Multiple People Identification Through Walls Using Off-the-Shelf WiFi

Belal Korany, *Student Member, IEEE*, Hong Cai, *Member, IEEE*, and Yasamin Mostofi, *Fellow, IEEE*

***Abstract*—In this article, we are interested in through-wall gait-based identification of multiple people who are simultaneously walking in an area, using only the WiFi magnitude measurements of a small number of transceivers. This is a considerably challenging problem as the gait signatures of the walking people are mixed up in the WiFi measurements. In order to solve this problem, we propose a novel multidimensional framework, spanning time, frequency, and space domains, that can separate the signal reflected from each walking person and extract its corresponding gait content, in order to identify multiple people through walls. To the best of our knowledge, this is the first time that WiFi signals can identify multiple people in an area. We extensively validate our proposed system with 92 test experiments conducted in four different areas, where the WiFi transceivers are placed behind walls, and where two or three people (randomly selected from a pool of six test subjects) are walking in the area. Our system achieves an overall average accuracy of 82% in correctly identifying whether a person walking in the test experiment (referred to as a query) is the same as a candidate person, based on 6404 query-candidate test pairs. It is noteworthy that none of the test subjects/areas has been seen in the training phase.**

***Index Terms*—Gait analysis, multiperson identification, radio-frequency (RF) sensing, through-wall sensing, WiFi.**

I. Introduction

DUE TO the emergence of smart homes and Internet of Things, recent years have seen a rapid growth in the number of devices with wireless capabilities. These devices emit radio-frequency (RF) signals (e.g., WiFi), which interact with people and objects in the environment. As such, there has been considerable interest in utilizing these RF signals to extract various types of information about the people and/or the environment, e.g., occupancy estimation [1], [2], activity detection/recognition [3], localization [4], [5], tracking [6], and imaging [7]. In particular, identifying a person based on their gait signature (i.e., the way they walk), using RF signals, has recently started to gain a considerable attention [8]–[11]. Using RF signals for identification is appealing since it can recognize a person from a distance, without requiring them to perform any specific active task (e.g., fingerprint scanning). In addition, unlike cameras, it is more versatile as it does not require a clear, unobstructed view of the walking person, and is not affected by lighting conditions. Utilizing off-the-shelf WiFi for person identification is, in addition, useful due to the low cost and ubiquity of WiFi transceivers. In general, using existing RF signals for identification can open up new possibilities in the areas of security, personalized service provisioning, crowd analytics, and public health. For instance, an existing home WiFi network can identify the house members who just entered the house and provide personalized service accordingly. In addition, the WiFi network in a mall can identify the people (while anonymizing the data) and track where they go, which will be helpful for analyzing people's shopping preferences and the traffic flow. Furthermore, thanks to the wide coverage of WiFi nowadays, RF-based identification can also be used for contact tracing to help slow down the spread of diseases (e.g., COVID-19).

All the existing RF-based gait identification approaches, however, can only be used in situations where there is only one person walking in the area, which is an overly restrictive condition that significantly limits the practical use of this technology. In addition, most of the existing work on person identification with RF signals require training the system with extensive measurements of the same subjects to be identified, as well as training in the same operation area. This further limits the applicability of these systems since they cannot be used for new people or in new areas, without additional data collection and retraining.

In this article, we propose a novel multiperson identification system, using off-the-shelf WiFi devices, which is capable of recognizing the identities of multiple people when they are simultaneously walking in an area behind a wall. More specifically, given only the WiFi channel state information (CSI) magnitude signal received when a number of unknown people are walking in an unknown area behind a wall, our proposed system is able to extract the gait information of each individual person and identify them from the overall WiFi magnitude measurement. By identifying a person, we mean determining if this person is the same as the person that was walking in another area where a WiFi receiver measured its received power. In other words, given a WiFi sample measurement where a person was walking in an area (candidate sample) and WiFi measurements where a number of people are walking in another area behind the wall (query samples), our system can determine if the candidate person is among those behind the wall, without any need for collecting training data of any of these people. Moreover, our proposed system only uses the

Manuscript received June 25, 2020; revised September 29, 2020; accepted November 6, 2020. Date of publication November 16, 2020; date of current version April 7, 2021. This work was supported in part by NSF under Grant 1816931, and in part by the Office of Naval Research under Grant N00014-20-1-2779. *(Corresponding author: Belal Korany.)*

The authors are with the Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: belalkorany@ece.ucsb.edu; hcai@ece.ucsb.edu; ymostofi@ece.ucsb.edu).

Digital Object Identifier 10.1109/JIOT.2020.3037945

CSI magnitude measurements of a WiFi link, which is more preferable in practice since the phase measurements are either unavailable or not reliable on off-the-shelf devices [12]. To the best of our knowledge, this is the first time that WiFi-based identification is made possible for the setting of multiple simultaneously walking people, which can lead to a significantly broader applicability of WiFi-based gait identification systems. Furthermore, the ability to identify people, without any prior training data of them, can have a tremendous impact on this technology.

The main challenge in identifying multiple simultaneously walking people is that their gait information will be all mixed up in the received WiFi signal magnitude measurement. We then propose a novel technique that jointly processes the received signal across time, frequency, and the Angle-of-Arrival (AoA) domains, in order to separate, extract, and track the parts that correspond to each person's gait from the overall received WiFi signal magnitude. We then extract key features from each separated part of the signal for the purpose of identification. We extensively evaluate our proposed approach using a large test set, where all the test subjects and test areas are completely disjoint from the training data, thus allowing us to demonstrate the applicability of our system to new people and new areas. In the test set, there are six subjects and four areas where the transceivers are placed behind a wall. In each test experiment, two or three randomly selected test subjects walk simultaneously in a test area, amounting to a total of 92 such test experiments. Given a candidate spectrogram generated from the received WiFi signal measured when a person was walking in another area, our system then determines if a person behind the wall, among the multiple walking people, is the same as the candidate person. For the two-people case, given a pair of query and candidate spectrograms, our proposed system achieves an identification accuracy of 82%, in determining whether the two data samples belong to the same person, based on 4892 test pairs. Additionally, for the test experiments with three simultaneously walking people behind a wall, our system achieves an identification accuracy of 83%, based on 1512 test pairs. Overall, our results demonstrate the performance of the proposed approach for multiple people identification.

II. Related Work and Preliminaries

In this section, we provide a brief overview of the related work on gait-based person identification using RF signals and a brief primer on spectrogram-based gait analysis.

A. RF-Based Gait Identification

Early work on gait analysis utilize dedicated radar hardware and/or wideband signals to analyze a single person's walking pattern. For instance, Tahmoush and Silvious [13] and Orović *et al.* [14] used radar to estimate the motion parameters of different body parts. Vignaud *et al.* [15] discussed the challenge of multiperson gait analysis, stating that it was not possible to interpret the individual gait signatures of simultaneously walking people without sophisticated radar equipment with a very high range resolution.

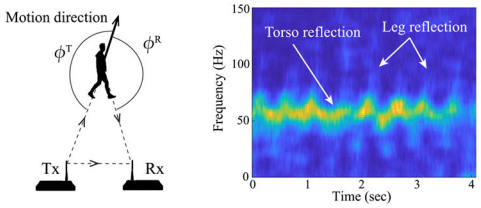

Fig. 1. (Left) Single-person gait-based identification scenario, where the person walks near a WiFi link. (Right) Spectrogram of the received WiFi magnitude measurements capturing the reflections off of different body parts.

Recently, there has been considerable interest in using off-the-shelf WiFi devices for gait-based person identification. WiFiU [8] uses WiFi CSI to generate spectrogram-based gait features, which are then used to classify the identities of a predefined set of people. WiWho [16] uses the time-domain signals, measured when a person is walking, for identification. Similarly, [11], [17]–[20] identify a person from a predefined set of people. In our previous work [10], we have proposed XModal-ID, a cross-modal person identification system from video to WiFi. More specifically, XModal-ID identifies whether a single walking person measured by WiFi is the same as a walking person in a video footage [10].

All the existing gait-based identification systems, however, are only applicable to situations where there is only one person walking in the WiFi area, and will fail when multiple people are simultaneously walking. In addition, with the only exception of our previous work [10], all the existing RF-based work on person identification need to collect training data of the same person that they want to identify. Furthermore, they need the training and test areas to be the same.

In this article, we propose a novel through-wall gait-based multiperson identification system using off-the-shelf WiFi devices. Our system is able to extract the gait information of each individual person from the aggregate WiFi magnitude measurement when multiple people walk simultaneously in an area. Furthermore, it does not require training with prior measurements of the subjects to be identified, does not need prior knowledge of the test/operation areas, and can identify people through walls. Finally, it only uses the CSI magnitude measurements of the received signal.

