2024 IEEE 30th International Conference on Parallel and Distributed Systems (ICPADS) | 979-8-3315-1596-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICPADS63350.2024.00021

Cover More with Less: Eliminating Blind Spots for Surveillance Camera via Passive WiFi Sensing

Khairul Mottakin, Jinhua Guo, and Zheng Song
Dept. of Computer and Information Science, University of Michigan-Dearborn, MI, USA
{*khairulm, jinhua, zhesong*}*@umich.edu*

***Abstract*—Surveillance cameras, even with Pan, Tilt, and Zoom (PTZ) capabilities, can only cover a limited directional range at a time, leading to inevitable blind spots. To mitigate these blind spots, users typically need to deploy additional cameras or sensors (such as motion sensors and microphones) to guide the PTZ platform's movement for capturing intruders, incurring additional deployment and maintenance costs. To address this challenge, this paper introduces a novel approach that utilizes the camera's built-in WiFi module to detect potential intruders and direct the PTZ platform's movements. The approach involves collecting WiFi Channel State Information (CSI) samples with humans positioned at various locations, and training a machine learning model to infer the real-time location of intruders. Given the intensive human effort required for collecting sample data, we developed an algorithm to optimize the selection of locations for collecting CSI samples. The algorithm assesses each location's contribution to the overall success rate of capturing intruders, thereby achieving optimal sample distribution. Our evaluation demonstrates that our approach achieves a capture rate of 78.24%, which is up to 24% higher than baseline methods, despite being trained with data collected from only 13% of the locations.**

***Index Terms*—Channel State Information (CSI), Blind Spot, WiFi Sensing, Machine Learning, Video Surveillance Systems**

I. Introduction

Video surveillance systems are crucial for safeguarding both individuals and properties, finding extensive use in various indoor environments like offices, factories, supermarkets and train stations. These intelligent systems deter and notify authorities about unwanted intruders posing threats within the monitored areas. Modern cameras have advanced features that significantly enhance surveillance capabilities. First, their IP and network connectivity enable the efficient transmission, storage, and analysis of video data over networks such as WiFi and Ethernet. Second, these cameras include PTZ (Pan, Tilt, and Zoom) capabilities, allowing for remote control to adjust the field of view and specifically focus on areas of interest, which is essential for monitoring dynamic environments. Lastly, they possess robust computing capabilities that facilitate real-time video data analysis and playback, supporting advanced functions like motion detection and the tracking of objects or intruders, thereby elevating them from mere recording devices to intelligent security systems.

Despite the capability of PTZ cameras to pan and tilt, they still encounter unavoidable blind spots, as not every area can be covered by surveillance [1]. Addressing these blind areas cost-effectively is vital for enhancing the efficiency of video surveillance systems. To eliminate the blind spots, a common approach is to sense the location of intruder using additional sensors, and redirect the camera by controlling PTZ accordingly. Examples of this approach include using ultrasonic [2], light sensor [3], acoustic sensor [4], and infrared sensor [5]. However, these improvements require deploying extra hardware, leading to increased costs and potentially more complex integration and maintenance requirements.

In this paper, we present a novel surveillance camera system that eliminates blind spots by sensing the intruder via the camera's built-in WiFi module, which is the first attempt of its kind to the best of our knowledge. Our approach is based on the well-established Channel State Information (CSI) sensing, which utilizes the fine-grained changes of WiFi signals measured as CSI to estimate the location of humans. As shown in Fig. 1, the WiFi module periodically captures CSI signal and uses a pre-trained CSI sensing model to estimate the presence and location of intruder, and control the PTZ to redirect the camera to its estimated location.

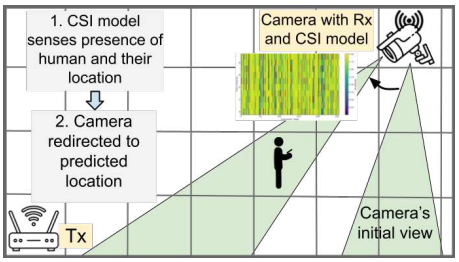


Fig. 1: Illustration of Camera Redirection

Implementing this system involves an offline phase, which collects a set of CSI data using the camera's WiFi module with a human subject moving to different locations, and trains a machine learning model. In our approach, human stands at various locations to capture CSI data, a process we refer to as "Sample Distribution" throughout this paper. When the system works online, low location estimation accuracy may lead to the camera being directed to a wrong direction, failing to capture the intruder. As such CSI sample collection is very labor intensive, we develop an approach that optimizes the sample collection process so that the failure can be minimized with limited CSI samples.

In particular, we introduced a new metric called **Successful Capture Rate** (SCR), which quantifies the likelihood that our system can successfully detect an intruder. We observed that a sample collected at various locations contributes differently to

979-8-3315-1596-6/24/$31.00 ©2024 IEEE
DOI 10.1109/ICPADS63350.2024.00021

SCR, making evenly distributing samples in space a less efficient solution. For each location, considering how its distance to the camera impacts SCR and how its nearby WiFi sensing-signal-to-noise-ratio (SSNR [6]) impacts CSI estimation accuracy, we develop a model for predicting how collecting CSI sample with human positioned at various locations impacts SCR. Utilizing the model, we formulate the distribution of samples as an optimization problem for maximizing SCR, and solve it using a greedy-based algorithm.

The contribution of this paper is three folds:

- We design and develop a novel system based on the well-established WiFi sensing that efficiently guides the camera to cover the intruder as well as eliminates the blind spots.
- We analyze the problem of CSI sample distribution and its impact on the camera accuracy and propose a solution to maximize the camera accuracy. We formulate the distribution of samples as an optimization problem for SCR, a metric we introduce for measuring system's reliability, and develop an algorithm to solve it.
- We implement and deploy our system in real environment and compare our sample distribution algorithm with 2 baseline approaches. Our results show that our approach exhibits higher accuracy than the baseline approaches.

