Counting Human Objects Using Backscattered Radio Frequency Signals

Han Ding, *Member, IEEE*, Jinsong Han,
Alex X. Liu, *Senior Member, IEEE*, Wei Xi, Jizhong Zhao,
Panlong Yang, and Zhiping Jiang, *Member, IEEE*

Abstract—In this paper, we propose a system called R# to estimate the number of human objects using passive RFID tags but without attaching anything to human objects. The idea is based on our observation that the more human objects are present, the higher the variation in the RSS values of the tag backscattered RF signals. Thus, based on the received RF signals, the reader can estimate the number of human objects. R# includes an RFID reader and some (say 20) passive tags, which are deployed in the region that we want to monitor the number of human objects, such as the region in front of a supermarket shelf. The RFID reader periodically emits RF signals to identify all tags and the tags simply respond with their IDs via EPCglobal Class 1 Generation 2 protocol. We implemented R# using commercial Impinj H47 passive RFID tags and Impinj reader model R420. We conducted experiments in a simulated picking aisle area of the supermarket environment. The experimental results show that R# can achieve high estimation accuracy (more than 90 percent with up to ten human objects).

Index Terms—RFID, crowd counting

1 Introduction

Estimating the number of human objects at a certain location is the basis for many applications such as quantifying the popularity of certain items among customers in a supermarket, correlating the items that customers visit and those that customers purchase, and finding crowded locations in a museum. Currently human object estimation uses mechanical barrier, binary sensor, imager, or pressure/vibration based technologies. The mechanical barrier based technology uses a turnstile or baffle gate to construct a one-way gate so that at any time there is only one person can pass through; thus, the number of human objects passing through the gate can be mechanically counted. The binary sensor based technology uses a break-beam (such as the infrared, laser, and ultrasound) sensor at a one-way gate so that each time a human object passing through the gate can be detected as the beam is blocked [1]. Binary sensors and mechanical barriers are often used together and are typically deployed at the entrance and exit of a building. The key limitation of both technologies is that it requires human objects to pass through a physical gate and therefore cannot be used to count free moving human objects, such as those moving around on a floor of painting displays. The imager based technology uses cameras or thermal imagers to first capture images/videos and then uses pattern recognition techniques to identify the number of human objects in the images/videos. The key limitations of this technology are that cameras require good lighting conditions (e.g., cannot operate in the dark) and the thermal imagers are too expensive (although it can operate in the dark); furthermore, cameras and thermal images often are deployed with fixed orientation and thus limiting the human object detection to a specific region. The pressure/vibration based technology embeds pressure or vibration sensors on the floor to detect human objects [2]. The key limitations of this technology lies in the high deployment cost and the interference among multiple people.

In this paper, we propose a system called R# to estimate the number of human objects using passive Radio Frequency Identification (RFID) tags but without attaching anything to human objects. The idea is based on our observation that *with different number of human objects in presence, the RF signals that the tags backscatter to the reader demonstrates different patterns*; thus, based on the patterns of the received RF signals, the reader can estimate the number of human objects. R# includes an RFID reader and some (say 20) passive tags, which are deployed in the region that we want to monitor the number of human objects, such as the region in front of a supermarket shelf. The RFID reader periodically emits RF signals to

- *H. Ding is with the School of Electronic and Information Engineering, Xi'an Jiaotong University, China, and State Key Laboratory for Novel Software Technology, Nanjing University, P.R. China. E-mail: dinghanxjtu@gmail.com.*
- *J. Han is with the Shaanxi Province Key Laboratory of Computer Network, School of Electronic and Information Engineering, Xi'an Jiaotong University, China. E-mail: hanjinsong@xjtu.edu.cn.*
- *W. Xi and J. Zhao are with the School of Electronic and Information Engineering, Xi'an Jiaotong University, China. E-mail: {xiwei, zjz}@xjtu.edu.cn.*
- *A. X. Liu is with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824. E-mail: alexliuxy@yahoo.com.*
- *P. Yang is with the School of Computer Science and Technology, University of Science and Technology of China, China. E-mail: panlongyang@gmail.com.*
- *Z. Jiang is with the School of Software Engineering, Xi'dian University, China. E-mail: jiangzp.cs@gmail.com.*

Manuscript received 24 Mar. 2017; revised 6 May 2018; accepted 18 June 2018. Date of publication 3 July 2018; date of current version 1 Apr. 2019. (Corresponding authors: Jinsong Han and Alex X. Liu.) For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TMC.2018.2852627

identify all tags and the tags simply respond with their IDs via protocols in the EPCGlobal Class-1 Generation-2 (C1G2) RFID standard [3].

Our R# system has two key features that make it easy to deploy. First, R# does not attach any device to human objects. Second, R# uses off-the-shelf commercial RFID readers and passive tags, which have already been widely deployed at places such as supermarkets.

There are three key challenges to correlate between the number of human objects and the patterns of backscattered RF signals. First, the information that we can extract from RF signals is limited. From commodity RFID readers, other than tag IDs, we can only retrieve three types of information: Phase, Doppler shift, and Radio Signal Strength (RSS). The phase value changes periodically and Doppler shift value is too noisy. Our initial attempts suggested that they are not suitable for our human object counting purpose. Thus, we have to resort to RSS information. Second, the RSS values of the same tag at different locations vary significantly (i.e., location diversity), and the RSS values of different tags at the same location also vary significantly (i.e., tag diversity). Third, the RF signal pattern is difficult to model and quantify.

Our R# system addresses these challenges based on our refined observation that *the more human objects are present, the larger the variation in the RSS values of the tag backscattered RF signals*. To correlate the number of human objects and the RSS value variations, to count a maximum of k human objects, we use machine learning techniques to build $k+1$ classifiers corresponding to $0, 1, \cdots, k$ human objects, respectively. Given a test case, we extract features and then use them to classify the case into one of the $k+1$ classes. Our classifier is based on the following three features extracted from the RSS values of the multiple tags during a certain time period. The first feature is the entropy of observed RSS values. The intuition is that the more human objects are present, the more random the distribution of RSS values reported from tags is, and thus the higher the entropy of the RSS values is. The second feature, extracted using image processing techniques, is the area size of the connected white pixels dilated from the points correlated to the RSS values. The intuition is that the more human objects are present, the higher the dispersion of the RSS values, simultaneously enlarging the area size of the connected white pixels after dilation. The third feature is the mean squared error (MSE) between the deflated image and original image. The intuition is that the more human objects are present, the higher the loss of fidelity of the image, resulting in a larger MSE.

We implemented R# using commercial Impinj H47 passive RFID tags and Impinj reader model R420. We conducted experiments in a simulated picking aisle area of the supermarket environment. Ten volunteers participated in the experiments as shopping customers. The experimental results show that R# can achieve high estimation accuracy. For example, with 0~10 human objects in the monitoring area, in nearly 93 percent of our tests, the estimated value of R# exactly matches the real value; and in nearly 98 percent of our tests, the estimated value of R# deviates from the real value by at most one person.

2 Related Work

In this section, we briefly review the related literature in human counting.

Mechanical Barrier/Binary Sensor-Based. Traditionally, people count or estimate individuals present in the scene via mechanical barriers. Later, the binary sensor use breakbeams, such as the infrared, laser, and ultrasound, to detect the number of human objects passing through a gate or specific line [1]. Deploying pressure/vibration sensors on the floor is also able to detect and count human objects, but suffering from high deployment cost [2].

Camera-Based. In recent decades, crowd counting or estimation has been widely studied in the computer vision literature [4], [5]. These approaches usually leverage the pattern recognition technique to detect the individuals in presence, based on either their shapes or motions. There are many challenges when using computer vision based methods for human object estimation. For example, cameras require line-of-sight, which means that blocked areas cannot be monitored. In addtion, environmental factors, such as smoking and dim lighting conditions, will severely degrade the image quality, and further decrease the estimation performance. Last but not the least, the computational complexity is another issue. Image based approaches usually depend on complicated algorithms, suffering from long system latency.

RF-Based. Besides utilizing the image or video, wireless signal is also used for crowd counting and estimation. The signal used for this purpose includes Wi-Fi [6], [7], [8], backscattered RF signals [9], LTE [10], Bluetooth [11], FM [12], and audio [13], [14], [15], *etc*. The recent efforts along this trend fall into two categories: device-based approaches and device-free approaches. Device-based approaches [11], [13] estimate the crowd density by counting the number of devices, e.g. mobile phones [11] or RFID tags [16], carried by the individuals. On the other hand, wireless signal is susceptible to environment variations or object movements [17], [18], [19], [20], [21], providing a potential way to device-free crowd estimation. Yuan et al. [22] deploy a 4 × 4 TelosB node grid and estimate the crowd density (0~3 persons) in each grid (3.6 m × 3.6 m) by utilizing k-means algorithm on RSS measurements. SCPL [23] analyzes the RSS mean difference derived from the wireless links and proposes a successive cancellation based algorithm to iteratively determine the number of multiple human objects. The authors deploy 12 transmitters and 8 receivers in a 400 m^2 open floor and can count up to four coexisting subjects. FreeCount [7] estimates the crowd density (up to 7 persons) in an area of 4.5 m × 6.5 m using WiFi CSI data. The authors adopt an information theory based scheme to select and extract features from CSI that are sensitive to human motions, then build a classifier to estimate the crowd. Domenico et al. [10] investigate the possibility to use LTE signal for crowd density estimation. Their approach exploits the correlation between the variations of received LTE signals and the number of people (up to 5 persons). To our knowledge, using the passive tag for human object estimation has been scarcely seen in the literature, although the passive tag has shown its effectiveness in localization and motion detection [24], [25], [26], [27], [28], [29]. In addition, our system follows the extensible RFID model [28], [29], where the RFID reader used for R# can be reused for other applications such as identification of items, tracking of objects, etc. It is because R# only uses a limited capacity of a reader, hence it can easily be deployed with existing RFID system and further reduces hardware cost and deployment difficulty.