B. Spectrogram-Based Gait Analysis

Spectrograms are commonly utilized in existing work on gait analysis and gait-based person identification to extract the attributes of a person's gait [8], [10], [19], [21]. In this part, we provide some preliminaries on spectrograms and discuss the gait information that they capture.

Consider the scenario shown in Fig. 1 (left), where one person is walking near a pair of WiFi transmitter (Tx) and receiver (Rx). The squared magnitude of the received signal can be written as follows, as a function of time [10]:

$$s(t) = P_{\mathrm{DC}} + \sum_m \alpha_m \cos\left(\frac{2\pi}{\lambda}\psi v_m(t)t + \Omega\right) \tag{1}$$

where P_{DC} is the DC component that does not carry motion information about the gait, α_m is the magnitude of the reflection of the mth body part, v_m is the speed of the mth

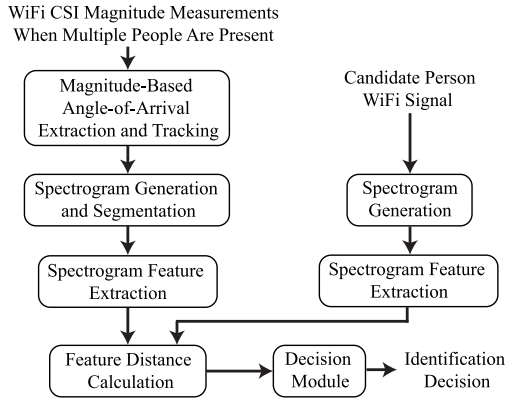

Fig. 2. Flowchart showing the modules of our proposed system for multiple people identification.

body part as a function of time, $\psi = \cos\phi^R + \cos\phi^T$ with ϕ^R and ϕ^T defined in Fig. 1 (left), Ω is a random initial phase, and λ is the wavelength of the wireless signal. It should be noted that ψ can, in general, be time varying. Equation 1 shows that the received signal consists of multiple sinusoids whose frequencies are related to the speeds of different body parts. Hence, by utilizing a time-frequency analysis technique (e.g., short-time Fourier transform), one can obtain the frequency content of the signal (i.e., the speed profile of the body parts) as a function of time, which is called the spectrogram of the signal. Fig. 1 (right) shows a sample spectrogram of the signal measured when a person is walking near the WiFi link. In the spectrogram, for instance, the stronger signal (indicated by the brighter colors) corresponds to the strong reflection from the torso of the person, while higher frequency weaker signals represent the reflections from the legs, which have higher speeds as compared to the average torso speed.

III. System Overview

In this article, we propose a system that can perform through-wall person identification on multiple people walking simultaneously in an area, all of whom have not been seen during the training phase of the system. The overall system architecture is outlined in the flowchart of Fig. 2, and summarized as follows.

1) Given the CSI magnitude of a WiFi signal measured at a small number of receivers when multiple people are walking simultaneously behind the wall in an area, we first detect and track the AoAs of the signal reflections from these people, using a combination of 2-D spectrum analysis, joint probabilistic data association filter (JPDAF), and track management techniques.
2) Given the AoAs corresponding to the different walking people in the area as a function of time, we extract the gait signal of each walking person by projecting the total received signal to the AoA of this person. Next, given the separated WiFi signal of one person (referred to as the query), we calculate the spectrogram of this signal to capture the time-frequency characteristics of this person's gait.
3) We then detect the segments from each spectrogram that are informative for identification. For each such segment, we extract several features to capture the person's gait signature.
4) In order to determine whether a detected spectrogram segment (i.e., a query segment) belongs to the same person as the candidate person, we calculate the feature distances between a query segment and a candidate spectrogram, and feed them to a neural network which determines whether these two spectrograms correspond to the same person. As there can be multiple informative segments from the query spectrogram of a person's entire track during a measurement period, we use an algorithm based on maximum likelihood estimation to fuse the segment-based decisions. The combined decision then indicates whether the query person of a walking track is the same as a candidate person.

IV. Proposed Methodology

In this section, we describe our proposed methodology for multiple people identification. We first describe the multiperson identification scenario and the corresponding signal model. We then show how to estimate and track the AoAs of the reflections from the walking people, based on only the received signal magnitude measurements, and subsequently separate the signals relevant to each walking person by projecting the overall signal to the respective AoAs. Given the separated signal that contains the walk of one person, a frequency-time spectrogram is generated and the informative segments of it are extracted for identification.

A. Signal Model for Joint Processing Based on Only Magnitude Measurements

Consider the scenario shown in Fig. 3, where a WiFi transmitter (Tx) transmits WiFi signals that are reflected off of N people walking in an area with S static objects, and then received by a receiver array (Rx). The received signal is then a combination of the direct signal from the Tx to the Rx, the reflected signals off of the S static objects, and the reflected signals off of the N walking people. More specifically, the time-domain received signal as a function of the distance ℓ along the array, where ℓ is measured with respect to the first antenna, can be written as follows [6]:

$$c(t,\ell) = \overbrace{\alpha_D e^{-j\frac{2\pi}{\lambda}\ell\cos\theta_D}}^{\text{Direct signal from Tx to Rx}} + \sum_{s=1}^{S} \overbrace{\tilde{\alpha}_s e^{-j\frac{2\pi}{\lambda}\ell\cos\tilde{\theta}_s}}^{\text{Reflected signal off of the s-th static scatterer}} + \underbrace{\sum_{n=1}^{N}\left(\sum_m \alpha_{n,m} e^{-j\frac{2\pi}{\lambda}v_{n,m}\psi_n^T t}\right) e^{-j\frac{2\pi}{\lambda}\ell\cos\theta_n}}_{\text{Reflected signals off of the nth person}} + \eta(t,\ell) \quad (2)$$

where α_D, $\tilde{\alpha}_s$, and $\alpha_{n,m}$ are, respectively, the complex values of the direct path from Tx to Rx, the sth static path (path reflected off of a static object), and the reflected path off of the mth body part of the nth person at $\ell = 0$, while θ_D, $\tilde{\theta}_s$, θ_n are their AoAs (measured with respect to the x-axis). S is the total number of static paths, $v_{n,m}$ is the speed of the mth body part of the nth person, $\psi_n^T = \cos\phi_n^T + \cos\phi_n^R$ is a parameter

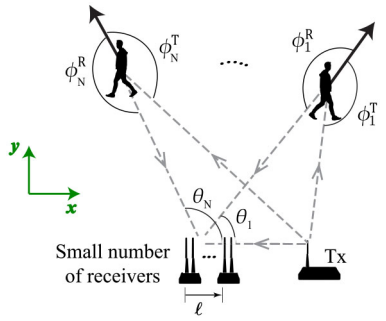

Fig. 3. Multiple people are walking simultaneously in an area, where a WiFi transmitter and a small number of receivers (placed behind a wall, without a direct view of the people) are used to collect WiFi measurements. The black arrows in the figure indicate the respective walking directions of the people.

related to the motion direction of the nth person (see Fig. 3 for definitions of ϕ_n^T and ϕ_n^R), $\eta(t,\ell)$ is additive noise, and λ is the wavelength of the WiFi signal.

Traditional AoA estimation relies on phase measurements across the receiver array. In our case, however, we only have the received magnitude measurements. We next show how the AoA information can be preserved in the magnitude measurements as well. The squared magnitude of the received WiFi signal can be calculated from (2), as shown in (3), at the bottom of the page. In (3), $P = |\alpha_D|^2 + \sum_n \sum_m |\alpha_{n,m}|^2 + \sum_s |\tilde{\alpha}_s|^2$ denotes the total power of the signal, $\psi_n^A = \cos\theta_D - \cos\theta_n$ is the difference between the cosine of the AoA of the direct path and the cosine of the AoA from the nth person, $\tilde{\psi}_s^A = \cos\theta_D - \cos\tilde{\theta}_s$ is the difference between the cosine of the AoA of the direct path from Tx to the Rx and the cosine of the AoA from the sth static object, and $\eta'(t,\ell)$ is additive noise. Since the Tx and the Rx are close to each other, the direct path from Tx to Rx will be much stronger than the other reflected paths (i.e., $|\alpha_D| \gg |\alpha_{n,m}|$ and $|\alpha_D| \gg |\tilde{\alpha}_s|$). Hence, the first three terms in (3) dominate the rest of the terms, which can then be neglected. Furthermore, by subtracting the temporal mean of the received signal at each of the receiver antennas (which can be easily implemented in practice), all the time-independent terms of (3) are zeroed out. More specifically, after subtracting the temporal mean, the squared magnitude of the received signal can be approximated as

$$\begin{aligned} c_{\text{mov}}(t,\ell) &= |c(t,\ell)|^2 - \mathbb{E}_t\{|c(t,\ell)|^2\} \\ &\approx \sum_{n=1}^{N}\sum_{m} 2|\alpha_D\alpha_{n,m}| \cos\left(\frac{2\pi\left(\psi_n^A + v_{n,m}\psi_n^T t\right)}{\lambda} + \mu\right) \\ &\quad + \eta'(t,\ell) \end{aligned} \tag{4}$$

where $\mu = \angle\alpha_D - \angle\alpha_{n,m}$, and $\mathbb{E}_t\{.\}$ denotes the temporal mean of the argument. The first term in (4) is a superposition of N terms, each of which corresponds to one of the moving persons, and contains a combination of sinusoids whose frequencies depend on the AoA of that person to the receiver (ψ_n^A) and the speeds of that person's body parts ($v_{n,m}\psi_n^T$). This shows that $c_{\text{mov}}(t,\ell)$ carries vital information about the AoAs of the reflected signals off of the walking people.

