GaitSense: Towards Ubiquitous Gait-Based Human Identification with Wi-Fi

YI ZHANG, Tsinghua University, China
YUE ZHENG, Tsinghua University, China
GUIDONG ZHANG, Tsinghua University, China
KUN QIAN, University of California San Diego, USA
CHEN QIAN, Tsinghua University, China
ZHENG YANG*, Tsinghua University, China

Gait, the walking manner of a person, has been perceived as a physical and behavioral trait for human identification. Compared with cameras and wearable sensors, Wi-Fi-based gait recognition is more attractive because Wi-Fi infrastructure is almost available everywhere and is able to sense passively without the requirement of on-body devices. However, existing Wi-Fi sensing approaches impose strong assumptions of fixed user walking trajectories, sufficient training data, and identification of already known users. In this paper, we present *GaitSense*, a Wi-Fi-based human identification system, to overcome the above unrealistic assumptions. To deal with various walking trajectories and speeds, *GaitSense* first extracts target specific features that best characterize gait patterns and applies novel normalization algorithms to eliminate gait irrelevant perturbation in signals. On this basis, *GaitSense* reduces the training efforts in new deployment scenarios by transfer learning and data augmentation techniques. *GaitSense* also enables a distinct feature of illegal user identification by anomaly detection, making the system readily available for real-world deployment. Our implementation and evaluation with commodity Wi-Fi devices demonstrate a consistent identification accuracy across various deployment scenarios with little training samples, pushing the limit of gait recognition with Wi-Fi signals.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Gait Recognition, Channel State Information, Commodity Wi-Fi

ACM Reference Format:
Yi Zhang, Yue Zheng, Guidong Zhang, Kun Qian, Chen Qian, and Zheng Yang. 2021. GaitSense: Towards Ubiquitous Gait-Based Human Identification with Wi-Fi. *ACM Trans. Sensor Netw.* 1, 1 (May 2021), 25 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

---

A preliminary version [40] of this article appeared in International Conference on Wireless Algorithms, Systems, and Applications (WASA 2020).
*Corresponding author.

Authors' addresses: Yi Zhang, zhangyithss@gmail.com, Tsinghua University, Beijing, China; Yue Zheng, cczhengy@gmail.com, Tsinghua University, Beijing, China; Guidong Zhang, zhanggd18@gmail.com, Tsinghua University, Beijing, China; Kun Qian, qiank10@gmail.com, University of California San Diego, San Diego, California, USA; Chen Qian, chen.cronus.qian@gmail.com, Tsinghua University, Beijing, China; Zheng Yang, hmilyyz@gmail.com, Tsinghua University, Beijing, China.

1550-4859/2021/5-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

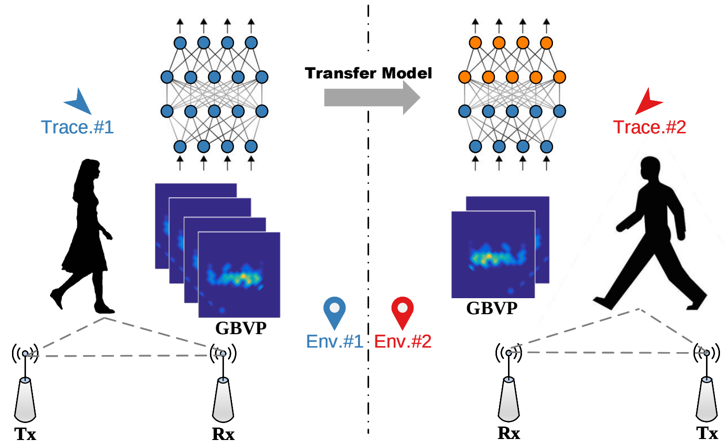


Fig. 1. System overview.

1 INTRODUCTION

Person identification is an important prerequisite that triggers many applications concerning convenience, security and, privacy, such as smart building [45], intruder detection [15] and area access [5]. Various biometric signs, e.g., fingerprint [5], iris [7], voice [8], vital signs [19, 29] and gait [30], have been widely studied and exploited for person identification. Among these modalities, gait acts as one of the most convenient signatures as it is not required to be actively inputted by a person, and can be measured within a wide range that enables room-scale and even home-scale person identification.

Various sensor modalities, e.g., cameras [10], inertial sensors on wearables [39] and Wi-Fi signals emitted by wireless devices [30], have been certificated to be able to extract the gait of a person for identification. Among these sensors, video requires a deliberate deployment of surveillance cameras in monitoring areas and has the risk of privacy leakage. Inertial sensors require users to actively carry mobile devices. In contrast, Wi-Fi signals have become a more attractive carrier for gait-based person identification since Wi-Fi infrastructure is ubiquitously available [34], and recent research has shown that it can recognize one's gait passively without wearing any devices.

