Human Object Estimation via Backscattered Radio Frequency Signal

Han Ding*, Jinsong Han*, Alex X. Liu†‡, Jizhong Zhao*, Panlong Yang§, Wei Xi* and Zhiping Jiang*
*School of Electronic and Information Engineering, Xi'an Jiaotong University, China
†Department of Computer Science and Engineering, Michigan State University, USA
‡National Key Laboratory for Novel Software Technology, Nanjing University, China
§Institute of Communication Engineering, PLA University of Science and Technology, China
Email: dinghan.331@stu.xjtu.edu.cn, {hanjinsong, zjz}@mail.xjtu.edu.cn,
alexliu@cse.msu.edu, {panlongyang, weixi.cs, jiangzp.cs}@gmail.com

***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 variance in the RSS values of the tag backscattered RF signal. Thus, based on the received RF signal, 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 painting. The RFID reader periodically emits RF signal to identify all tags and the tags simply respond with their IDs via C1G2 standard protocols. 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%).**

I. 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 signal that the tags backscatter to the reader demonstrates different patterns*; thus, based on the patterns of the received RF signal, 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 painting. The RFID reader periodically emits RF signal to identify all tags and the tags simply respond with their IDs via protocols in the EPCGlobal Class-1 Generation-2 (C1G2) RFID standard [3]. The human object estimation module runs on RFID readers.

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 signal. First, the information that we can extract from RF signal 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. 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, and the RSS values of different tags at the same location also vary significantly. Third, the RF signal pattern is difficult to model and quantify.

978-1-4799-8381-0/15/$31.00 ©2015 IEEE

Our R# system addresses these challenges based on our refined observation that *the more human objects are present, the higher the variance in the RSS values of the tag backscattered RF signal*. To correlate the number of human objects and the RSS value variances, 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. Note that for better accuracy and larger detection coverage, we use multiple (say 20) passive RFID tags. 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 larger the area size of the connected white pixels after dilation is. 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 larger the MSE is.

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% of our tests, the estimated value of R# exactly matches the real value; and in nearly 98% of our tests, the estimated value of R# deviates from the real value by at most one person.

The rest of the paper proceeds as follows. In Section II, we explain our observation and analysis on the RSS value of tags. In Section III, we present the system design of R#. In Section IV, we describe our implementation and evaluation results. In Section V, we discuss the deployment issues of R# in practice. In Section VII, we discuss related work. We conclude the paper in Section VII.

II. Observation and Analysis

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

A. Backscatter Communication and RF Signal

Passive RFID readers utilize the backscatter communication to interact with tags. Other than tag IDs, the backscattered RF signal also contains 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.

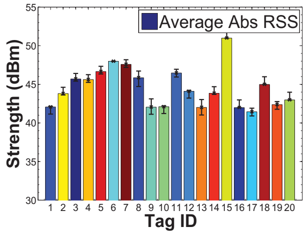


Fig. 1. The RSS collected from different tags at a given location.

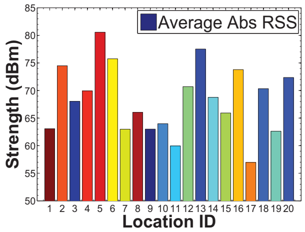


Fig. 2. The RSS collected from a tag at 20 different locations.

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. Indeed, the CW emitted from readers induces a current within the tag such that this tag can accumulate sufficient energy for above modulation.

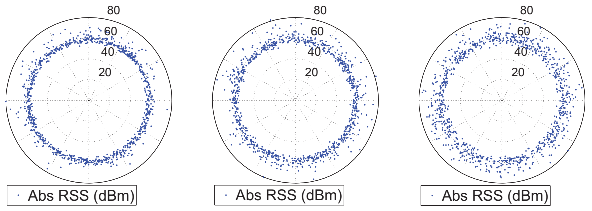
The Received Signal Strength (RSS) is a measurement on the power of the tag backscattered signal. Let s represents the signal the reader receives from the tag. The power can be expressed as:

$$P = s^2/R \tag{1}$$

where R is a coefficient from Ohm's law. The RSS is measured in dB relative to 1 mW: $RSS = 10\log(P/1mW)$.

B. Observation

Our goal is to correlate the number of human objects to the features of RF signal in the region that we want to monitor (region for short in the following). However, there are two challenges in realizing this goal, sufficiency and stability. As aforementioned, the information we obtain from COTS passive readers are only the RSS, phase, Doppler Shift, and tag IDs. Compared to the fine-grained information, such as the Channel State Information (CSI) used in Wi-Fi [4], the information reported by RFID readers is in low resolution and noisy. On the other hand, the information itself, unfortunately, is not stable 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 distance from the reader to tag is also fixed. As shown in Fig. 1, the color bars represent the Abs RSS values 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



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

Fig. 3. The RSS distribution vs. # of moving human objects.

caused by the tag diversity may introduce uncertainty to the correlation. (2) Location sensitivity. In practice, the RSS of tags is 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.

The above challenges motivate us to consider the problem in a different way. Intuitively, ignoring the differences among individual tags or locations, 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 signal. We conduct a series of *proof-of-concept* experiments to validate this ieda. As shown in Fig. 4, we deploy 8 passive tags in a line, with an distance $20cm$ in between. We invite some (say 0∼4) volunteers to walk in the region between the reader antenna and tags, and collect the RSS values of tags within a short period, $e.g.$, $5s$. The results are shown in Fig. 3. 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. Note that we use the absolute value as a measurement on the collected RSS in the following text. 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 variance and the number of human objects. To prove its feasibility, we try to analyze this phenomenon based on the multipath effect of RF signal.

C. Analysis

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

With the multipath effect, 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 signal. This can be modeled as [5]:

$$R_i(t) = \sum_{n=1}^{mp} H_n x(t - \tau_n), \tag{2}$$

where i is the number of human objects in the region at the time point t, mp is the number of distinct multipaths 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 an random variable $\Re_0 = \{R_0(t)|t > 0\}$.

We then assume that there are certain human objects walking around in the region, as illustrated in Fig. 4. When one human object enters the region, he or she may introduce some variations to the RF signal in the channel. We denote the signal variation incurred by this human object as $\Re_1 = \{R_1(t)|t > 0\}$. The wireless channel between the reader and tags can be modeled as a linear time-varying system [5]. Based on the additivity in such systems, the RF signal can be expressed as: $\Re_1^{sum} = \Re_0 + \Re_1 = \{R_1^{sum}(t)|R_1^{sum}(t) = R_0(t) + R_1(t), t > 0\}$. Due to the independence of $\Re_0$ and $\Re_1$, we can get the variance of $\Re_1^{sum}$:

$$\begin{aligned} Var(\Re_1^{sum}) &= Var(\Re_0 + \Re_1) = Var(\Re_0) + Var(\Re_1) \\ &\geq Var(\Re_0) \end{aligned} \tag{3}$$

The equality holds iff $\Re_1$ is a constant. This 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. So the variance in the RF signal is monotonically increasing when the number of human objects increases.

Similarly, we can infer the variance relationship between any two random variables of RF signal as:

$$\forall i, j \in N^+, if\ j > i, Var(\Re_j^{sum}) > Var(\Re_i^{sum}) \tag{4}$$

Given that the number of multipaths is large in the indoor experiment, the central limit theorem implies that if all of the path gains do not vary greatly in their distribution, then the resultant complex channel gain will converge to a Gaussian random variable [5]. In the wireless channel, the phase of each multipath gain is equally likely distributed at anywhere in $[-\pi, \pi]$, which indicates that $E(\Re_i^{sum})$ = 0.