By taking the 2-D Fourier Transform of $c_{\text{mov}}(t,\ell)$ with respect to time t and distance ℓ, we then have

$$\begin{aligned} C(f_T, f_D) &= \left| \iint c_{\text{mov}}(t,\ell) e^{-j\frac{2\pi}{\lambda}(f_T t + f_D \ell)} dt\, d\ell \right| \\ &= \sum_{n=1}^{N}\sum_{m} |\alpha_D\alpha_{n,m}| \delta\left(f_T - v_{n,m}\psi_n^T, f_D - \psi_n^A\right) \\ &\quad + \sum_{n=1}^{N}\sum_{m} |\alpha_D\alpha_{n,m}| \delta\left(f_T + v_{n,m}\psi_n^T, f_D + \psi_n^A\right) + N_\circ \end{aligned} \tag{5}$$

where f_T is the temporal frequency, f_D is the spatial frequency, $\delta(*,*)$ is the 2-D Dirac Delta function, and $N_\circ$ is the noise floor.

It can be seen from (5) that each reflector in the operation area results in 2 deltas in the 2-D spectrum $C(f_T, f_D)$. By placing the Tx such that the AoA of the direct path from Tx to Rx is zero (i.e., $\cos\theta_D = 1$), the values of $\psi_n^A = 1 - \cos\theta_n$ are restricted to the interval $[0, 2]$, which is disjoint from the range of values of $-\psi_n^A$ [6]. Therefore, by constraining the search space for ψ_n^A to $[0, 2]$ (e.g., using only positive values for f_D), we obtain only one delta function for each reflector in the area at $(f_T, f_D) = (v_{n,m}\psi_n^T, \psi_n^A)$ in the 2-D spectrum. For the delta functions of the body parts of the same person, the one with the largest coefficient corresponds to the strongest reflecting part of the body, which is the torso. We refer to this strongest delta of a person as a peak and denote its location in the 2-D spectrum by $(\bar{\psi}_n^T, \psi_n^A)$. Fig. 4 shows an example of a 2-D spectrum $C(f_T, f_D)$ generated from WiFi measurements collected using an 8-element antenna array in a real experiment, where two people are walking simultaneously in an area. The ground-truth AoAs of the two people are 90° and 125° (i.e., $\psi_1^A = 1$ and $\psi_2^A = 1.57$), respectively. It can be seen that there are two peaks at $(f_T, f_D) = (-25, 1)$ and $(f_T, f_D) = (54, 1.6)$ in the 2-D spectrum, which correspond to the strong torso reflections of the two walking people.

$$\begin{aligned} |c(t,\ell)|^2 = c(t,\ell)c^*(t,\ell) &= P + \sum_{n=1}^{N}\sum_{m} 2|\alpha_D\alpha_{n,m}| \cos\left(\frac{2\pi}{\lambda}\left(\psi_n^A \ell + v_{n,m}\psi_n^T t\right) + \angle\alpha_D - \angle\alpha_{n,m}\right) \\ &+ \sum_{s=1}^{S} 2|\alpha_D\tilde{\alpha}_s| \cos\left(\frac{2\pi}{\lambda}\tilde{\psi}_s^A \ell + \angle\alpha_D - \angle\tilde{\alpha}_s\right) + \sum_{(n,m)}\sum_{s} 2|\alpha_{n,m}\tilde{\alpha}_s| \cos\left(\frac{2\pi}{\lambda}\left(v_{n,m}\psi_n^T t + (\cos\theta_n - \cos\tilde{\theta}_s)\ell\right) + \angle\alpha_{n,m} - \angle\tilde{\alpha}_s\right) \\ &+ \sum_{s}\sum_{s'\neq s} \tilde{\alpha}_s\tilde{\alpha}_{s'}^* e^{j\frac{2\pi}{\lambda}(\cos\tilde{\theta}_s - \cos\tilde{\theta}_{s'})\ell} + \sum_{(n,m)} \sum_{\substack{(n',m') \\ (n',m')\neq(n,m)}} \alpha_{n,m}\alpha_{n',m'}^* e^{j\frac{2\pi}{\lambda}\left((v_{n,m}\psi_n^T - v_{n',m'}\psi_{n'}^T)t + (\cos\theta_n - \cos\theta_{n'})\ell\right)} + \eta'(t,\ell) \end{aligned} \tag{3}$$

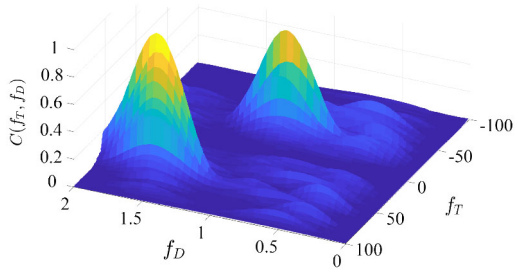

Fig. 4. Sample 2-D spectrum in the (f_T, f_D) domain for the case where two people are walking simultaneously in an area, based on a 0.4-s WiFi measurement. It can be seen that there are two peaks appearing in the 2-D spectrum, corresponding to the strong torso reflections of the two walking people.

Our new 2-D model can additionally detect whether a person is walking toward or away from the link. As can be seen from Fig. 3, the reflected signal from a person walking away from the link has a negative ψ^T, while that from a person walking toward the link has a positive ψ^T. The peak location in the 2-D spectrum then indicates the walking direction, as can be seen in Fig. 4, where the first person (whose $\bar{\psi}^T = -25$) is walking away from the link while the second person (whose $\bar{\psi}^T = 54$) is walking toward it. Hence, even if two people have the same AoA at the receiver array but are walking in different directions, their reflections are separable in the 2-D spectrum $C(f_T, f_D)$. On the other hand, in the traditional 1-D signal model of (1), the peak corresponding to a walking person appears twice, at $\pm v_m \psi^T$ at each time instance, which makes the motion direction of the walking person ambiguous, unless the information of all the subcarriers are exploited.

Based on the properties of the Fourier Transform, we summarize the following properties about the shape of $C(f_T, f_D)$ in the f_D dimension.

1) The width of a peak in the f_D dimension is inversely proportional to the length of the Rx antenna array. Thus, to increase the resolution (or separability) of the measurements in the AoA domain, one can increase the length of the array.
2) Based on the Nyquist sampling theorem, the maximum allowable antenna separation in the antenna array is determined by the maximum possible ψ^A. More specifically, the maximum antenna separation is given by: $\Delta\ell = \lambda/(2\psi^A_{\max})$. Therefore, if the AoAs of the reflected signals from the walking people span the range of [0, 180°], the maximum allowable antenna separation is $\lambda/4$.

Let $C_t(f_T, f_D)$ be the 2-D spectrum generated from a time window of the WiFi measurements from t to $t + T_{\text{win}}$. We then measure the temporal sequence of peaks appearing in consecutive snapshots of $.C_t(f_T, f_D)|_{t=0:\Delta t:T_{\max}}$ (where $T_{\max}$ is the total measurement duration and Δt is the time step) to keep track of the AoAs of different people, as we show next.

B. AoA Tracking

Given the 2-D spectrum $C_t(f_T, f_D)$ generated from a time window of the received WiFi signal from t to $t + T_{\text{win}}$, we extract a set of J_t peaks located at $\psi_j = (\bar{\psi}^T_j, \psi^A_j)$, for $j \in \{1, \ldots, J_t\}$. We denote the set of peak locations at this time instance by $\Psi_{\mathbf{t}} = \{\psi_j : j = 1, \ldots, J_t\}$. In order to separate and extract the part of the received signal that is relevant to each of the walking people in the area, the system needs to perform two main tasks as described as follows.

