

Electronic Frog Eye: Counting Crowd Using WiFi

Wei Xi[†], Jizhong Zhao[†], Xiang-Yang Li^{*}, Kun Zhao[†], Shaojie Tang[‡], Xue Liu[§], Zhiping Jiang[†]

[†]School of Electronic and Information Engineering, Xi'an Jiaotong University

^{*} Department of Computer Science, Illinois Institute of Technology

[‡] Department of Computer and Information Science, Temple University

[§] School of Computer Science, McGill University

Abstract—Crowd counting, which count or accurately estimate the number of human beings within a region, is critical in many applications, such as guided tour and crowd control. A crowd counting solution should be scalable and be minimally intrusive (*i.e.*, device-free) to users. Image-based solutions are device-free, but cannot work well in a dim or dark environment. Non-image based solutions usually require every human being carrying device, and are inaccurate and unreliable in practice. In this paper, we present FCC, a device-Free Crowd Counting approach based on Channel State Information (CSI). Our design is motivated by our observation that CSI is highly sensitive to environment variation, like a frog eye. We theoretically discuss the relationship between the number of moving people and the variation of wireless channel state. A major challenge in our design of FCC is to find a stable monotonic function to characterize the relationship between the crowd number and various features of CSI. To this end, we propose a metric, the Percentage of nonzero Elements (PEM), in the dilated CSI Matrix. The monotonic relationship can be explicitly formulated by the Grey Verhulst Model, which is used for crowd counting without a labor-intensive site survey. We implement FCC using off-the-shelf IEEE 802.11n devices and evaluate its performance via extensive experiments in typical real-world scenarios. Our results demonstrate that FCC outperforms the state-of-art approaches with much better accuracy, scalability and reliability.

I. INTRODUCTION

Robust crowd counting is an important yet challenging task. It is of great interest in a number of potential applications, such as guided tour, and crowd control, *etc.* Crowd behaviors, however, are often unpredictable. Thus, crowd counting (or estimation) may face various challenges, including, but are not limited to, reliable observation collection, object occlusions, and real-time processing requirement. Traditional approaches that tackle these issues are mainly classified into two categories: video (or image) based recognition and non-image based localization.

Video (or image) based recognition has been widely deployed in many public places [1]. However, these methods have inherent drawbacks. First, cameras can only work in a line-of-sight pattern, leading many blind areas to the monitoring. Second, the environmental contribution of smoke or lacking of light will severely degrade the visual quality of cameras. Third, objects overlapping further deteriorates the performance. Furthermore, the use of cameras poses privacy concerns.

Non-image based solutions typically leverage radio devices to locate objects, such as RFID tags, mobile phones, sensor nodes, *etc.* These device-based approaches [2], [3], [4], [5],

[6], [7], [8] require people to carry certain devices for surveillance, which significantly constrains the usage scope. For a public area with mass people, distributing the device to each person is impractical and costly, and may not be doable under emergent events.

Some device-free approaches have been proposed recently. Most approaches employ RSS fingerprints for localization [9], [10], [11], [12], which can be easily obtained for most off-the-shelf wireless devices. However, site survey is time-consuming, labor-intensive, and easily affected by environmental dynamics. To avoid site survey, the model based localization approaches use RSS as an indication of signal propagating distance to locate objects. Unfortunately, attenuation models perform poorly due to multipath propagation in complex indoor environments. One of the main trends recently is to explore the fine-grained physical layer information to locate or count people in multipath environments [13], [14], which are usually based on USRP.

Channel State Information (CSI) from OFDM-based system has new potential to overcome above drawbacks. Different from RSS, CSI, a fine-grained value derived from the physical layer, refers to known channel properties of a communication link. This information describes how a signal propagates from the transmitter to the receiver and represents the combined effect of, for example, scattering, fading, and power decay with distance. It consists of the attenuation and phase shift experienced by each spatial stream on every subcarrier in the frequency domain. Therefore, CSI is more sensitive to environmental variance owing to moving object. In practice, we can acquire these information quickly and easily from some off-the-shelf 802.11n devices such like Intel 5300 NIC.

In this paper, we propose FCC, a device-Free Crowd Counting approach based on CSI measurements. The rationale behind FCC is that crowd motions can be accurately inferred from the change of CSI readings. The contributions of this work are summarized as follows.

- We theoretically study and experimentally validate the relationship between the number of moving people and the variation of wireless channel. The results suggest that CSI is sensitive to the environment influences, and there is a monotonic relationship between CSI variations and the number of moving people. This lays a solid foundation for crowd counting.
- We propose a metric PEM, the Percentage of non-zero Element, in the dilated CSI Matrix, to indicate the crowd size. PEM can adaptive reflect the changes of CSI in a

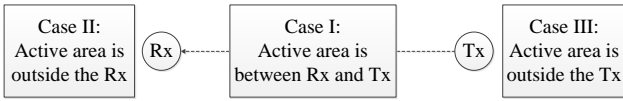


Fig. 1. Moving people affect the signal transmission in 3 cases

short period of time. The value of PEM is convex and increases with an increasing number of active people, and it reaches the saturated state when crowd density reached a certain threshold. We apply the Grey Verhulst model to directly estimate the number of people from PEM measurements without fingerprinting.

- To count/estimate the number of people in a large monitoring area, we deploy multiple FCC devices, organized in a grid array. The challenge here is that as CSI has a high sensitivity to environment, people moving in one grid will cause CSI fluctuations in other adjacent grids. To address this challenge, we propose an interference cancelation method to adaptively adjust the sensing range of each receiver for improving the counting accuracy in a large monitoring area.
- We implement our FCC scheme with off-the-shelf 802.11n device and evaluate it with extensive real-world experiments. The results demonstrate that our approach outperforms existing approaches in terms of higher accuracy, scalability, and reliability.