According to Equation 1, the received power at the reader antenna is: $P_i = (\Re_i^{sum})^2$. Let σ_i denote the signal variance with i human objects. According to above analyses, we have $\Re_i^{sum} \sim N(0, \sigma_i^2)$. We normalize $\Re_i^{sum}$ as $RN_i = \frac{\Re_i^{sum} - 0}{\sigma_i} \sim N(0, 1)$. Then we have

$$(RN_i)^2 = \frac{(\Re_i^{sum})^2}{\sigma_i^2} = \frac{P_i}{\sigma_i^2} \sim \chi^2(1) \tag{5}$$

We apply the variance operation on both sides of Equation 5,

$$Var(\frac{P_i}{\sigma_i^2}) = Var(\chi^2(1)) = 2 \tag{6}$$

Because σ_i^2 is a constant when i is fixed, Equation 6 can be transformed to

$$Var(P_i) = 2\sigma_i^2 \tag{7}$$

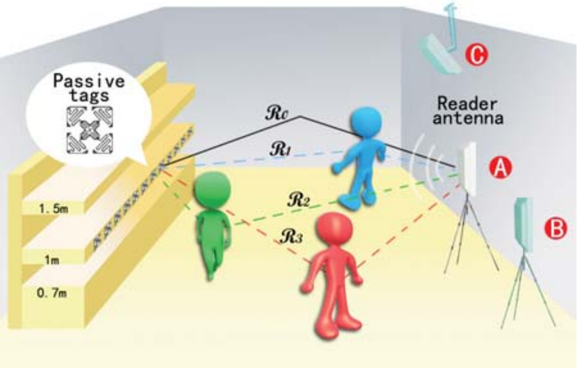


Fig. 4. The multipath propagation of RF signal in R#.

Therefore, according to our hypothesis, for $\forall i, j \in N^+$, if $j > i$, we have the relationship between the variance of powers,

$$Var(P_j) = 2\sigma_j^2 > 2\sigma_i^2 = Var(P_i) \tag{8}$$

It is worth noting that RSS is an average measurement on the power of received signal. Based on above analysis, we know that when the number of human objects grows, the power of received signal varies in a larger range, resulting in an increase on the variance in the RSS values.

It indicates us that we can use the variance in the RSS values of tags to estimate the number of human objects. However, directly using the variance of RSS cannot achieve accurate estimations. We will demonstrate this in Section IV-B.

III. System Design

In this section, we first overview our system R#, then present the three features and detail the design of R#.

A. 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 (say 20) of passive tags using protocols in EPC C1G2 [3]. In particular, the tags are deployed in a line, as illustrated in Figure 4. The reader periodically collects the tag IDs and RSS values. For simplicity, we call an RSS sequence collected from such 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*. In the first phase, R# preprocesses the raw RF signal data collected from the reader 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 signal. Finally, R# employs machine learning techniques for human object estimation.

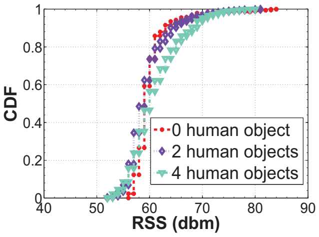


Fig. 5. CDF of RSS

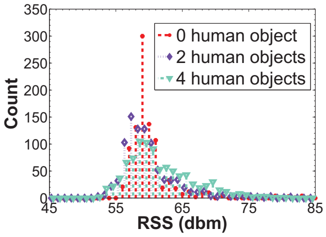


Fig. 6. Distribution of RSS

B. Data Preprocessing

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

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

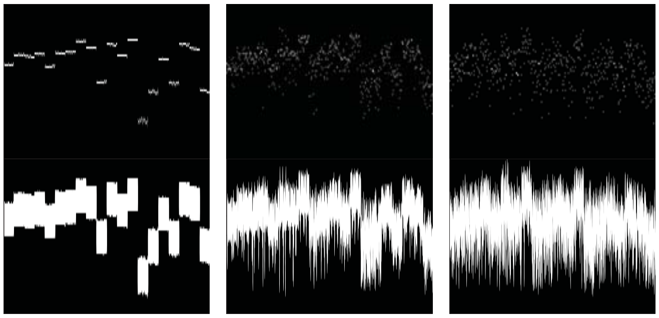
Again due to the slotted Aloha mechanism, the slot in which a tag replies is randomly and uncontrollably distributed, leading various numbers of samples from different tags in one observation. In addition, ambient factors, such as the shadowing, interference, and tag location, also influence the backscattered signal, resulting in different sampling periods to these tags. Suppose K is the maximum number of samples collected from a tag within one observation. To avoid the unfairness in the sampling and later processes, for each tag we use a Linear Interpolation method to virtually increase the number of its samples to K.

C. Feature Extraction

In R#, our expectation on a feature of the RF signal is that it monotonously increases if enlarging the number of human objects, or vice versa. To this end, we select three features, *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 (32dBm in our implementation) and fixed location for each tag, the collected RSS values vary within a range, denoted as $[R_{mi}, R_{ma}]$. As we analyzed in Section II, the human interference may either strengthen or offset the overlapped signal at the reader antenna. Correspondingly, the R_{mi} and R_{ma} may change as well.

We setup 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 Fig. 5 and Fig. 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#.



(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.

According to the information theory, the 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 [6]. 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 (BL), where $N = \lceil (R_{ma} - R_{mi})/BL \rceil$. 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:

$$E(X) = -\sum_{i=1}^{N} p_i \cdot \log(p_i) \qquad (9)$$

• ***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 the points could covered. To formalize this observation,

we first binary-visualize an observation with the following operations.

To simplify the subsequent processes, we extend the range of $[R_{mi}, R_{ma}]$ to $[R'_{mi}, R'_{ma}]$. Since the RSS resolution of our RFID readers is 0.5 dBm, we then normalize each RSS value x with the operation $(x - R'_{mi}) * 2$. Every x is within $[0, (R'_{ma} - R'_{mi}) * 2]$.

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 $(R'_{ma} - R'_{mi}) * 2$. We then set all elements $R(k, x_i)$ to 1, where $1 \leq k \leq l, 1 \leq i \leq c$. 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. 7(a), (b), and (c) show the results of performing the binary visualization operation on the observations with 0, 2, and 4 human objects, respectively.

The core operation of MIPS is the morphological dilation, 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: $A \bigoplus B = \{z | ((\hat{B})_z \bigcap A) \neq \emptyset\}$, 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. 7 (a), (b), and (c) demonstrate the dilation results of upper original binary images. We can see 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 in the surveillance region, the area of white pixels becomes larger in the dilated image.

After the dilation, we find 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 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 independent feature, the Mean Squared Error (MSE). By compressing $f(x, y)$ using Discrete Cosine Transform (DCT) followed by an de-compression, we obtain an recovered image $\hat{f}(x, y)$. We find that the *fidelity* between f and $\hat{f}$ can be used for characterizing the observations. Specifically, the Mean Squared Error (MSE) represents the fidelity feature. MSE is normally used to measure the information loss. We notice that when there are more human objects moving in the area, the image of the observation will be more complicated. Correspondingly, when doing DCT compression, more information will lose.

Considering a $M \times N$ image $f(x, y)$, the forward DCT $T(u, v)$ can be expressed as:

$$T(u, v) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} f(x, y) g(x, y, u, v)$$

$$g(x, y, u, v) = a(u) a(v) k(u, v)$$

$$k(u, v) = \cos[\frac{(2x+1)u\pi}{2M} \cos[\frac{(2y+1)v\pi}{2N}]]$$

$where,$

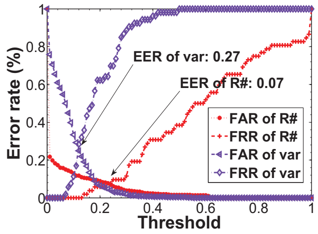
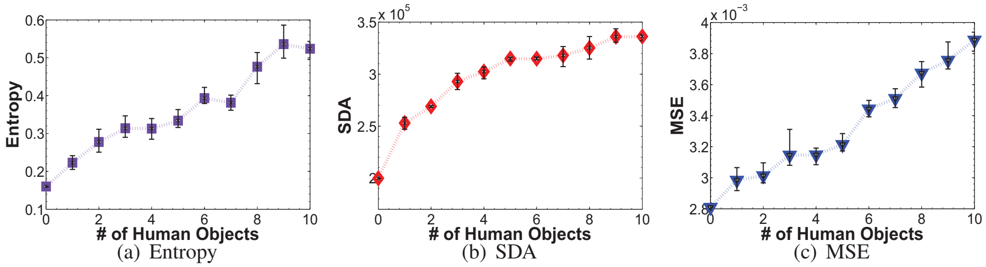


Fig. 8. The error rate of R# vs. the error rate of variance-based method.



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

Fig. 9. Three features used by R#.

$$a(u) = \begin{cases} \sqrt{\frac{1}{M}} & \text{u = 0} \\ \sqrt{\frac{2}{M}} & \text{u = 1, 2, ..., M-1} \end{cases}$$

$$a(v) = \begin{cases} \sqrt{\frac{1}{N}} & \text{v = 0} \\ \sqrt{\frac{2}{N}} & \text{v = 1, 2, ..., N-1} \end{cases}$$

DCT compression is an invertible transform and can be easily recovered. A brief procedure of DCT compression is as follows: we first divide the original image into some 8×8 subimages. We then use the DCT to represent these subfigures, and discard a part of the coefficients achieved (85% in our implementation). When recovering the image, we run the reverse DCT on the intercepted coefficient matrix. Although the discarded coefficients have little visual influence on the recovered image, they still incur MSE. The incurred MSE between $f(x, y)$ and $\hat{f}(x, y)$ is:

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

D. Machine Learning based Estimation

Before the design of estimation mechanism, we exam the effectiveness of three features.

We arrange 0∼10 volunteers to move arbitrarily in the region and check the trend of features with varying the number of volunteers. The result is reported in Fig. 9 (a), (b), (c). 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∼4 persons for SDA and 6∼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 IV.

IV. Implementation and Evaluation

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

A. 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 a gain of 8dbi, and 20 commodity passive tags ($i.e.$, Impinj H47 with size of $44mm \times 44mm$). Those tags are attached to a number of cartons arranged in a line. The space between two adjacent tags is $20cm$ and the distance from the reader antenna to the line of tags is $3.5m$. Fig. 10 shows our experiment scenario. To avoid the influence caused by frequency hopping, we fix the communication frequency of the reader on 924.375 MHz, which is a frequency conforming to the specification in the EPC Global C1G2. We develop a software for the data collection. The software communicates with the reader through LLRP (Low Level Reader Protocol) toolkit released by the Impinj company.

B. Parameter tuning

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

First, 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 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

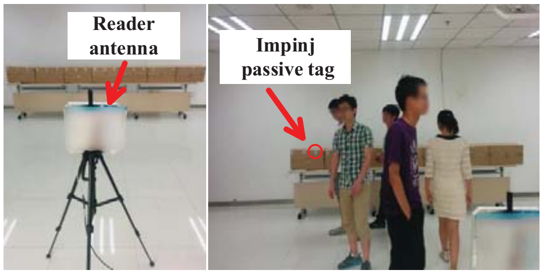


Fig. 10. The scenario in our implementation.

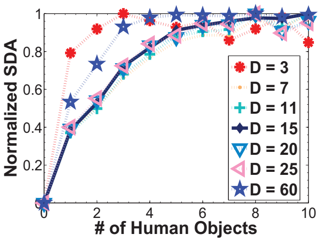


Fig. 11. SDA vs. element size

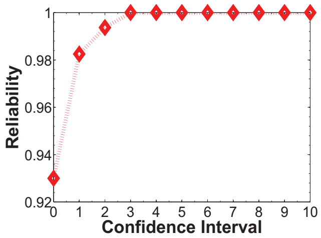


Fig. 12. Reliability of R#.

T, we investigate the relationship between the FAR and FRR of R# and report the EER of R# in Fig. 8. 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%. Recall that directly using the variance of RSS as a feature is also possible. Similar to above process, we use the variance of RSS values as the feature and compare the classification accuracy with that of R#. The result clearly shows that both the FAR and FRR are higher than those of R#, as shown in Fig. 8. In particular, the EER is 0.27 if using the variance as the feature, implying a significant estimation error compared to R#. Thus, we do not recommend to directly use the variance of RSS for estimation.

Second, 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 with a radius D. Thus, we only evaluate the impact from different size of structuring elements. Since under different Ds, the dilated area of white pixels varies considerably, we normalize the area of every test into the range of [0, 1], to examine the influence of various Ds on the numerical discrimination together. Fig. 11 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. We can see 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 an 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. From the results, we learn that only with a proper setting of D, the dilated area is able to show an analogical increasing relationship to the number of human objects. To this end, we compare different settings of D, and select the proper one as $D = 15$ in our implementation.

C. Evaluation on the estimation accuracy

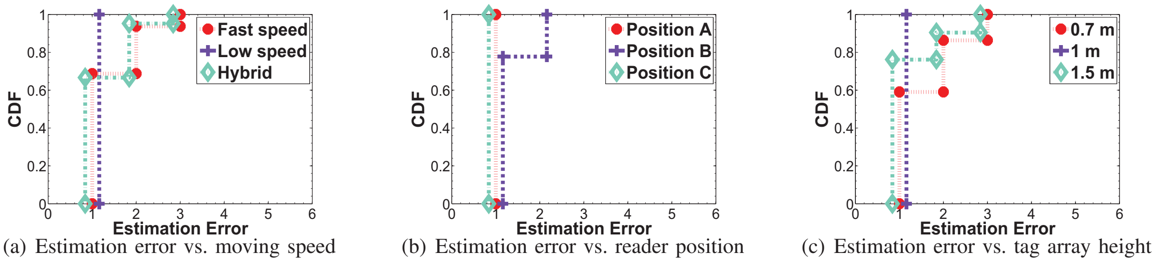
After determining the system parameters, we use the following evaluation strategy: given a *confidence interval* $\beta \in [0, n]$, some human objects within the monitoring region of unknown number k ($k \leq n$), and an R#'s estimation on the number of human objects is $\tilde{k}$, we use the *reliability*, $\alpha = P\{|\tilde{k} - k| \leq \beta\}$, for evaluating the accuracy. Note that the confidence interval is an integer, representing the estimation error range of R#. An 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 above tests is 80% with the confidence interval as 0, 90% with the confidence interval as 1. 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. 12. The result shows that R# achieves highly accurate estimations on human objects. For example, the reliability of R# is 93% and 98%, with the confidence interval as 0 and 1, respectively.

V. Deployment

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

A. The speed of human objects

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 IV, 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. 13(a) shows the CDF of the estimation error. At the low speed mode, in 100% 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% 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 accurately estimations.



(a) Estimation error vs. moving speed (b) Estimation error vs. reader position (c) Estimation error vs. tag array height

Fig. 13. CDF of estimation error of different deployments.

B. The position of reader antennas

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. 4. Observing the result shown in Fig. 13 (b), 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% 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* pattern in practice.

C. The height of tags

We also evaluate R# when changing the height of tags. Fig. 13 (c) compares the deviation of estimated values of R# with the height of tags as $0.7m$, $1m$, and $1.5m$, respectively. We see that when with the height of $0.7m$, R# has the worst performance and only 60% 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 $0.7m$ 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#.

D. Detection region

The maximum number of human objects R# estimates is 10 in our experiments. This is highly correlated with the size of the detection region ($13m^2$) in this paper, which is limited by the reader transmitted power and tag sensitivity. When too many people stay in the region, the influence on the backscattered signal of tags will reach a saturation point. Hence the features of R# become indistinguishable for estimation. To improve the estimation capability of R#, we can adjust the position of reader antenna, like the choice of A position in our deployment, or make the detection region as large as possible by augmenting the transmission power of readers.

The maximum number of human objects R# estimates is 10 in our experiments. This is highly correlated with the size of the detection region ($13m^2$) in this paper, which is limited by the reader transmitted power and tag sensitivity. When too many people stay in the region, the influence on the backscattered signal of tags will reach a saturation point. Hence the features of R# become indistinguishable for estimation. To improve the estimation capability of R#, there are two possible ways. First, we can adjust the position of reader antenna, like the choice of A position in our deployment, or make the detection region as large as possible by augmenting the transmission power of readers. However, doing so may be constrained in the real deployment and not allowed by the regulation institution, such as the Federal Communications Commission (FCC). In fact, merely increasing the transmission power is not proper, since it may involve the movement of human objects in adjacent regions. The second way is to increase the number of tags for collecting more detail information to extract the features of RF signal, which is helpful for differentiating the numbers of human objects. For example, one possible solution is to deploy a tag grid instead of a tag array. In our future work, we will design more efficient scheme for processing the RF signal with more tags.

VI. Related Work

Traditionally, people count or estimate individuals present in the scene via mechanical barriers. Later, the binary sensor use break-beams, 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]. In recent decades, crowd counting or estimation has been widely studied in the computer vision literature. Computer vision based techniques usually leverage the pattern recognition technique to detect the individuals in presence, based on either their shapes or motions. Z. Ma and A. B. Chan propose an integer programming method for estimating the pedestrians crossing a line of interest using video cameras [7]. They use a regression function to map the local features to a count. H. Idrees et al. leverage multiple sources of information to estimate dense crowd [8]. A. B. Chan et al. present a privacy-preserving crowd counting method by segmenting the crowd and analyzing their holistic properties [9] in the video. There are many challenges when using computer vision based methods for human object estimation, such as the cross trajectories, lighting variations, shadows and so on. Dealing with other concerns, such as the deployment cost and recognition accuracy, is also non-trivial for computer vision techniques.

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 [4], backscattered RF signal [10], Bluetooth [11], FM [12], and audio [13] [14], 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 or RFID tags [15] [16], carried by the individuals. On the other hand, wireless signal is susceptible to environment variations or object movements [17] [18] [19] [20], providing a potential way to device-free crowd estimation. Nakatsuka et al. [21] towards the experimentally demonstrate the feasibility of using the RSS mean and variance to count the number of people moving between two sensor nodes. Wei et al. [4] propose to use Channel State Information (CSI) for crowd estimation using commercial 802.11n devices. Their work presents a feature named as the Percentage of nonzero Elements (PEM) from CSI information and a Grey Verhulst Model to characterize the crowd density. In [22], Xu et al. utilize the RSS mean difference derived from the wireless links to count device-free objects in large-scale deployments.

Considering the extremely constrained capacity of passive RFID tags, it is impossible to adopt above sensing devices or techniques. 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 or motion detection [23], [24], [25], [26].

VII. Conclusions

In this paper, we propose R# for estimating human objects. R# is based on the observation on a phenomenon in our empirical study, more human objects incur a larger variance to the RF signal 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.

Acknowledgment

This work was supported in part by NSFC under Grant No. 61190112, 61325013, 61373175, 61402359, 61472184, 61321491, 61272546; the Natural Science Basic Research Plan in Shaanxi Province of China under Grant No. 2014JQ832; the Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 20130201120016; the Fundamental Research Funds for the Central Universities of China (Xian Jiaotong University); the Jiangsu Future Internet Program under Grant Number BY2013095-4-08, and the Jiangsu High-level Innovation and Entrepreneurship (Shuangchuang) Program. Jinsong Han is the corresponding author.

References

[1] T. Teixeira, G. Dublon, and A. Savvides, “A Survey of Human-sensing: Methods for Detecting Presence, Count, Location, Track, and Identity,” *ACM Computing Surveys*, vol. 5, 2010.

[2] M. Valtonen, J. Maentausta, and J. Vanhala, “Tiletrack: Capacitive Human Tracking using Floor Tiles,” in *Proceedings of IEEE PerCom*, 2009.