Algorithm 1 AoA Tracking Algorithm

Input: Set of peaks in the 2-D spectrum over time: $\Psi_{\mathbf{t}}|_{t=0:\Delta t:T_{\max}}$
Output: Set of active tracks and finished tracks: $\mathcal{A}$, $\mathcal{F}$

1: Initialize the sets of active tracks $\mathcal{A}$, potential tracks $\mathcal{P}$, and finished tracks $\mathcal{F}$ to empty sets: $\mathcal{A} = \emptyset$, $\mathcal{P} = \emptyset$, $\mathcal{F} = \emptyset$.
2: **for** $t = 0 : \Delta t : T_{\max}$ **do**
3: Find the association between the current set of peaks $\Psi_{\mathbf{t}}$ in the 2-D spectrum $C_t(f_T, f_D)$ and the current active and potential tracks using the JPDAF.
4: For each track $\mathcal{T} \in \mathcal{A} \cup \mathcal{P}$ that is associated with a peak in $\Psi_{\mathbf{t}}$, update its $\bar{\psi}^T$ and ψ^A using the associated peak.
5: For each peak $\psi_j \in \Psi_{\mathbf{t}}$ that is not associated with any active or potential tracks in $\mathcal{A} \cup \mathcal{P}$, initialize a new potential track $\mathcal{T}_{\text{new}}$ with ψ_j and add it to the current set of potential tracks: $\mathcal{P} = \mathcal{P} \cup \{\mathcal{T}_{\text{new}}\}$.
6: **for** each $\mathcal{T} \in \mathcal{P}$ **do**
7: **if** the potential track $\mathcal{T}$ meets confirmation criteria **then**
8: $\mathcal{T}$ is confirmed as an active track and is then moved to the set of active tracks: $\mathcal{P} = \mathcal{P} \backslash \mathcal{T}$, $\mathcal{A} = \mathcal{A} \cup \{\mathcal{T}\}$.
9: **else**
10: $\mathcal{T}$ is a false alarm and is then deleted from the set of potential tracks: $\mathcal{P} = \mathcal{P} \backslash \mathcal{T}$.
11: **end if**
12: **end for**
13: **for** each $\mathcal{T} \in \mathcal{A}$ **do**
14: **if** the active track $\mathcal{T}$ meets the termination criteria **then**
15: remove $\mathcal{T}$ from $\mathcal{A}$ and add it to the set of finished tracks: $\mathcal{A} = \mathcal{A} \backslash \mathcal{T}$, $\mathcal{F} = \mathcal{F} \cup \{\mathcal{T}\}$.
16: **end if**
17: **end for**
18: **end for**

1) AoA Track Management: A track $\mathcal{T}$ is defined as a temporal sequence of peaks in the 2-D spectrum resulting from the same moving person. As we do not assume that the system has any prior knowledge about the number of people in the area or when each person starts walking, it is necessary to automatically initialize a track when a new person starts walking in the area, or terminate a track when the corresponding person stops walking or leaves the area. We refer to a track for a person currently walking in the area as an *active track*, a newly initialized but not yet confirmed track (due to the possibility of false alarm) as a *potential track*, and a terminated track for which the person does not walk in the area anymore as a *finished track*.

2) AoA Data Association: At each time instance, given a set of active and potential tracks and the set of new peaks in the 2-D spectrum, the system needs to automatically associate each peak with the track of the corresponding walking person.

In order to perform these tasks, we propose an AoA tracking algorithm, as described in Algorithm 1. The algorithm takes as input the measurement sequence, in the form of the set of peaks in the 2-D spectrum over time. We then utilize the JPDAF to associate the peaks with the current active or potential tracks (lines 3 and 4 in Algorithm 1), with filter parameters

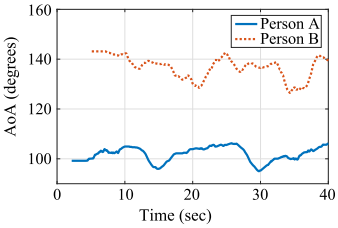

Fig. 5. Sample output of the AoA tracking algorithm. Measurements of person A are first detected at $t = 2.2$ s (blue solid curve). Measurements of person B are first detected at $t = 5.1$ s (red dashed curve). Both tracks remain active till the end of the experiment.

pertaining to the association process estimated from the training measurements. We refer the reader to [22] for more details on JPDAF. During the association, a measured AoA closer (in terms of Euclidean distance) to the current AoA of a track is more likely to belong to that track, as compared to an AoA that is farther away.

Once the JPDAF associates the peaks/measurements to the current set of tracks, our proposed algorithm enforces the following set of rules for track management.

1) *Initialization Criterion (Line 5 in Algorithm 1):* A potential track is initialized if a peak in the 2-D spectrum is not associated with any active or potential tracks at the current time instance.
2) *Confirmation Criterion (Line 7 in Algorithm 1):* A potential track is confirmed active if it is associated with some peaks in the 2-D spectrum for more than a percentage p_{T_C} of the time during the first T_C seconds after initialization. Otherwise, it is considered as a false alarm and discarded.
3) *Termination Criterion (Line 14 in Algorithm 1):* An active track is finished if it is not associated with any peaks for more than a percentage p_{T_F} of the time during the past T_F seconds.

Then, the total number of people in the area, N, will be estimated as the total number of active and finished tracks at the end of the experiment. We can then obtain the AoA time series of the nth person: $\hat{\theta}_n(t) = \cos^{-1}(1 - \psi_n^A(t))$, where $\psi_n^A(t)$ is the time series of ψ^A in the nth track $\mathcal{T}_n$. Fig. 5 shows a sample output of the AoA tracking algorithm for a 40-s experiment of two people walking. The track of person A is initialized at $t = 2.2$ s. The AoA time series of person A is shown by the blue solid curve in Fig. 5. Similarly, the track of person B is initialized at $t = 5.1$ s, with the corresponding AoA time series shown by the red dashed curve in the figure. For each track, the AoA sequence $\hat{\theta}_n(t), n = 1, \ldots, N$, is utilized to generate the spectrogram that only contains the gait information of the corresponding person, isolated from the signals of the other walking people, as we show next.

C. Spectrogram Generation and Segmentation

Given the time series of the AoA of the reflected path off of a walking person, we project the received signal at the Rx array, $c_{\text{mov}}(t, \ell)$, to the AoA of this person. The projected signal contains only the reflections from this person, and thus makes it possible to analyze this person's gait and perform identification. We then perform time-frequency analysis on the projected signal. More specifically, given $\hat{\psi}_n^A(t)$, the spectrogram for the nth walking person is generated by

$$S_n(t,f) = \left| \underbrace{\int_\tau \left(\int_\ell c_{\text{mov}}(\tau, \ell) e^{-j\frac{2\pi}{\lambda}\hat{\psi}_n^A(\tau)\ell} d\ell \right)}_{\substack{\text{projecting the signal} \\ \text{to the nth person's AoA}}} \chi(\tau; t) e^{-j2\pi f\tau} d\tau \right|^2 \quad (6)$$

where $\hat{\psi}_n^A(t) = 1 - \cos\hat{\theta}_n(t)$ and $\chi(\tau; t)$ is a time window function starting at time t. In other words, the measured signal $c_{\text{mov}}(\tau, \ell)$ across all the Rx antennas as a function of time is first projected to the nth person's AoA, using the operation inside the inner brackets in (6). This results in a 1-D signal (as a function of time) which contains only the reflected signal off of the nth person. This 1-D signal is then time-windowed using a window $\chi(\tau; t)$ and the frequency content of the time-windowed signal (which contains that person's gait information) is obtained using FFT. Different window functions $\chi(t)$ result in different favorable characteristics in the spectrogram. For instance, as shown in the literature, a rectangular window [in which case $S_n(t, f)$ is the short-time Fourier transform of the projected signal] produces a spectrogram that has clearer information on the speeds of the limbs, whereas the Hermite window functions produce spectrograms with better time-frequency resolution [10]. As such, we utilize a multiwindow spectrogram by additively combining the spectrograms generated using different window functions, as follows:

$$S_n(t,f) = \sum_{k=0}^{K} \left| \int_\tau \left(\int_\ell c_{\text{mov}}(\tau, \ell) e^{-j\frac{2\pi}{\lambda}\hat{\psi}_n^A(\tau)\ell} d\ell \right) \chi_k(\tau; t) e^{-j2\pi f\tau} d\tau \right|^2$$

where $\chi_0(\tau; t)$ is the rectangular window function starting at time t, and $\chi_k(\tau; t)$ is the kth Hermite function [14].