II. Related Works and Background

This section introduces the background for our work, as well as existing solutions.

A. Blind Area in Surveillance Camera and Mitigation

The limited view range of surveillance cameras, typically spanning 90 to 120 degrees, restricts their coverage and creates blind spots that can be exploited by intruders [7], [8]. Although PTZ cameras offer a solution by enabling dynamic repositioning to capture images of intruders once their location is known, identifying the precise location of an intruder remains a challenge. Existing strategies to mitigate the issue of blind areas in surveillance systems primarily rely on the integration of additional sensors, specifically: 1) **Motion Sensing:** Utilizing sensors that detect changes in the physical environment to infer the presence and movement of intruders [5]; 2) **Sound Sensing:** Employing audio detection mechanisms to capture noise or vocal disturbances that may indicate unauthorized access or activities [2], [4].

These methods typically require attaching extra sensors to the surveillance setup, enhancing the system's ability to detect activities beyond the limited view of the camera. Nevertheless, such enhancements come with increased costs due to the need for additional hardware and possibly more complex integration and maintenance requirements. Unlike traditional methods, such as motion and sound sensing, our method leverages the built-in WiFi capabilities of IP cameras, offering a cost-effective solution to the blind area problem. Although using WiFi to detect humans in blind spots has been explored previously [9], our work is the first to apply it to guide the surveillance camera to capture video of the intruder.

Despite advanced capabilities of smart cameras, they have inherent limitations that prevent them from fully eliminating blind spots in surveillance systems. Primarily, their field of view is restricted to where they are positioned and the areas their lenses can cover. Even with PTZ functionalities discussed before, there are inevitable gaps that remain unmonitored unless multiple cameras are strategically placed [10].

B. Sensing-Signal-to-Noise-Ratio (SSNR)

The SSNR, introduced by [6], quantifies the quality of a sensing signal relative to background noise within a sensing zone, where background noise includes any disruptive signals. A sensing zone, typically an ellipse, is defined by the distance between the transmitter(Tx) and receiver(Rx), which affects its size, shape, and coverage. The placement of Tx relative to Rx influences the ellipse's elongation and coverage, impacting the CSI model's accuracy, particularly in distinguishing between Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) areas, as depicted in Fig.2. LoS represents a clear path between Tx and Rx, allowing for direct signal transmission, whereas NLoS involves obstructions like walls or furniture that cause signal alterations before reaching the receiver. A higher SSNR indicates the wireless signals at a location have higher sensing capability [11], [12]. SSNR for location i is defined as:

$$SSNR_i = \frac{d_{TxRx}^2}{(d_{iTx} \cdot d_{iRx})^2} \quad (1)$$

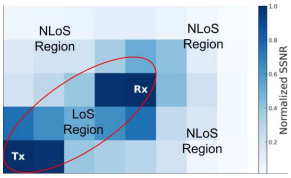


Fig. 2: SSNR Illustration

where d_{TxRx} is the distance between the Tx and Rx; d_{iTx} and d_{iRx} are the distance from location i to Tx and Rx respectively.

C. Channel State Information (CSI)

Fine-grained CSI is frequently used to characterize the propagation of a WiFi signal as it comes in contact with obstacles [13], [14]. While RSSI (Received Signal Strength Indicator) averages the signal strength across all sub-carriers and provides one reading for a received WiFi signal, CSI catches changes occurring at each sub-carrier and provides more fine-grained measurements. The number of sub-carriers differs based on the hardware and the channel bandwidth. For *m* data sub-carrier, CSI is expressed as complex number h_m, containing both amplitude($|h_m|$) and phase($\angle h_m$) values. Due to multi-path effects such as phase shift and amplitude attenuation, the CSI amplitude and phase values are affected by human movements [15], which is often used to accurately sense the behavior and location of the human subject.

III. System Overview and Problem Formulation

The basic workflow of our system is depicted in Fig.3. The system comprises two main phases: 1) Offline Phase: determining the likelihood of sampling from each location, collecting the samples with human standing at locations, and training an offline model; 2) Online Phase: using the offline

model to estimate intruder localization, guiding camera focus to the predicted location. Next we introduce the two phases in details:

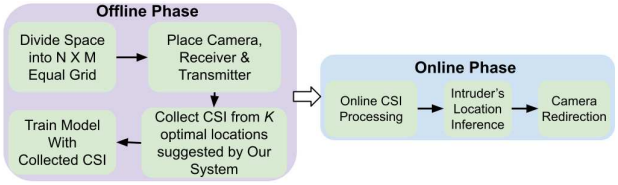


Fig. 3: Basic Workflow

1) Offline phase: We first divide the indoor room into certain number of grids, each represents a location with a two dimensional coordinate. We then place the devices, including the WiFi-enabled camera and the transmitter inside the room, and record their location coordinates. Knowing all locations in the room and the locations of the devices, our system gives instructions on where to collect CSI samples. A human subject, following the instructions, stands still on the selected locations for a certain period of time, for the system to collect a sufficient amount of CSI samples. We also collect samples when there is nobody in the room. Utilizing these CSI samples, our system trains an ML model for location estimation.

2) Online phase: During the online phase, the camera periodically samples the CSI and uses the CSI estimation model to determine if a person is present in the room and identify their location. Upon detecting an individual and pinpointing their location, our system redirects the camera to focus on that area. Considering that a camera covers a fixed **Field of View** (FoV) and the CSI estimation suffers from errors, we point the center of the camera's FoV to the estimated location, to maximize the likelihood of capturing the intruder via the camera.

A. System Model

For simplicity, we consider a rectangular-shaped indoor room. A single camera, denoted as C, is positioned at the centroid or geometric center of the room, with $FoV(C)$ denoting its FoV (for IP cameras, the general range of FoV is 20.6 degrees to 90 degrees [16]). Furthermore, a transmitter Tx, functioning as an Access Point (AP), is positioned at one of the four corners of the room. To build the coordinate of our system, we divide the indoor room into $N \times M$ equal grids, with the camera's location as $(0, 0)$. We use $L_i = (X_i, Y_i)$ to denote the location of each grid, where $i = \mathcal{I} = \{0, 1, ...M \times N\}$, $X_i \in [-M/2, M/2]$ and $Y_i \in [-N/2, N/2]$. The location of Tx is denoted as L_{Tx}.

B. Formulating the Sample Distribution Problem

Collecting CSI data during the offline phase is labor-intensive and requires human involvement. For the system to be applicable in real world scenarios, the total number of samples to be collected, denoted as K, should be small (i.e., $K << M * N$).

However, collecting a smaller number of CSI impacts the localization accuracy, and further impacts our system's capability in capturing intruders, as depicted in Fig.4. If the intruder is at a location $L = (X, Y)$ and the location estimation model estimates the intruder's location as $L' = (X', Y')$, the surveillance camera will turn to (X', Y'), with keeping the location (X', Y') in the middle of its FoV. However, there is a chance that the actual location (X, Y) falls outside the camera's FoV, resulting in the camera failing to capture intruder's footage. In Fig.4, the camera is supposed to cover location $(-3, -5)$, but failed to do so, since $\angle LCL' > FoV/2$.

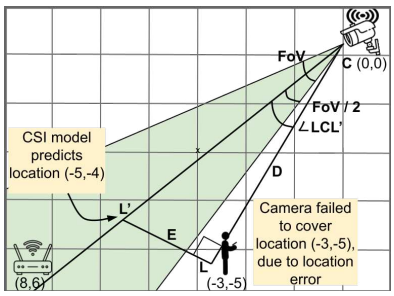


Fig. 4: Problem Illustration - Failing to Capture Intruder (*FoV is the camera's Field of View, L is the real location of intruder, L' is the location estimated by our system, and C is the location of camera.*)

To describe the problem, this paper introduces a new metric named **Successful Capture Rate (SCR)**. SCR measures the overall probability that an intruder can be successfully captured by our system. For each location $L_i, \forall i \in \mathcal{I}$, the camera can successfully capture the intruder if the angle $\angle L_i C L_i'$ formed by the points L_i, camera C, and the estimated location L_i' is smaller than $FOV/2$. Therefore, we define $SCR = \sum_{\forall i \in \mathcal{I}} P(\angle L_i C L_i' < FoV/2)$, where P is the probability.

SCR depends on the locations where the K samples are collected. For example, compared to distributing samples evenly throughout the space, collecting samples only within a sub-region (e.g., the top right corner of the room) leads to poor SCR because it fails to capture intruders appearing in other sub-regions. Without losing generality, we formulate the sample distribution problem as to find **whether or not to collect a sample (denoted as $s(i) \in [0, 1]$) at each location i, so that SCR can be maximized**:

$$\mathcal{K}* = \{s(i) | \forall i \in \mathcal{I}\} = \arg\max SCR$$
$$= \arg\max \sum_{\forall i \in \mathcal{I}} P(\angle L_i C L_i' < FoV/2)$$
$$s.t., s(i) = 0 \text{ or } 1,$$
$$\sum_{s(i) \in \mathcal{I}} s(i) = K$$

IV. Solution

One core contribution of this work is the development of an optimal sample distribution strategy that considers two important factors associated with sensing: (1) how CSI localization accuracy and the distance between the camera and the human's location impact SCR? (2) how the sensing capability of a sample location and its nearby sample locations impact CSI localization accuracy?

A. Key Idea Behind Our Solution

1) How CSI localization accuracy and the distance between the camera and the human's location impact SCR?: By understanding the relationship between localization errors, FoV constraints, and the probability of accurate location estimation, we can evaluate how effectively the system captures intruders within the camera's coverage area. We define E_{max_i} as the maximum CSI localization error that still allows the camera to capture the intruder at location i. As depicted by Fig.4, we can calculate E_{max_i} using the distance between the camera and the sample location i, and the view range of the camera $\angle FoV/2$, as follows:

$$E_{max_i} = D_i.\tan(\angle FoV/2) \tag{2}$$

Considering the CSI localization error is a distribution function $y = f(x)$, where x is the error distance and y is the probability that the estimated location L_i' falls within that error distance, we can determine the likelihood of an estimated location being within a specific error distance. For example, given error distances such as $x = \{3m, 2.5m, 2m, 1.5m, 1m, 0.5m\}$, the corresponding probabilities can be $y = \{85\%, 75\%, 65\%, 55\%, 45\%, 35\%\}$. Using this error distribution function, we can calculate the probability that the estimated location will fall within any maximum tolerable error distance E_{max_i}. For instance, if E_{max_i} is derived from the camera's field of view and distance D_i, the function $y = f(E_{max_i})$ provides the probability that the estimated location L_i' is within E_{max_i} meters of the actual location L_i. We can therefore build a relationship with the CSI localization error and the SCR and can be alternatively expressed as:

$$SCR = \sum_{\forall i \in \mathcal{I}} f(E_{max_i}) \tag{3}$$

For each location L_i, $f(E_{max_i})$ provides the probability that an intruder falls within the camera's field of view, and summing these probabilities across all locations yields the SCR. However, SCR varies by location even with the same error distance. As shown in Fig. 5, both locations, L_1 and L_2 have a CSI localization error of 3m(circles indicate CSI localization accuracy), equating to an 85% probability that the estimated location is within 3m. Since $D_1 < D_2$, $SCR1$ is lower than $SCR2$, suggesting a reduced capture probability at shorter distances. This indicates that an intruder at L_1 is more likely to evade detection compared to L_2 due to its closer proximity to the camera. Therefore, placing more samples near the camera correlates with higher SCR, prompting our approach to increase sample density near the camera to maximize SCR.

2) How sensing capability of a sample location and its nearby sample locations impact CSI localization accuracy?: The sensing capability of a specific sample location, along with that of its nearby locations, significantly influences the accuracy of CSI localization. When a human collects a sample from a location, it alters the error distribution and affects the localization accuracy of neighboring locations, emphasizing

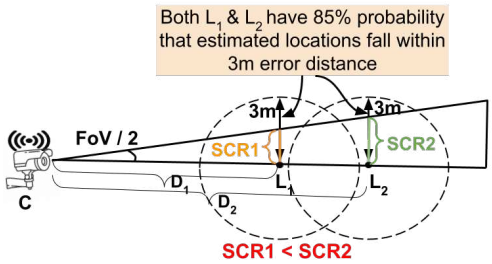


Fig. 5: Importance of assigning more samples near the camera

interdependence of sensing capabilities in environment. We mathematically define the sensing capability of location i:

$$L_{SC_i} = W_1 \cdot SSNR_i + W_2 \left(\frac{\sum_{q \in N_8(i)} SSNR_q \cdot \delta_q}{8} \right) + W_3 \left(\frac{\sum_{r \in N_{16}(i)} SSNR_r \cdot \delta_r}{16} \right) \tag{4}$$

where $N_8(i)$ represents 8 locations immediately adjacent to location i, fewer if i is on edge or corner. $N_{16}(i)$ denotes 16 locations in next outer layer, varying in number near edges. $SSNR_q$ and $SSNR_r$ are the SSNR values for locations q and r respectively, while δ_q and δ_r are indicators (1 if sample has been collected from the location, 0 otherwise). W_1, W_2, and W_3 are weights assigned to SSNR of location i itself, and sum of SSNR values for the 8 and 16 surrounding locations, respectively. The Eq. 4 considers impacts on the error distribution by integrating various levels of spatial signal information while ensuring sparse sampling. The term $W_1.SSNR_i$ directly influences the error distribution by narrowing it around location i, as higher SSNR values at i reduce localization errors. The term associated with W_2 incorporates the signal quality from immediately adjacent locations, capturing local spatial correlations and reducing error variance through neighboring signal data, while sparse sampling ensures these samples are not overly concentrated, avoiding overfitting to local conditions. Lastly, W_3 term provides broader spatial context considering the next layer of surrounding locations, which helps balance the error distribution by accounting for wider-area signal variations and ensuring robust localization accuracy, with sparse sampling enhancing the model's generalizability.

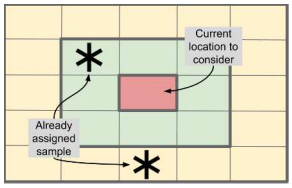


Fig. 6: Impacts of Adjacent Locations

Explanation of Weight Assignment: The weights W_1, W_2, and W_3 are assigned based on spatial relevance, variance, and stability, as well as signal quality correlation [17], [18]. Spatial relevance ensures that closer locations ($N_8(i)$) have a stronger influence due to their proximity, but sparse sampling moderates this effect to prevent overfitting. As depicted in Fig.6, variance and stability considerations lead to higher weights for immediate neighbors (W_2) because they typically exhibit

more stable signal qualities, while more distant neighbors (W_3) capture broader signal variations. Finally, signal quality correlation means that SSNR values of adjacent locations are more closely related to the SSNR at location i, and the weights reflect this by emphasizing nearby locations while ensuring a balanced contribution from the broader spatial context.

B. Solution Workflow

Considering the impacts outlined in IV-A, we formulated an empirical equation that fits our requirements, which are: prioritizing gathering samples closer to the camera and with greater sparsity relative to neighboring samples.

$$SCR_i \leftarrow L_{SC_i} * \frac{1}{\mathcal{D}_i + \epsilon} \tag{5}$$

The term L_{SC_i} is responsible for distributing samples sparsely comparing with adjacent locations. $\frac{1}{\mathcal{D}_i+\epsilon}$ term (ϵ prevents division by 0 when $\mathcal{D}_i$ is 0) is accountable for assigning samples more closer to the camera. Although L_{SC_i} term encourages sparsity, we can optionally include penalty term in Eq. 5 to further prevent densely clustered sample locations. Algorithm 1 presents the pseudo-code of optimal sample distribution. Lines 4 - 8 calculate SSNR and distance from camera to all locations. From lines 9-28, we define a function that employs a greedy approach to select optimal sampling locations. The function initializes a matrix for tracking samples and an empty list for the chosen locations. It begins by calculating SCR for all potential locations, selecting the one with the highest SCR as the initial point, adding it to the list, and updating the matrix. For the remaining $K-1$ locations, it iteratively recalculates SCR and chooses new points that increase the total SCR. This strategy ensures the locations selected maximize the sum of SCR values, thereby optimizing the sampling effectiveness. The function returns the list of selected locations after K iterations. The $COMPUTE\_SCR$ function (lines 29-33) calculates the SCR for all potential sampling locations. It first computes the sensing capability L_{SC} for each location using the Eq. 4. We experimentally set $W_1 = 1$, $W_2 = 0.3$ and $W_3 = 0.1$. Then, it calculates the SCR for each location by multiplying the sensing capability with the inverse distance from the camera to that location.

Next, we train an LSTM model with the data collected from K locations. We then use the trained model to estimate intruder's location. Finally, our system calculates the minimum rotation angle required for the camera to cover the predicted location and determines the direction of rotation—either clockwise or counterclockwise.

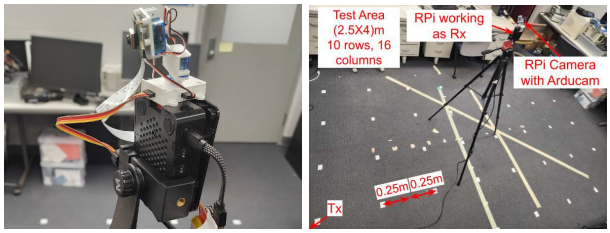
Algorithm 1 Optimal Sample Distribution

1: **Input:** Set of all locations $\mathcal{L}$ in $N \times M$ equal grid and their coordinates (X_i, Y_i)
2: **Input:** K, camera and receiver's location $C(0,0)$ and $Rx(0,0)$ respectively
3: **Output:** Subset of locations $\mathcal{S}$, where $\mathcal{S} \subseteq \mathcal{L}, |\mathcal{S}| = K$
4: Calculate SSNR Matrix $\mathcal{SS}$ and distance matrix $\mathcal{D}$:
5: **for** each (i, j) in the indices of $N \times M$ **do**
6: $\quad \mathcal{D}[i][j] \leftarrow \sqrt{(X_i)^2 + (Y_i)^2}$
7: $\quad \mathcal{SS}[i][j] \leftarrow \frac{D^2_{Tx,Rx}}{(\sqrt{(X_{Tx}-X_i)^2+(Y_{Tx}-Y_i)^2}\cdot\sqrt{X_i^2+Y_i^2})^2}$
8: **end for**
9: **function** GREEDY_SELECT_LOCATIONS($\mathcal{SS}$, (X_i, Y_i), K)
10: $\quad$ Initialize *collected_samples* and *selected_locations* as empty lists
11: $\quad$ $SCR \leftarrow$ *COMPUTE_SCR*($\mathcal{SS}$, (X_i, Y_i), collected_samples)
12: $\quad$ Initialize *best_initial_location* to None
13: $\quad$ **for** each (i, j) in the indices of *SCR* **do**
14: $\quad\quad$ Find maximum SCR sum and store (i, j) to *best_initial_location*
15: $\quad$ **end for**
16: $\quad$ Add *best_initial_location* to *selected_locations*
17: $\quad$ Set *collected_samples*[*best_initial_location*] to 1
18: $\quad$ **for** each k from 2 to K **do**
19: $\quad\quad$ $SCR \leftarrow$ *COMPUTE_SCR*($\mathcal{SS}$, (X_i, Y_i), collected_samples)
20: $\quad\quad$ Initialize *best_location* to None
21: $\quad\quad$ **for** each (i, j) in the indices of *SCR* **do**
22: $\quad\quad\quad$ Find maximum SCR sum and store (i, j) to *best_location*
23: $\quad\quad$ **end for**
24: $\quad\quad$ Add *best_location* to *selected_locations*
25: $\quad\quad$ Set *collected_samples*[*best_location*] to 1
26: $\quad$ **end for**
27: $\quad$ **return** *selected_locations*
28: **end function**
29: **function** COMPUTE_SCR($\mathcal{SS}$, (X_i, Y_i), collected_samples)
30: $\quad$ Compute sensing capability L_{SC} for all (i, j) locations using Eq.4
31: $\quad$ Compute SCR for all (i, j) locations using Eq.5
32: $\quad$ **return** *SCR*
33: **end function**

V. Implementation and Evaluation

A. System Implementation and Deployment

System Implementation: Our system involves 1 Raspberry Pi (RPi) camera (with horizontal FoV of 62.2 deg) attached with a RPi 4B, working as a receiver to collect CSI data and 1 transmitter. The RPi camera component is unable to pant and tilt by itself. Therefore, we attached the camera with Arducam component, which is capable of rotating horizontally and vertically, as shown in Fig.7a. We selected the RPi 4B for its widespread availability and cost-effectiveness, as well as its capability to support 80 MHz bandwidth for WiFi networks. However, the default firmware of its Broadcom WiFi chip does not support CSI data capture. To overcome this, we utilized a modified firmware developed by Nexmon [19], which enables the capture of CSI data by transmitting it from the link layer to the host system within frames embedded as transport layer payloads. Additionally, a TP-Link Archer A7 router, operating in the 5GHz band, served as a transmitter. We configured the system to ping the router every 5ms, prompting a pong packet response, which the RPi receives.



(a) System Implementation (b) Testbed Deployment

Fig. 7: System and Testbed Setup

System Deployment: We conducted experiments in a real indoor setting to thoroughly assess our system's performance. The environment was a lab office measuring 7.5m by 6m, equipped with few tables, chairs, several desktops and monitors. This office is characterized by rich multi-path reflections due to its furnishings. We defined a test area of 4m by 2.5m,

which we subdivided into 160 equidistant grids, each spaced 0.25m apart, as shown in Fig.7b. We obtained the ground truth locations using tape-measured (X, Y) coordinates. We placed C and Rx at $(0, 0)$ and Tx at $(-8, -5)$(Refer to Fig.8).

B. Data Collection and Processing

Deploying our system, we recruited three volunteers who each stood at one of the 160 designated locations, and collected 5,000 CSI samples per volunteer from each location, yielding total of 15,000 CSI samples per location. Complex numbers representing CSI phase and amplitude are extracted from the data sub-carriers [20]. The phase value of CSI was deemed highly sensitive and was therefore discarded, with only the absolute amplitude of CSI being retained for analysis. The CSI data collected by the Raspberry Pi is inherently noisy. To mitigate this, we employed the Least-square smoothing filter [21], which effectively smooths the CSI data while preserving the integrity of the waveform. After smoothing amplitude values of CSI, we averaged every five consecutive samples to produce a total of 3,000 CSI samples for each 160 locations with further reduced noise levels. Then the denoised and smoothed data were used to train and test LSTM models.

C. Baseline Sample Distribution Approaches

To the best of our knowledge, no existing state-of-the-art WiFi-based technique exists for guiding the camera's PTZ component to eliminate blind spots. Therefore, we implemented 2 strawman baseline approaches for comparison.

Baseline 1: Uniform distribution sampling approach: Baseline 1 represents the scenario where samples are collected at equal or regular intervals throughout the space, a method most commonly employed for data collection in indoor spaces for localization purposes [22]. In this baseline, we evaluate the performance of detecting intruders using a uniformly distributed sampling approach against our method of uneven sampling. Our system provides the flexibility to the user to collect samples from any number of K locations. For our sampling approach, we utilized Algorithm 1 to select the K optimal locations that achieve the highest SCR. Fig.8a, 8b and 8c show the uniform distribution of $K = 10, 15$ and 20 samples respectively across the 160 location grids. Similarly, Fig.8g, 8h and 8i exhibit sampling using our approach.

Baseline 2: SSNR based sampling approach: Baseline 2 involves selecting sample locations based on their sensing capabilities relative to the positions of Rx and Tx, without considering the impact on nearby areas. Similar to Baseline-1, we compare the accuracy of our system with this baseline. Fig.8d, 8e and 8f exhibit sample distribution approach of Baseline-2, clearly showing a tendency to allocate more samples in and around the LoS regions.

D. Model Training

During data collection phase, we collected data from all 160 locations with volunteers standing on each location. For Baseline-1, 2 and our approach, we trained and tested 3 LSTM models using data from only $K = 20$ locations, as illustrated in Fig.8c, 8f and 8i respectively. The reason behind selecting $K = 20$ is detailed in the next subsection. We divided the samples from these 20 locations into 3 parts: 60% for training, 10% for validation and 30% for testing the model. Each LSTM model includes 3 stacked layers with 256 hidden units each, a linear layer with inputs of 256 and outputs of 2, and a Softmax layer. We trained the models with a learning rate of 0.001 for up to 265 epochs, using Mean Square Error (MSE) for loss and the Adam optimizer. A 0.5 dropout rate is applied in the LSTM layers to prevent overfitting.

E. Performance Results

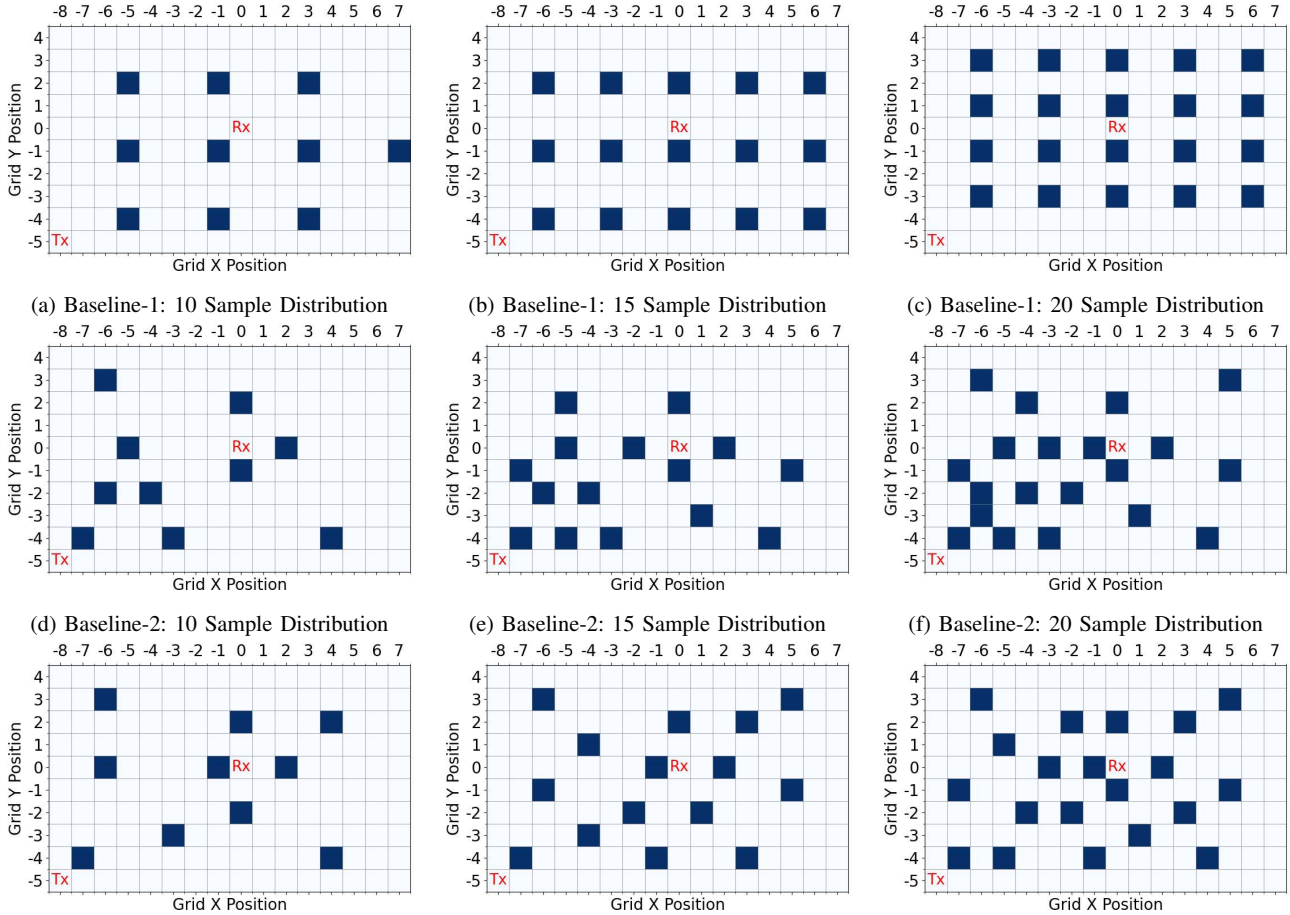
We emphasize that achieving high accuracy in the camera capturing model is possible with limited sample collection. This is due to the camera's FoV, which allows a certain degree of localization error to capture the intruder. From our extensive experiments, we found that $K = 20$ is the least number among the 160 locations to achieve a good camera capturing accuracy (SCR accuracy). Fig.10a illustrates how the greedy approach, as detailed in Algorithm 1, incrementally maximizes the $Total\ SCR$ through iterations to select the K best locations, resulting in a normalized $Total\ SCR$ of 54.2147. This value signifies that, by collecting data from only 20 (only 13%) of the 160 sample locations, our system can achieve a maximum normalized $Total\ SCR$ of 54.2147%.

Localization Error Comparison: We tested the 3 LSTM models by calculating location error distribution. The CDF shown in Figure 10b reveals that our approach achieved an 80th percentile error of 0.88m, compared to 1.41m for the Baseline-1 model. This represents a nearly 37% reduction in location error, significantly enhancing the camera's ability to capture intruders. Fig.10b indicates that Baseline-2 exhibits an 80th percentile error of 1.07m, while our system demonstrates a lower error of 0.88m. This represents a 17% reduction in error compared to Baseline-2.

Accuracy Comparison: We assessed and compared the accuracy of our system with both the baselines by having a volunteer stand at each location 10 times and recording how often the camera successfully detected them. The heatmaps in Figures 9a and 9c display the accuracy at each location, demonstrating that our system achieved approximately 24% higher accuracy than Baseline-1. Fig.9b and 9c demonstrate that our system can capture an intruder with an accuracy that is 18.6% higher than that of Baseline-2. This improvement is due to our system taking into account both the distance from the camera and the sensing impacts on surrounding locations.

F. Sensitivity Analysis of Variable Sampling Size

In our experiments and analyses so far, we started with an equal number of CSI samples (15,000) from optimally selected locations. To explore how changes in the number of CSI samples affect system accuracy, we partitioned our test environment into two areas: (1) LoS and (2) NLoS and camera-nearby regions, as depicted in Fig.11a (with dark and light grids representing NLoS and LoS regions, respectively). We collected 25,000 CSI samples from NLoS regions and 15,000

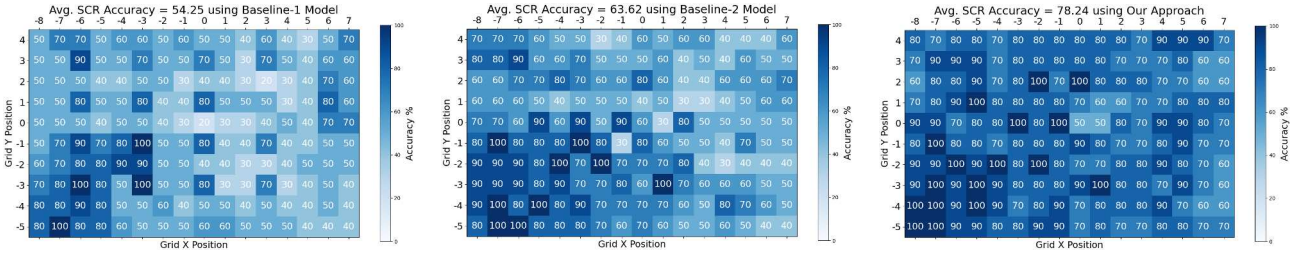


(a) Baseline-1: 10 Sample Distribution (b) Baseline-1: 15 Sample Distribution (c) Baseline-1: 20 Sample Distribution

(d) Baseline-2: 10 Sample Distribution (e) Baseline-2: 15 Sample Distribution (f) Baseline-2: 20 Sample Distribution

(g) Our Approach: 10 Sample Distribution (h) Our Approach: 15 Sample Distribution (i) Our Approach: 20 Sample Distribution

Fig. 8: 10, 15 and 20 Sample Distribution for Baseline-1 (a,b,c), Baseline-2 (d,e,f) and Our Approach (g,h,i)



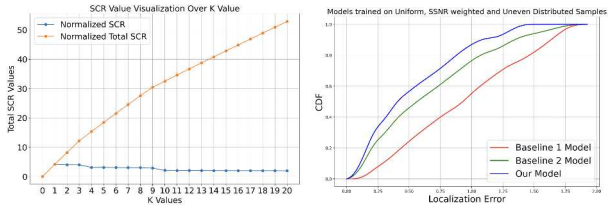
(a) Baseline-1 Avg. SCR Accuracy = 54.25 (b) Baseline-2 Avg. SCR Accuracy = 63.62 (c) Our Model Avg. SCR Accuracy = 78.24

Fig. 9: Location-wise SCR Accuracy for Models Trained with 20 Samples. Rx at (0,0) and L_{Tx}=(-8,-5)

from LoS regions to train an LSTM model (LM_1). Conversely, we gathered 15,000 CSI samples from NLoS regions and 25,000 from LoS regions to train another LSTM model (LM_2). After testing both models across all 160 locations, we found that LM_1 achieved an accuracy of 85.06%, which is around 5% higher than that of LM_2, as shown in Fig.11b. These results suggest that collecting a greater number of CSI samples from NLoS and camera-nearby locations, as opposed to LoS regions, can significantly enhance accuracy.

G. Discussion and Limitation

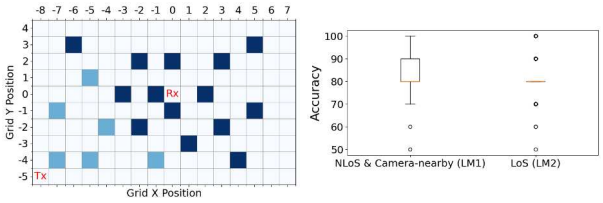
As far as we know, our approach represents the first attempt to guide a camera's PTZ component using passive WiFi sensing to eliminate blind spots. Our evaluations indicate that our method achieves high accuracy compared to other baseline models. Additionally, it significantly reduces the extensive manual effort required for data collection. We also recognize some of its limitations. Similar to other passive WiFi sensing applications like gesture recognition and patient monitoring,



(a) Maximizing SCR for K=20; Total SCR = 54.2147

(b) Comparison of Localization Error

Fig. 10: Normalized SCR and Location Error Comparison



(a) Higher Sampling Rate for NLoS & Camera Nearby Locations

(b) LM_1 Accuracy = 85.06, LM_2 Accuracy = 80.81

Fig. 11: Model Accuracy for LoS, and NLoS and Camera-nearby Locations

our approach is also affected by rich multipath environments. Its accuracy significantly decreases when multiple persons stand closely together at the same time. Our approach is limited to horizontal camera rotation, which can sometimes pose challenges in fully capturing the intruder. For future work, we aim to integrate sensing and communication using advanced protocols to optimize data transfer and enhance real-time surveillance coverage. Additionally, we will thoroughly explore the impact of a location on surrounding areas to better understand spatial relationships, applying these insights to scenarios where indoor location data is crucial.

VI. Conclusion

In this paper, we introduced a novel approach to eliminate blind spots in surveillance cameras utilizing WiFi sensing to guide their PTZ movements. Our method optimally selects a minimal number of sample locations from the surveillance area, informed by our empirical analysis, thereby achieving high accuracy in intruder detection. Implemented using Raspberry Pi and a Pi camera, our system demonstrated a significant improvement in efficiency, attaining a 78.24% accuracy rate in intruder capture while collecting data from only 13% of the locations—up to 24% more accurate than baseline methods. This success underscores our claim of ”covering more areas with less data,” marking a substantial progress in surveillance technology.

Acknowledgment

This research is supported by NSF through grant 2104337.

References

[1] Y. Yao, C.-H. Chen, B. Abidi, D. Page, A. Koschan, and M. Abidi, “Can you see me now? sensor positioning for automated and persistent surveillance,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 1, pp. 101–115, 2009.

[2] G. Lulla, A. Kumar, G. Pole, and G. Deshmukh, “Iot based smart security and surveillance system,” in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2021, pp. 385–390.

[3] M. Kim and T. Suh, “A low-cost surveillance and information system for museum using visible light communication,” *IEEE Sensors Journal*, vol. 19, no. 4, pp. 1533–1541, 2019.

[4] J. Tejedor, J. Macias-Guarasa, H. F. Martins, J. Pastor-Graells, P. Corredera, and S. Martin-Lopez, “Machine learning methods for pipeline surveillance systems based on distributed acoustic sensing: A review,” *Applied Sciences*, vol. 7, no. 8, p. 841, 2017.

[5] C. Perra, A. Kumar, M. Losito, P. Pirino, M. Moradpour, and G. Gatto, “Monitoring indoor people presence in buildings using low-cost infrared sensor array in doorways,” *Sensors*, vol. 21, no. 12, p. 4062, 2021.

[6] X. Wang, K. Niu, J. Xiong, B. Qian, Z. Yao, T. Lou, and D. Zhang, “Placement matters: Understanding the effects of device placement for wifi sensing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 1, pp. 1–25, 2022.

[7] V. Tsakanikas and T. Dagiuklas, “Video surveillance systems-current status and future trends,” *Computers & Electrical Engineering*, vol. 70, pp. 736–753, 2018.

[8] A. S. Olagoke, H. Ibrahim, and S. S. Teoh, “Literature survey on multi-camera system and its application,” *IEEE Access*, vol. 8, pp. 172 892–172 922, 2020.

[9] L. Gui, W. Yuan, and F. Xiao, “Blind-area elimination in video surveillance systems by wifi sensing with minimum qos loss,” in *2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*. IEEE, 2022, pp. 1–10.

[10] R. Wolniak and W. Grebski, “The usage of smart cameras in smart home.” *Scientific Papers of Silesian University of Technology. Organization & Management/Zeszyty Naukowe Politechniki Slaskiej. Seria Organizacji i Zarzadzanie*, no. 188, 2023.

[11] K. Adamkiewicz, P. Koch, B. Morawska, P. Lipiński, K. Lichy, and M. Leplawy, “Improving uwb indoor localization accuracy using sparse fingerprinting and transfer learning,” in *International Conference on Computational Science*. Springer, 2021, pp. 291–302.

[12] Z. Yu, Z. Chaczko, and J. Shi, “A novel algorithm modelling for uwb localization accuracy in remote sensing,” *Remote Sensing*, vol. 14, no. 19, p. 4902, 2022.

[13] N. Zhou, W. Sun, and M. Liang, “Human activity recognition based on wifi signal using deep neural network,” in *International Conference on Smart City and Informatization (iSCI)*. IEEE, 2020, pp. 26–30.

[14] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, “A survey on behavior recognition using wifi channel state information,” *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[15] T. Xin, B. Guo, Z. Wang, P. Wang, J. C. K. Lam, V. Li, and Z. Yu, “Freesense: A robust approach for indoor human detection using wi-fi signals,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–23, 2018.

[16] SafeTrolley, “Angle of coverage of cctv camera and ip camera,” 2022, accessed: 2023-06-25. [Online]. Available: https://www.safetrolley.com/angle-coverage-cctv-ip-camera/

[17] P. A. Burrough, R. A. McDonnell, and C. D. Lloyd, *Principles of geographical information systems*. Oxford University Press, USA, 2015.

[18] J.-P. Chiles and P. Delfiner, *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons, 2012, vol. 713.

[19] F. Gringoli, M. Schulz, J. Link, and M. Hollick, “Free your csi: A channel state information extraction platform for modern wi-fi chipsets,” in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, 2019, pp. 21–28.

[20] M. S. Gast, *802.11 ac: a survival guide: Wi-Fi at gigabit and beyond*. ” O’Reilly Media, Inc.”, 2013.

[21] R. W. Schafer, “What is a savitzky-golay filter?[lecture notes],” *IEEE Signal processing magazine*, vol. 28, no. 4, pp. 111–117, 2011.

[22] T. Yang, A. Cabani, and H. Chafouk, “A survey of recent indoor localization scenarios and methodologies,” *Sensors*, vol. 21, no. 23, p. 8086, 2021.