

Indoor people density sensing using Wi-Fi and channel state information

Mohan Liyanage^{1*}, Chii Chang¹, Satish Srirama¹, Seng Loke²

¹ Mobile & Cloud Lab, Institute of Computer Science, University of Tartu, Tartu 50090, Estonia

² School of Information Technology, Deakin University, 221 Burwood Highway, Burwood VIC 3125, Australia

Corresponding Author Email: liyanage@ut.ee

https://doi.org/10.18280/ama_b.610107

ABSTRACT

Received: 15 March 2018

Accepted: 29 March 2018

Keywords:

channel state information, crowd estimation, device-free, RF sensing, Wi-Fi

Device-free passive crowd estimation technologies are capable of measuring the density of people in an area, using existing wireless network infrastructure. It has been applied in various application domains such as pedestrian control, crowd management in subways, guided tours and so forth. In this work, we have designed, implemented and validated a device-free indoor human crowd density sensing method based on Channel State Information (CSI) captured by a single Wi-Fi receiver. We investigate the behaviour of the CSI amplitude variance of each receiving stream over the different subcarriers and propose a method to aggregate the CSI amplitude over time without losing critical information. Further, we evaluated the method using three different machine learning algorithms. The result shows the proposed method achieves an estimated accuracy of 99.8% with the Weighted K-Nearest Neighbour.

1. INTRODUCTION

The process of estimating the number of people in a given area has become a significant research area over the past years. Robust crowd counting in either an open or closed environment is an important and also a challenging task. A lot of research has been done in the development of crowd density estimation systems and they have been applied in a wide range of applications, such as counting people in a festival [1], pedestrian control, crowd management in subways [2], and customer count estimation in retail shops and so forth.

The most common and traditional approaches are vision-based systems that analyse a video or images to estimate the number of people in the scene [2-5]. Mostly, those image processing techniques are able to estimate the population effectively. However, these methods suffer from some inherent drawbacks. For example, environmental factors such as light, fog, and dust greatly affect the quality of the video/image. Further, installing cameras in public areas raises privacy concerns and also requires additional costs.

Emerging smartphones have been utilised in many sensing scenarios such as interacting with proximal Internet of Things (IoT) to assist ubiquitous computing applications [6], utilising the inbuilt sensing mechanism to provide environmental sensor data to remote clients [7-8] or brokering the environmental sensor data from proximal sensors to remote data centres on the move [9]. Furthermore, establishing an ad-hoc network among smartphones is not limited to performing data routing or distributed processing [10], with the inbuilt proximal networking mechanisms (e.g. Bluetooth) or audio processing mechanisms, the smartphones can sense the density of people in proximity [1, 11-12].

Although the smartphone-based sensing approaches are promising, they rely on the participants to carry the specific devices. Different from the device-specific approaches, device-free Radio Signal Strength (RSS)-based human density

sensing approaches, which use RSS as an indication of the signal propagation strength, do not require people to carry devices. The related frameworks have been proposed in various applications such as localisation [13], motion detection [14], human activity recognition [15], and crowd estimation [16-17]. Moreover, researchers also estimate the crowd density based on analysing RSS values of RFID tags [18-19] or ZigBee wireless nodes [20]. However, most existing RSS measurement-based models face a critical challenge in accuracy because of the fundamental problems of RF wave propagation in the indoor environment. There are possibly multiple signals arriving at the receiver through multiple paths, and also attenuated by reflection when the signal hits the surface of an obstacle. Consequently, the time-varying nature and, one Receiving Signal Strength Indication (RSSI) value per packet cannot establish an accurate prediction model in complex environments.

In contrast to having only one RSSI value per packet, current widely used Orthogonal Frequency Division Multiplexing (OFDM) systems explore the fine-grained physical layer information in multipath environments. Different from RSS, Channel State Information (CSI) is a complex matrix of values from the physical layer where data are modulated on multiple subcarriers at different frequencies and simultaneously transmitted over the IEEE 802.11n Wi-Fi link [21].

The CSI consists of amplitude and phase shift information that describes how the propagated signal experienced by the different effects of scattering, fading, and power decay for each spatial stream on every subcarrier. According to this characteristic, CSI is more stable and has more information than RSSI and also we can get more information that is sensitive to the environmental variance.

Recently, due to the robust nature of the CSI in complex environments, there are many passive-sensing (device free sensing) frameworks that have been proposed in several

application domains such as indoor localisation [22-23], activity recognition [24-25], gesture recognition [26], scheduling algorithm for base stations [27], counting people [28-29], etc.

In this paper, we present the design and implementation of a CSI based crowd density estimation framework for the indoor environment using one Wi-Fi router and a receiver with off-the-shelf 802.11n Intel 5300 NIC. Our method is based on amplitude measurements of the CSI and a normalisation of the received data.

We have observed that in some CSI streams, the effect of human presence is more than in other streams. Therefore, unlike previous approaches such as [25] and [30], we do not calculate the mean value of all streams, which will lose the distinctive sensitivity of streams and also causes a false prediction. Instead, our approach divides the captured CSI datasets into samples that contain 10 packets per sample. Thereafter, we average the CSI values within the sample for each antenna separately to minimise the information loss and also filter noises due to sudden changes in the environment.

After extraction of the relevant features of the captured CSI data, our system uses three different classification methods which are available with the MATLAB Classification Learner to estimate the density. The experimental classification results have shown that the proposed method has the accuracy of 99.8% with the Weighted K-Nearest Neighbour (K-NN), 99.6% with Linear Support Vector Machine (SVM), and 94.3% with the Complex Tree.

The rest of the paper is organised as follows. Section 2 provides an overview of related work and Section 3 presents theoretical background information on wireless signal propagation. Section 4 describes the architecture and the methodology of the proposed crowd density estimation method followed by the performance evaluation. The paper concludes in Section 5 along with future research directions.

2. RELATED WORK

Crowd density estimation has attracted significant attention from both industrial and academic researchers in last decade. In this section, we review a number of existing approaches in the density estimation field based on a classification of their technologies.

2.1 Computer vision-based crowd estimation

In recent years, researchers drastically used computer vision related techniques to estimate and monitor crowds automatically. These techniques are mainly focused on detecting and counting people's presence in images and video scenes.

Li et al. [2] proposed a computer vision-based crowd estimation scheme. Specifically, the scheme combines both the MID (Mosaic Image Difference)-based on foreground segmentation algorithm and the Histograms of Oriented Gradients (HOG)-based head-shoulder detection algorithm to count people in surveillance scenes. Furthermore, this scheme is capable of locating the position of each individual in the image scene.

Chan et al. [3] proposed a privacy-preserving system for estimating the size in a homogeneous crowd. Preserving privacy is an essential task in most of the video-based crowd counting systems. In contrast to the approaches that use

explicit object segmentation or tracking, Chan et al. track the direction of movement of each and every single pedestrian while preserving privacy. Further, their system extracted a set of simple holistic features from a region which segmented into components of homogeneous motion. Based on this, they could estimate the number of people per segment using Gaussian Process regression method.