As an illustrative example for our spectrogram generation method, Fig. 6 (left) shows the spectrogram of the overall measured WiFi signal squared magnitude in a real experiment when two people are walking simultaneously in an area. It can be seen that the gait signals of the two people severely interfere with each other, making it extremely challenging to identify them. On the other hand, by utilizing our AoA tracking and spectrogram generation approach, we obtain two separate spectrograms [Fig. 6 (right)], where each spectrogram carries the gait information of only one individual.

When a person is walking toward/away from the WiFi link, the value of ψ^T is approximately constant when the Tx and Rx are close to each other. Hence, the frequency components of the spectrogram $S_n(t, f)$ directly correspond to the speeds of different body parts of the nth person [see (5)] and we do not have to consider the walking route to compensate for ψ^T.[1] Furthermore, in our previous work on cross-modal person identification [10], we have shown that the *most informative segments* of a spectrogram are those that correspond to the person walking away from the WiFi link as the reflection from the back of the body results in a cleaner signal. We

[1] See [10, Secs. 4 and 8] for more detailed discussions on ψ^T.

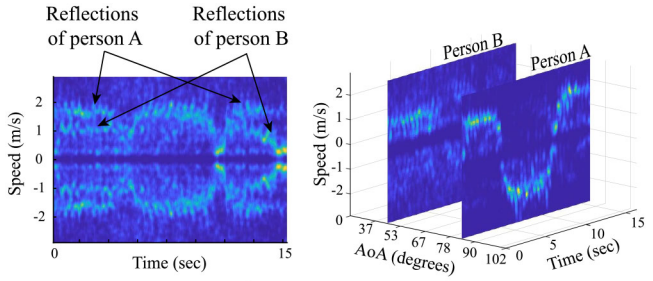

Fig. 6. (Left) Sample spectrogram of the received WiFi signal when two people are walking simultaneously in an area, showing that their gait signatures are mixed up in the overall signal. (Right) The output of our 2-D signal processing pipeline, showing two separate spectrograms each carrying the gait information of only one person.

then utilize the spectrogram segmentation algorithm in [10] to extract these informative segments (where ψ^T is constant and the person is walking away from the link) from the spectrogram of the entire walking duration of a person, with the same algorithm parameters as in [10]. The subjects can walk on any general path. The segmentation algorithm will then extract only the informative segments of each person's walk. Note that there can be more than one such informative segments from a person's track, depending on the duration and the path.

In case the reflected signals from two (or more) people are not separable for a long time, e.g., due to them always walking very close to each other, a corresponding extracted spectrogram segment could contain the mixed gait information of both people. Such segments, which are not useful for identification, can be automatically detected and removed, since their energy spread over frequency is very large due to the combined gait patterns. More specifically, we consider the energy spread of a segment obtained during operation to be abnormally large if it is 20% more than the maximum energy spread observed in the training spectrograms generated from single-person walking scenarios. The energy spread of a spectrogram is measured by the 80th percentile of the difference between the 60th and 40th percentiles of the frequency distribution across time. Additionally, when two people are walking very closely, it is also possible that a spectrogram segment of one person gets dominated by the gait information of the other person (e.g., due to a possibly stronger reflection from the person who is close-by). Such a segment can be automatically detected and discarded, based on its very high similarity to the segments of another person/track. The final remaining spectrogram segments are then considered as the *valid segments* to be used for identification.

V. Feature Extraction and Person Identification

Given a spectrogram segment of a person's walk derived from the separated WiFi signal, our proposed system identifies this person by comparing this query segment with a candidate spectrogram. More specifically, we extract several features from both the query segment and the candidate spectrogram, and then compute a set of distances between them. Given these feature distances, we train a neural network to determine whether a pair of query segment and candidate spectrogram belong to the same person. In this way, the neural network is able to identify people who have not been seen during training. Furthermore, if multiple valid segments are obtained from a person's track, we can fuse the neural network's decisions on these segments via a maximum likelihood approach.

A. Spectrogram Features and Distances

We extract 10 key features from a spectrogram to capture a person's gait. More specifically, we look at the frequency and time dimensions of the spectrogram, which carry different types of gait information that are useful for identification. The frequency dimension carries information about the speeds of different body parts, as shown in (4). We use the following features to capture various aspects of the frequency information.

1) *Frequency Distribution:* This feature is obtained by averaging the spectrogram over time. This feature captures the average distribution of the frequency components, i.e., the speed distribution of the body parts, during the person's walk.
2) *Frequency Distribution in Four Gait Phases:* These are the time averages of the spectrogram for each of the 4 phases of the gait cycle [8]. This feature provides more detailed information on the speed profile of the body parts of the walking person in each gait phase.
3) *Average Leg and Torso Speeds:* We calculate the averages of the torso speed curve and the leg speed curve, respectively, over time. These speed curves can be extracted from the spectrogram using the method in [8].

In addition, we use several time-domain features to capture the temporal information on a person's gait (e.g., the gait cycle) as follows.

1) *Average Autocorrelation:* Given a spectrogram, we compute the autocorrelation across time (with a maximum lag of 2 s) for each frequency bin. We then compute a weighted sum of the autocorrelation curves of the frequency bins, based on the energy distribution over the frequencies. This autocorrelation-based feature captures the periodic pattern of the speed profile of the person's body during walking.
2) *Autocorrelation of Percentile Curves:* Given a spectrogram, we extract the 50th and 70th percentiles of the frequency distribution over time. We then compute the autocorrelation of these two percentile curves, respectively. These features also capture the periodicity of the gait and are more robust to noise.

In order to quantitatively compare the similarity between the query spectrogram segment and the candidate spectrogram, for each of the 10 spectrogram features, we compute the distance between the corresponding features of the query spectrogram segment and the candidate spectrogram, which results in a vector of 10 feature distances. More specifically, for the frequency distributions, we use the Kullback–Leibler Divergence (KLD) as the distance metric, which is commonly used to compare distributions. For the autocorrelation-based features, we use the cosine similarity, which compares both the values and patterns of the autocorrelations. For the average speed features, we use the Euclidean distance.

B. Identification

Given a pair of a query spectrogram segment and a candidate spectrogram, we compute the 10 feature distances as described previously. We then utilize a simple fully connected neural network, which has 1 hidden layer with 30 units, to combine these distances into a binary identification decision. More specifically, the neural network takes as input a 10-D vector consisting of the feature distances and outputs a binary classification decision indicating whether or not the person of the query segment from the multiple people scenario is the same as the person of the candidate spectrogram. The neural network is trained on spectrogram pairs of subjects and locations disjoint from those of the test set (more details on the training set in Section VI-C). During training, these 10 distances between a pair of spectrograms of the training subjects are fed into the neural network, along with a binary label indicating whether these two spectrograms belong to the same person. A positive label of 1 indicates that the pair of spectrograms belong to the same person and a negative label of 0 indicates otherwise. Utilizing the softmax operation, the neural network outputs a scalar soft decision $\mathbf{s} \in [0, 1]$ that resembles the probability of declaring a same-person spectrogram pair given the input, i.e., $\mathbf{s} = p(1|\text{input})$. The neural network is trained using the cross-entropy loss and does not suffer from overfitting due to its simple structure. After training, we estimate the distribution of the values of the output $\mathbf{s}$ on all the positive-label training data, $p(\mathbf{s}|1)$, as well as its distribution on all the negative-label training data, $p(\mathbf{s}|0)$. We then estimate a Likelihood Ratio function: $\text{LLR}(\mathbf{s}) = p(\mathbf{s}|1)/p(\mathbf{s}|0)$.

During the testing phase, we first compute the distances between the previously described 10 features of the query segment (a valid spectrogram segment of a person among the group of walking people) and the candidate spectrogram. This 10-D distance vector is then fed into the trained neural network, for identifying whether or not the query spectrogram segment and the candidate spectrogram belong to the same person. The neural network outputs a soft decision $\mathbf{s}$, for which we calculate $\text{LLR}(\mathbf{s})$ using the LLR function estimated from the training set. If $\text{LLR}(\mathbf{s}) > 1$, a positive binary decision is declared that the person of the query spectrogram segment is the same as the person of the candidate spectrogram. Otherwise, the system declares that the query segment and the candidate spectrogram belong to two different people.