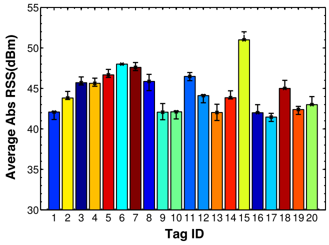


Fig. 1. The RSS collected from different tags at a given location.

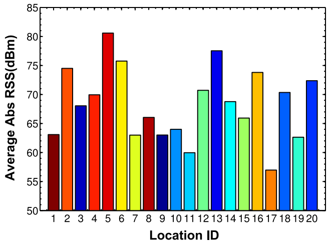


Fig. 2. The RSS collected from a tag at 20 different locations.

3 Observation and Analysis

In this section, we first introduce the technical background of RFID techniques. Then we report our observation and analysis of the empirical studies for estimating the number of human objects.

3.1 Backscatter Communication and RF Signal

Passive RFID tags utilize the backscatter communication to interact with readers. Other than tag IDs, the backscattered RF signals also contain certain features about the path along which it travels. Leveraging those features, we are able to detect the ambient change in the region that the passive RFID system is deployed.

Backscattering communication in passive RFID systems requires the reader to emit Continuous Waves (CW) for interrogating tags. There are two opposite directional links in the backscattering communication. The one associated with the reader-to-tag communication is the *forward link*, in which the reader modulates its messages, such as the Query, ACK, or other commands, into the CW. The other one associated with the tag-to-reader communication is the *reverse link*. Due to the cost concern, the passive RFID tag does not have a radio transmitter. Instead, the tag modulates its information, e.g., Response and ID, into the backscattered CW waves by changing the impedance of its antennas. The reader receives the tag reply and can report its ID, Received Signal Strength, and Phase.

3.2 Observation

Our goal is to correlate the number of human objects to the features of RF signals in the region that we want to monitor (region for short in the following). Compared to the fine-grained information, such as the Channel State Information (CSI) used in Wi-Fi [6], the information reported by RFID readers is in low resolution and noisy. On the other hand, the information itself is inconsistent from the following aspects. (1) *Tag diversity*. Due to the imperfection of manufacture, different tags (even from the same model) have diverse noise levels. We conduct a number of experiments in a static environment. We put 20 tags in a fixed location, and collect their RSS values for $30s$. The orientation and the distance from the reader to tag are fixed. As shown in Fig. 1, the color bars represent the absolute values of RSS from 20 tags, and the black lines mark the maximum and minimum values. We observe that although each tag RSS value remains relatively stable, the measured values in different tags are diversely distributed from 41 to 60 dBm. This disparity caused by the tag diversity may introduce uncertainty to the correlation. Note that the signal strength of UHF passive tag is always negative in the unit of dBm. Since the sign has no effect to the phenomenon, we use the absolute RSS value as RSS in the following for the convenience of representation and signal processing. (2) *Location sensitivity*. In practice, the RSS values of tags are sensitive to their location diversity. We put a tag 3m away from the reader antenna, and vary its position among 20 locations in a line. The average RSS value reported at each location is shown in Fig. 2. We observe that the value of RSS is highly dynamic, even between two adjacent locations.

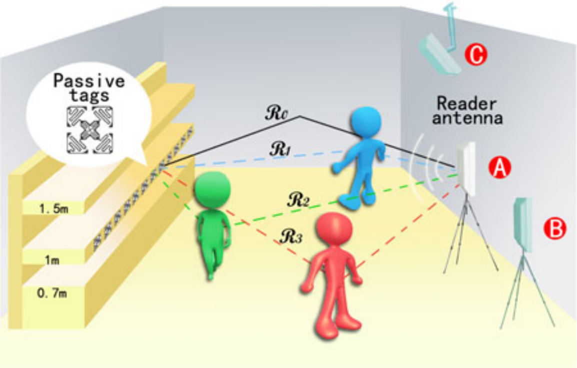
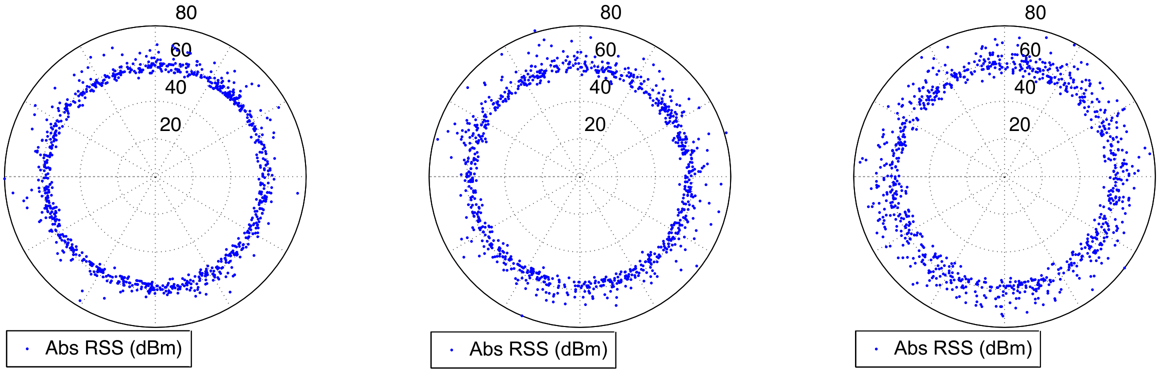


Fig. 3. The multipath propagation of RF signals in R#.

The above challenges motivate us to consider the problem in a different way. Intuitively, ignoring the differences among individual tag or location, we treat a group of tags as an entire detector. We then attempt to capture the overall impact of human movements on the tag backscattered signal to establish a stable correlation between the number of human objects and RF signals. We conduct a series of *proof-of-concept* experiments to validate this idea. As shown in Fig. 3, we deploy 8 passive tags in a line, with a distance 20 cm in between. We invite some (say 0$\sim$4) volunteers to walk in the region between the reader antenna (i.e., antenna A) and tags, and collect the RSS of tags within a short period ($5s$). The results are shown in Fig. 4. In each subfigure, 1000 RSS values are plotted in a polar form with random directions, and the radius of each points indicates an RSS value. It is obvious that: (1) when no human object moves in the region, the distribution of RSS values is centralized along a thin circle and relatively stable; (2) when the number of moving human objects grows, the RSS values become increasingly dispersive and the range of their distribution becomes wider. This result implies the potential correlation between the RSS variation and the number of human objects. To prove its feasibility, we try to analyze this phenomenon theoretically.

3.3 Analysis

We validate the following hypothesis: *the more moving human objects are present, the higher the variation in the RSS values of the tag backscattered RF signals.*



(a) 0 human object (b) 2 human objects (c) 4 human objects

Fig. 4. The RSS distribution vs. # of moving human objects.

The propagation of the tag backscattered signal in indoor environments is complicated. First, we consider a multipath propagation environment, the received signal at the reader antenna is the superposition of the effects of all multipath signals. This can be modeled as [30]:

$$R_i(t) = \sum_{n=1}^{m} H_n x(t - \tau_n), \tag{1}$$

where i is the number of human objects in the region at the time point t, m is the number of distinct paths in the channel, H_n is the multipath multiplicative distortion, x is the transmitted signal, and τ_n is the corresponding path time delay. Note that each path has a complex H_n, influenced by multiple factors, such as the distance between the reader and tag, the types of media that the radio signal propagates through, and the surfaces of objects scattering the signal, etc..

Assume that there is no moving human object in the region. We define a variable $\Re_0$, which represents the received singal in this scenario, and we have $\Re_0 = \{R_0(t)|t > 0\}$. When one human object enters the region, he or she may introduce difference to the RF signals in the channel. We denote the signal variation incurred by this human object as $\Re_1$. The wireless channel between the reader and tags can be modeled as a linear time-varying system [30]. Based on the additivity property in such systems, the RF signal can be expressed as: $R_1 = \Re_0 + \Re_1 = \{R_1(t)|R_1(t) = R_0(t) + \Re_1(t), t > 0\}$. Due to the independence of $\Re_0$ and $\Re_1$, we can get the variance of R_1:

$$\begin{aligned} Var(R_1) &= Var(\Re_0 + \Re_1) = Var(\Re_0) + Var(\Re_1) \\ &\geq Var(\Re_0) = Var(R_0). \end{aligned} \tag{2}$$

The equality holds iff $\Re_1$ is a constant, which means that the human object either has no influence on the RF signal propagation or keeps unmoved. However, this is either impossible in practice or inconsistent with our assumption. Thus the variance in the RF signals is monotonically increasing when the number of human objects increases.

It is worth noting that RSS is an average measurement on the power of received signal. Based on above analysis, we can deduce that when the number of human objects grows, the variance of received signal power increases, resulting in distribution variation in the RSS values. It indicates that we can utilize above correlation to estimate the number of human objects.

4 System Design

Inspired by the analysis in previous section, we propose to extract three features for reflecting the RSS distribution variation and use machine learning based methods to estimate the number of human object in the corresponding region. In this section, we first overview our system R#, then elaborate the system design.

4.1 System Overview

We assume that in a supermarket with a pre-deployed passive RFID system, passive tags are attached to items. COTS readers with their antennas can successfully interrogate those passive tags in the region that we want to monitor the number of customers, e.g., corridors or picking aisles. For a given region, a reader identifies a number of passive tags using protocols of EPCglobal C1G2 [3]. In the implementation of R#, the tags (say 20 tags) are deployed in a line, as illustrated in Fig. 3. The reader periodically collects the tag IDs and RSS values. For simplicity, we call an RSS sequence collected from a period as an *observation* and each collection from a tag as a *sample* in the following sections.

R# works in three phases, *data preprocessing*, *feature extraction*, and *estimation* (i.e., *classification*). In the first phase, R# conducts a series of preprocesses on the raw RF signals collected from the reader and prepares the data for later operations. In the second phase, R# extracts three features from the RSS values to establish a correlation between the number of human objects and the backscattered RF signals. Finally, R# employs machine learning techniques for human object estimation.