The rest of this paper is organized as follows. Section II presents the theoretical analysis and real-world observations of the relationship between the number of moving people and the variation of wireless channel. The design of FCC is elaborated in Section III, followed by performance evaluation in Section IV. Section V briefly reviews the related work. We conclude this paper in Section VI.

II. THEORETICAL ANALYSIS AND OBSERVATION

In radio communications, the emitted electromagnetic waves often do not reach the receiving antenna directly due to obstacles blocking the line-of-sight path. In fact, the received waves are a superposition of waves coming from all directions due to reflection, diffraction, and scattering caused by furniture, people, and other obstacles. This effect is known as multipath propagation.

In current widely used Orthogonal Frequency Division Multiplexing (OFDM) systems, Channel State Information (CSI) refers to known channel properties of each subcarrier. This information describes how a signal propagates from the transmitter to the receiver and represents some other combined effect. In a narrowband flat-fading channel, the OFDM system in the frequency domain is modeled as

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

where Y and X are the receive and transmit vectors, respectively, and H and N are the channel matrix CSI and the noise vector, respectively. Therefore, the estimated value of H can be expressed as

$$\hat{H} = \frac{Y}{X} \quad (2)$$

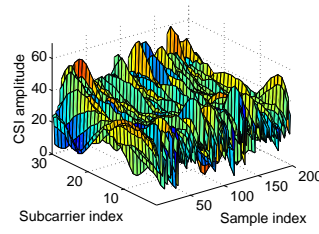


Fig. 2. Measured CSI values from all subcarriers

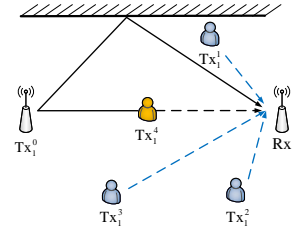


Fig. 3. Simple Scenario of Multipath Effect

The variation of CSI amplitude from 30 subcarriers is shown in Figure 2. Figure 3 illustrates how the people influence the signal propagation. The black solid lines demonstrate the transmission paths from Tx_1^0 to Rx in a stationary scene. The receive vector is denoted as Y^0 and the corresponding CSI is denoted as H^0 . The H^0 can be considered as a constant in a certain scenario with no person. Each person may reflect or block signal and cause the receive vector being strengthened or weakened. It can be regarded as a virtual antenna, and the receive vectors from virtual antennas are denoted as Y^i ($i = 1, 2, 3, \dots$). The virtual antennas have the following properties:

- The Rx cannot distinguish the signals from Tx_1^0 and Tx_1^1 , because the signals transmitted from all the virtual antennas actually originate from Tx_1^0 , and the transmit vectors X^i are the same and equal to X^0 from Tx_1^0 .
- Y^i obeys normal distribution $N(\mu_n, \sigma_n^2)$, where μ_n and σ_n^2 are functions that depends on the number of Tx_1^1 . In other words, Y^i is independent and identically distributed with the condition of the same number of Tx_1^1 .

Thus, the CSI between Rx and Tx can be estimated as

$$\hat{H} = \frac{Y}{X} = \frac{Y^0 + \sum_{i=1}^n Y^i}{X} = \hat{H}^0 + \frac{1}{X} \sum_{i=1}^n Y^i \quad (3)$$

Applying variance operator to the two sides of Equation 3, we get

$$D(\hat{H}) = \frac{n\sigma_n^2}{X^2} \quad (4)$$

σ_n^2 decreases as the number of Tx_1^1 grows, which is caused by intercepting mutually. The blocking inhibits effect of virtual antenna, and the range of Y^i is reduced. The probability of blocking, $S(n)$, can be calculated as following:

$$S(n) = \frac{K_n}{C_m^n} \Rightarrow \begin{cases} S(n+1) > \frac{K_n C_m^1 - n}{C_m^{n+1}} \\ \frac{S(n+1) - S(n)}{S(n)} > n \end{cases} \quad (5)$$

where m is the number of Y^i that the area can contain, K_n is the number of choices that there is blocking when the number of Y^i is n . $S(1) = 0$, $S(n)$ with the number of Y^i increases exponentially. The σ_n^2 will convex reduce with the $S(n)$ growth until $S(n)$ equals to 1. In other words, $D(\hat{H})$ will convex increase with growing number of moving people.

We use Intel 5300 wireless net card and Linux 802.11n CSI Tool to obtain 30 pairs of amplitude and phase CSI values of