C. Multisegment Decision Fusion

As previously discussed in Section IV-C, there can be multiple valid spectrogram segments for the same track/person in an experiment. Let G_n be the number of valid segments for the nth person from a group of walking people. Let $\mathbf{s}_i$, $i \in \{1, \ldots, G_n\}$, be the soft decision outputs when the ith valid segment is tested against the candidate spectrogram. To optimally combine these soft decisions, in terms of maximum likelihood, we adopt the following decision rule: $p(\mathbf{s}_1, \ldots, \mathbf{s}_{G_n}|1) \underset{0}{\overset{1}{\gtrless}} p(\mathbf{s}_1, \ldots, \mathbf{s}_{G_n}|0)$. Furthermore, assuming independent decisions by the neural network on the different G_n inputs, the decision rule becomes: $\prod_{i=1}^{G_n} LLR(\mathbf{s}_i) \underset{0}{\overset{1}{\gtrless}} 1$, where the query person is declared to be the same as the one of the candidate spectrogram if the product of the LLR values corresponding to the multiple segments is greater than 1.

VI. Experimental Setup and Data Collection

In this section, we discuss our experiments used to validate our proposed system.

A. Experiment Setup

For the WiFi data collection, we use laptops equipped with the Intel 5300 WiFi Card to act as WiFi transmitter and receiver. For the transmitter, we use a tripod-mounted antenna connected to one port of the Intel card in one laptop, which transmits 1000 WiFi packets per second on WiFi channel 36 (with a center frequency of 5.18 GHz). For the receiver array, we use 4 laptops. More specifically, we use 8 antennas in a linear placement (with antenna separation of $\lambda/4$, except for one area, for which we use $\lambda/2$ separation), mounted to a tripod 1 m away from the Tx and connected to the ports of the Intel WiFi cards of 4 laptops, where each laptop provides 2 antenna ports. Since we rely only on the CSI magnitude measurements of the received WiFi signals, no calibration or synchronization between the multiple NICs is needed. This also facilitates increasing the antenna array length if needed. The CSI measurements of the receiver laptops are logged using Csitool [23] and processed offline using our proposed algorithm in Section IV. For spectrogram generation, we use a time window of length $T_{\text{win}} = 0.4$ s and a time shift of 4 ms.

The Intel 5300 WiFi cards report the CSI measurements on 30 different subcarriers. We utilize the extra measurements in the subcarrier domain to denoise the signals. More specifically, we utilize principal component analysis (PCA) to find the top 5 PCA components of the subcarrier measurements, and sum up their spectrograms in order to generate the denoised spectrogram. As opposed to traditional PCA denoising [8], where the measurement on each subcarrier is a 1-D temporal signal, the measurement on each subcarrier in our case is a 2-D spatiotemporal signal $c(t, \ell)$. Hence, we utilize the multidimensional PCA to find the top 5 PCA components [24].

For the AoA track management criteria (discussed in Section IV-B), we set the parameters as follows: $p_{T_C} = 50\%$, $T_C = 2$, $p_{T_F} = 90\%$, and $T_F = 10$ s.

B. Training Experiments

Training Subjects: We have recruited a total of 13 training subjects (ten male subjects and three female subjects) to collect data for training the neural network described in Section V-B. The training subjects have an average walking speed of 1.22 m/s (standard deviation: 0.19 m/s) and an average gait cycle of 1 s (standard deviation: 0.1 s).

Training Areas and Data Collection: The training data are collected in the two areas shown in Fig. 7(a) and (b), in a line-of-sight (non-through-wall) setting. For each training subject, we collect nine WiFi measurements of their walk in each

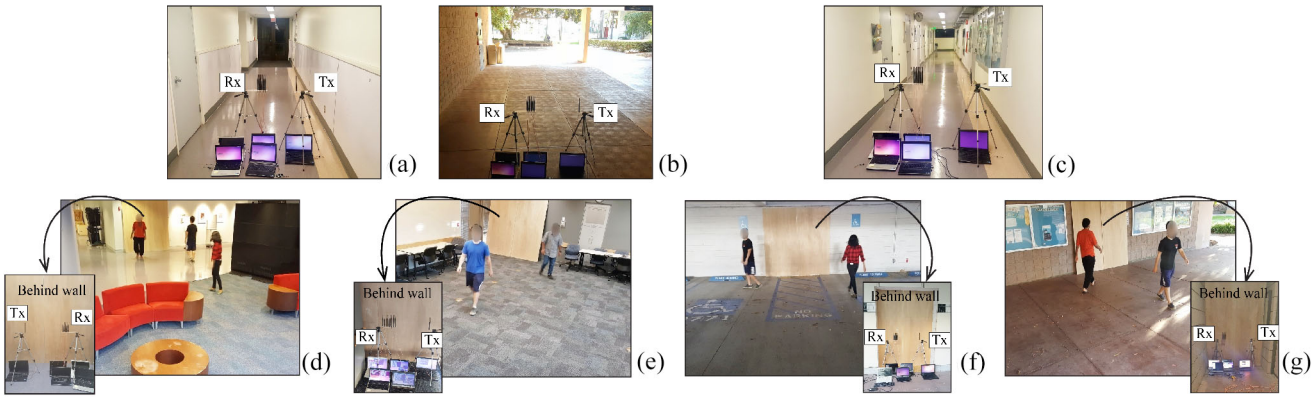

Fig. 7. (a) and (b) WiFi training areas: We collect the WiFi measurements of the training subjects in a line-of-sight setting in these two areas. (c) Test candidate area: We collect the measurements of the test subjects in this hallway area, in a line-of-sight setting, to be used for the candidate spectrograms. (d)–(g) Test areas: We conduct the test experiments in these four behind-wall areas (lounge, conference room, parking lot, and outdoor area), which represent a variety of real-world scenarios. The WiFi transmitter and receiver are placed behind the wall in each area, as shown by the black arrow.

of the two training areas. For each measurement, the training subject walks away from the link for about 10 m. We then transform each WiFi measurement into a corresponding spectrogram, which results in a total of 234 spectrograms for training.

Training Data Set: As our neural network of Section V-B is designed to predict whether a query spectrogram segment and a candidate spectrogram belong to the same person, we construct the training set in the form of spectrogram pairs. More specifically, based on the 234 training spectrograms, we generate a total of 27 261 spectrogram pairs. For each pair, we calculate the distances between the corresponding features of the two spectrograms and use this distance vector as one training sample. A training sample is given a positive label if the pair of spectrograms belong to the same training subject and a negative label otherwise. Since we have different numbers of positive and negative training samples, we utilize oversampling to obtain a balanced training set [25].

C. Test Experiments

Given the WiFi measurements when a person was walking in an area (referred to as the candidate person), and the WiFi measurements when multiple people are walking behind the wall in another area (test area), our system can separate the gait information of each person in the test area and determine if he/she (referred to as the query person) is the same as the candidate person. Both query and candidate people were not seen during training. In this part, we describe the details of our evaluation setup, including the test subjects, the data collection of the candidate spectrograms, and the test experiments.

Test Subjects: Our test set has 6 subjects (5 male subjects and 1 female subject), none of whom is part of the training set. Their heights range from 162 cm to 186 cm. They have an average speed of 1.31 m/s (standard deviation: 0.27 m/s) and an average gait cycle of 1.01 s (standard deviation: 0.15 s).

Candidate Spectrograms: We collect 10 WiFi measurements for each test subject in the area of Fig. 7 (c) in a line-of-sight (non-through-wall) setting to serve as candidate spectrograms. For each measurement, one test subject walks away from the link for about 10 m. The WiFi measurements are transformed into spectrograms, which then serve as the candidate spectrograms for identification. As a WiFi measurement in this setting only involves one person walking away from the link, each candidate spectrogram translates into one valid segment that continuously spans the entire walking duration. Thus, instead of using the term "candidate spectrogram segment", we use the shortened form "candidate spectrogram" in this article.

Test Experiments: We conduct the test experiments in four different test areas, as shown in Fig. 7(d)–(g), which are disjoint from both the training areas of Fig. 7(a) and (b) and the area where the candidate spectrogram was generated. These test areas represent several real-world scenarios with a variety of area size, geometry, and clutterness. More specifically, the area of Fig. 7(d) is a lounge area with couches and coffee tables. The area of Fig. 7(e) is a conference room with several tables and chairs. The area of Fig. 7(f) is located inside a parking structure. The area of Fig. 7(g) is a roofed outdoor area near a building. We conduct a total of 92 test experiments in these test areas, where the WiFi Tx and Rx are placed behind walls, with no direct view of the walking subjects. The walls consist of wooden panels that attenuate the wireless signals by 4 to 5 dB, which is similar to or larger than those of common nonconcrete and nonmetal building materials [26]. It is worth noting that wood is used for the walls of 90% of the residential and small commercial buildings in the U.S. [27].