Although various researchers have proposed various approaches that use video data to estimate the density of the human crowd, the challenge still exists. In general, the computer vision-based approaches face issues in (1) the limitation in the resolution quality of the images; (2) changes of illumination and weather conditions in the area; (3) the speed of the objects' movements; (4) privacy protection. For instance, Subburaman et al. [5] mentioned that it is difficult to detect heads in far views with low-resolution frames.

2.2 Smartphone-based crowd estimation

Today, smartphones are popular and powerful. Without using pre-installed infrastructure, one can estimate the crowd density from the number of smart devices presented in an area. For example, Kannan et al. [12] proposed audio-tones based crowd estimation solution that utilises the mobile phone's microphones and speaker.

Another common approach is using Bluetooth scans to estimate the crowd densities. For example, Weppner and Lukowicz [1] estimated the crowd density of the famous Munich Oktoberfest beer festival using Bluetooth scanning. Specifically, they designed two Bluetooth scan methodologies - stationary and dynamic. Similarly, Versichele et al. [11] also utilised a Bluetooth-based proximal tracking methodology. Their framework analyses the complex spatiotemporal dynamics such as visitor counts, returning visitors, and visitor flow maps in a very large crowd (around 1.5 million people).

Utilising the Bluetooth function in a smartphone provides several advantages such as ease of use, no need for pre-installed infrastructure, and less cost. But they also have numerous limitations. For instance, significant variations in the average number of people carrying a discoverable Bluetooth device can occur depending on the people who use the smart device, because some people may not use any extra features of the smart devices such as Bluetooth or Wi-Fi [1].

2.3 RF-based human crowd density estimation

Some recently proposed frameworks have analysed the property of radio signal strength changing in the crowded areas. For instance, Fadhlullah and Ismail [20] investigated the differences in signal attenuation between dynamic and static crowds towards identifying the significant crowd properties that affect wireless signal propagation with 2.4GHz ZigBee wireless nodes. Their proposed system was able to estimate with 75.00% and 70.83% accuracy the low and medium human crowd densities, respectively.

Based on the variance of Radio Signal Strength (RSS) values of the receiving signal, Ding et al. [18] estimated the people density using passive RFID tags. Accordingly, they deployed around 20 RFID passive tags in the target area and utilised the RFID reader to estimate the people density based on analysing the backscattered signal from the passive tags.

Yuan et al. [19] proposed an RSS device-free crowd counting approach, which exploits the space-time relativity of crowd distribution in order to minimise the estimation errors.

Further, Xu et al. [31] proposed SCPL, which uses RSS data to estimate people density in different environments. Particularly, they introduced a scheme that uses the profiling data collected by a single subject for tracking multiple subjects, which also reduced training overhead. Moreover, they designed a successive cancellation based algorithm to iteratively resolve the people count.

2.4 Wi-Fi signal-based people counting applications

The phenomenon of the large deployment of Wi-Fi access points has motivated researchers to investigate the use of RSSI to estimate people density. To enumerate, Yoshida and Taniguchi [32] estimated crowd density by applying a support vector regression-based approach to the RSSI data from the existing Wi-Fi access points; Depatla et al. [33] also utilised the RSSI from Wi-Fi access points. Specifically, the proposed simple motion model probabilistically characterise the influence of the total number of people based on analysing the scattering effects and multi-path fading.

Although the RSSI-based approaches are promising, they face performance challenges in the diverse indoor environment. Particularly, the challenges derive from the multi-path propagation of the received signal. Further, external factors such as humidity, water amount of the human body, or some other particles in the environment can influence the received signal and hence, makes the estimation process more complex.

Recently, Channel State Information (CSI) from OFDM-based systems has become popular in several passive-sensing application domains. CSI provides channel properties of a communication link from the physical layer that is more stable and more capable of providing fine-grained information in complex environments. Therefore, due to the robust nature in complex environments, researchers have applied CSI-based approaches in various application domains such as motion detection [24-25, 34-36], localisations [22-23], human identification [37] and crowd counting [28, 29].

Among the recent CSI-based approaches, WFID [37] is a passive device-free indoor human identification approach based on learning the feature pattern of subcarrier CSI matrix that changes due to the human body curve; WiFall [25], is a device-free automatic fall detection system which achieved 87% of fall detection precision. Moreover, the authors applied a weighted moving average to smooth the CSI matrix that is slightly influenced by environmental noise. Further, they classified the fall from other human activities based on using a one-class support vector machine algorithm. Electronic Frog Eye [28] is a CSI-based device-free crowd counting approach that presents a metric called Percentage of nonzero Elements (PEM), which characterises the relationship between the crowd density and deviation of CSI features.

Most of the frameworks discussed above require a training phase in order to predict the unknown activity. If the original environment has changed, the system needs to train again with the new conditions/settings. In contrast, Domenico et al. [29] introduced a crowd counting method that only requires one-time training for different environments. Moreover, the authors presented the use of a sum of the Davies-Bouldin indexes to select the most effective specific features related to CSI variations.

In contrast to previous work, our approach is based on the normalised amplitude measurements of the CSI in the received data. Moreover, to preserve the distinctive sensitivity

information of receiving streams, we divided the captured CSI datasets into samples and average the CSI amplitude within the sample for each stream and for each subcarrier separately. This process also helps to minimise the information loss and filter the noise due to sudden environmental changes.

3. TECHNICAL OVERVIEW

In this section, we provide a background on wireless signal propagation and Channel State Information, which form the theoretical basis of the proposed method.

3.1 Radio signal propagation in indoor environments

In conventional wireless communication systems, various factors such as scattering, reflection, or diffraction by the physical environment can influence the transmitted radio signal that reaches the receiver. Moreover, the transmitted signal reaches the receiving antenna by many paths (also known as multi-path propagation) due to some obstacles that block the Line-Of-Sight (LOS) path and also the reflections from the physical environment (see Figure 1). These multi-path components contain different time delay, amplitude attenuation, and phase shift information, which makes it possible to identify the various situations in the environment such as human presence, activity, gesture and so on.

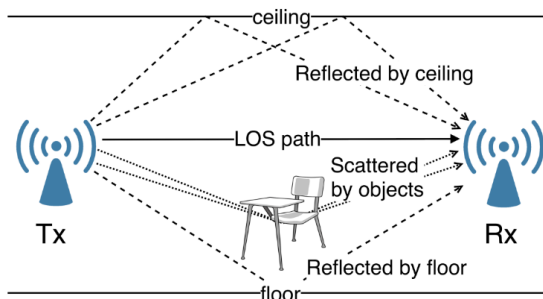


Figure 1. Radio signal propagation in indoor environment

Commonly, a typical Wi-Fi (IEEE 802.11a/g/n) system increases the data transfer capacity of the link by multiplexing data streams over the multiple transmit antennas to multiple receive antennas. Specifically, this type of arrangement also called as a Multiple-Input and Multiple-Output (MIMO) communication system, which typically contains M antennas at the transmitter and N antennas at the receiver, each receiver antenna receives the LOS signal and also a fraction of the signal from other propagation paths.

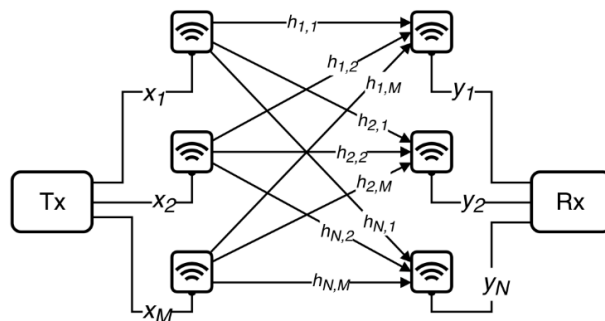


Figure 2. Example of a MIMO system

As shown in Figure 2, each receiver antenna receives not only the direct signal intended for it but also the signals from other propagation paths.

Based on the incoming signal that contains amplitude and phase, the receiver measures discrete Channel Frequency Response (CFR) in two domains: CFR of time and CFR of frequency. Afterwards, the system can generate Channel State Information (CSI) as a matrix with the dimension of $N \times M$.

3.2 Channel state information

CSI describes how the radio signal propagates from the transmitter to the receiver and discloses the channel properties of the wireless link with respect to the reflection, fading, and scattering. Commonly, Wi-Fi (IEEE 802.11a/g/n) systems use Orthogonal Frequency Division Multiplexing (OFDM) to divide the overall spectrum band into many small and partially overlapped frequency bands called subcarriers for high performance wireless communications [38]. In the OFDM systems, CSI contains the complex values of the CFR which represents the channel properties of each subcarrier.

In general, for the narrowband flat fading channel (a Wi-Fi channel in the 2.4 GHz/5GHz band) with MIMO, the CSI is represented in terms of the channel transmission matrix H by

$$Y = HX + N \quad (1)$$

where Y and X are the received and transmitted vector. N is the noise vector, which denotes the noise that corrupts the received data. Overall, the vectors can be expressed as follows:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} H = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,m} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n,1} & h_{n,2} & \cdots & h_{n,m} \end{bmatrix} \quad (2)$$

For example, for a 2×2 MIMO configuration, the received signal vector is expressed as

$$y_1 = h_{1,1}x_1 + h_{1,2}x_2 + n_1$$

$$y_2 = h_{2,1}x_1 + h_{2,2}x_2 + n_2$$

(3)

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

Moreover, the estimated CSI for a single sub-carrier can be represented as:

$$H_{n,i} = \| A_{n,i} \| e^{j\angle P_{n,i}} \quad (4)$$

where $\| A_{n,i} \|$ is the amplitude, $\angle P_{n,i}$ is the phase in which n is the subcarrier and i is the number of the received stream.

4. METHODOLOGY AND EXPERIMENTS

In this section, we describe the methodology of the proposed people density sensing system along with the experiments. As shown in Fig. 3, the system consists of three main stages: data collection, data processing, and classification.

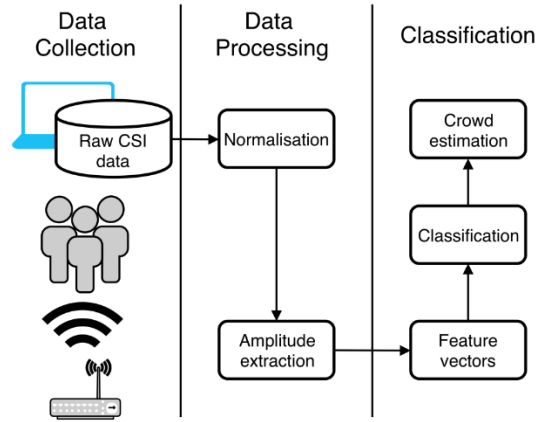


Figure 3. Architecture of the proposed system

4.1 Data collection

In the data collection phase, the main task is collecting data from different arrangements (empty room, a single person, two persons etc.) that can be used as the training dataset. We deployed TP-Link TL-R940N Wi-Fi router with three antennas as a transmitter and the receiver is an Intel Ultimate N WiFi Link 5300 IEEE 802.11a/b/g/n wireless network adapter with Linux 802.11n CSI Tool [21] installed on a laptop (HP Elitebook G3). The laptop sends ICMP packets to the Wi-Fi router in which the packet transmission rate is 100 packets per second and extracts the CSI information from the received ICMP echo reply packets.



Figure 4. Experiment area

The experiment area is a small classroom with the size of 5×7 meters and the distance between the transmitter and the receiver is approximately 3.3 meters (see Figure 4). In our experimental setup, we have used 3 transmit antennas (M) and 3 receive antennas (N), and so 9 RF streams are available for CSI processing. Moreover, the 802.11n CSI Tool can capture 30 OFDM subcarriers at the receiver end and therefore we have $3 \times 3 \times 30$ CSI data points (270) in each received packet for the processing. So, we can simply represent these CSI data points with relevance to the RF streams as follows.

$$\begin{aligned} CSI_1 &= \{ CSI_{1,1}, CSI_{1,2}, CSI_{1,3}, \dots, CSI_{1,30} \} \\ CSI_2 &= \{ CSI_{2,1}, CSI_{2,2}, CSI_{2,3}, \dots, CSI_{2,30} \} \\ CSI_9 &= \{ CSI_{9,1}, CSI_{9,2}, CSI_{9,3}, \dots, CSI_{9,30} \} \end{aligned} \quad (5)$$

where in $CSI_{i,n}$, i and n is the RF stream number and the OFDM subcarrier number respectively.

We collected data for one minute in each situation with the rate of 100Hz, which indicates that we acquired over 5000 packets to extract the unique features from each scenario. During the data collection, firstly, we collected the first CSI data set of the empty room. Thereafter, we asked participants to come into the room one by one. When the first object is in the room, we recorded CSI data for one minute, and then ask the second object to come in, and recorded CSI in a minute, and so on. Finally, with this manner, we have collected CSI data sets for 16 participants in total.

4.2 Data processing

In the data processing stage, the first task is to clean the raw CSI data. In general, wireless signals are influenced by three main factors—path loss, multi-path fading and shadowing—due to various environmental factors such as humidity, other wireless devices, furniture, etc. Therefore, as the first step, the system has to normalise the recorded raw CSI data as mentioned in [39] and also design an exponential filter to remove the noise.

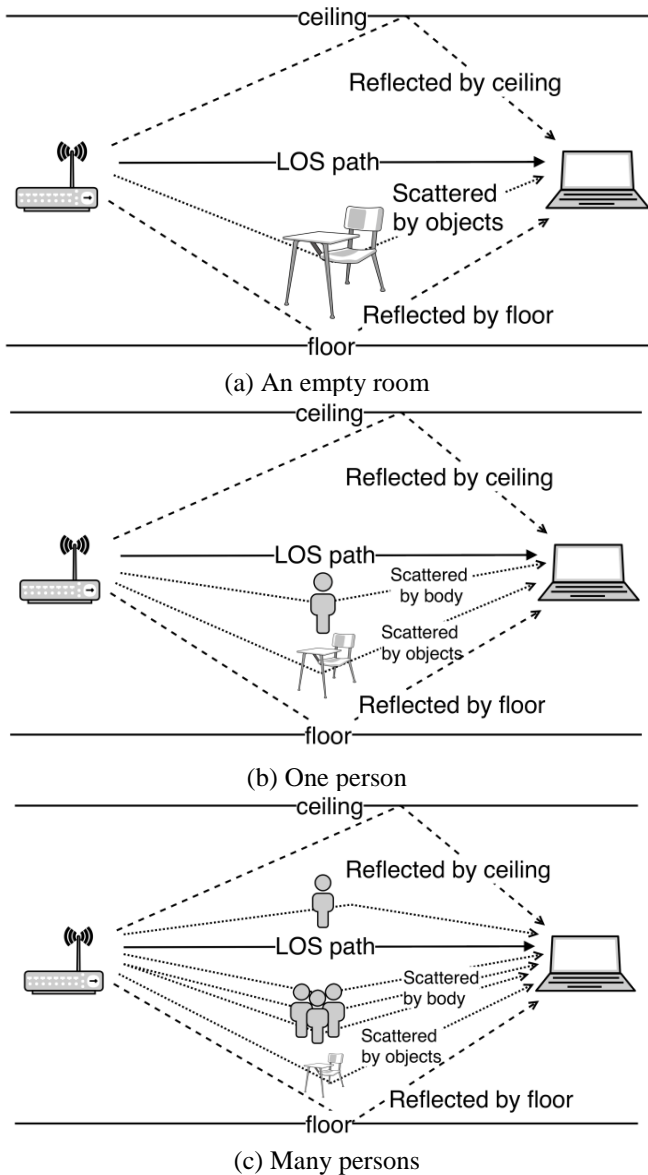
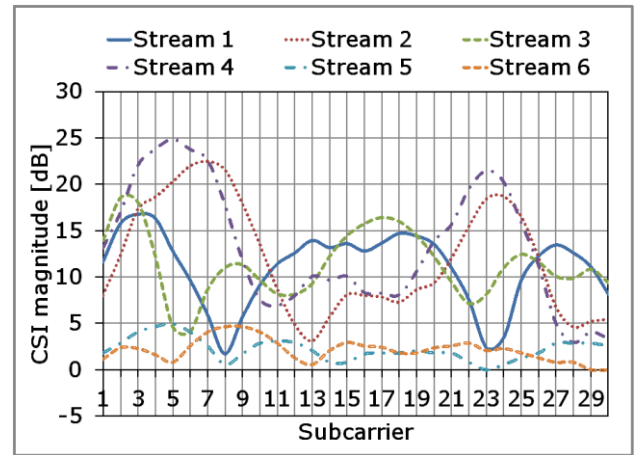


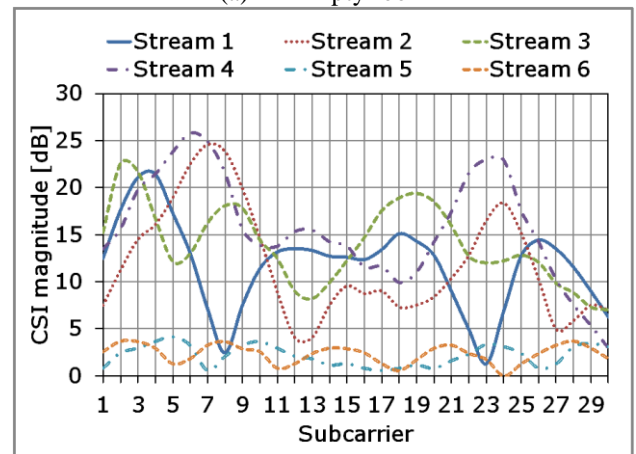
Figure 5. Multi-paths and scattering caused by humans

After the normalisation, the CSI data moves to the feature extraction step. The main idea of feature extraction step is extracting as much information as possible towards showing how the corresponding CSI reflect the number of people in the environment.

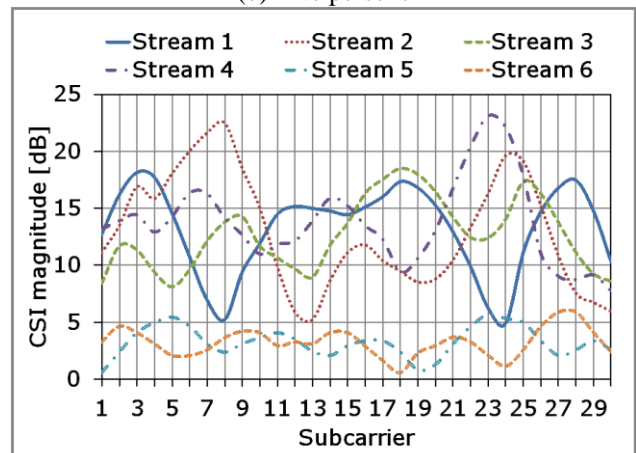
During the experiment, we have noticed that in some time slots, although there are three antennas at the transmitter, the receiver receives data only from the two transmit antennas. After investigation, we have realised that this is due to the behaviour of the rate adaptation algorithm at the Wi-Fi router. Because of this reason, we have some datasets with nine streams (3 Tx and 3 Rx) and some data sets with six streams (2 Tx and 3 Rx). This is a kind of technical issue which is beyond our control and therefore, we decided to continue the data processing with six streams.



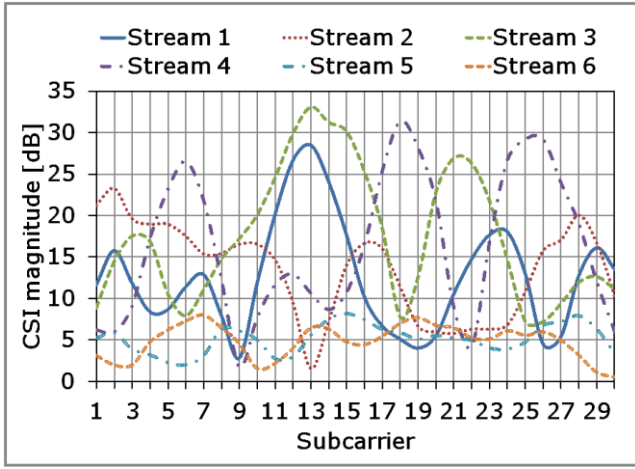
(a) An Empty room



(b) Five persons



(c) Ten persons



(d) Fifteen persons

Figure 6. Extracted CSI streams for different crowd densities

When there are more people in the room, the more scattering of signals from the human bodies occur [40]. For example, when there is no one in the room, the distortion of the signal due to the multi-path propagation will be stable. On the other hand, as shown in Figure 5, when the number of people increases, the multi-path- induced fading will increase too.

With the captured CSI of the empty room, five persons, ten persons and fifteen persons, Figure 6 clearly shows that when there are more persons in the environment, the greater is the CSI variation. Moreover, we have observed that the presence of the humans affects all streams independently and differently. For instance, if we consider the amplitude levels of all streams, in Figure 6a (empty room), stream 2 has a higher amplitude level in the subcarriers 6-8 and 23,24, and the stream 4 has a higher amplitude level in the subcarriers 4-6 and 22-24. When there are five persons (see Figure 6b), stream 2 has a higher amplitude level in the subcarriers 7,8, and the stream 4 has a higher amplitude level in the subcarriers 5-7 and 23,24. Also the stream 3 has a higher amplitude level in the streams 2,3 and 18-20. However, when there are fifteen persons, we can see the significant affection on the stream 2 and 3 compared to other density levels. As shown in Figure 6d, the amplitude level of the stream 2 has reduced and the stream 3 has higher amplitude levels, which is a significant difference in the density of fifteen people.

4.3 Classification

The next step is classifying the captured CSI amplitude series to estimate the number of people in the room. Preparing collected datasets for the classification is done as follows. As mentioned earlier, we have observed that each antenna has a unique CSI feature in particular subcarriers. As mentioned in the literatures [25, 30] taking the average CSI across all subcarriers will lose some important information. Instead, in this experiment, taking the average CSI of each antenna in each subcarrier for 100ms (for 10 received packets) is more informative and also tolerates the sudden changes of environmental factors that may affect only a part of the dataset.

For each situation (empty room, a single person, 2 persons, etc.), for the classification task, there should be a separate dataset which can provide as many significant features from the CSI for the training and testing purposes. The preparation of the dataset is as follows.

As the first step, 500 packets of the raw dataset from the empty room was divided into fifty samples in which each sample contains ten packets. Each packet has a CSI matrix size of 180 data points (3 Rx antennas, 2 Tx antennas, 30 subcarriers), and with all ten packets, we have a total of 1800 data points. After that, we create a list of CSI amplitudes of the first stream of the first subcarrier from all ten packets. Moreover, we remove the sudden noises by applying the exponential filter on the values of the list, and then take the average amplitude value from the list. This value refers to the average amplitude value of the first stream of the subcarrier one. In this manner, we calculate the list of average amplitude values for each stream for each subcarrier for one sample. According to the experimental setup, we have fifty samples per scenario.

We follow the same method to prepare datasets for one person, two persons, and so on, up to sixteen persons. Thereafter, from each dataset, 60% of records are used to train the classifier and 40% of records use as a test set.

4.4 Density estimation

In order to estimate the people density from the prepared dataset, the first part is training the machine learning algorithm with training data. In this experiment for each scenario, we have a dataset which contains 850 records and we consider 510 records as the training set and 340 records as the testing set.

Table 1 shows the sample of the training dataset. As mentioned earlier, we have 180 data points per record and we need 180 columns to keep those observations. The first column is a categorical variable called ‘‘Population’’, which is used as a label for each density class. In this training dataset, we have 17 density classes as ‘‘Empty’’ representing no person, ‘‘One’’ representing one person, ‘‘Two’’ representing two persons, etc., up to ‘‘Sixteen’’ for sixteen persons. Also, each density class contains 30 records. These density classes are considered as the response classes by the Classification Learner. Rests of the columns contain the values of each antenna regard to each subcarrier. Since there are six antennas and thirty subcarriers, there are 180 columns to keep data from each antenna and each subcarrier. Therefore, the second column is labelled as ‘‘S1A1’’ that denotes subcarrier-1 antenna-1, the third column is labelled as ‘‘S1A2’’ that denotes subcarrier-1 antenna-2, and so on until the last column ‘‘S30A6’’, subcarrier-30 antenna-6.

Data classification is done with a MATLAB Classification Learner tool by applying three different classification models as: a) Complex Tree b) Support vector machine c) Weighted K-Nearest Neighbour. Moreover, the training dataset consists of 180 predictors (variables), 17 response classes, and 510 observations.

Table 2 shows a detailed comparison of classification models with their corresponding configuration parameters. With the Complex Tree classifier, the training model achieves the maximum of 94.3% accuracy with the configuration: maximum number of splits as 100 and the split criterion as the Twoing rule. These are the best configuration parameters for this model. The second model is trained with the Medium Gaussian SVM classifier with the parameters of Gaussian as the Kernel function, 13 as the Kernel scale, and 1 as the Box constraint level. This classifier achieves the maximum of 99.4% accuracy. The best classifier is the Weighted K-NN which achieves 99.8% accuracy. The configuration parameters of the model are: 10 neighbours, Euclidean distant metric, squared in-verse distant weight, and standardise data distance.

Table 1. Sample of the training dataset

Population	S1A1	S1A2	S1A6	S2A1	S2A6	S30A5	S30A6
Empty	8.791110648	2.817050387	7.006465943	0.66009879	5.201691782	15.01303821	5.305919547
Empty	8.761823618	2.834738854	7.295316224	0.154014817	5.104513236	16.24097758	5.727554083
Five	13.15708808	8.332728406	2.510295517	18.73988295	3.54976989	3.351741948	2.00481519
Ten	14.21931672	12.09246283	3.448009984	18.37839203	5.157626068	3.084589767	2.654696582
Sixteen	17.55739829	17.81347935	5.556473994	22.36546911	6.130249828	3.214668408	1.982171616

Table 2. Comparison of classification models

Classification algorithm	Configuration	Accuracy
Complex Tree	Maximum amount of split: 100, Split criterion: Twoing rule	94.3%
Medium Gaussian SVM	Kernel function: Gaussian Kernel scale: 13 Box constraint level: 1	99.4%
Weighted K-NN	Number of neighbours: 10 Distance Metric: Euclidean Distance weight: Squared inverse Standardised data: true	99.8%

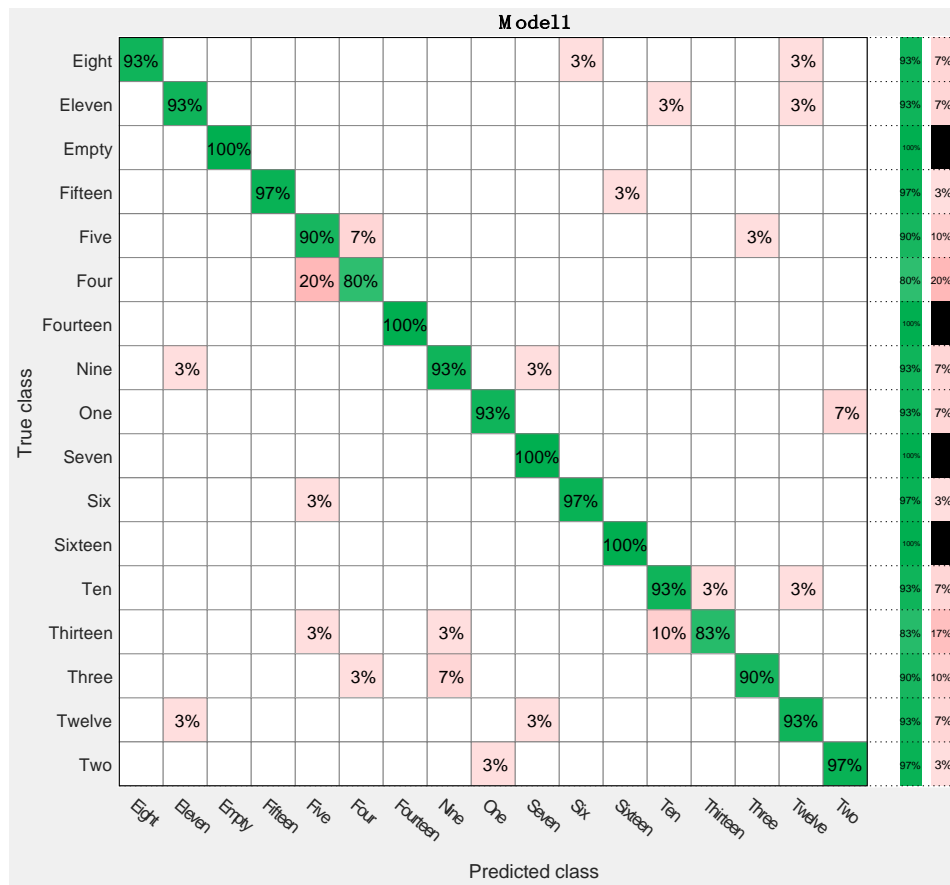


Figure 7. Confusion matrix - complex tree

Figure 7 shows the confusion matrix of the Complex Tree classifier. The columns show the predicted classes and the rows show the true classes. In the top row, 93% of the records from “Eight” density class are correctly classified and other records in the “Eight” rows are misclassified. For example, 3% of the records are incorrectly classified as from the density class “Six”, and 3% are classified as from “Twelve”. So, 93%

is the true positive rate and 7% is the false negative rate for this density class that is shown in the last green and red columns. Moreover, the Complex Tree classifier can classify “Empty”, “Fourteen”, “Seven” and “Sixteen” density classes with 100% of accuracy and the class “Four” with the least accuracy (true positive rate 80%, false negative rate 20%).