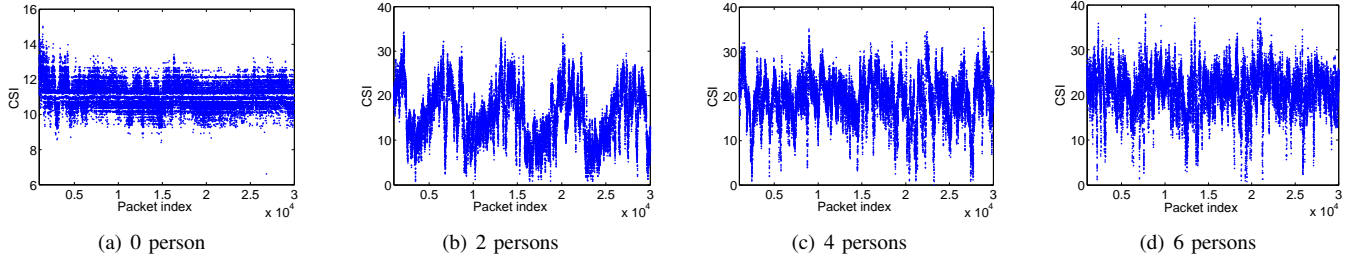


Fig. 4. CSI measurements of one subcarrier with different moving people

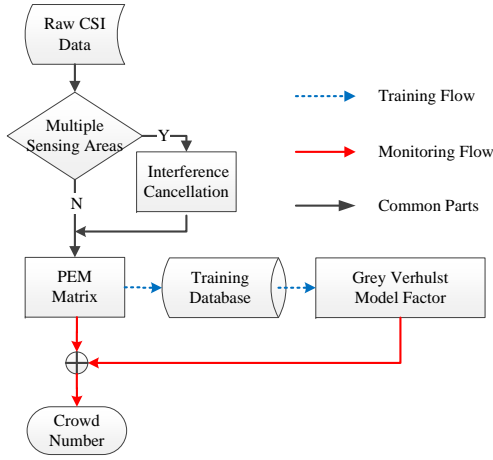


Fig. 5. The work flow

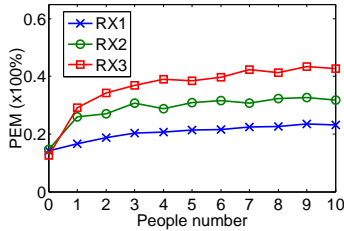


Fig. 6. PEM with different number of people

each packet. We conduct a set of experiments to check the relationship between CSI measurements and the number of moving people. We place a pair of laptops 5 meters away from each other as shown in Figure 1. One continues to send packets to the other, and there are some people walking in the active area.

Figure 4 (a), (b), (c) and (d) respectively demonstrate the variation of CSI amplitude of one subcarrier when 0, 2, 4, 6 people are walking in Case I as shown in Figure 1. Figure 4 clearly indicates that CSI values distribute more widely and change more drastic when there are more moving people. If we find a proper quantifiable index to characterize the variation of CSI measurement, it becomes possible to use CSI to count crowd.

III. METHODOLOGY

In this section, we give the details of our FCC system. We start by an overview of the system architecture followed by the details of the system modules.

Algorithm 1: Dilatation-based crowd profiling

Input: $C_d; S; P; C_U; C_V; M_c; D$.

Output: \mathbb{P} .

```

1 for  $i = 1 : S$  do
2   for  $j = 1 : P$  do
3      $k = \lfloor \frac{C_d[i][j] - C_l}{C_U - C_l} \cdot (M_c - 1) \rfloor + 1$ ;
4     for  $u = -D : D$  do
5       for  $v = -D : D$  do
6         if  $1 \leq j + u \leq P \& 1 \leq k + v \leq M_c$  then
7            $\mathbb{M}[i + u][j + v] = 1$ ;
8         end
9       end
10    end
11  end
12  for  $l = 1 : P$  do
13    for  $m = 1 : M_c$  do
14       $Ones = Ones + \mathbb{M}[l][m]$ ;
15    end
16  end
17   $\mathbb{P}[i] = Ones / (P \times M_c)$ ;
18 end

```

A. System Design

Figure 5 gives an overview of the system architecture. The modules of the proposed system are implemented in the application server that collects samples from the monitoring points and processes them. The system works in two phases: 1) A short offline phase, during which the system studies the CSI values when no or a few of people are present inside the area of interest to construct what we call a "training profile" for each stream. The profiles of all streams are constructed concurrently in that short phase. 2) A monitoring phase, in which the system collects readings from the monitoring points and decides whether there is human activity. It also updates the stored training profile so that it can adapt to environmental changes. Finally, a decision refinement procedure is applied to further enhance the accuracy.

In our basic FCC system, there are 4 laptops with Intel 5300 NICs at each four corner. One laptop continues to broadcast beacon messages, and the other three ones work as receivers to measure the CSI values of the channel fadings. In the next three subsections, we elaborate on the design of FCC.

B. CSI Profile Construction

Training phase and monitoring phase contain a common component to learn the behavior of the signal readings in the monitoring area. The selected feature for system operation should be resistant to possible environmental changes that may affect the stored data. In addition, the selected feature should also be sensitive to the human motion to enhance the detection accuracy.

As mentioned before, moving people can influence the transmission channel, change the signal paths and cause CSI variation. The more people who are moving in the area, the more signal transmission paths will be changed. Figure 4 illustrates that the change of time-adjacent CSI amplitude becomes larger when the number of moving people increases. When the CSI measurements during a time period are plotted as a two-dimensional diagram, the distance of time-adjacent points indicates the variation of CSI profiling. A major challenge is to formulate the relationship between CSI variation and the number of crowd. The biggest obstacle is how to describe the variance of CSI. If we select a large size sample space, the system performance would suffer the impact of the long sampling time. Furthermore, the average variance is helpless when the crowd number changes very frequently. On the other hand, the variance is not statistically significant when the sample size was too small. Therefore, we should find a metric which has the capability of mirroring the current CSI variance. We observe that when the points are expanded a certain size, they will overlap with their neighboring points. The overlapped area varies inversely to the intensity of CSI variation. We propose a metric PEM, the Percentage of non-zero Elements in the dilated CSI Matrix, to indicate variation of CSI profiling as shown in Algorithm 1. It is generally divided into 3 steps: transform the CSI amplitude values into two-dimensional matrix, dilate the matrix, and count the non-zero element. In Algorithm 1, C_d is the barriered CSI results after time domain filtering (see in Algorithm 2), S is the number of subcarriers, P is the number of packets, C_u and C_l are the maximum and minimum value of CSI measurements, M_c is the number of rows of matrix, D is the dilatation coefficient.

Firstly, each element in the $M_c \times P$ matrix \mathbb{M}_0 is initialized to "0". The CSI reading $C_d[i][j]$ are converted into integers k by $k = \left\lfloor \frac{C_d[i][j] - C_l}{C_u - C_l} \cdot (M_c - 1) \right\rfloor + 1$, then the elements in row k and column j in \mathbb{M}_0 is set to "1". There is a "1" in each column, and the rest are "0"s. Obviously, the variance of the row numbers of non-zero elements between adjacent columns becomes larger, when the CSI readings take dramatic turns.

Secondly, the elements around "1" are set to "1"s, which is called matrix dilatation. After dilatation, the CSI matrix \mathbb{M}_0 is transformed into dilated CSI matrix \mathbb{M} . There are less overlap of dilated elements when the CSI reading are changing more

sharply. That is, significant change of CSI usually comes along with more 1's.