First, consider the test experiments with two people (total of 80 experiments). In each such experiment, 2 of the 6 test subjects are randomly selected to walk simultaneously in the area. Each subject then walks casually on general paths. During the experiment, the AoAs of the walking subjects are tracked at the receiver, and their respective individual spectrograms are separated and extracted from the aggregate WiFi signal as described in Section IV. The informative and valid segments of the spectrograms of the tracks are automatically detected as discussed in Section IV-C, which are then used as the query segments. In the test experiments with three people, three of the six test subjects are randomly selected in a total of 12 experiments. The AoAs of the three walking people are then tracked to extract individual spectrograms, whose valid segments then serve as the query segments.

TABLE I
SEGMENT-BASED AND TRACK-BASED IDENTIFICATION ACCURACIES OF OUR PROPOSED SYSTEM ON THE TEST SET, IN 4 DIFFERENT THROUGH-WALL AREAS. THE LAST ROW SHOWS THE AVERAGE PERFORMANCE OVER ALL THE AREAS

Area	Two people		Three people	
	Segment based	Track based	Segment based	Track based
Area of Fig. 7 (d)	79%	81%	79%	73%
Area of Fig. 7 (e)	83%	82%	79%	83%
Area of Fig. 7 (f)	84%	86%	81%	87%
Area of Fig. 7 (g)	77%	78%	79%	83%
Average	**81%**	**82%**	**80%**	**83%**

Test Data Set for the Case of Two People: This test set consists of the test subjects' candidate spectrograms and the query spectrogram segments from the test experiments when two people were present, in the form of query-candidate pairs. More specifically, each test pair consists of a candidate spectrogram and a query spectrogram segment, and is assigned a positive label if they belong to the same person and a negative one otherwise. Our test set has in total 15 792 such test pairs.

Since there can be more than one query spectrogram segment obtained from the same person's track in an experiment, we also look at the test cases where the system identifies whether the person of a track is the same as the person of a candidate spectrogram, by fusing the decisions of all the valid segments from this track as discussed in Section V-C. In this setting, we have a total of 4892 pairs of a query track and a candidate spectrogram. We refer to the first setting where each query is one single segment as the *segment-based* setting and the second setting where the query consists of all the valid segments from a track of a person as the *track-based* setting.

Test Data Set for the Case of Three People: For the test experiments with three people walking simultaneously, the corresponding test set has 2416 query-candidate test pairs in the segment-based setting and 1512 test pairs in the track-based setting.

VII. SYSTEM EVALUATION

In this section, we present extensive experimental evaluations of our proposed system in four different through-wall areas (see Fig. 7), and for both cases of two and three people in the area. Furthermore, the subjects and areas in the test experiments have never been seen during the training phase.

A. Evaluation Criteria

We evaluate our proposed system by using pairs of a query spectrogram segment and a candidate spectrogram. Given such a test pair, the system identifies whether they belong to the same person or not. The resulting binary classification accuracy is used as the evaluation metric. As we have different numbers of test pairs with positive (same-person) labels and negative (different-people) labels, we report the balanced classification accuracy, i.e., the average of the respective accuracies over the same-person and different-people pairs.

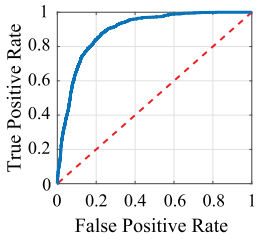

Fig. 8. ROC curve for identification in the track-based setting.

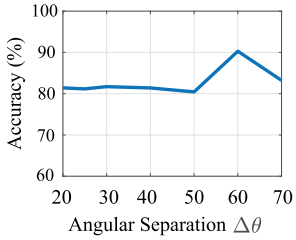

Fig. 9. Track-based identification accuracy as a function of the angular separation between the subjects' tracks in the two-people experiments.

B. Evaluation With Two Walking People

We evaluate our proposed system on our extensive test set, as described in Section VI-C, which contains only areas and subjects not seen during training. The results are summarized in Table I. For the track-based metric, our proposed system achieves identification accuracies of 81%, 82%, 86%, and 78% for the four areas, respectively, with an overall average accuracy of 82%. This indicates that given the spectrogram segments of a person's track in an experiment, our system is able to correctly identify whether this person is the same as the test subject of a candidate spectrogram for 82% of the time. As for the segment-based setting, our system achieves identification accuracies of 79%, 83%, 84%, and 77% for the 4 areas, respectively, with an overall average of 81%.

Finally, we show the receiver operating characteristic (ROC) curve for the track-based identification in Fig. 8. It can be seen that the ROC curve is significantly above the 45-degree line. In particular, the area under curve (AUC) is 0.89 in the track-based setting (and 0.88 in the segment-based setting), indicating a good performance. The AUC is equal to the probability that a randomly drawn positive sample has a higher score [e.g., the LLR($\mathbf{s}$) in Section V-B] than a randomly drawn negative sample.[2] It should be noted our threshold is automatically given by the neural network as described in Sections V-B and V-C. These results further confirm the performance of our system.

C. Angular Separation

We analyze the system's performance with respect to the angular separation between two walking people (difference between their AoAs) in the test areas. Fig. 9 shows the average identification accuracy as a function of the angle. It can be seen that even when the angle between the two people is

[2]See [28] for more details on ROC and AUC.

as small as 20°,[3] our system still performs robustly with an accuracy similar to those when the angular separation is larger.

D. Evaluation With Three Walking People

For the three-people test experiments, our proposed system achieves overall segment-based and track-based identification accuracies of 80% and 83%, respectively, averaged over the 4 through-wall test areas. It can be seen that even with three people simultaneously walking in the area, our system still achieves a good identification accuracy.

Overall, these extensive test results show that our proposed system can successfully perform identification when multiple people are simultaneously walking in the area, even though the people and areas have never been seen during training, and in a through-wall setting. They thus demonstrate the efficacy and applicability of our system in various real-world scenarios.

VIII. Discussion and Future Extensions

In this section, we provide discussions on various aspects of our proposed framework, as well as possible future extensions.

Gait as a Unique Signature: While some studies suggest that gait identification errors are inevitable for large groups (e.g., 100 people) [29], other studies have shown that RF-based gait identification can be effective for small to medium group sizes (e.g., up to 50 people) [8], [16]. It should be noted, however, that these papers have the same subjects in training and test, which makes the learning problem considerably simpler. They nonetheless show that gait-based identification has a good potential for applications that involve a small to medium pool of subjects (e.g., in a residential or an office setting).

Resolution of Peaks in the AoA Domain: The maximum number of people that the system can simultaneously detect is determined by the resolution (or width) of a peak in the AoA domain, which depends on length of the RX array. In our experiments, we used an antenna array of total length 2λ, resulting in a first-null peak resolution of 0.23, i.e., two peaks are completely separated in the AoA domain if the difference of their ψ^A values is greater than 0.23. This resolution can further be improved by extending the array, which does not add any synchronization/calibration overhead since we use only the magnitude of the received WiFi signal. Different processing techniques (e.g., MUSIC) can also improve the resolution of the peaks, at the expense of much higher computational costs.

Increasing the Number of People in the Area: While we have tested our system for up to three people walking simultaneously in an area, the proposed approach can be used for a larger number of people. The maximum number of people that the system can identify is determined by the resolution of a peak in the AoA domain, which depends on the length of the RX antenna array, as discussed above, as well as how crowded the area gets, how close to each other the people walk, and their walking directions. For instance, in our current setup, the AoA resolution is 10 degrees. This means that if two people consistently walk with AoA separation of less than 10 degrees in the same direction, our current setup cannot separate them (if they walk in different directions, then it is still separable). As such, as long as the area is large enough such that each person's AoA is not consistently less than 10 degrees from someone else who is going in the same direction, then the current system is able to identify all the people. As the area gets more crowded for its size, then one can increase the length of the array to allow for a better AoA resolution, which would allow our system to identify people even if they are consistently walking close to each other in the same direction. One can also utilize more resources, such as positioning more transceivers in different locations in the space in order to capture different views of the people, and/or utilizing different frequency subcarriers to capture the time-of-flight data.

Tracks of the Walking People: In Section IV-C, we have utilized the fact that when people move away/toward the link, ψ^T is constant, excluding the need to directly estimate it via tracking people's routes. In the rare case when no such segment is detected in the entire spectrogram (e.g., if the person's whole walking path is parallel to the link), the constantly varying value of ψ^T can be estimated by existing RF-based tracking systems, and then incorporated in the proposed approach to identify a person walking on any general track.

Impact of Misalignment of the Rx Array: When the Rx array is rotated by θ_e, the AoA of a person changes from θ to $\theta + \theta_e$, where θ is the AoA before the rotation. The separability of the reflected signals from two people depends on the difference between their respective ψ^A values, where $\psi^A = \cos(\theta_e) - \cos(\theta + \theta_e)$. As such, given a small θ_e, the separation between two people in the 2-D spectrum can slightly increase or decrease due to the nonlinear cosine operation. For instance, consider two people whose original AoAs are 120° and 150°. When the array is rotated by 5°, the separation between their ψ^A values decreases from 0.37 to 0.33. Note that given our current hardware setup, the two people's signals are fully separable in the 2-D spectrum as long as their ψ_A values are at least 0.23 apart. Our system is also robust to small interantenna spacing errors, since they only slightly affect the width of the main beam and the side lobe levels, but not the location/AoA of the main beam [30].