[3] EPCglobal, *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] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, “Electronic Frog Eye: Counting Crowd Using WiFi,” in *Proceedings of IEEE INFOCOM*, 2014.

[5] F. Giorgio and S. Sabatino, *Wireless Networks: From the Physical Layer to Communication, Computing, Sensing and Control*. Academic Press Inc, 2006.

[6] J. Zhang, Z. Qin, L. Ou, P. Jiang, J. Liu, and A. Liu, “An Advanced Entropy-based DDOS Detection Scheme,” in *Proceedings of IEEE ICINA*, 2010.

[7] Z. Ma and A. B. Chan, “Crossing the Line: Crowd Counting by Integer Programming with Local Features,” in *Proceedings of IEEE CVPR*, 2013.

[8] H. Idrees, I. Saleemi, C. Seibert, and M. Shah, “Multi-source Multi-scale Counting in Extremely Dense Crowd Images,” in *Proceedings of IEEE CVPR*, 2013.

[9] A. Chan, Z.-S. Liang, and N. Vasconcelos, “Privacy Preserving Crowd Monitoring: Counting People Without People Models or Tracking,” in *Proceedings of IEEE CVPR*, 2008.

[10] Y. Zheng, M. Li, and C. Qian, “PET: Probabilistic Estimating Tree for Large-Scale RFID Estimation,” in *Proceedings of IEEE ICDCS*, 2011.

[11] J. Weppner and P. Lukowicz, “Collaborative Crowd Density Estimation with Mobile Phones,” in *Proceedings of ACM PhoneSense*, 2011.

[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 *Proceedings of WASA*, 2013.

[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 *Proceedings of ACM SenSys*, 2012.

[14] A. Musa and J. Eriksson, “Tracking unmodified smartphones using wi-fi monitors,” in *Proceedings of ACM SenSys*, 2012.

[15] X. Liu, K. Li, H. Qi, B. Xiao, and X. Xie, “Fast Counting the Key Tags in Anonymous RFID Systems,” in *Proceedings of IEEE ICNP*, 2014.

[16] L. Kong, L. He, Y. Gu, M. Y. Wu, and T. He, “A Parallel Identification Protocol for RFID Systems,” in *Proceedings of IEEE INFOCOM*, 2014.

[17] Y. Guo, L. Yang, B. Li, T. Liu, and Y. Liu, “RollCaller: User-Friendly Indoor Navigation System Using Human-Item Spatial Relation,” in *Proceedings of IEEE INFOCOM*, 2012.

[18] D. Ma, C. Qian, W. Li, J. Han, and J. Zhao, “GenePrint: Generic and Accurate Physical-Layer Identification for UHF RFID Tags,” in *Proceedings of IEEE ICNP*, 2013.

[19] 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 *Proceedings of ACM MobiCom*, 2014.

[20] J. Han, H. Ding, C. Qian, D. Ma, Z. Wang, W. Xi, Z. Jiang, and L. Shangguan, “CBID: A Customer Behavior Identification System Using Passive Tags,” in *Proceedings of IEEE ICNP*, 2014.

[21] M. Nakatsuka, H. Iwatani, and J. Katto, “A Study on Passive Crowd Density Estimation using Wireless Sensors,” in *Proceedings of IEEE ICMU*, 2008.

[22] 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 *Proceedings of ACM IPSN*, 2013.

[23] 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 *Proceedings of ACM MobiCom*, 2014.

[24] J. Wang and D. Katabi, “Dude, Where’s my card?: RFID Positioning that Works with Multipath and Non-line of Sight,” in *Proceedings of ACM SIGCOMM*, 2013.

[25] L. Shangguan, Z. Li, Z. Yang, M. Li, Y. Liu, and J. Han, “OTrack: Order Tracking for Luggage in Mobile RFID Systems,” *Transactions on Parallel and Distributed Systems*, vol. 25, pp. 2114–2125, 2014.

[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,” in *Proceedings of IEEE INFOCOM*, 2014.