Finally, the percentage $\mathbb{P}[i]$ (which is the PEM of i_{th} subcarrier) of non-zero elements in the dilated CSI matrix of each subcarrier can be calculated. The larger overlap areas of dilated points are, the low percentage of non-zero elements in the dilated CSI matrix. So the $\mathbb{P}[i]$ can indicate the number of moving people. Figure 6 shows that there exist monotonous relations between $\mathbb{P}[i]$ and the people count. $\mathbb{P}[i]$ increases with the growing number of moving people.

Based on the monotonous relation, we can gather CSI fingerprint to determine the function between PEM and people count when a few people walk in the area, and then estimate numbers when there are more moving people.

C. CSI Profile Fitting

The purpose of the CSI Profile Construction Module is to construct a normal profile and capture the CSI characteristics. These characteristics can be used by other modules to count crowd. As shown in Figure 6, we can see that there are some quasi-monotonous relationships between the crowd number and the corresponding PEM. PEM is non-negative, and its growth is a saturation process. We introduce the Grey Theory [15] to describe them.

The Verhulst model [16] was first introduced in 1838 by Pierre Franois Verhulst. The main purpose of Verhulst model is to limit the whole development for a real system. The raw data series is assumed to be

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\} \quad (6)$$

where n is the number characteristic data sequence.

The Grey Verhulst Model can be constructed by establishing a first order differential equation for $X^{(1)}(k)$ as:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = b(X^{(1)})^2 \quad (7)$$

where parameters a and b are the developing coefficient and grey action quantity, respectively. In practice, parameters a and b can be obtained by using the least square method.

$$[a, b]^T = (B^T B)^{-1} B^T Y \quad (8)$$

$$B = \begin{bmatrix} z^{(1)}(2) & (z^{(1)}(2))^2 \\ z^{(1)}(3) & (z^{(1)}(3))^2 \\ \vdots & \vdots \\ z^{(1)}(n) & (z^{(1)}(n))^2 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \quad (9)$$

where $Z^{(1)}$ is the Mean Generation of Consecutive Neighbours Sequence (MGCNS) defined as following:

$$z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k-1)) (k = 2, 3, \dots, n) \quad (10)$$

The $\hat{X}^{(1)}$ turns to

$$\hat{x}^{(1)}(k) = \begin{cases} \frac{ax^{(1)}(1)}{bx^{(1)}(1) + (a - bx^{(1)}(1))e^{a(k-1)}} & k = 2, 3, \dots \\ x^{(1)}(1) & k = 1 \end{cases} \quad (11)$$

where $\hat{x}^{(1)}(k) (k \leq n)$ are fitted sequence, and $\hat{x}^{(1)}(k) (k > n)$ are the forecast values. Since the initial sequence $X^{(0)}$ is a saturated sequence, it can be used instead of $X^{(1)}$ [16].

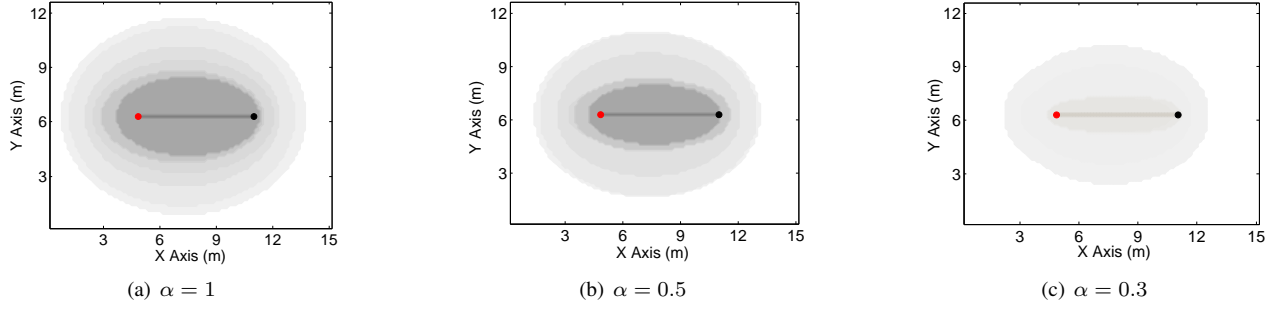


Fig. 7. CSI sensing range with different α

D. Crowd Counting

According to Grey Theory, smaller $|a|$ is better for forecast. So the crowd estimation can be decomposed into 2 steps: 1) each Rx generates its own developing coefficient a_i and grey action quantity b_i , and calculate its estimating number:

$$\widehat{num}_i = \ln \frac{\frac{a_i x^{(1)}(1)}{PEM_i} - b_i x_i^{(1)}(1)}{(a_i - b_i x_i^{(1)}(1))e^{a_i}} + 1 \quad (12)$$

2) obtain the final estimation through weighted average algorithm if there are N RXs

$$Enumber = \frac{\sum_{i=1}^N \widehat{num}_i \times \frac{1}{|a_i|}}{\sum_{i=1}^N \frac{1}{|a_i|}} \quad (13)$$

E. Scalability

Since each receiver can only estimate the number of moving people within a certain range, the scalability of FCC should be considered to work well in large scale scenes. Intuitively, more devices can be deployed as the grid to cover the larger monitoring regions. It leads to a new problem that there are each mutual interference between two neighboring grid. Therefore, we need to eliminate distractions.

For narrow-band systems, these reflections will not be resolvable by the receiver when the bandwidth is less than the coherence bandwidth of the channel. Fortunately, the bandwidth of 802.11n waveforms is 20MHz (with channel bonding, the bandwidth could be 40MHz), which provides the capability of the receiver to resolve the different reflections in the channel. We propose a multipath mitigation mechanism that can distinguish the LOS signal or the most closed N-LOS from other reflections in the expectation of eliminating distractions from a distant.

The commonly used profile of multipath channel in the time domain is described as follow:

In practice, OFDM technologies are efficiently implemented using a combination of fast Fourier Transform (FFT) and inverse fast Fourier Transform (IFFT) blocks. The 30 groups of CSI represent the channel response in frequency domain, which is about one group per two subcarriers. With IFFT processing of the CSI, we can obtain the channel response in the time domain, *i.e.*, $h(t)$. Due to the bandwidth limitation, we

may not be able to distinguish each signal path, but multiple signal clusters. Therefore, we keep the first part of $h(t)$ which includes the LOS and a few of NLOS signal paths, and filter out the residual clusters using a truncation window. The first part time duration is determined by setting the truncation threshold α as shown in Algorithm 2. C_r means the raw CSI measurement of i_{th} subcarriers and j_{th} packet.

Algorithm 2: Truncation Denoise Procedure

Input: $C_r; S; P; \alpha$.
Output: C_d .

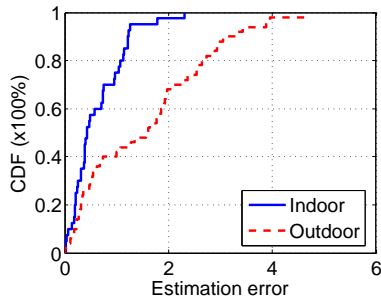
- 1 $np = 2^{\lceil \log_2 \text{length}(C_r) \rceil}$;
- 2 **for** $i = 1 : P$ **do**
- 3 $tmp[S] = IFFT(C_r[S][i], np)$;
- 4 **for** $j = \lfloor S * \alpha \rfloor : S$ **do**
- 5 $tmp[j] = 0$;
- 6 **end**
- 7 $C_d[S][i] = FFT(tmp[S], np)$
- 8 **end**

In doing so, we expect to mitigate the estimation error introduced by the mutual interference between two neighboring grid. After the time domain signal processing, we reobtain the CSI using FFT.

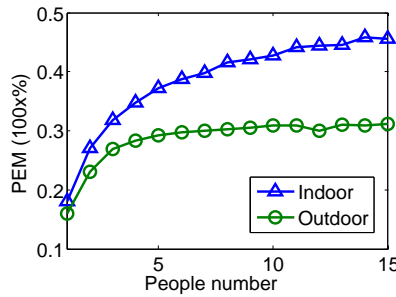
We conduct a set of experiments to evaluate the impact of α in all three cases shown in Figure 1. Figure 7 shows the contour of PEM. The darker color indicates a larger PEM. The red point and black point are Tx and Rx respectively. It's interesting that the CSI sensing range is asymmetrical. From the figure 7 we can see that it is more sensitive around Tx than Rx, and the sensing range of Tx and larger than Rx. With the decreased α , the sensing range is reduced. When α is 3, the effective sensing range is reduce to the area between Rx and Tx. In this case, the moving people in adjacent regions cannot effect the PEM, which can efficiently eliminate distractions among adjacent regions.

IV. EXPERIMENT RESULTS

In this section, we illustrate the implementation and do various experiments under different conditions to evaluate the performance of FCC. We conduct real-world experiments to show the performance and robustness of FCC in different scenarios as shown in Figure 8.



(a) CDF of estimation error



(b) The relationship between PEM and crowd number

Fig. 9. Overall performance

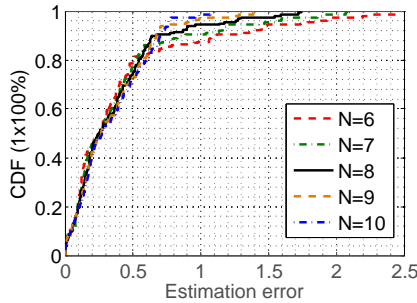


Fig. 11. Performance of Verhulst Model

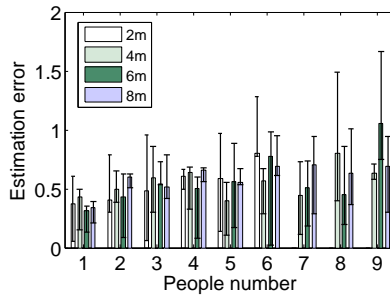


Fig. 12. Impact of communication distance between Rx and Tx

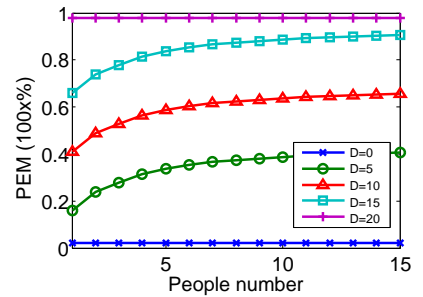


Fig. 10. The impact of dilatation coefficient

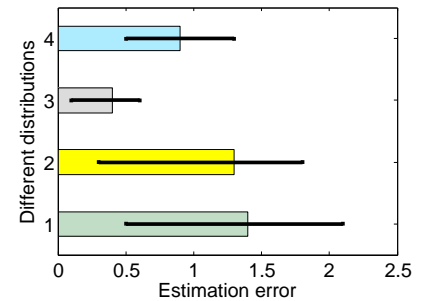


Fig. 13. Impact of crowd distribution



Fig. 8. The experiments scenes and device

A. Overall performance

We summarize the results we have obtained from both indoor (*i.e.* meeting room, office, laboratory) and outdoor settings.

As explained in previous section, FCC works well in different scenario because the CSI variance is sensitive to the change of scenario but not to the scenario itself. Figure 9(a) compares the cumulative distribution of crowd counting errors in indoor and outdoor environments. The X axis represents the difference between the actual number of moving people

and the estimation value, while the Y axis represents the CDF percentage. It is easy to see FCC works well in both indoor and outdoor environments, and better in former. More than 98% estimation errors are less than 2 persons in indoor environment, while about 70% errors is 2 persons in outdoor environment. It is because that there are more obvious multi-path effects in indoor environments than outdoors. Figure 9(b) plots the PEM in two kinds of scenes. The PEM is increasing with the increasing numbers of people, especially in indoor environments. In outdoor environments, the curve becomes gentler when the number of people is more than 10.