IX. Conclusion

In this article, we proposed a gait-based identification system that can, for the first time, identify multiple simultaneously walking people through walls, using CSI magnitude measurements of a small number of off-the-shelf WiFi devices. In order to do so, our system first estimates the AoA of the reflected WiFi signals from the walking people, and uses this information to separate their gait signatures. Given the extracted signal of an individual person's walk, our system generates a spectrogram to capture the frequency-time features of the gait for identification. We have extensively validated our proposed system with 92 test experiments in four test areas. In each experiment, two or three test subjects walk simultaneously in an area. Overall, our system achieves an

[3]Note that the minimum angular separation resolvable by the system depends on the length of the Rx array. For instance, given our current hardware setup, the system will not be able to fully resolve two angles that are 10° (or less) apart. See Section VIII for a detailed discussion on AoA resolution.

overall average accuracy of 82% in identifying whether the person of a query data sample (extracted from a multiperson walking experiment) is the same as the person of a candidate spectrogram.

REFERENCES

[1] S. Depatla and Y. Mostofi, "Crowd counting through walls using WiFi," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2018, pp. 1–10.

[2] S. Kianoush, S. Savazzi, V. Rampa, and M. Nicoli, "People counting by dense WiFi MIMO networks: Channel features and machine learning algorithms," *Sensors*, vol. 19, no. 16, p. 3450, 2019.

[3] T. Wang, D. Yang, S. Zhang, Y. Wu, and S. Xu, "Wi-Alarm: Low-cost passive intrusion detection using WiFi," *Sensors*, vol. 19, p. 2335, May 2019.

[4] C. R. Karanam, B. Korany, and Y. Mostofi, "Magnitude-based angle-of-arrival estimation, localization, and target tracking," in *Proc. ACM/IEEE Conf. Inf. Process. Sens. Netw.*, 2018, pp. 254–265.

[5] I. Bisio, F. Lavagetto, M. Marchese, and A. Sciarrone, "Performance comparison of a probabilistic fingerprint-based indoor positioning system over different smartphones," in *Proc. Int. Symp. Perform. Eval. Comput. Telecommun. Syst.*, 2013, pp. 161–166.

[6] C. R. Karanam, B. Korany, and Y. Mostofi, "Tracking from one side: Multi-person passive tracking with WiFi magnitude measurements," in *Proc. ACM/IEEE Int. Conf. Inf. Process. Sens. Netw.*, 2019, pp. 181–192.

[7] A. Gonzalez-Ruiz and Y. Mostofi, "Cooperative robotic structure mapping using wireless measurements—A comparison of random and coordinated sampling patterns," *IEEE Sensors J.*, vol. 13, no. 7, pp. 2571–2580, Jul. 2013.

[8] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proc. ACM Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 363–373.

[9] A. Seifert, M. Amin, and A. M. Zoubir, "Toward unobtrusive in-home gait analysis based on radar micro-Doppler signatures," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 9, pp. 2629–2640, Sep. 2019.

[10] B. Korany, C. R. Karanam, H. Cai, and Y. Mostofi, "XModal-ID: Using WiFi for through-wall person identification from candidate video footage," in *Proc. ACM Int. Conf. Mobile Comput. Netw. (Mobicom)*, 2019, pp. 1–15.

[11] C. Wu, F. Zhang, Y. Hu, and K. R. Liu, "GaitWay: Monitoring and recognizing gait speed through the walls," *IEEE Trans. Mobile Comput.*, early access, Feb. 19, 2020, doi: 10.1109/TMC.2020.2975158.

[12] Y. Zhuo, H. Zhu, and H. Xue, "Identifying a new non-linear CSI phase measurement error with commodity WiFi devices," in *Proc. IEEE Int. Conf. Parallel Distrib. Syst.*, 2016, pp. 72–79.

[13] D. Tahmoush and J. Silvious, "Stride rate in radar micro-Doppler images," in *Proc. IEEE Conf. Syst. Man Cybern.*, 2009, pp. 4218–4223.

[14] I. Orović, S. Stanković, and M. Amin, "A new approach for classification of human gait based on time-frequency feature representations," *Signal Process.*, vol. 91, no. 6, pp. 1448–1456, 2011.

[15] L. Vignaud, A. Ghaleb, J. Le Kernec, and J. Nicolas, "Radar high resolution range & micro-doppler analysis of human motions," in *Proc. Int. Radar Conf.*, 2009, pp. 1–6.

[16] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in *Proc. ACM/IEEE Int. Conf. Inf. Process. Sens. Netw.*, 2016, pp. 1–12.

[17] Y. Li and T. Zhu, "Using Wi-Fi signals to characterize human gait for identification and activity monitoring," in *Proc. IEEE Conf. Connected Health Appl. Syst. Eng. Technol.*, 2016, pp. 238–247.

[18] T. Xin, B. Guo, Z. Wang, M. Li, and Z. Yi, "Freesense: Indoor human identification with Wi-Fi signals," in *Proc. IEEE Global Commun. Conf.*, 2016, pp. 1–7.

[19] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "WiFi-ID: Human identification using WiFi signal," in *Proc. Int. Conf. Distrib. Comput. Sens. Syst.*, 2016, pp. 75–82.

[20] J. Lv, W. Yang, D. Man, X. Du, M. Yu, and M. Guizani, "Wii: Device-free passive identity identification via WiFi signals," in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–6.

[21] B. Vandersmissen *et al.*, "Indoor person identification using a low-power FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3941–3952, Jul. 2018.

[22] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE J. Ocean. Eng.*, vol. 8, no. 3, pp. 173–184, Jul. 1983.

[23] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[24] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.

[25] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data Mining and Knowledge Discovery Handbook*. Boston, MA, USA: Springer, 2009.

[26] R. Wilson, "Propagation losses through common building materials 2.4 GHz vs. 5 GHz," Magis Networks, Inc., San Diego, CA, USA, Rep. E10589, 2002.

[27] A. Sinha, R. Gupta, and A. Kutnar, "Sustainable development and green buildings," *Drvna Industrija*, vol. 64, no. 1, pp. 45–53, 2013.

[28] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[29] D. Gafurov, E. Snekkenes, and P. Bours, "Spoof attacks on gait authentication system," *IEEE Trans. Inf. Forensics Security*, vol. 2, pp. 491–502, 2007.

[30] C.-C. Yu, "Sidelobe reduction of asymmetric linear array by spacing perturbation," *Electron. Lett.*, vol. 33, no. 9, pp. 730–732, Apr. 1997.

Belal Korany (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical and computer engineering from Cairo University, Giza, Egypt, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering, in the area of communications and signal processing with the University of California Santa Barbara, Santa Barbara, CA, USA.

His research interests include RF sensing, wireless communications, and array signal processing. His research has appeared in several reputable news outlets, such as BBC, Yahoo, CNET, Science Daily, Tech Xplore, and others.

Hong Cai (Member, IEEE) received the B.E. degree in electronic and computer engineering from Hong Kong University of Science and Technology, Hong Kong, in 2013, and the M.S. degree in electrical and computer engineering from the University of California Santa Barbara, Santa Barbara, CA, USA, in 2015, where he is currently pursuing the Ph.D. degree in electrical and computer engineering, in the area of communications, control, and signal processing.

His research interests include robotic visual understanding and robot decision optimization.

Yasamin Mostofi (Fellow, IEEE) received the B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1997, and the M.S. and Ph.D. degrees in wireless communications from Stanford University, Stanford, CA, USA, in 1999 and 2004, respectively.

She is currently a Professor with the Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA, USA. Her research is on mobile sensor networks. Current research thrusts include X-ray vision for robots, RF sensing, communication-aware robotics, occupancy estimation, see-through imaging, and human–robot collaboration. Her research has appeared in several reputable news venues, such as BBC, Huffington Post, Daily Mail, Engadget, and NSF Science360, among others.

Dr. Mostofi is a recipient of the 2016 Antonio Ruberti Prize from IEEE Control Systems Society, the Presidential Early Career Award for Scientists and Engineers, the National Science Foundation CAREER award, and the IEEE 2012 Outstanding Engineer Award of Region 6, among other awards. She currently serves on the Board of Governors for IEEE Control Systems Society and is also a Senior Editor for the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS.