behind our training-free approach based on the analysis of the mean Doppler spectrum is the following: As the number of mobile scatterers and also their speed increase, the mean Doppler spectrum tends to broaden. This effect can be noticed from the mean Doppler spectrum shown in Fig. 1 as a first result of the experiments described in the next sections.

The Doppler spectrum is hence correlated to the presence/motion of people, but it is independent of the background environment, which is stationary. Therefore, it is expected to be a good candidate for reducing the need for training.

Doppler Spectrum Estimation

Since the WiFi packet transmission is not continuous, the sequence of estimated CSI provides a sampling of the human movements with a low rate, and, as a consequence, the Doppler spectrum must be computed over a sliding window, which allows a certain number of movements to be observed. We set this sliding window to W = 1024 CSI, corresponding to 10 s of time. For each received packet, a number of N_{ch} CSI vectors are collected (i.e., one CSI for each RF channel). Each CSI $\mathbf{H}^{\mathbf{i,j}} = [H_1^{i,j}\ H_2^{i,j} \dots H_{30}^{i,j}]$ is a complex-valued vector of $N_{sub} = 30$ elements providing the baseband channel gain for the ith Tx spatial stream, $i = 1, \dots, N_{ss}$, and for the jth Rx antenna, $j = 1, \dots, N_{rx}$. The received signal after the discrete Fourier transform (DFT), at the jth Rx antenna, and for the kth subcarrier, $k = 1, \dots, N_{sub}$, can be written as

$$Y_k^j = \sum_{i=1}^{N_{ss}} H_k^{i,j} X_k^i + N_k^j \qquad (2)$$

where X_k^i is the transmitted symbol and N_k^j is the additive white Gaussian noise (AWGN). Then a single CSI is obtained by concatenating $N_{ch} = N_{ss}N_{rx} = 6$ CSI vectors $\mathbf{H}^{\mathbf{i,j}}$, resulting in a global CSI vector $\mathbf{H}$ of length $N_{el} = N_{sub}N_{ch} = 180$. In order to remove the power fluctuations of the AP, which are correlated over the subcarriers and the channels, each global CSI vector is normalized by its mean value.

The Doppler spectrum is obtained by performing the fast Fourier transform (FFT) of the sequence of the CSI magnitude for each of the N_{el} subcarriers over a sliding window of size W. Then a single mean Doppler spectrum is computed by averaging the Doppler spectrum over the N_{el} subcarriers. This approach is based on the concept of pulsed radar and allows a total Doppler bandwidth of 50 Hz to be achieved. It is worth noting that during the collection of CSI in one sliding window, the speed of the human movement is not constant in both magnitude and direction.

Extraction of Features from the Doppler Spectrum

In order to quantify the Doppler spectrum broadening with the aim of recognizing human motion, a set of features charactering the shape of the mean Doppler spectrum should be defined. A non-exhaustive list of spectral features that can be computed on the mean Doppler spectrum is shown in Table 1. See [6] for a more detailed description of these features.

The selection of the best set of features depends on the training approach:

- Trained: In this approach the set of features is selected only on the basis of accuracy without having any constraint on the number of features (besides complexity constraints), and generally that set is different for each environment.
- Trained-once: In this approach the set of features is driven by the need to only consider good features that are also independent from the environment. In this case, the same set of features is used for every environment.

Feature	Definition
RMS Doppler bandwidth	Square root of the second moment of the Doppler spectrum.
Mean	Mean of the Doppler spectrum.
Standard deviation	Standard deviation of the Doppler spectrum.
Doppler ratio	Ratio between the third and the first frequency component of the Doppler spectrum.
Spectral spread	Second central moment of the Doppler spectrum with log frequencies.
Spectral rolloff	80th percentile of the Doppler spectrum distribution across frequencies.
Spectral slope	Rate of falloff of the Doppler spectrum.
Spectral flatness	Measure of the flatness of the Doppler spectrum.
Spectral skewness	Measure of the asymmetry of the Doppler spectrum about its mean.
Spectral kurtosis	Measure of the tailedness of the Doppler spectrum.
Spectral centroid	Measure of the center of mass of the Doppler spectrum.
Spectral entropy	Measure of the information contained in the Doppler spectrum.

Table 1. List of features extracted from the mean Doppler spectrum.

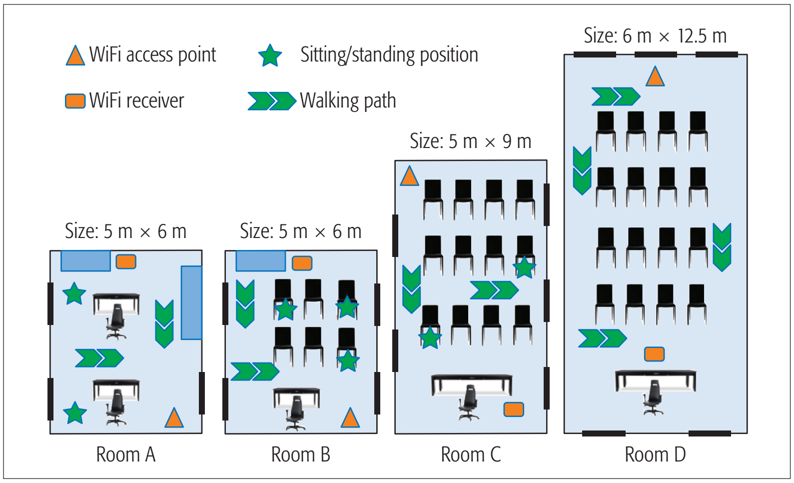

Figure 2. Experimental environments.

- Training-free: In this approach the set of features should be constrained to a few good features that are independent of the environment and can also be directly correlated (i.e., through a monotonic relationship) to the extent of the human movement and the number of people.

Experimental Setup

The system includes a commercial 2.4 GHz WiFi AP acting as a double-antenna transmitter and a laptop with Ubuntu 10.04 LTS with a WiFi card Intel 5300 with 3 antennas acting as a receiver. CSI is extracted every T_p = 10 ms from the data packets received from the AP by using a customized firmware and an open source Linux wireless driver for the Intel 5300 WiFi card [10, 11]. A single WiFi AP and a single WiFi receiver have been used.

CSI vectors have been collected under the following conditions:

- Concerning the activity recognition scenario, the activity of the volunteer (i.e., sitting, standing, and walking) was carried in different positions/patterns.
- Concerning the people counting scenario, volunteers could move as they wanted (randomly, repeating a path, etc.), or they could simply stand.

We carried out experiments in four different rooms (Fig. 2):

- Room A: small office room (5 m × 6 m)
- Room B: small office room (5 m × 6 m)
- Room C: medium sized meeting room (5 m × 9 m)
- Room D: large meeting room (6 m × 12.5 m)

The considered rooms are quite different in terms of size, furniture, and location of the transmitter and receiver. The distance between the AP and the WiFi receiver was larger than 5 m in any room. In order to fairly compare the performance of the system, the number of people during the experiments in the environments under test was the same. Due to the space constraints of Room A, which is the smallest one, the crowd size was limited to N_c = 7 people. A number of N_{CSI} = 5000 CSI complex vectors has been collected for each class and for each room.

In the following sections, we focus on the trained-once and training-free approaches, while the trained approach is considered as a reference/benchmark method. All these methods are based on the mean Doppler spectrum as previously described.

Trained-Once Approach

Concerning the trained and trained-once options, the set of classes used for activity recognition and for people counting scenarios are defined as follows:

- Activity recognition classes:
  - Class 1: empty
  - Class 2: sitting
  - Class 3: standing
  - Class 4: walking
- People counting classes:
  - Class 1: 0 people
  - Class 2: 1 person
  - Class 3: 2 people
  - Class 4: 3, 4 people
  - Class 5: 5, 6, 7 people

A naive Bayes classifier was used for both trained and trained-once options; starting from the overall set of features listed in Table 1, a feature selection process based on the maximization of the average accuracy resulted in the selection of three features for both trained and trained-once options.

Concerning the trained-once case, the classification results are achieved performing a rotation of the training and test environments. The results in terms of average classification accuracy are shown in Table 2. As expected, a trained-once approach always leads to lower accuracy. However, the performance degradation is strictly related to the room chosen for the training phase. It is worth outlining that, for instance, choosing the medium sized room (room C in the case of people counting) for the training is not always the best choice. On the other hand, this is expected as the characteristics of the propagation channel depend not only on the size of the environment, but also on the geometry/materials of the furniture, the position of Tx/Rx, and the position of doors/windows.

Scenario (Environment) / Training option	Trained	Trained-once	Training-free (empirically modeled)
Activity recognition (room A)	81% (4 classes)	79% (train room B) 70% (train room C) (4 classes)	95% (3 classes)
Activity recognition (room B)	94% (4 classes)	79% (train room A) 53% (train room C) (4 classes)	89% (3 classes)
Activity recognition (room C)	72% (4 classes)	59% (train room A) 59% (train room B) (4 classes)	66% (3 classes)
People counting (room A)	94% (5 classes)	84% (train room C) 73% (train room D) (5 classes)	99% (3 classes)
People counting (room C)	92% (5 classes)	82% (train room A) 77% (train room D) (5 classes)	97% (3 classes)
People counting (room D)	72% (5 classes)	52% (train room A) 46% (train room C) (5 classes)	80% (3 classes)

Table 2. Comparison of classification accuracy for different training options and environments.

Training-Free

In a training-free approach, the training data for a particular environment is not available, and the recognition method cannot be based on a classifier, but it should be based on a physical-mathematical model. The main difference between a training-free and a trained or trained-once approach is that when using a training-free approach you must select a feature that is a monotone function of the degree of motion, while this is not needed using a classifier as in the trained or trained-once approach. In the training-free approach, the model should lead to a very simple recognition process that can permit distinguishing the different classes of interest by using a set of thresholds.

We have first wondered how to recognize the presence of a moving person in a room, and we have chosen the spectral rolloff as a feature that can be directly connected to the speed of movement (this is not true for other features such as the spectral kurtosis, for instance). The spectral rolloff is a specific estimation of the cutoff frequency of the mean Doppler spectrum, as it represents the frequency below which most of the accumulated magnitude of the mean Doppler spectrum is concentrated.

In order to develop a simple recognition method based on the spectral rolloff, a threshold should be defined only on the basis of the model assumptions. In this case we considered a simplified set of classes: "empty" and "1 moving person." Considering that the empty scenario does not include any important movement, the threshold on the spectral rolloff may be set to 0 Hz. Figure 3, which shows the spectral rolloff over time in the case of a moving person (three curves represent three different walking paths), proves that the spectral rolloff is a good choice for a fully modeled approach as in the case of empty room, the spectral rolloff, regardless of the specific environment, is exactly zero. Therefore, using the spectral rolloff and the considered

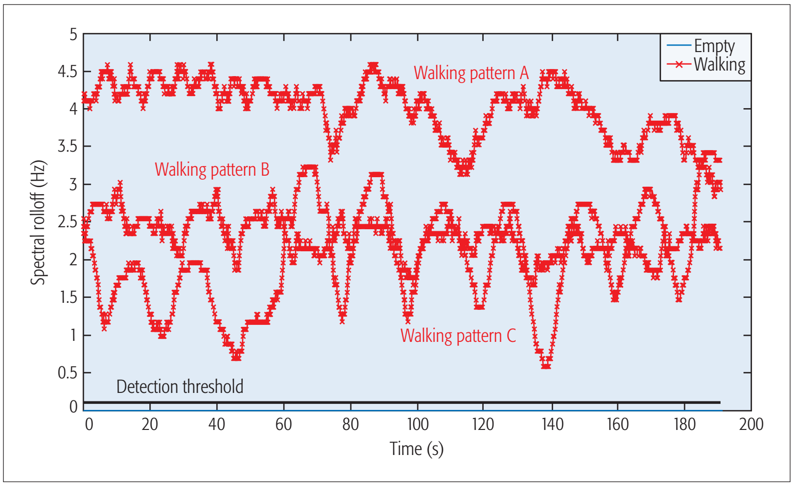

Figure 3. Spectral rolloff vs. time.

threshold, the average accuracy of the proposed recognition method is 100 percent (i.e., the method is able to clearly distinguish the presence of one moving person with respect to the empty room).

We have then tried to consider more "complex" recognition tasks such as activity recognition with three classes (empty, sitting/standing, walking) and people counting with three classes (empty, 1 person, more than one person). Going for a fully modeled approach in these cases is very challenging. As already mentioned, more models would be needed, and a multistep recognition process should be used. Therefore, we wondered if it is possible to perform the above-mentioned complex tasks by using an empirically modeled approach, where the time-consuming training phase is replaced by the calibration of some thresholds. Figure 4 shows the scatter plots for the three classes concerning both the activity recognition and the people counting scenarios. For the scatter plots, the following two features, computed on the mean

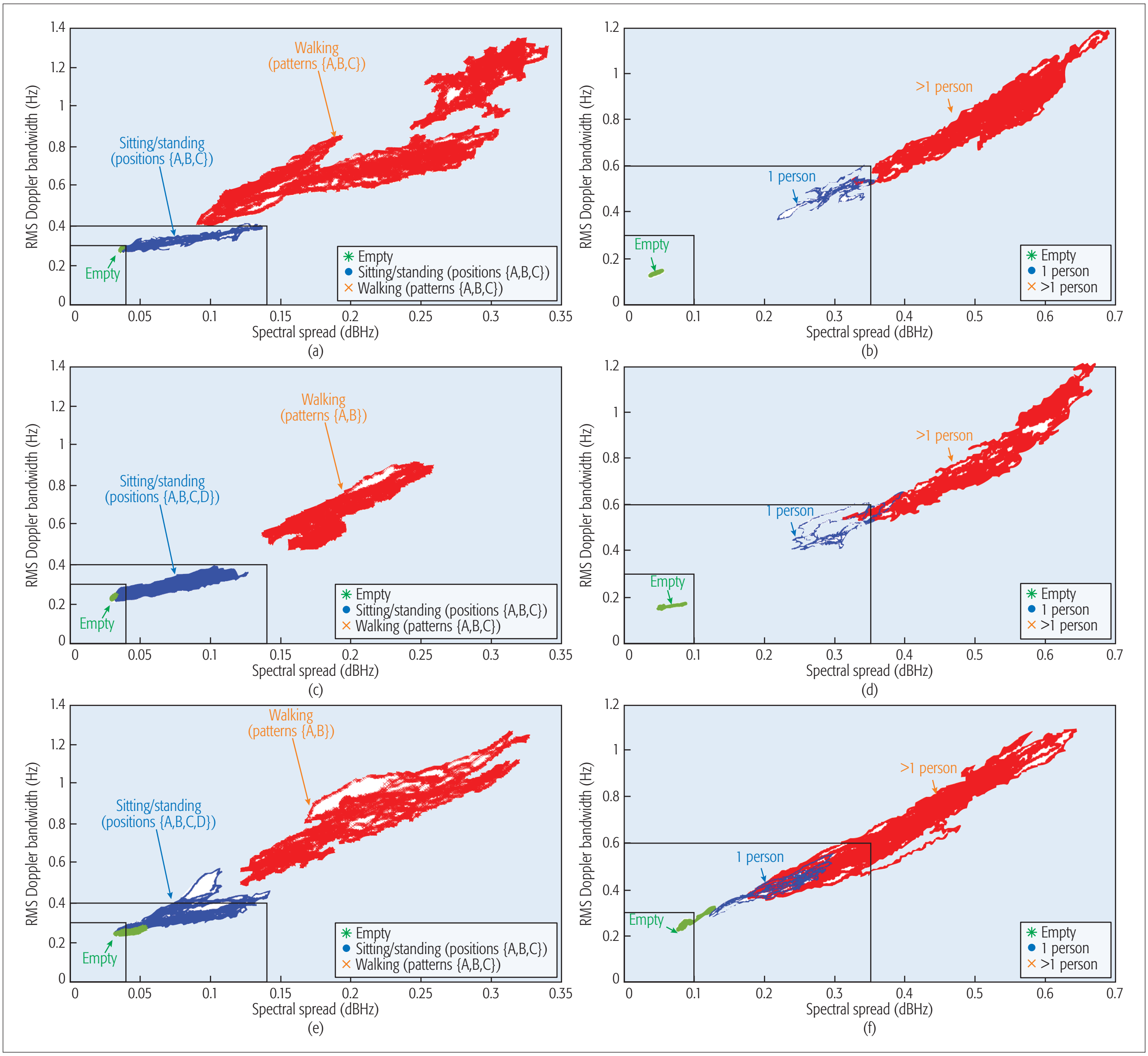

Figure 4. Scatter plot of RMS doppler bandwidth vs. spectral spread for both activity recognition (a), c), e)) and people counting (b), d), f) scenarios and for each considered environment. Empirical thresholds for each scenario are also shown.

Doppler spectrum, have been considered: root mean squared (RMS) Doppler bandwidth and spectral spread. These features have been identified as good metrics as they should be independent from the background environment and monotonically increasing as the degree of motion increases. It is evident from these scatter plots that it is possible to identify two thresholds to clearly distinguish the three classes. The fundamental result is that these thresholds are more or less the same in the three different environments, proving that these thresholds, even if they must be empirically evaluated, do not depend on the specific environment. Thus, once they are calculated, they can be used in any environment even if not previously considered. Table 2 shows the average accuracy for each training option, for each scenario and for each environment. These results show the effectiveness of the training-free approach that has been presented.

Conclusion

This article discusses three possible training options that can be used by a device-free RF sensing system using CSI: trained, trained-once, and training-free.

The trained approach allows us to achieve the best accuracies because it does not have any limitations on the number of features or on the training data, and it may follow very accurately any specific behavior of the CSI. However, the limitations of this approach arise in practical use in many real application scenarios. Moreover, attention should be paid to the overfitting problem, which limits the reproducibility of the results.

The reduction of the need for training can be achieved by using:

A trained-once approach: In this framework an important step is the choice of type of measurements on the received signals (e.g., phase or amplitude of CSI, Doppler shift), but much more important is the choice of features. Features related to the variations of CSI amplitude/phase, rather than the absolute values, are less sensitive to the surrounding environment. Moreover, it can be observed that when trying to reduce the training, the use of a lower number of features is preferable.

A training-free approach: In this case a complete fully modeled approach is rather challenging. In the article a fully modeled approach has been proposed and experimentally validated for the simpler case of presence detection. In this framework, much effort is still needed in both:

- Identifying proper physical-mathematical models that correlate the activity/gesture or the number of people with one descriptor that is independent from the environment
- Understanding how to better estimate this descriptor from noisy CSI measurements that are mainly correlated to the environment

In both cases, useful hints can be derived from passive radar systems. However, a multisteps approach is likely to be needed, where the recognition is performed by making more and more refined recognition steps, using different models in each step. However, the need for training could be effectively removed by still using a model-based option, but following an empirically modeled approach, where there is the need to empirically calibrate some thresholds. In particular, we have shown that by using features such as the RMS Doppler bandwidth and the spectral spread, the same thresholds can be used in different environments for classifying three classes.

The trained approach allows us to achieve the best accuracies because it does not have any limitations on the number of features or on the training data, and it may follow very accurately any specific behavior of the CSI. However, the limitations of this approach come from practical use in many real application scenarios.

References

[1] Q. Chen *et al.*, "Activity Recognition Based on Micro-Doppler Signature with In-Home Wi-Fi," *2016 IEEE 18th Int'l. Conf. e-Health Networking, Applications and Services*, Sept 2016, pp. 1–6.

[2] W. Li *et al.*, "Opportunistic Physical Activity Monitoring via Passive Wifi Radar," *2016 IEEE 18th Int'l. Conf. e-Health Networking, Applications and Services*, Sept 2016, pp. 1–6.

[3] K. Chetty, G. E. Smith, and K. Woodbridge, "Through-the-Wall Sensing of Personnel Using Passive Bistatic WiFi Radar at Standoff Distances," *IEEE Trans. Geoscience and Remote Sensing*, vol. 50, no. 4, Apr. 2012, pp. 1218–26.

[4] J. B. Andersen *et al.*, "Doppler Spectrum from Moving Scatterers in a Random Environment," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, June 2009, pp. 3270–77.

[5] C. Wu *et al.*, "Non-Invasive Detection of Moving and Stationary Human with WiFi," *IEEE JSAC*, vol. 33, no. 11, Nov. 2015, pp. 2329–42.

[6] S. Di Domenico *et al.*, "A Trained-Once Crowd Counting Method Using Differential WiFi Channel State Information," *Proc. 3rd Int'l. Wksp. Physical Analytics*, ser. WPA '16, 2016, pp. 37–42.

[7] S. Di Domenico *et al.*, "Trained-Once Device-Free Crowd Counting and Occupancy Estimation Using Wifi: A Doppler Spectrum Based Approach," *2016 IEEE 12th Int'l. Conf. Wireless and Mobile Computing, Networking and Commun.*, 2016, pp. 1–8.

[8] L. Gong *et al.*, "Wifi-Based Realtime Calibration-Free Passive Human Motion Detection," Sensors, vol. 15, no. 12, 2015, pp. 32213–229.

[9] M. Skolnik, *Radar Handbook*, 3rd ed., McGraw-Hill, 2008.

[10] W. Xi *et al.*, "Electronic Frog Eye: Counting Crowd Using WiFi," *IEEE INFOCOM 2014*, Toronto, Canada, 2014, pp. 361–69.

[11] D. Halperin *et al.*, "Predictable 802.11 Packet Delivery from Wireless Channel Measurements," *Proc. 2nd Int'l. Wksp. Wireless Network Measurement*, ser. ACM SIGCOMM 2010, pp. 159–70.

Biographies

Simone Di Domenico received both Bachelor's and Master's degrees in Internet technology engineering from the University of Roma "Tor Vergata," Italy, in 2012 and 2014, respectively. Currently, he is a Ph.D. student in electronic engineering at the University of Roma "Tor Vergata," and his main research interests are RF device-free human activity recognition and RF device-free people counting.

Mauro De Sanctis (mauro.de.sanctis@uniroma2.it) received his Ph.D. degree in telecommunications and microelectronics engineering in 2006 from the University of Roma "Tor Vergata." He is an assistant professor at the Department of Electronics Engineering of the same university. He is serving as Associate Editor for *IEEE Aerospace and Electronic Systems Magazine*. His main areas of interest are wireless terrestrial and satellite communication networks, human activity recognition, data mining, and information theory. He has co-authored more than 80 papers published in journals and conference proceedings.

Ernestina Cianca is an assistant professor at the Department of Electronic Engineering, University of Rome "Tor Vergata" and Vice-Director of the Interdisciplinary Center CTIF. Her research interests cover wireless access technologies (CDMA, OFDM, SC-FDE) and equalization for both terrestrial and satellite communications, device-free RF-based activity recognition/crowd density estimation and localization, and microwave medical imaging (i.e., stroke detection). She has relevant experience in national and European projects. She has published about 100 papers in international journal and conference proceedings.

Fabrizio Giuliano received his engineering Master's degree in telecommunications in 2009. He received his Ph.D. in telecommunication engineering in 2014. He is currently a researcher for Consorzio Interuniversitario per le Telecomunicazioni (CNIT). His research activity has focused on wireless networks, with experience in IEEE 802.11 and cognitive radio at the physical and MAC layers. He has been involved in several national and European research projects.

Giuseppe Bianchi has been a full professor of networking at the School of Engineering of the University of Roma "Tor Vergata" since 2007. His research activity, documented in more than 200 peer-reviewed international journal/conference papers having received to date more than 14,000 citations, focuses on wireless networks (his WLAN modeling work received the ACM SIGMOBILE Test Of Time award in 2017), programmable network systems, and privacy and security monitoring in both the wireless and wired domains.