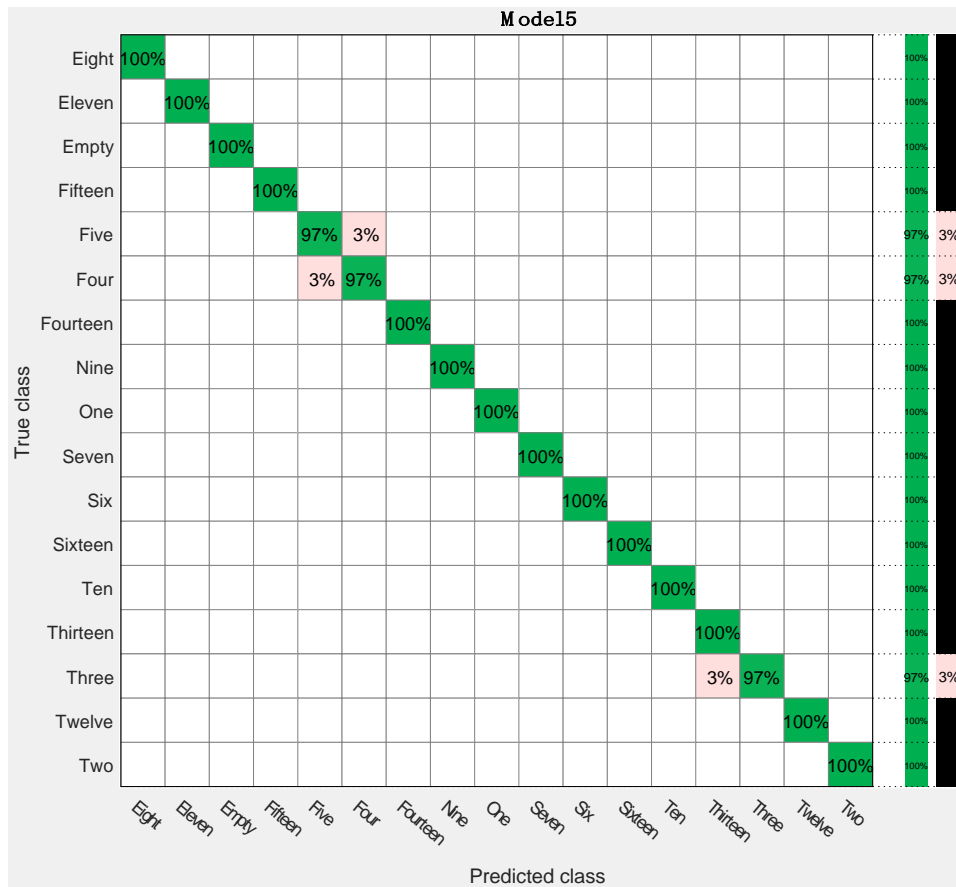


Figure 8. Confusion matrix - medium gaussian SVM

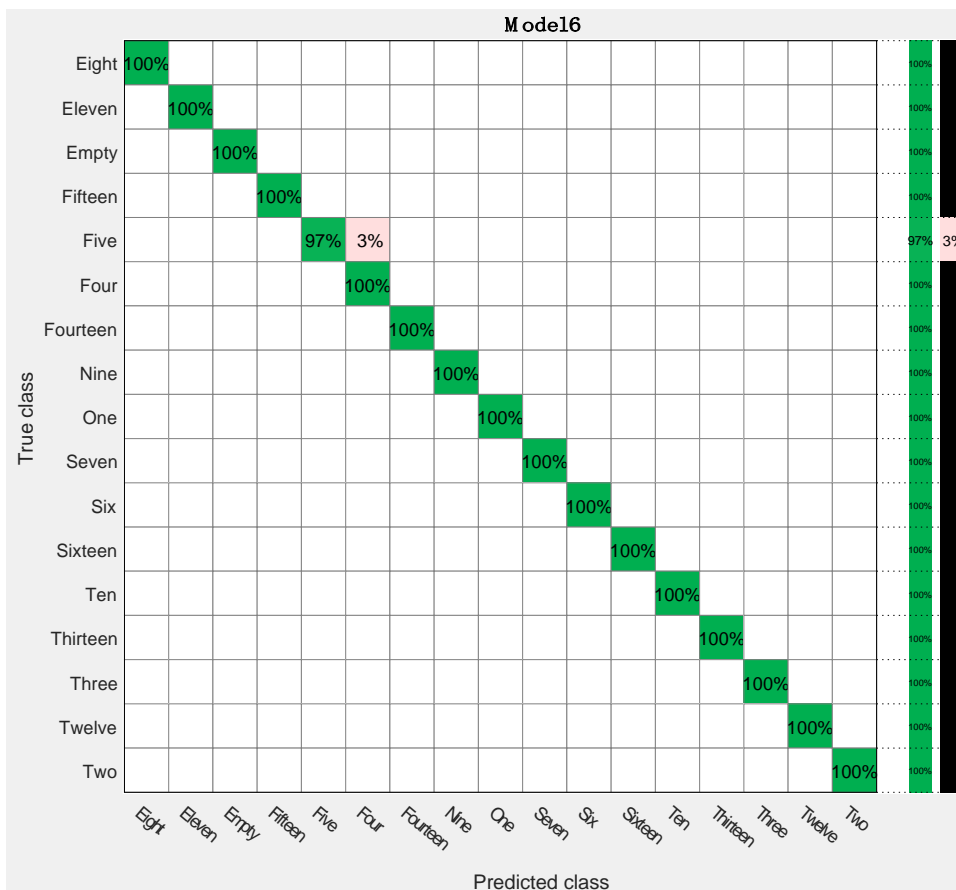


Figure 9. Confusion matrix - weighted KNN

Figure 8 shows the confusion matrix of the Medium Gaussian SVM classifier. This classifier can correctly classify most of the density classes with 100% true positive rate except for classes “Five”, “Four” and “Three” which are classified with 97% true positive rate and 3% false negative rate.

The confusion matrix in figure 9 represents the Weighted K-NN classifier. Here, almost all the density classes are classified with 100% true positive rate except the density class “Five”, which has 97% true positive rate and 3% false negative rate.

We also noticed that a certain amount of records from the density class “Five” is misclassified in all classifiers due to some reason. This may be due to a kind of radio signal interference affected to the received signal amplitude that suddenly occurred at the data collection.

According to the performance testing, the Weighted K-NN classification model shows the best accuracy (99.8%) with the training dataset. Therefore, we have selected this model for the future crowd density estimations.

5. CONCLUSION

In this paper, we proposed a CSI-based device free framework for crowd estimation. Instead of averaging the CSI information into a smaller dimension, we proposed to extract the CSI amplitude of each stream (antenna) for a particular subcarrier separately, which is able to preserve more CSI information. We evaluated the proposed method in the real environment and the experimental results show that it has 99% accuracy with the Weighted KNN classification model. However, the proposed model assumes that there are no major changes in the environment during the training and the testing phase.

As for the future research, we would like to investigate a predictive model that including the following features.

- To recognise a situation if all people in the room are sitting or standing.
- Predict actual people count in the different environments. For example, estimate the people count in a much bigger or much smaller room or in the entire floor of a building.
- In the current implementation, we filtered out some high-frequency components that caused by the movement of people. We would like to add a feature that can recognise if people in the room are moving or stationary.

We also upload the raw datasets, training and testing datasets, and Matlab codes that used in this paper into a GitHub repository (see <https://github.com/ldmohan/People-Density-Sensing-using-CSI>) for the readers’ further reference.

ACKNOWLEDGMENT

This research is supported by Estonian Research Council grant IUT20-55. Special thanks to Amnir Hadachi for his contribution on collecting datasets.

REFERENCES

[1] Weppner J, Lukowicz P. (2013). Bluetooth based collaborative crowd density estimation with mobile phones. In *Pervasive Computing and Communications (PerCom)*, 2013 IEEE International Conference on, pp. 193-200.

<http://dx.doi.org/10.1109/PerCom.2013.6526732>

[2] Li M, Zhang Z, Huang K, Tan T. (2008). Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4. <https://doi.org/10.1109/ICPR.2008.4761705>

[3] Chan AB, Liang ZSJ, Vasconcelos N. (2008). Privacy preserving crowd monitoring: Counting people without people models or tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–7. <https://doi.org/10.1109/CVPR.2008.4587569>

[4] Idrees H, Saleemi I, Seibert C, Shah M. (2013). Multi-source multi-scale counting in extremely dense crowd images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2547–2554. <https://doi.org/10.1109/CVPR.2013.329>

[5] Subburaman VB, Descamps A, Carincotte C. (2012). Counting people in the crowd using a generic head detector. In *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pp. 470–475. <https://doi.org/10.1109/AVSS.2012.87>

[6] Chang C, Srirama SN, Mass J. (2015). A middleware for discovering proximity-based service-oriented industrial internet of things. In *Proceedings of the 2015 IEEE International Conference on Services Computing (SCC)*, pp. 130–137. <https://doi.org/10.1145/3012000>

[7] Liyanage M, Chang C, Srirama SN. (2016). Energy-efficient mobile data acquisition using opportunistic internet of things gateway services. In *Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2016 IEEE International Conference on*, pp. 217–222. <https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2016.60>

[8] Chang C, Srirama SN, Liyanage M. (2015). A service-oriented mobile cloud middleware framework for provisioning mobile sensing as a service. In *Parallel and Distributed Systems (ICPADS 2015), 21st IEEE International Conference on*. <https://doi.org/10.1109/ICPADS.2015.24>

[9] Liyanage M, Chang C, Srirama SN. (2015). Lightweight mobile web service provisioning for sensor mediation. In *Mobile Services (MS), 2015 IEEE International Conference on*, pp. 57–64. <https://doi.org/10.1109/MobServ.2015.18>

[10] Liyanage M, Chang C, Srirama SN. (2016). MePaaS: mobile-embedded platform as a service for distributing fog computing to edge nodes. In *17th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT-16)*. <https://doi.org/10.1109/PDCAT.2016.030>

[11] Versichele M, Neutens T, Delafontaine M, de Weghe N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Appl. Geogr.* 32(2): 208–220. <https://doi.org/10.1016/j.apgeog.2011.05.011>

[12] Kannan PG, Venkatagiri SP, Chan MC, Ananda AL, Peh LS. (2012). Low cost crowd counting using audio tones.

- In Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, pp. 155–168. <https://doi.org/10.1145/2426656.2426673>
- [13] Yuan Y, Qiu C, Xi W, Zhao J. (2011). Crowd density estimation using wireless sensor networks. In Mobile Ad-hoc and Sensor Networks (MSN), 2011 Seventh International Conference on, pp. 138–145. <https://doi.org/10.1109/MSN.2011.31>
- [14] Moussa M, Youssef M. (2009). Smart cevicees for smart environments: Device-free passive detection in real environments. In Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on, pp. 1–6. <https://doi.org/10.1109/PERCOM.2009.4912826>
- [15] Pu Q, Gupta S, Gollakota S, Patel S. (2013). Whole-home gesture recognition using wireless signals. In Proceedings of the 19th annual international conference on Mobile computing & networking, pp. 27–38. <https://doi.org/10.1145/2500423.2500436>
- [16] Nakatsuka M, Iwatani H, Katto J. (2008). A study on passive crowd density estimation using wireless sensors. In The 4th Intl. Conf. on Mobile Computing and Ubiquitous Networking (ICMU 2008).
- [17] Schauer L, Werner M, Marcus P. (2014). Estimating crowd densities and pedestrian flows using wi-fi and bluetooth. In Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp. 171–177. <https://doi.org/10.4108/icst.mobiquitous.2014.257870>
- [18] Ding H, Han J, Liu AX, Zhao J, Yang P, Xi W, Jiang Z. (2015). Human object estimation via backscattered radio frequency signal. In Computer Communications (INFOCOM), 2015 IEEE Conference on, pp. 1652–1660. <https://doi.org/10.1109/INFOCOM.2015.7218545>
- [19] Yuan Y, Zhao J, Qiu C, Xi W. (2013). Estimating crowd density in an RF-based dynamic environment. IEEE Sens. J. 13(10): 3837–3845. <https://doi.org/10.1109/JSEN.2013.2259692>
- [20] Fadhlullah SY, Ismail W. (2016). A statistical approach in designing an rf-based human crowd density estimation system. Int. J. Distrib. Sens. Networks. <https://doi.org/10.1155/2016/8351017>
- [21] Halperin D, Hu W, Sheth A, Wetherall D. (2011). Tool release: gathering 802.11n traces with channel state information. ACM SIGCOMM CCR 41(1): 53. <https://doi.org/10.1145/1925861.1925870>
- [22] Li X, Li S, Zhang D, Xiong J, Wang Y, Mei H. (2016). Dynamic-music: accurate device-free indoor localization. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 196–207. <https://doi.org/10.1145/2971648.2971665>
- [23] Gao L. (2015). Channel State Information Fingerprinting Based Indoor Localization: a Deep Learning Approach. Auburn University. <http://dx.doi.org/10.1109/TVT.2016.2545523>
- [24] Li F, Al-qaness MAA, Zhang Y, Zhao B, Luan X. (2016). A robust and device-free system for the recognition and classification of elderly activities. Sensors 16(12): 2043. <https://doi.org/10.3390/s16122043>
- [25] Wang Y, Wu K, Ni LM. (2017). Wifall: Device-free fall detection by wireless networks. IEEE Trans. Mob. Comput. 16(2): 581–594. <https://doi.org/10.1109/TMC.2016.2557792>
- [26] Sun L, Sen S, Koutsonikolas D, Kim KH. (2015) Widraw: Enabling hands-free drawing in the air on commodity wifi devices. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp. 77–89. <https://doi.org/10.1145/2789168.2790129>
- [27] Jayanthi K, Dananjayan P. (2006). Prediction based smart scheduler using neural network for wireless networks. Adv. Model. Anal. B 49(3-4): 13–24.
- [28] Xi W, Zhao J, Li XY, Zhao K, Tang S, Liu X, Jiang Z. (2014). Electronic frog eye: Counting crowd using wifi. In Infocom, 2014 proceedings ieee, pp. 361–369. <http://dx.doi.org/10.1109/INFOCOM.2014.6847958>
- [29] Di Domenico S, De Sanctis M, Cianca E, Bianchi G. (2016). A trained-once crowd counting method using differential wifi channel state information. In Proceedings of the 3rd International on Workshop on Physical Analytics, pp. 37–42. <https://doi.org/10.1145/2935651.2935657>
- [30] Palipana S, Agrawal P, Pesch D. (2016). Channel state information based human presence detection using non-linear techniques. In Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, pp. 177–186. <https://doi.org/10.1145/2993422.2993579>
- [31] Xu C, Firmer B, Moore RS, Zhang Y, Trappe W, Howard R, Zhang F, An N. (2013). Scpl: Indoor device-free multi-subject counting and localization using radio signal strength. In Information Processing in Sensor Networks (IPSN), 2013 ACM/IEEE International Conference on, pp. 79–90. <https://doi.org/10.1145/2461381.2461394>
- [32] Yoshida T, Taniguchi Y. (2015). Estimating the number of people using existing wifi access point in indoor environment. In Proceedings of the 6th European Conference of Computer Science (ECCS'15), pp. 46–53.
- [33] Depatla S, Muralidharan A, Mostofi Y. (2015). Occupancy estimation using only wifi power measurements. IEEE J. Sel. Areas Commun. 33(7): 1381–1393. <https://doi.org/10.1109/JSAC.2015.2430272>
- [34] Gong L, Yang W, Man D, Dong G, Yu M, Lv J. (2015). WiFi-based real-time calibration-free passive human motion detection. Sensors 15(12): 32213–32229. <https://doi.org/10.3390/s151229896>
- [35] Liu J, Wang Y, Chen Y, Yang J, Chen X, Cheng J. (2015). Tracking vital signs during sleep leveraging off-the-shelf wifi. In Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 267–276. <https://doi.org/10.1145/2746285.2746303>
- [36] Xiao J, Wu K, Yi Y, Wang L, Ni LM. (2012). Fimd: Fine-grained device-free motion detection. In Parallel and Distributed Systems (ICPADS), 2012 IEEE 18th International Conference on, pp. 229–235. <https://doi.org/10.1109/ICPADS.2012.40>
- [37] Hong F, Wang X, Yang Y, Zong Y, Zhang Y, Guo Z. (2016). WFID: Passive device-free human identification using wifi signal. In Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp. 47–56. <https://doi.org/10.1145/2994374.2994377>
- [38] Malathi P, Vanathi PT. (2008). Orthogonal Frequency Division Multiplexing (OFDM) for wireless Local Area Network (WLAN) systems. Adv. Model. Anal. B 51(1-

- 2): 1–16.
- [39] Halperin. (2011). Linux 802.11n CSI Tool-FAQ, Things to Know, and Troubleshooting.
- [40] Wang W, Liu AX, Shahzad M, Ling K, Lu S. (2015). Understanding and modeling of wifi signal based human activity recognition. In Proceedings of the 21st annual international conference on mobile computing and networking, pp. 65–76. <https://doi.org/10.1145/2789168.2790093>