B. Impact of dilatation coefficient

Since PEM is calculated based on the dilated CSI matrix \mathbb{M} , the impact of dilatation coefficient should be evaluated. Figure 10 shows the variation of PEM under different dilatation coefficient D . When $D = 0$, the PEM keeps constant with the increase in the number of people. It is because that no matter how the CSI changes over time, there are none overlapping parts or components since each column only has one non-zero element before dilatation. When $D = 20$, the PEM also keeps constant. It is because that almost all the elements are set to "1", when dilatation coefficient is too big. In other words, the dilatation covers the variance. When D is equal to 5, 10, 15, there is a monotone increasing relationship between PEM and the number of people. In follow-up experiments, we set $D = 10$.

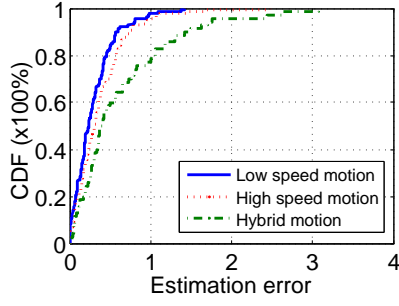
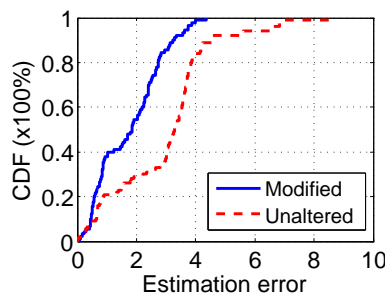
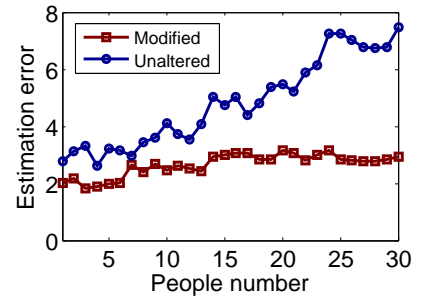


Fig. 14. Impact of moving speed



(a) CDF of estimation error



(b) Estimation errors with different number of people

Fig. 15. Scalability in a large area

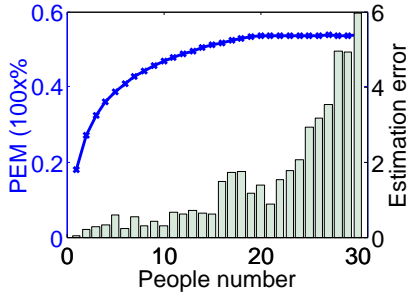
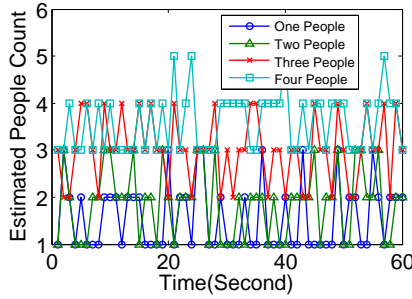
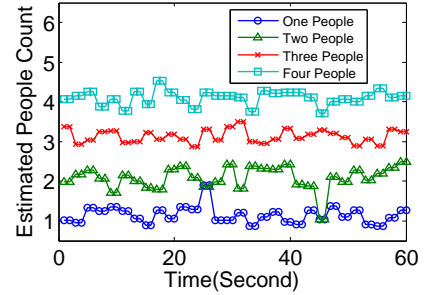


Fig. 16. The maximum distinguished number of people



(a) SCPL



(b) FCC

Fig. 17. Comparison

C. Performance of Verhulst Model

We take a set of experiments to evaluate the performance of Verhulst Model with different number of characteristic data sequence. The estimation results are presented in Figure 11. We use 0- N people's characteristic data sequence in training phase to estimate the number of moving people in monitoring phase where the people is from 0 to 30. The Verhulst model is an ad-hoc model. It is an equation chosen to fit data, which is a very good model for the number of people growth. The maximum estimation error is about 2.5, and it reduces to 1.1 with N growth. Since the parameters a and b is calculated by the least square estimate method, a bigger N means the more equations, which will obtain a better a and b , and have more accurate fitted sequences. Generally speaking, the Verhulst model can take a long term forecast when $|a| < 0.3$, and take a short term forecast when $|a| < 0.7$.

D. Impact of communication distance

We evaluate how the distance between Tx and Rx influences the accuracy of our FCC system. We estimate the number of people walking between a pair of laptops with different distances apart. Due to space limitation, the communication distance between two laptops is increased from 2m to 8m.

Figure 12 shows the estimation error with different communication distances. It's very interesting that the estimation error is unrelated to communication distance, as long as the communication distance is less than a certain threshold. The experiment shows that we can provide more flexible

deployment mechanism in order to adapt to various indoor layouts without reducing the estimation accuracy.

E. Impact of distribution

In this subsection, we explore how the crowd distribution affects estimation number. There are four cases: case 1, 12 people split into 2 groups, ones move near Rx_1 , and others move near Rx_3 ; case 2, one group move near Tx, and others move near Rx_2 ; case 3, 12 people move random in the scenario; case 4, 12 people split evenly into 3 groups with 4 people, and they move near Tx, Rx_1 , Rx_3 respectively.

Figure 13 shows the impact of distribution. The X and Y axis represent the estimation error and four cases respectively. The estimation error in case 3 is less than other three cases. It is because that the crowd is most similar to stochastic distribution in case 3. Although the distribution affects estimation accuracy, the the maximum average error is less than 1.5.

F. Impact of moving speed

In actual environment, different people has different walking speed, therefore, we should consider that whether the moving speed would influence estimation accuracy.

Figure 14 shows the CDF of estimation error in different moving speed. The low speed motion and high speed motion has a similar estimation accuracy, while hybrid motion has larger estimation errors. It is because that FCC estimate the moving people based on the observation in training phase. For low speed motion and high speed motion, the training

phase and monitoring phase have the same CSI feature. For the hybrid motion, different people have different moving speed, the relationship between PEM and the number of moving people is unstable, which will affect the estimation accuracy. This could be addressed in our future work.

G. Scalability in large area

Since each receiver can only estimate the number of moving people within a certain range, the scalability of FCC should be considered to work well in large scale scenes. In order to evaluate the scalability of FCC, we implement 5 laptops in a hall. 4 laptops is in 4 corner as Rx's, and 1 laptop is in center as a Tx.

Figure 15 compares the performance between unaltered and modified FCC. Figure 15(a) shows the CDF of estimation error. More than 50% estimation errors are less than 2 using modified method, while more than 70% errors more than 2 using unaltered method. Figure 15(b) shows the variation of average estimation error with people increasing from 1 to 30. It is worth noting that the estimation errors remain stable for different number of people using modified method, while the error is increasing as the number of people increase using unaltered method, especially when people is more than 22. It is because that modified method uses Truncation Denoise Procedure to eliminate distractions among adjacent regions. It is feasible, flexible and extensible. FCC works well when 30 people in the scenario, while recent similar work [17], [13] can only distinguish no more than 4 people.

H. The resolution of crowd counting

Figure 16 shows the maximum distinguished number of people in indoor environment. The experiments are implemented in a hall with four laptops in four corners. The blue curve plots the PEM variation with the increasing people. Obviously, the curve becomes slower gradually when the number of people is more than 20. The green bars illustrates the average estimation errors. At first, the estimation error is small. When people is more than 15, the estimation error increases to 2. When people is more than 27, the estimation error increases to more than 5. It is because that the PEM almost stops growing when people is more than 22. In a word, if people is less than 30, FCC can work well in the indoor environment. By contrast, to the best of our current knowledge, the distinguishability of exist non-vision device-free approach [17] remains the single digits level using radio link, using ISAR [13] or a dozen people by excessively deploying dense sensors [11].

I. Comparison with other approach

There is few methods on device-free crowd counting. SCPL [17] is one of the latest device-free technique to count and localize multiple subjects in indoor environments. We do a set of experiments to compare the performance of FCC and SCPL. There are 5 devices deployed on the both sides of the room with area of 40 m². Each device is both a transmitter and receiver. Figure 17(a) and (b) show the people counting results of SCPL and FCC respectively. In all four cases, the estimated number always fluctuates over time for both SCPL and FCC. The errors caused by temporally overlapping trajectories and

the environmental disturbance. For SCPL, about 45% results are accurate. It is because that SCPL is a link-based scheme, which requires more intensive devices to provide adequate links. For FCC, the estimated results are relative stable. The number of estimated moving people is jitter with a range of plus or minus about 0.4. CSI is information which represents the state of a communication link from the transmit source to the receiver source. It is more sensitive to the diversity of transmission channel than RSS, which is more suitable than RSS at counting moving people.

V. RELATED WORK

Recently, indoor localization and counting problem has attracted much attention. The solutions can be classified into two categories by whether the object needs to carry equipments or not: Device-based Approach and Device-free Approach.

A. Device-based Approach

Device-based approach utilizes device carried by users to locate or count objects. Cricket system [2] calculates the position relying on coupled RF and ultrasonic signals. RFID-based localization [3], [18] has gained much attention in recent years. RSS-based approaches can be categorized into the Radio Propagation Model [4], [19] and the Fingerprint Method [20], [21], [5]. Recently, Channel State Information (CSI) was used for localization [22], [23], [24].

With the sharply increasing scale of mobile computing, using built-in sensors and modules of smart phones for locating, tracing or counting became a wider trend. Jens *et al.* [6] exploited Bluetooth collecting environment data to estimate crowd density. Kannan *et al.* [7] utilized the headset to estimate the crowd density(the number of phones exactly). Phone inertial sensors [8] can also be used for localization.

Device-based approach is essentially counting devices. It is more suitable for object localization than crowd counting. In many crowd counting scenarios, it is hard to require each person taking a device with him. If a person takes more devices or dose not carry the device, it will import estimated errors.

B. Device-free Approach

Device-free approach [9] has gained much more attention because this approach doesn't require targets to equip any devices. The most intuitive device-free solution exploit the machine vision technology. Most of vision-based crowd density estimations have a considerable overhead of computing though, and can only work in a line-of-sight pattern. Moreover, because of privacy protection and some other reasons, more and more non-vision-based approaches appeared in recent year. Most of them are built on radio frequency techniques. These solutions can be broadly categorized into Location-based scheme and Link-based scheme.

Location-based scheme: This kind of scheme [12], [11] divides locating procedure into two phases: training phase and operating phase. Xu *et al.* [12] propose to formulate the localization problem into a probabilistic classification problem in order to mitigate the error caused by the multipath effect in cluttered indoor environments. These approaches aimed

at locating only one person, and Yuan *et al.* [11] employ classification algorithm to estimate the number of people. Masaya Arai *et al.* [25] attempted to get the relationship between the feature of radar chart and the crowd movement patterns, but they only got some qualitative results.

This kind of scheme require site survey over areas of interests to build a fingerprint database. The considerable manual cost and efforts, in addition to the inflexibility to environment dynamics are the main drawbacks of this methods. For crowd counting, the training cost is too much heavy for large-scale scenarios and it's hard to obtain the ground truth when the number of people is large.

Link-based scheme: This kind of scheme [26], [27] is unique in device free approach. Radio link is a radio communication system, N nodes can construct $N(N-1)/2$ links. If objects active nearby Link $L_{i,j}$, the RSS of $Node_i$ and $Node_j$ would change obviously. But if object moves far away from any link, the performance decrease sharply. Nakatsuka *et al.* [10] first demonstrated the feasibility of using static profile (average and variance) of RSS to estimate the crowd density. Patwari *et al.* [28] provides a statistical model for the RSS variance as a function of ONE person's position with respect to the transmitter (TX) and receiver (RX) locations. Xu *et al.* [17] exploited link-based scheme to count the number and locate the people using RSS.

Obviously, this kind of scheme is more sensitive when the obstacle moves near by the links. Users must deploy exorbitantly intensive nodes to provide adequate links, which causes high cost issues.

Utilizing fine-grained physical layer information in localization and counting draws increasing attentions recently. Dina *et al.* [13] exploit inverse synthetic aperture radar (ISAR) technique to track and count moving objects. Shwetak *et al.* [14] extract Doppler shifts from Wireless Signals for gesture recognition. The USRP plays an important role in obtaining the physical layer information such as signal angle and frequency shift, *etc.*

VI. CONCLUSION

Crowd counting is an important service needed by many applications. Previous crowd counting approaches either rely on smart phone or RFID tag carried by users or need to collect abundant fingerprint which may not be suitable for large-scaled scene. This paper proposed a novel device-free crowd counting method called FCC based on CSI measurements. It can be obtained from off-the-shelf 802.11n devices. We propose a new metric PEM to extract the feature of CSI variation. The relation between the PEM and the number of people can be leveraged by the Grey Verhulst Model to count crowd. Extensive real-world experiments results demonstrate that our approach perform well in accuracy, scalability and reliability.

REFERENCES

[1] M. Li, Z. Zhang, K. Huang, and T. Tan, "Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection," in *Proceedings of ICPR2008*, pp. 1–4.

[2] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proceedings of MobiCom 2000*, pp. 32–43.

[3] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "Landmarc: indoor location sensing using active rfid," *Wireless networks*, vol. 10, no. 6, pp. 701–710, 2004.

[4] Y. Ji, S. Biaz, S. Pandey, and P. Agrawal, "Ariadne: a dynamic indoor signal map construction and localization system," in *Proceedings of MobiSys 2006*, pp. 151–164.

[5] M. Azizyan, I. Constandache, and R. Roy Choudhury, "Surroundsense: mobile phone localization via ambience fingerprinting," in *Proceedings of MobiCom 2009*, pp. 261–272.

[6] J. Weppner and P. Lukowicz, "Collaborative crowd density estimation with mobile phones," in *Proceedings of Sensys 2011*.

[7] 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 SenSys 2012*, 2012, pp. 155–168.

[8] F. Li, C. Zhao, G. Ding, J. Gong, C. Liu, and F. Zhao, "A reliable and accurate indoor localization method using phone inertial sensors," in *Proceedings of UbiComp 2012*.

[9] M. Youssef, M. Mah, and A. Agrawala, "Challenges: device-free passive localization for wireless environments," in *Proceedings of MobiCom 2007*, pp. 222–229.

[10] M. Nakatsuka, H. Iwatani, and J. Katto, "A study on passive crowd density estimation using wireless sensors," in *Proceedings of ICMU 2008*.

[11] Y. Yuan, C. Qiu, W. Xi, and J. Zhao, "Crowd density estimation using wireless sensor networks," in *Proceedings of MSN 2011*, pp. 138–145.

[12] C. Xu, B. Firner, Y. Zhang, R. Howard, J. Li, and X. Lin, "Improving rf-based device-free passive localization in cluttered indoor environments through probabilistic classification methods," in *Proceedings of IPSN 2012*, pp. 209–220.

[13] F. Adib and D. Katabi, "See through walls with wi-fi!" to appear at SIGCOMM 2013.

[14] S. G. S. P. Qifan Pu, Sidhant Gupta, "Whole-home gesture recognition using wireless signals," to appear at Mobicom 2013.

[15] J. Deng, "Introduction to grey system theory," *The Journal of grey system*, vol. 1, no. 1, pp. 1–24, 1989.

[16] Y. L. Sifeng Liu, *Grey information: theory and practical applications (advanced information and knowledge processing)*. Springer-Verlag New York, Inc, 2005.

[17] 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 IPSN 2013*.

[18] C. Wang, H. Wu, and N.-F. Tzeng, "Rfid-based 3-d positioning schemes," in *Proceedings of INFOCOM 2007*, pp. 1235–1243.

[19] H. Lim, L.-C. Kung, J. Hou, and H. Luo, "Zero-configuration, robust indoor localization: Theory and experimentation," in *Proceedings of IEEE INFOCOM 2006*, pp. 1–12.

[20] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proceedings of INFOCOM 2000*, vol. 2, pp. 775–784.

[21] M. Youssef and A. Agrawala, "The horus location determination system," *Wireless Networks*, vol. 14, no. 3, pp. 357–374.

[22] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "Spot localization using phy layer information," in *Proceedings of MobiSys 2012*, vol. 12.

[23] X. Jiang, W. Kaishun, Y. Youwen, W. Lu, and M. N. Lionel, "Pilot : Passive device-free indoor localization using channel state information," in *Proceedings of ICDCS 2013*.

[24] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "Fila: Fine-grained indoor localization," in *Proceedings of INFOCOM 2012*, pp. 2210–2218.

[25] M. Arai, H. Kawamura, and K. Suzuki, "Estimation of zigbee's rssi fluctuated by crowd behavior in indoor space," in *Proceedings of SICE 2010*, pp. 696–701.

[26] D. Zhang, Y. Liu, and L. M. Ni, "Rass: A real-time, accurate and scalable system for tracking transceiver-free objects," in *Proceedings of PerCom 2011*, pp. 197–204.

[27] O. Kaltiokallio, M. Bocca, and N. Patwari, "Enhancing the accuracy of radio tomographic imaging using channel diversity," in *Proceedings of MASS 2012*.

[28] N. Patwari and J. Wilson, "Spatial models for human motion-induced signal strength variance on static links," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 791–802, 2011.