4.2 Data Preprocessing

There are two operations in the preprocess phase, regrouping and interpolation.

Following the specification of EPCglobal C1G2[3], tags are identified using a slotted ALOHA mechanism. That is, each tag randomly selects a time slot for reporting to the reader. Hence the RSS values in one observation are ‘timeline-based’ but not grouped according to the tag ID, and cannot be used by the feature extraction scheme of R#. Thus, we regroup the RSS values by tag IDs in each observation.

Again due to the slotted ALOHA mechanism, the slot in which a tag replies is randomly and uncontrollably distributed, leading various numbers of samples from different

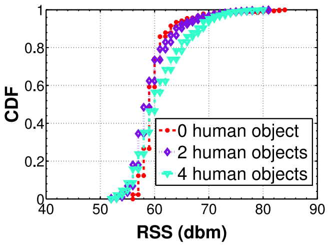


Fig. 5. CDF of RSS.

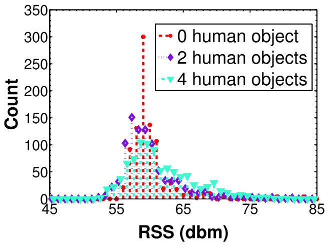


Fig. 6. Distribution of RSS.

tags in one observation. In addition, ambient factors, such as the shadowing, interference, and tag location, also influence the backscattered signal, resulting in different sampling periods to these tags. Suppose K is the maximum number of samples collected from tags within one observation. To avoid the unfairness in the sampling and later processes, for each tag we use *Linear Interpolation* method to virtually increase the number of its samples to K.

4.3 Feature Extraction

In R#, our expectation on a feature of the RF signals is that it monotonously increases if enlarging the number of human objects, or vice versa. To this end, we select three complementary features, i.e., *Entropy*, *size of dilated area* (*SDA*), and *mean squared error* (*MSE*). We present their extraction schemes as follows.

- ***Entropy.*** Given a constant transmission power of the reader (30 dBm in our implementation) and fixed location for each tag, when there is a given number of human objects in the region, the collected RSS values shall vary within a range, denoted as $[R_{mi}, R_{ma}]$. As we analyzed in Section 3, the human interference may either strengthen or weaken the signal power received at the reader antenna. Correspondingly, the R_{mi} and R_{ma} may change with the number of human objects as well.

We make a mathematical abstraction on the RSS values in an observation. If we use a random variable to represent them, those values are actually a distribution of this variable. As shown in Figs. 5 and 6, the distributions with different numbers of moving human objects are distinguishable. Inspired by this, we attempt to apply an information entropy based scheme to reflect the distribution of this variable, and hence yield the first *Entropy* feature for R#.

According to the information theory, entropy is a measurement of the uncertainty for a random variable. By monitoring the entropy of different observations, we can measure the degree of their disparity or concentration [31]. To calculate the entropy of an observation, we first establish the empirical distribution for this observation. The RSS range is divided into N bins with an equalized length (L), $N = \lceil(\hat{R_{ma}} - \hat{R_{mi}})/L\rceil$, where $\hat{R_{mi}}$ is a constant lower than the sensitivity of the reader and $\hat{R_{ma}}$ is a constant beyond the possible maximum RSS value. Let i denote the bin ID ($i < N$), x_i denote the number of RSS values falling into the ith bin, and $p_i = x_i / \sum_{i=1}^{N} x_i$ denote the probability that x_i RSS values fall into the ith bin. According to entropy theory, the discrete entropy of an observation can be calculated as:

$$E(X) = -\sum_{i=1}^{N} p_i \cdot \log(p_i). \tag{3}$$

- ***SDA.*** We extract the second feature from the collected RSS values by using a Morphological Image Processing based Scheme (MIPS). This feature is firstly discovered from the RSS data visualization. Intuitively, the higher the RSS variance, the larger the area that the points could cover. To formalize this observation, we first binary-visualize an observation with the following operations.

Since the RSS resolution of our RFID reader is 0.5 dBm, we first normalize each RSS value x with the operation $(x - \hat{R_{mi}}) * 2$. Every x is within $[0, (\hat{R_{ma}} - \hat{R_{mi}}) * 2]$. We introduce a two-dimensional array $R_{l\times c}$ and initialize its elements as zero. Here l is the length of the observation and c is $(\hat{R_{ma}} - \hat{R_{mi}}) * 2$. We then set all elements $R(k, x_i)$ to 1, where $1 \le k \le l, 1 \le i \le c$. Note that there is only one '1' in every column of $R_{l\times c}$. Each element in this array represents a pixel. Let '1' denote a white pixel and '0' denote a black pixel. Then we obtain a binary image. For example, the upper half parts of Fig. 7a, 7b, and 7c show the results of performing the binary visualization operation on the collected RSS values with 0, 2, and 4 human objects coexisting, respectively.

To amplify above distinction and make it quantizable, we then conduct the morphological dilation operation to the three images. Dilation is a core opeation of MIPS, which makes a target in an image 'grown' or 'fatten'. This operation is based on two fundamental morphological operations, *reflection* and *translation*. The reflection of a set A is defined as: $\hat{A} = \{w | w = -a, a \in A\}$. The translation from a set A to a point $z = (z_1, z_2)$ is defined as: $(A)_z = \{c | c = a + z, a \in A\}$. To formalize, dilating a set A by using a set B is expressed as: $A \oplus B = \{z | ((\hat{B})_z \bigcap A) \neq \varnothing\}$, where B is the *structuring element*, which decides the degree of dilation. Generally, B is a set of '1's with a specific shape, such as a 'line', 'diamond', and the like. When conducting a dilation operation, B will translate on the entire image region of A, and exam which position is overlapped with its '1's. The dilation result is a set with all these overlapped positions. The lower parts of Fig. 7a, 7b, and 7c demonstrate the dilation results of upper original binary images. They suggest that after dilation, white pixels get connected and the area of them is extended. We find that the area of white pixels can well represent the feature of observations. That is, when there are more human objects within the surveillance region, the area of white pixels becomes larger in the dilated image.

After the dilation, we notice that the image has many burrs, which are probably derived from some outliers. To eliminate these thin protrusions, we conduct an *open operation*. The open operation is a combination of dilation and erosion operations. Different from the dilation, erosion can shrink or diminish a target in an image. In the open operation of R#, an erosion is followed by a dilation. An open

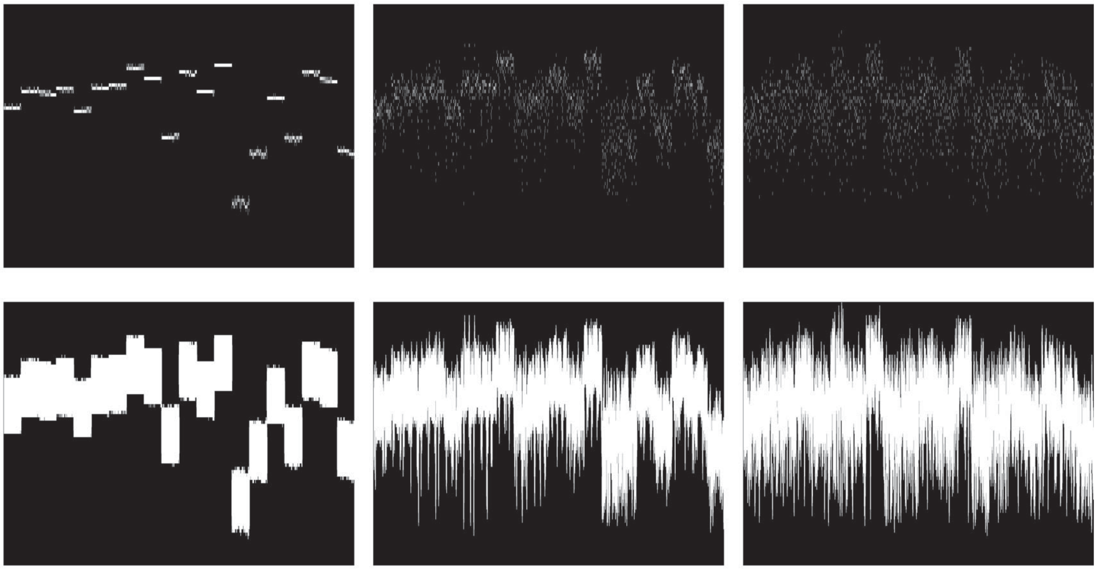

(a) 0 human object (b) 2 human objects (c) 4 human objects

Fig. 7. Binary visualization and dilation results of the observation of 0, 2, 4 human objects. When the number of human objects increases, the size of dilated area is enlarged.

operation often makes the edge of an object smooth, fills the gap, or eliminates the burr. Hence, the purpose of performing open operations for R# is to remove the noise and outliers from an observation. We define an erosion on a set A by using a set B as: $A \ominus B = \{z|(\hat{B})_z \subseteq A\}$. Hence, the open operation is symbolically expressed as: $A \circ B = (A \ominus B) \oplus B$, where B is a structuring element. After the open operation, we calculate the area of white pixels as the second feature.

- ***MSE***. Based on the dilated image $f(x, y)$, we introduce another feature, the Mean Squared Error. By realizing the compression (i.e., Discrete Cosine Transform (DCT)) and decompression algorithms, we obtain an recovered image $\hat{f}(x, y)$. Mean Squared Error is usually used to represent the *fidelity* between f and $\hat{f}$, since it measures the information lose. We notice that when there are more human objects moving in the area, the image of the observation will be more complicated. Correspondingly, more information will be lost if performing compression. Hence, MSE is eligible to characterize the observations.

We choose DCT for the compression since it is an invertible transform and can be easily recovered. The procedure of DCT compression is as follows: First, the original image is divided into some 8×8 subimages. Then discrete cosine transform is conducted to transform these subfigures and derive a transform coefficient matrix (C). We realign C in a zigzag pattern (denote the result as C'). In this case, the (0,0) element (i.e., top-left) of C' is the DC (zero-frequency) component. Elements with increasing vertical and horizontal index values represent higher frequencies. For DCT-represented image, the energy of the image is mainly concentrated in the lower-frequency coefficients because they contain more components of the original signals. On the other hand, the higher-frequency coefficients represent the details of the image. Given the DCT coefficient matrix C', we then discard a part (85 percent in our implementation) of the coefficients by multiplying C by a *mask*. The *mask* is a matrix composed with '0' and '1', an example is shown in (4). All '1's in the mask determines how many and which coefficients will be kept.

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

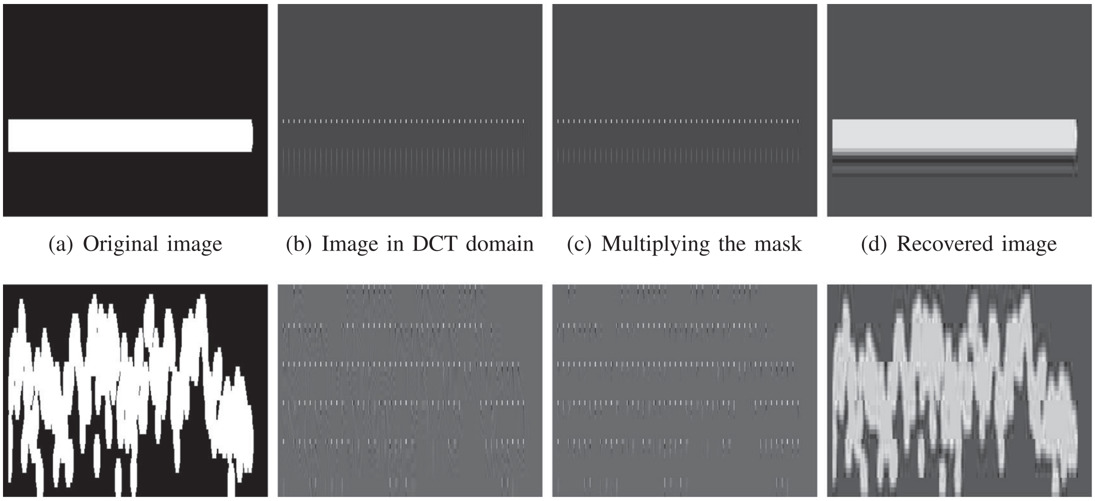
We can utilize different *mask*s to change the compression rate. When recovering the image, we run the reverse DCT on the intercepted coefficient matrix.

As examples, we conduct the above procedure on one tag's data in two cases, i.e., there is no human object and six human objects in the detection region respectively. We plot the results in Fig. 8 for comparison. Subfigure (a) is the original image of the tag's RSS sequence with no human object; subfigure (b) shows the DCT domain of (a); subfigure (c) is the result after multiplying (b) by the mask M (Eq. (4)), which means we discard 85 percent of the coefficients; and subfigure (d) is the recovered image. We observe that by DCT, the main structure of the image can be clearly recovered and the discarded coefficients have little impact. The only impact is that the edges of recovered images become blurred due to the loss of the details in the high-frequency part of the original image by the effect of the mask. In addition, there is no doubt that the impact is more noticeable when the number of human objects is larger. We use MSE to reflect this impact. The incurred MSE between $f(x, y)$ and $\hat{f}(x, y)$ is:

$$MSE = \left[\frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\left[\hat{f}(x, y) - f(x, y)\right]^2\right]^{1/2}. \tag{5}$$

4.4 Machine Learning Based Estimation

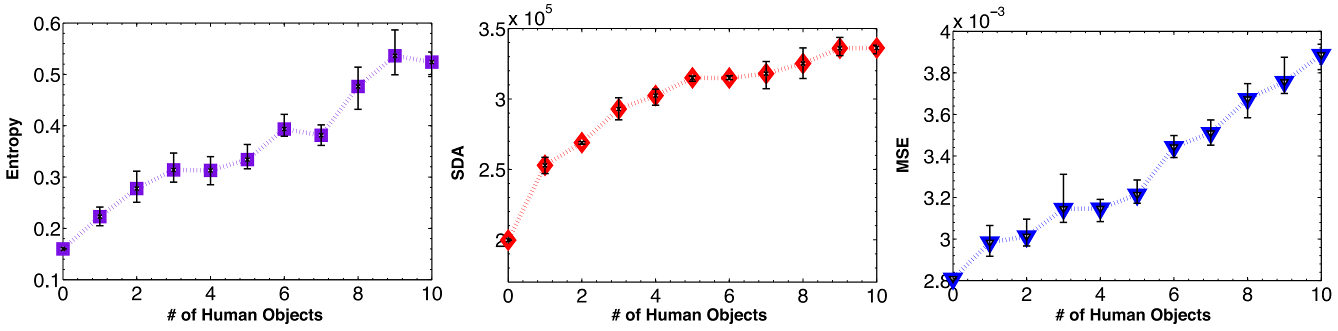
Before the design of estimation mechanism, we first exam the effectiveness of three features. We arrange 0∼10 volunteers to move arbitrarily in the surveillance region and



(a) Original image (b) Image in DCT domain (c) Multiplying the mask (d) Recovered image

(e) Original image (f) Image in DCT domain (g) Multiplying the mask (h) Recovered image

Fig. 8. The procedure of DCT compression. (a)-(d) There is no human object. (e)-(h) There are six human objects.



(a) Entropy (b) SDA (c) MSE

Fig. 9. Three features adopted by R#.

check the trend of features with varying the number of volunteers. The result is reported in Fig. 9a, 9b, 9c. It reveals an interesting insight. Each of the three features shows a good ability of distinguishing different numbers of human objects in some ranges, e.g., 0$\sim$4 persons for SDA and 6$\sim$10 persons for MSE. While the entropy feature shows an instable variance upon certain numbers of persons. Fortunately, the three features exhibit the complementarity with each other, which inspires us to jointly use them in the classifier of R#.

We adopt Naive Bayes method from WEKA as our machine learning classifier. In real implementation, it is necessary to tune the classifier parameters to optimize the estimation accuracy. To achieve this goal, we organize above experiment results into two datasets, one for training and another for test. We then perform a 10-fold cross validation on the datasets to determine the key parameter of classification. We will present the tuning procedure in Section 5.1.

5 Implementation and Evaluation

In this section, we present the implementation of R# and evaluate its performance via extensive experiments.

5.1 Implementation

In the implementation, we employ Commercial-Off-The-Shelf (COTS) Impinj readers, i.e., Impinj R420, one directional antenna, i.e., Laird A9028R30NF with the gain of 8dbi, and 20 commodity passive tags (i.e., Impinj H47 with size of 44 mm × 44 mm). Those tags are attached to a number of cartons arranged in a line. The space between two adjacent tags is 20 cm and the distance from the reader antenna to the line of tags is 3.5 m. Fig. 10 shows our experiment scenario.

Configurations. We developed software for data collection. The software communicates with the reader through LLRP (Low Level Reader Protocol) toolkit released by the Impinj company. The corresponding configuration of the software is detailed in Table. 1. To avoid the influence caused by frequency hopping, we fix the communication frequency of the reader on the 16th available channel, i.e., 924.375 MHz. The tag modulation type is FM0. The reader mode is set to 'MaxThroughput', which supports the highest data rate. The reader search mode is 'DualTarget'. Note that reader search

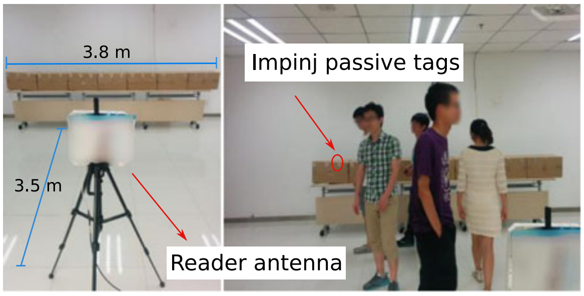


Fig. 10. The scenario in our implementation.

TABLE 1
Configurations in R# Implementation

Configuration parameters	Status
ChannelList	16
ReaderMode	Maxthroughput
SearchMode	DualTarget
Tag modulation	FM0
PeakRSSIMode	enabled

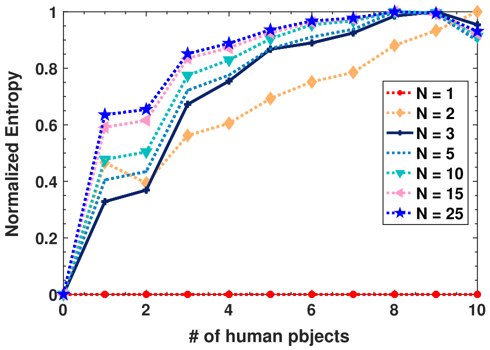


Fig. 11. Entropy versus # of bins.

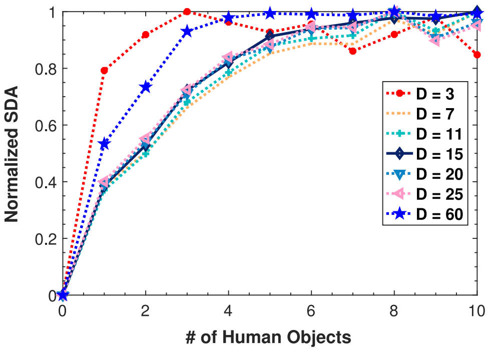


Fig. 12. SDA versus element size.

mode controls which tags respond to the reader in the read field. In the 'DualTarget' mode, the tag will be interrogated continuously regardless of the tag state ('A' or 'B') or Session number (0, 1, 2, 3), ensuring sufficient sampling rates for the system. The whole system runs on a PC with an Intel Core i7-4600U 2.10 GHz CPU and 8 GB RAM.

5.2 Parameter Tuning

It is necessary to tune the parameters of R# for performance optimization. We focus on four key parameters, number of bins, dilation degree, different masks, and the threshold of classifier.

Impact of the Bin Size. As aforementioned in Section 4.3, the entropy feature is highly related to the number of bins (N). When N varies, the distribution of an observation changes correspondingly, leading an impact to the entropy feature. In this part, we evaluate the impact of N. We conduct 11 groups of experiments by varying the number of human objects from 0 to 10, and observe the changes of entropy values. Fig. 11 shows the results. When we set N to 1, the entropy remains zero all the time. When increasing N, we observe that the entropy increases and its slope becomes gentle. We found that $N = 3$ is enough for accurate estimation. Note that in some tests the entropy of 10 human objects is smaller than that of 9 human objects. It implies that when the region is densely distributed by many human objects, the entropy feature does not work well. This is expected because more RF signals are absorbed or shadowed by the human objects in such cases. Due to the shielding effect, the neighbor-based interpolation, as mentioned in Section 4.2, may reduce the dispersion degree of RSS values in an observation and hence generate inaccurate estimations.

Impact of the Dilation Degree. Then, we pursue appropriate parameters for MIPS methods. When performing MIPS, the type and size of structuring elements have an significant impact on the SDA feature in the dilated image. In R#, the type of structuring elements is fixed, *i.e.*, a flat disk-shaped structuring element. Thus, we only evaluate the impact from different size (i.e., the radius D) of the structuring element. Since the dilated area of white pixels varies considerably with different Ds, we normalize the area of every test into the range of [0, 1] to examine the influence of D on the numerical discrimination together. Fig. 12 plots the relationship between the SDA feature and the number of human objects. For each setting of the number of human objects, we conduct 10 tests and calculate the average result. The result shows that, when $D = 3$, the normalized SDAs keep relatively stable even if increasing the number of human objects. In this case, the estimation may fail. The reason behind is that the dilation of R# dose not amplify the RSS characteristics. The RSS value becomes dispersive when the number of human objects increases, resulting in more white pixel regions. However, due to the sparsity of RSS, these regions are still small. Some of them may be shrunk after open operations, leading to a reduction of the whole area. On the other hand, if D is too large, *e.g.*, 60, the dilated area of white pixels will be extended so much that the SDA feature is virtually drowning in this region. To this end, we compare different settings of D, and set $D = 15$ in our implementation.

Impact of Masks. We also evaluate the impact of different masks to the MSE feature. The mask adopted in R# is a 8×8 (0,1)-matrix. Different masks mean different portions of the coefficient would be discarded. This controls the compression rate and the value of MSE. We conduct the compression and recover procedure mentioned in Section. 4.3 on Fig. 8e using different masks. The recovered images are illustrated in Fig. 14. The bottom-right square in each subfigure shows the corresponding $8{\times}8$ (0,1)-mask, in which the small green square represents element 1 and the white square represents element 0. There are six 1s in the mask of subfigure(a), which means that we discard $1 - \frac{6}{64} = 91\%$ percent of the coefficient and use only 9 percent of the coefficient to recover the image. We observe that the mask in subfigure (a) has the minimum number of 1s, leading that subfigure (a) is the most fuzzy one among all six subfigures. On the contrary, subfigure (d) has the highest reduction degree. In addition, the less number of 1s in the mask corresponds to less storage overhead. Considering that the DCT coefficient is stored in a zigzag pattern, we choose the mask in subfigure (b) in our implementation.

Impact of the Classifier Threshold. Last, we check the classification accuracy of R# based on a threshold-based metric, *i.e.*, Equal Error Rate (EER). Following the principle of 10-fold cross validation, we divide the experiment result to two datasets, the training set and the testing set. The

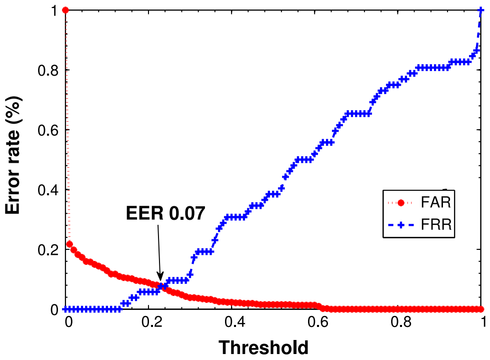


Fig. 13. Error rate versus threshold.

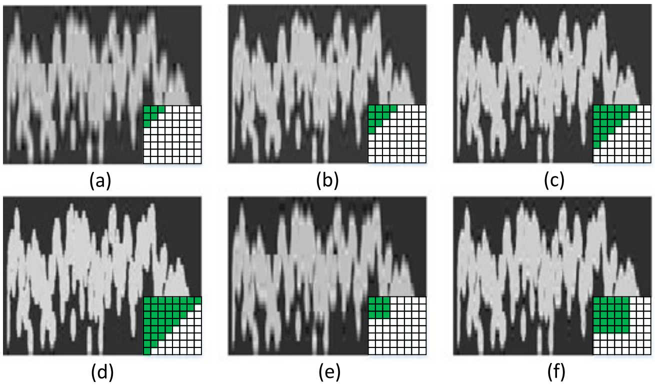


Fig. 14. Influence of different masks.

features are derived from the training set, and the testing set is used to analyze the performance of our classifier. We compute the similarity between all the training and testing data. Then we calculate False Reject Rate (FRR) and False Accept Rate (FAR) based on a predefined threshold (T). FRR is the percentage of cases that two tests belonging to one category have a similarity above T and then are classified into two categories (two different counting numbers). FAR is the percentage of cases that two tests coming from two categories have a similarity below T and are classified into one category. EER is the error rate when FAR and FRR are equal, indicating a good balance between the two types of errors. It is a common measurement on the accuracy of classification systems. With different settings of T, we investigate the relationship between the FAR and FRR and report the EER of R# in Fig. 13. The corresponding $T = 0.23$ is determined for the classifier to make an accept/reject decision. With this threshold, the EER of R# is 0.07, meaning the classification accuracy is around 93 percent.

5.3 Evaluation on Estimation Accuracy

After determining the system parameters, we adopt the following strategy for further evaluation: given a *confidence interval* $\beta \in [0, n]$, some human objects within the monitoring region of unknown number k ($k \leq n$), and an R#'s estimation on the number of human objects $\tilde{k}$, we use the *reliability*, $\alpha = P\{|\tilde{k} - k| \leq \beta\}$, for evaluating the system accuracy. Note that the confidence interval is an integer, representing the estimation error range of R#. A simple illustration on this strategy is as follows. Assume R# experiences 10 tests with 6 human objects. R# reports 6 in 8 tests, 5 in one test, and 3 in another test. Then we can conclude the reliability of R# in

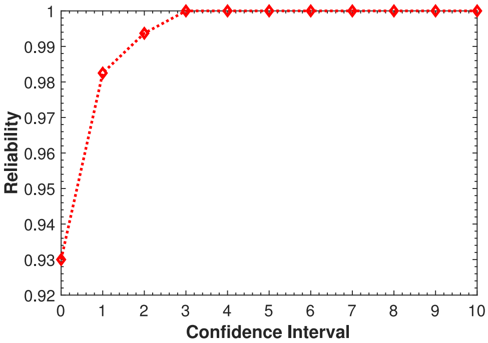


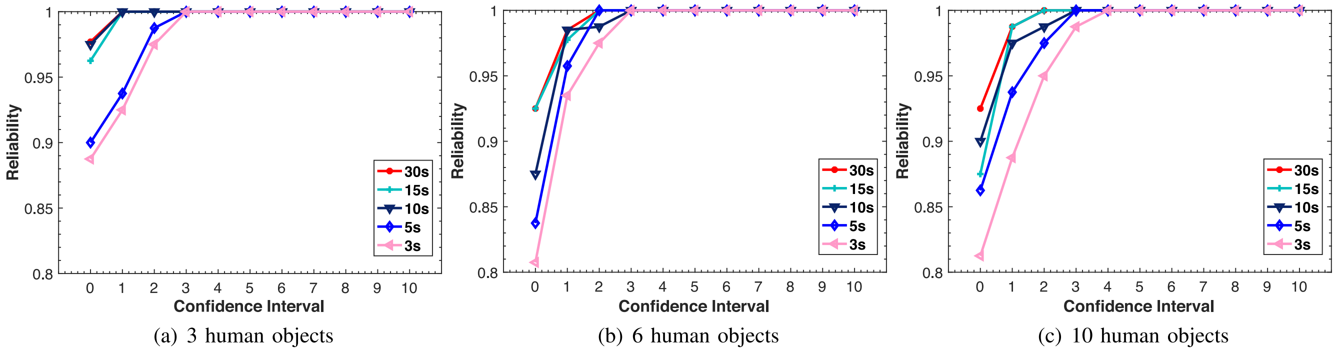
Fig. 15. Reliability of R#.

above tests is 80 percent with the confidence interval as 0, 90 percent with the confidence interval as 1. For validation, we conduct 600 tests with the number of human objects varying from 0 to 10, and plot the reliability of R# with different confidence intervals in Fig. 15. The result shows that R# achieves highly accurate estimations on human objects. For example, the reliability of R# is 93 and 98 percent, with the confidence interval as 0 and 1, respectively.

We note that human motion pattern varies significantly across measurements. However, above results demonstrate our proposed method has tolerance for this. The reasons are manifold. First, the modern COTS Impinj R420 reader can read more than 300+ tags per second. Such a high sampling rate allows us to monitor the test region very fast. In addition, we utilize a period of samplings for estimation, which will incorporate multiple snapshots of human motion patterns, and we have a good reason to consider that multiple snapshots can neutralize the signal variations caused by motion pattern variations. Second, we adopt the tag array to cover the whole region, which makes sure that human motions can be well captured by each tag. In addition, we conducted extensive experiments to test the system. During the experiments, the movement pattern of each volunteer is not restrictive, and the results prove the efficiency of our method.

5.4 Evaluation on Time Efficiency

Intuitively, a longer sampling period produces a larger RSS dataset during the data collection process, which potentially generates more precious estimations. However, extending the sampling period will produce a higher computational complexity and latency. Here, we analyze the tradeoff between the length of sampling period (ranging from 3s to 30s) and classification accuracy. Fig. 16 presents the reliability of R#. With the extension of sampling period, the reliability of R# increases, indicating an increase of estimation accuracy. For example, when the sampling period is around 5s, given the confidence of 1, the reliability of R# is 96 percent if 6 human objects are in the monitoring area. For the test with 10 human objects, the reliability of R# is still 93.75 percent. If we set the confidence to 0, the reliability of the above two tests becomes 83.75 and 86.25 percent, respectively. By analyzing the result, we observe that 5s is a good tradeoff. After the data collection, R# spends only about 2.3s to perform processing and estimation. The above results demonstrate that R# achieves high time efficiency. In addition, the size of data (consisting ID and RSS) gathered by the end of augmenting process (the 5th



(a) 3 human objects (b) 6 human objects (c) 10 human objects

Fig. 16. Reliability of R# versus sampling periods.

second) is approximately 4 Kb, implying that R# also has low storage overhead.

5.5 Comparison with Existing Approaches

We conduct a set of experiments to compare R# with two state-of-the-art device-free crowd counting approaches, i.e., VAR [32] and SCPL[23], in terms of accuracy. VAR first demonstrated the feasibility of using radio signal strength to estimate the crowd density. VAR shows that RSS variance increases as the number of human objects grows. On the other hand, SCPL utilizes the RSS mean difference derived from the wireless links to count device-free objects in indoor environments. We set up the same scenario as in previous experiments and plot the results of three approaches in Fig. 17. The results suggest that VAR achieves a overall accuracy of 82 percent. The accuracy of SCPL is nearly 100 percent when there is zero and only one human object. But its accuracy experiences a sharp decrease (down to 85 percent) when increasing of human objects. The reason of such a decrease might be that SCPL identifies the subject count by subtracting the individual's interference to each Tx-Rx link, which implicitly requires each subject disperses with each other. However, in practice, the surveillance region is sometimes not very large and does not meet this requirement. Compared with VAR and SCPL, R# can always achieve high accuracy (say 97 percent in average), while the number of human objects ranges from 0 to 4.

6 Deployment

We further evaluate R# with the purpose of enhancing practical deployments.

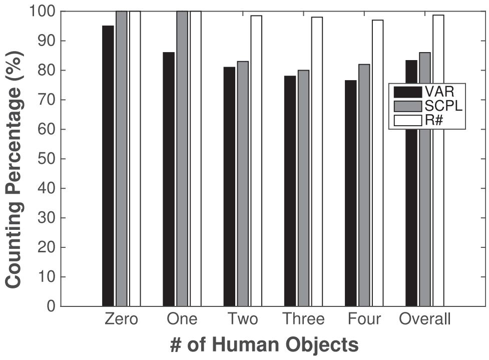


Fig. 17. Accuracy comparisons with prior approaches.

6.1 Impact of the Moving Speed

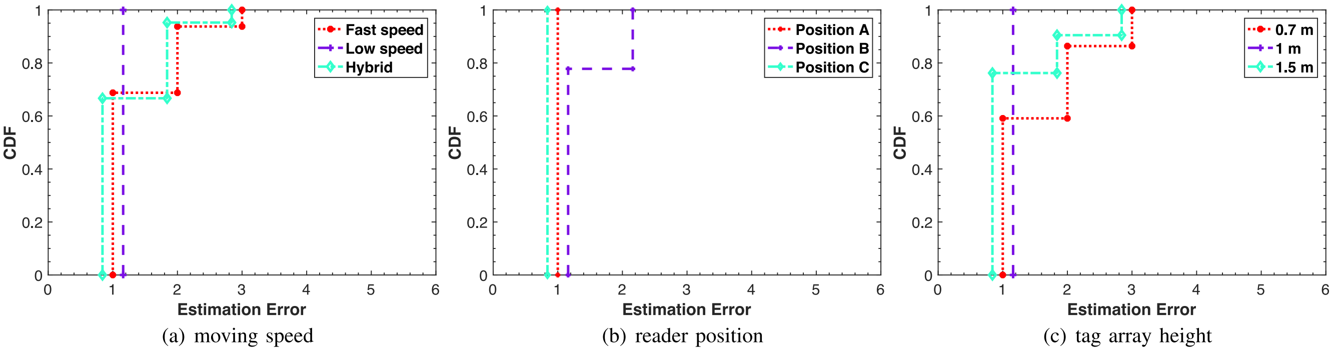
In practice, people may walk at different speeds. Therefore, it is necessary to check whether and how the moving speed of human objects affects the accuracy and robustness of R#. We repeat the experiment conducted in Section 5.1, but ask the volunteers to walk at 3 modes, namely the *high* (about 1.3 m/s), *low* (about 0.7 m/s) and *hybrid* speed modes. In the hybrid mode, the walking speed of each volunteer shifts between the high and low modes arbitrarily. We set the confidence interval as 0, which means we focus on the classification accuracy of R#. Fig. 18a shows the CDF of the estimation error. At the low speed mode, in 100 percent of tests the estimated value of R# deviates from the real value by at most 1 person, while for other two modes, in nearly 93 percent of tests the deviation is no larger than 2 persons. This is because in the low speed mode, sufficient samples can be obtained from the tags, due to a less shielding effect among human objects than in other two modes, which well supports the feature extraction of R#. Hence, R# is more suitable for estimating the crowd in the low speed scenario. In fact, most people are likely to walk in the low speed mode during their in-store shopping such that R# can make accurate estimations.

6.2 Impact of the Reader Position

In the real deployment, the antenna may be placed at different positions due to the management requirements or other considerations. Hence the influence of the antenna position should be investigated. We select three patterns for positioning the reader antenna, namely the *front*, *side*, and *top-view* positions, as the points A, B, and C shown in Fig. 3. Observing the result shown in Fig. 18b, we find that for the *front* (postion A) and *top-view* (position C) deployments, the estimated value of R# deviates from the real value by at most 1 person. For the *side* deployment, the maximum estimation deviation is 2 persons, and in 80 percent of tests the deviation is at most 1 person. The reason is that when the reader antenna is at the *side* position, tags are deployed spatially asymmetrical towards the reader antenna. In this case, the tags farther away from the reader antenna are more vulnerable to the ambient interference, resulting in unbalanced sampling rates among tags. Thus, we recommend a spatially symmetrical deployment for the reader antenna like *front* or *top-view* pattern in practice.

6.3 Impact of Tags Height

We also evaluate R# when changing the height of tags. Fig. 18c compares the deviation of estimated values of R#



(a) moving speed (b) reader position (c) tag array height

Fig. 18. CDF of estimation error of different deployments.

with the height of tags as 0.7 m, 1 m, and 1.5 m, respectively. We see that with the height of 0.7 m, R# has the worst performance and only 60 percent of the estimations with their confidence interval $\beta \geq 1$ can keep the deviation as 1 person. We notice that for most volunteers, the height of 0.7 m only reaches their thighs. The shielding or reflection effect induced by volunteers has less impact on the signal propagation than that with other height settings. The result indicates that the height of $1m$ is an appropriate option for the real deployment of R#.

In summary, we analyze the performance of our system using the Receiver Operating Characteristic (ROC) curves for the above deployment patterns in Fig. 19. The ROC is a curve that automatically computes False Reject Rate when the False Accept Rate is fixed at a desired level and vice versa. The vertical axis of ROC curve is False Reject Rate, and the horizontal axis is the False Accept Rate [33]. We adopt the True Accept Rate (TAR) (TAR = 1 - FRR) instead of FRR since TAR shows the rate of accepts of legitimate identities for target values of FAR, which better illustrates the accuracy of the system. From Fig. 19 we observe that R# achieves high estimation accuracy. If FAR is fixed to a value higher than 0.13, our system will correctly identify 100 percent of the number of human objects for all three cases. Even when FAR is required to operate at much lower degree (e.g., 0.05), R# can still achieve a high average TAR of 0.9.

6.4 Impact of Obstacles

In this series of experiments, we investigated the impact of obstacles within the test environment. We separated the experiments into three cases, as illustrated in Fig. 20a, 20b, 20c. *Scenario 1*: obstacles (two tables, two chairs) were put at the boundary of the test region; *Scenario 2*: obstacles (two tables, two chairs) were put at certain positions within the test region; *Scenario 3*: obstacles (one ladder, one shopping cart) were put at random positions within the test region; and we think ladders and shopping carts are representative obstacles in a supermarket environment. In all scenarios, the reader-to-tag distance was set to 3.5 m. We collected 30 observations (samples) for each case of human objects (i.e., 0, 1, 2, 3, and 4 persons). Note that the default setup in this figure is the same as described in Section. 5. In Scenario 3, to diversify the situations, we randomly change the locations of these two obstacles 5 times. The comparison of the counting results is shown in Fig. 20d. Interestingly, the results of Scenario 1 and 3 are even a little better than that of our default setting when the number of human objects is 3 or 4. This is because our system leverages the variation in tag signals to conduct the counting, and both scenarios (i.e., multipath-rich scenarios) incurs more difference in the signals and hence could potentially help to make more accurate estimation. However, Scenario 2 achieves a relative inferior performance, indicating that obstacles, which directly block the line-of-sight between the reader antenna and tags, would result in a decrease of the system accuracy.

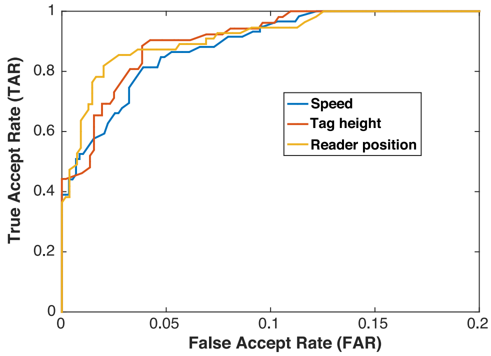
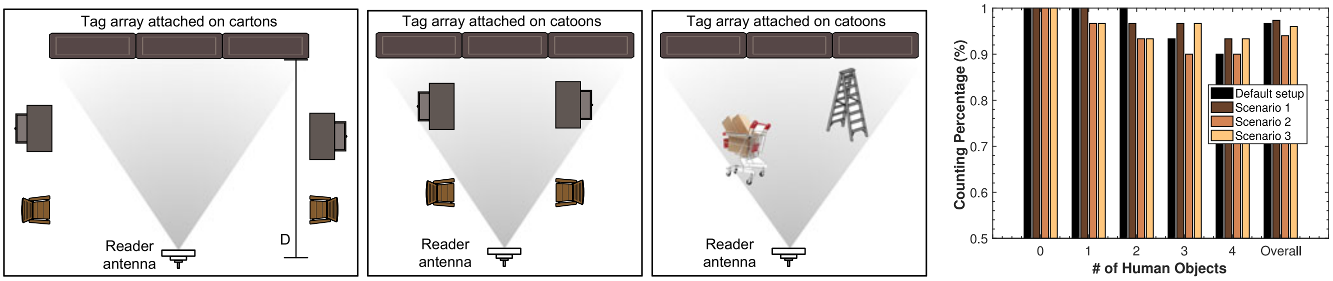


Fig. 19. ROC.

6.5 Impact of Reader-to-Tag Distance

In real applications, constrained by specific space limit, the distance between the reader and tags could be different. To check the impact of reader-to-tag distance, we conducted a group of experiments. The experimental scenario is the same as shown in Fig. 20a. We varied the reader antenna position to change the distance (D) from 1 m to 3.5 m with a step length of 0.5 m. Four volunteers were invited to participate in this experiment. We collected 30 observations for each case of human objects (i.e., 0, 1, 2, 3, and 4 persons), which means we have a total of 150 observations with each distance. The counting results are shown in Fig. 21. We observe that when the distance is only 1m, the accuracy is lower than 70 percent. And the estimation accuracy increases gradually and reaches above 90 percent when the distance becomes 2 m. This is because the reader-to-tag distance actually determines the effective region of the system. Since the reader antenna is a directional antenna (with a beam angle of 72 degree), if the distance is getting shorter, the tags at the edge of effective region would be apt to be unreadable. In this case, only a few tags facing the reader antenna can be used for estimation, which will implicitly reduce the system accuracy. Further analysis can be found in Section. 7.1.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3 (d) Accuracy vs. Scenarios

Fig. 20. Impact of obstacles.

7 Discussion

This section discusses limitations and practical ways to improve the capability of the system.

7.1 Upper Bound of the Number of Human Objects

It is important to pursue the maximum number (say 10 in current implementation) of human objects that R# can estimate for a given region. Intuitively, this upper bound is constrained by the detection region size. Furthermore, it is also related to the number of readers adopted, the readers' transmitted power, and the density of tags. As illustrated in Fig. 22, in our implementation we only use one antenna to monitor the area. The antenna is placed almost at the edge of the wall, and 20 tags are located more or less symmetrically to the antenna. The vertical distance between the antenna to the tag array is about 3.5 m. The space between two adjacent tags is 20 cm. The actual gain of our antenna is 8 dBi. Hence, the radiated power is concentrated in a beam propagating in an angle (nearly 72 degree) from the center of the antenna patch, according to the radiation pattern of the directional panel antenna. Ideally, the detection region can be approximately described as an irregular semiellipse, as the gray area shown in Fig. 22. Based on above parameters, the area (A) of the detection region should follow the relation:

$$\frac{1}{2}D * H < A < D * H. \tag{6}$$

Thus, we have $6.7\ \mathrm{m}^2 < A < 13.3\ \mathrm{m}^2$. We use the mean to simulate the area, which is A $\approx 10\ \mathrm{m}^2$. Splitting A into 10 subregions, when there are 10 human objects in the region, each of them occupies an 1 m^2 space. Their influence (i.e., absorption, reflection, blocking, etc.) to the tag backscattered signals will remain at a high level. We can utilize *Pigeonhole principle* to formalize this problem. The pigeonhole principle states that if n items are put into m containers ($n > m$), then at least one container must contain more than one item [34]. Approximatively, when there are more than 10 human objects staying in the area, there will be a subarea containing more than one human object. The variation introduced to the system from this subarea is minor because the impact will reach a saturation point. Hence, the features of R# used for estimation will become indistinguishable if there are more than 10 human objects coexisting. In addition, when the number of people is getting larger, during their movement, the persons might physically collide with each other in the crowed region. In this case, some people will always impede anothers movement. Apparently, such a crowed case is rare in practice. Since people usually prefer to keep a certain distance from others, we think for a 3.8 m *3.5 m region, the number of 10 is sufficient for people to feel comfortable. However, it is always desirable to scale the system to count more users. We envision that to count more users, one can deploy more reader antennas/tags to cover a larger surveillance region.

7.2 Resorting to Different Classifiers

We use different classifiers to evaluate the effectiveness of our proposed system. The goal is to evaluate whether we can improve the system reliability by using different classification techniques. We conducted a group of new experiments with five human objects. The features are extracted as described in Section. 4, and fifty instances are collected for each class (e.g., class k corresponds to k human objects). We run various classification algorithms using WEKA. Five classic classifiers are NaiveBayes, BayesNet, KNN, SVM, and PART. The accuracy is listed in Table. 2, and the reliability is shown in Fig. 23. BayesNet has the best accuracy. NaiveBayes achieves the

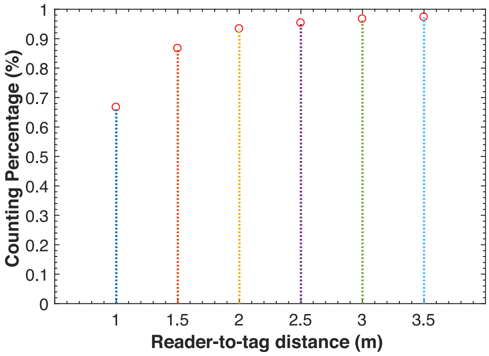


Fig. 21. Impact of reader-to-tag distance

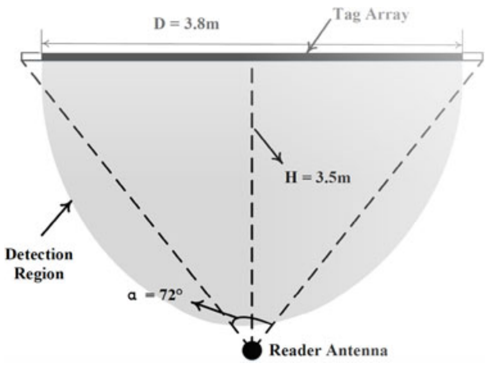


Fig. 22. Illustration of the detection region for R#.

TABLE 2
Accuracy versus Different Classifiers

Classifier	NaiveBayes	BayesNet	KNN	SVM	PART
Accuracy	95.54%	99.10%	95.54%	22.32%	93.75%

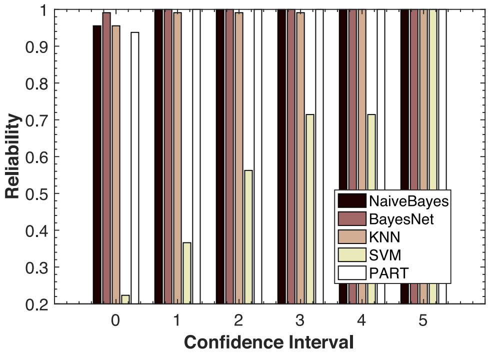


Fig. 23. Reliability comparisons with different classifiers.

same reliability as BayesNet when the confidence interval is larger than 0. KNN (k=1) or PART has comparable accuracy as NaiveBayes, but the reliability is a little inferior. SVM suffers from the worst performance both in accuracy and reliability. In current prototype, we implemented R# using Naive Bayes classifier for comparable accuracy and easy implementation. However, as the results suggested, the performance could be further improved by choosing a proper classifier.

7.3 Handling Static Users

Counting static people is challenging. Past works mostly utilize device-based methods, which means they estimate the crowd density by counting the number of devices carried by each individual user. One recent wireless signal based work that can localize and count static people is WiTrack 2.0 [15], which adopts a Frequency-Modulated Carrier Waves (FMCW) radio to measure the time-of-flight (TOF) of reflected signal from a user and localize him. The FMCW radio generates signals sweeping 5.46∼7.25 GHz with a step length of 2 MHz. WiTrack 2.0 can localize and count up to five static people in an area that spans 5 m × 7 m. It requires customized hardware, yet cost-inefficient for large-scale deployment. Our work is a device-free system that solely utilizes the reflections bounce off the moving human bodies to estimate the crowd. We purely use commercial off-the-shelf RFID products and can count 10 persons in an area that spans 3.8 m × 3.5 m in a furnished room (with tables, chairs, etc.). The adopted reader communicates with tags around 920 MHz with a bandwidth of 2 MHz, which provides coarse-grained sensing ability.

In the current implementation, R# is not able to estimate static users. We envision that to handle static users, one complementary method is to set up a tag array to monitor the entrance (or exit) of the aisle. When a person walks into (or off) the aisle, the system could estimate and record it. Then, compared with the measurement of R#, if there is a mismatch, we could conclude that with high probability there are static users within the region. Another possible way is to leverage the ultra-wideband techniques. Ultra-wideband RFID [28] provides far more fine-grained estimation for the signal transmitting time and distance, which could even be able to sense the chest movement of human breathing. This estimation can be utilized to distinguish static humans from static objects, and hence estimate the number of humans. However, ultra-wideband RFID requires customized devices or modification on the tag hardware, which is cost-inefficient and cannot be easily applied to existing RFID applications.

8 Conclusions

As RFID systems have been widely deployed in warehouses or supermarkets, R# provides a means for counting human objects within the surveillance area. R# is based on the observation on a phenomenon in our empirical study, more human objects incur a larger variation to the RF signals of passive tags. We extract three features to describe this phenomenon and analyze the feasibility of using them for estimating human objects. We adopt the entropy and morphological image processing along with machine learning techniques to effectively estimate the crowd density based on those features. Our experiments show that R# achieves high estimation accuracy and processing efficiency. We believe our work takes an important step for human objects estimation using backscatter signals.

Acknowledgments

This work is partially supported by NSFC Grants No. 61802299, 61572396, 61772413, 61672424, 61472184, 61321491, 61632010, Project funded by China Postdoctoral Science Foundation No. 2018M643663, the Jiangsu Innovation and Entrepreneurship (Shuangchuang) Program, and NSF of Jiangsu For Distinguished Young Scientist: BK20150030.

References

[1] T. Teixeira, G. Dublon, and A. Savvides, “A survey of human-sensing: Methods for detecting presence, count, location, track, and identity,” ENALAB Technical Report, vol. 1, 2010, pp. 9–10.

[2] M. Valtonen, J. Maentausta, and J. Vanhala, “Tiletrack: Capacitive human tracking using floor tiles,” in *Proc. IEEE PerCom*, 2009, pp. 1–10.

[3] GS1 EPCglobal Inc., *Specification for RFID Air Interface EPC: Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz-960 MHz*, 2008.

[4] Z. Ma and A. B. Chan, “Crossing the line: Crowd counting by integer programming with local features,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2539–2546.

[5] A. Chan, Z.-S. Liang, and N. Vasconcelos, “Privacy preserving crowd monitoring: Counting people without people models or tracking,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–7.

[6] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, “Electronic frog eye: Counting crowd using WiFi,” in *Proc. IEEE Conf. Comput. Commun.*, 2014, pp. 361–369.

[7] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, “Freecount: Device-free crowd counting with commodity WiFi,” in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–6.

[8] C. Yen-Kai and C. Ronald Y ., “Device-free indoor people counting using Wi-Fi channel state information for Internet of Things,” in *Proc. IEEE IEEE Global Commun. Conf.*, 2017, pp. 1–6.

[9] Y. Zheng, M. Li, and C. Qian, “PET: Probabilistic estimating tree for large-scale RFID estimation,” in *Proc. IEEE 31st Int. Conf. Distrib. Comput. Syst.*, 2011, pp. 37–46.

[10] S. Di Domenico, M. De Sanctis, E. Cianca, P. Colucci, and G. Bianchi, “LTE-based passive device-free crowd density estimation,” in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–6.

[11] J. Weppner and P. Lukowicz, “Bluetooth based collaborative crowd density estimation with mobile phones,” in *Proc. IEEE PerCom*, 2013, pp. 193–200.

[12] S. Shi, S. Sigg, and Y. Ji, "ActiviTune: A multi-stage system for activity recognition of passive entities from ambient FM-radio signals," in *Proc. Int. Conf. Wireless Algorithms Syst. Appl.*, 2013, pp. 221–232.

[13] P. G. Kannan, S. P. Venkatagiri, M. C. Chan, A. L. Ananda, and L.-S. Peh, "Low cost crowd counting using audio tones," in *Proc. ACM SenSys*, 2012, pp. 155–168.

[14] A. Musa and J. Eriksson, "Tracking unmodified smartphones using Wi-Fi monitors," in *Proc. ACM SenSys*, 2012, pp. 281–294.

[15] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," in *Proc. Proc. 12th USENIX Conf. Netw. Syst. Des. Implementation*, 2015, pp. 279–292.

[16] X. Liu, K. Li, H. Qi, B. Xiao, and X. Xie, "Fast counting the key tags in anonymous RFID systems," in *Proc. IEEE IEEE 22nd Int. Conf. Netw. Protocols*, 2014, pp. 59–70.

[17] Y. Guo, L. Yang, B. Li, T. Liu, and Y. Liu, "RollCaller: User-friendly indoor navigation system using human-item spatial relation," in *Proc. IEEE Conf. Comput. Commun.*, 2012, pp. 2840–2848.

[18] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. ACM MobiCom*, 2014, pp. 617–628.

[19] J. Han, H. Ding, C. Qian, W. Xi, Z. Wang, Z. Jiang, L. Shangguan, and J. Zhao, "CBID: A customer behavior identification system using passive tags," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2885–2898, Oct. 2016.

[20] K. Kleisouris, B. Firner, R. Howard, Y. Zhang, and R. P. Martin, "Detecting intra-room mobility with signal strength descriptors," in *Proc. ACM Mobihoc*, 2010, pp. 71–80.

[21] A. Rohit, K. Sudhir, and M. H. Rajesh, "Algorithms for crowd surveillance using passive acoustic sensors over a multimodal sensor network," *IEEE Sensor J.*, vol. 15, no. 3, pp. 1920–1930, Mar. 2015.

[22] Y. Yuan, C. Qiu, W. Xi, and J. zhao, "Crowd density estimation using wireless sensor networks," in *Proc. IEEE WSN*, pp. 138–145, 2011.

[23] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An, "Scpl: Indoor device-free multi-subject counting and localization using radio signal strength," in *Proc. 12th Int. Conf. Inf. Process. Sensor Netw.*, 2013, pp. 79–90.

[24] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proc. ACM MobiCom*, 2014, pp. 237–248.

[25] L. Shangguan, Z. Zhou, and K. Jamieson, "Enabling gesture-based interactions with objects," in *Proc. ACM MobiSys*, 2017, pp. 239–251.

[26] J. Han, C. Qian, W. Xing, D. Ma, J. Zhao, P. Zhang, W. Xi, and J. Zhiping, "Twins: Device-free object tracking using passive tags," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp.1605–1617, June 2016.

[27] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao, "FEMO: A platform for free-weight exercise monitoring with RFIDs," in *Proc. ACM SenSys*, 2015, pp. 3279–3293.

[28] Y. Ma, N. Selby, and F. Adib, "Minding the billions: Ultra-wideband localization for deployed RFID tags," in *Proc. ACM MobiCom*, 2017, pp. 248–260.

[29] T. Wei and X. Zhang, "Gyro in the air: Tracking 3D orientation of batteryless Internet-of-Things," in *Proc. ACM MobiCom*, 2016, pp. 55–68.

[30] F. Giorgio and S. Sabatino, *Wireless Networks: From the Physical Layer to Communication, Computing, Sensing and Control*. New York, NY, USA: Academic, 2006.

[31] J. Zhang, Z. Qin, L. Ou, P. Jiang, J. Liu, and A. Liu, "An advanced entropy-based DDOS detection scheme," in *Proc. IEEE Int. Conf. Inf. Netw. Autom.*, 2010, pp. V2–67–V2-71.

[32] M. Nakatsuka, H. Iwatani, and J. Katto, "A study on passive crowd density estimation using wireless sensors," in *Proc. IEEE ICMU*, pp. 1–7, 2008.

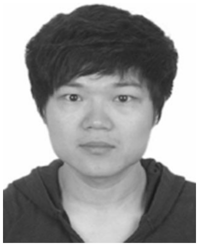
[33] R. M. Bolle, J. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, *Guide to Biometrics*. New York, NY, USA: Springer, 2003.

[34] R. A. Brualdi, *Introductory Combinatorics (5th Edition)*. Englewood Cliffs, NJ, USA: Prentice Hall, 2010.

Han Ding received the PhD degree in computer science and technology from Xi'an Jiaotong University, in 2017. She is currently an assistant professor with Xi'an Jiaotong University. Her research interests include RFID system and smart sensing. She is a member of the IEEE.

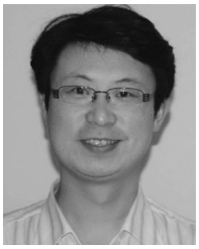
Jinsong Han received the PhD degree in computer science from Hong Kong University of Science and Technology, in 2007. He is now a professor at Xi'an Jiaotong University. His research interests include mobile computing, RFID, and wireless network. He is a senior member of the IEEE and ACM.

Alex X. Liu received the PhD degree in computer science from the University of Texas at Austin, in 2006. He received the IEEE & IFIP William C. Carter Award, in 2004, an NSF CAREER award in 2009, and the Withrow Distinguished Scholar Award in 2011 at Michigan State University. He is an associate editor of the *IEEE/ACM Transactions on Networking*, an editor of the *IEEE Transactions on Dependable and Secure Computing*, and an area editor of the *Computer Communications*. He received Best Paper Awards from ICNP-2012, SRDS-2012, and LISA-2010. His research interests include networking and security. He is a senior member of the IEEE.



Wei Xi received the PhD degree in computer science computer science and technology from Xi'an Jiaotong University, in 2014. He is now an associate professor with Xi'an Jiaotong University. His research interests include wireless networks, smart sensing, and mobile computing. He is a member of the IEEE, CCF, and ACM.

Jizhong Zhao received the PhD degree in computer science computer science and technology from Xi'an Jiaotong University, in 2001. His research interests include computer software, pervasive computing, distributed systems, network security. He is a member of the IEEE, CCF, and ACM.



Panlong Yang received the PhD degree in communication and information system from Nanjing Institute of Communication Engineering, in 2005. He was a visiting scholar in HKUST during September 2010 to September 2011. He is a professor in the School of Computer Science and Technology, University of Science and Technology of China. He is a member of the IEEE Computer Society and ACM SIGMOBILE Society. His research interests include mobile communication, wireless network, wireless sensor network, and social network analysis. He is a member of the IEEE.

Zhiping Jiang received the PhD degree in computer science and technology from Xi'an Jiaotong University in 2017. He is currently an assistant professor with Xi'dian University. His research interests focus on localization, smart sensing, wireless communication, and image processing.