The current state of Wi-Fi-based gait identification approaches [30, 37, 38], however, rely on extensive training efforts for *every* target person in *each* monitoring area. Such cumbersomeness stems from three limitations of the existing approaches. First, Wi-Fi signals reflected by a target person not only possess the gait signature of the person but also are distorted by the surrounding multipath environment. Thus, the recognition model directly trained with raw Wi-Fi features or their statistics, as WiWho [37], may overfit the environment where the data is collected and cannot be generalized to new environments without retraining. Second, besides the effect of environmental factors, features related to the gait of the person in Wi-Fi signals still depend on how the person moves relative to the Wi-Fi devices. WiFiU [30] derives parameters of gaits from Doppler Frequency Spectrum (DFS). However, it requires that the target person walks right towards or away from the Wi-Fi devices on fixed trajectories to ensure the consistency of the DFS profile, which limits the practicality of the approach. Third, as the learning model becomes more and more sophisticated, e.g., in terms of the number of parameters that need to be trained, a sufficiently large amount of training data is required when each new person is added. In a recent work [26], a 3-layer Deep Neural Network (DNN) requires approximately 300 training samples to be collected for each person

Table 1. A comparison of state-of-the-art works for Wi-Fi-based gait recognition

Properties	WifiU [30]	WiWho [37]	WiFi-ID [38]	AutoID [44]	**GaitSense**
Arbitrary track	No	No	No	No	**Yes**
Wi-Fi link(s)	1	1	1	1	**6**
Feature	DFS	CSI+DFS	CSI+DFS	CSI	**GBVP**
Training samples/user	∼50	∼20	∼20	∼10	**∼50**
Maximum user numbers	50	7	6	20	**11**
Accuracy[1] (#users)	79% (50)	75% (7)	77% (6)	90% (20)	**76% (11)**

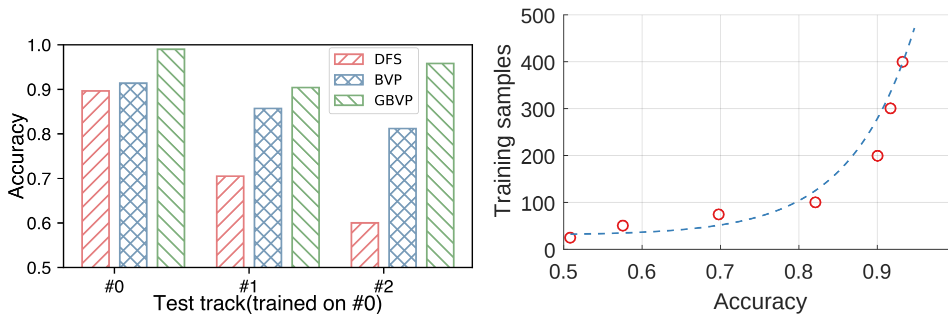
[1] Note that the reported accuracies in these works are for different numbers of users and we only present the results for the maximum number of users in these works.

for training. Moreover, existing approaches only focus on identifying legal users whose gait data has been collected, but make little effort on detecting unauthorized persons. With these limitations, existing approaches are hardly practical nor scalable. Table.1 compares recent Wi-Fi-based gait recognition systems from aspects of resource requirements and system performance. All the existing works impose strict restrictions on the walking tracks, undermining their application prospect. On the contrary, GaitSense is less intrusive to the users by performing recognition for arbitrary walking tracks, making it one of the most practical Wi-Fi-based gait recognition systems. Even though the number of Wi-Fi links used by GaitSense is more than other works, we believe the deployment overhead is acceptable given that more Wi-Fi infrastructures are on their way to our daily lives.

In this paper, we present GaitSense, a ubiquitous Wi-Fi-based person identification framework, which is robust to walking manners and environmental variance and reduces training efforts significantly as Fig.1 shows. GaitSense has three key characteristics that enhance the robustness of this system with limited training samples. First, GaitSense is immune to environmental variations and motion status (e.g., location and velocity) of target persons, and retains prominent generalizability between environments and trajectories. Second, GaitSense is capable of transferring the learned model of existing persons to newcomers with only a small amount of training data collected from the person. Third, GaitSense is designed to detect anomalies of gaits when unauthorized users appear for more robust identification. To support these features, we overcome three critical challenges.

The first challenge is to overcome the negative impact of environmental variations and the motion status of the target person during walking. GaitSense borrows the idea of body-coordinate velocity profile (BVP) [43], which represents the velocities of body parts during walking. The BVP is resilient to scenario factors, including environmental changes, and the location and orientation of walking trajectories. However, BVP can not be directly used as a gait feature, as it is dedicated to in-place activities where movements of the torso and legs are ignored. Besides, the BVP extraction algorithm is too complicated and can not be applied in real-time systems. To adapt BVP to real-time gait identification, we propose an agile algorithm to extract gait-specific feature GBVP and devise novel normalization algorithms to boost its robustness to gait-irrelevant factors. The designed feature is theoretically both environment and trajectory-independent, which mitigates gait-irrelevant components.

The second challenge is to effectively train the model with a small quantity of data collected from each new user. GaitSense adopts a deep neural network as the gait identification model, which is proved to be effective but has so sophisticated structure that requires a large number of data samples to get fully trained. To overcome the challenge, GaitSense exploits transfer learning



(a) Cross-track performance for different features. (b) Dataset scale demanded by DNN.

Fig. 2. Challenges of Wi-Fi-based gait identification.

to avoid retraining of the partial network which extracts high-level gait features from the input velocity profile and has the same network parameters shared by all persons.

The third challenge is to detect novel users for robust identification and intruder detection purposes. As the conventional recognition pipelines are meant to perform identification within the predefined set, which would output erroneous human identity if the test users are not included in the training set. GaitSense designs an anomaly detection algorithm to determine whether the test users have been seen during the training phase and perform identification afterward to avoid misidentification.

Putting it all together, we implement GaitSense on commodity off-the-shelf Wi-Fi devices and conduct experiments in typical indoor scenarios. We collect data from 11 subjects and overall 4,600 traces. GaitSense achieves the accuracy of 93.2% for 5 user identification and 76.2% for 11 user classification while reducing the training data of newcomers from 400 samples to 50 samples on arbitrary trajectories with arbitrary speeds. To the best of our knowledge, GaitSense is the first gait-based identification approach without any restriction on walking trajectory and speed.

In summary, we make the following contributions.

- We propose an agile algorithm to extract gait-specific feature GBVP that is resilient to environment and trajectory change and thus relieves restrictions on walking manners.
- The proposed gait identification approach requires little training efforts for various scenarios and persons and thus can be easily deployed and extended.
- The system is capable of detecting illegal users that are not included in the training set before the recognition process, which enables robust identification and intruder detection.
- We implement GaitSense on COTS Wi-Fi devices and extensive experiments have demonstrated the effectiveness and robustness of the proposed system.

2 MOTIVATION

GaitSense attempts to tackle three main challenges in Wi-Fi-based gait identification, which may hinder its steps towards ubiquitous sensing.

Immune to trajectory and speed variance. Both WiWho [37] and WiFiU [30] try to preserve human-specific information in their extracted features, e.g., DFS from Wi-Fi signals. Such features, however, are highly correlated with users' relative movements to Wi-Fi devices, thus impose stringent restrictions on their monitoring tracks. Widar3.0 [43] proposes a domain-independent feature BVP, which is mainly designed for in-place activities and sensitive to the moving speed

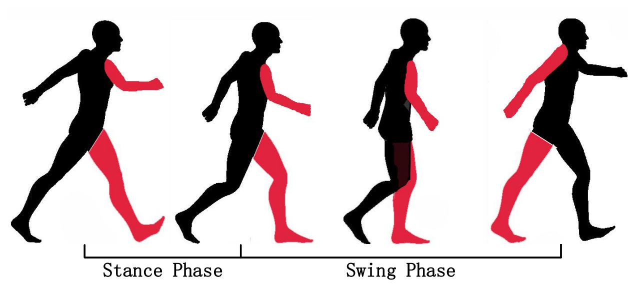


Fig. 3. Gait cycle for one leg.

of the target. As a brief example, two people are asked to walk on three tracks multiple times at different speeds. We train the simple human identification classifier composed of convolutional and recurrent layers with the features collected on track #0, including DFS, BVP, and our proposed GBVP, and test with those on the other trajectories. As shown in Fig.2a, while DFS and BVP indicate human identity from the same track with high accuracy, they fail to hold performance when testing and training datasets are from different tracks. Whereas, GBVP is robust to track and speed variance.

Reducing training data for newcomers. To fully exploit the spatial and temporal properties of motion features, existing works [9] leverage sophisticated deep neural networks to achieve high accuracy. However, a more complex structure usually means more parameters need to be trained, which leads to the requirement of a massive amount of training data. This problem becomes increasingly conspicuous when new users are added and the network should be re-trained. Fig.2b illustrates the exponential growth trend of required training samples needed to reach specific accuracy for a typical DNN network.

Detecting anomaly user. Current Wi-Fi-based gait identification approaches focus on enhancing classification accuracy within a pre-defined user set. A more realistic challenge is how to identify illegal intruders whose gait patterns have not been seen by the training set. However, WiWho [37] and WiFiU [30] either ignore this scenario or use threshold-based anomaly detection algorithms, which are sensitive to environmental changes.

Lessons learned. The deficiency of existing gait identification works demands to be relieved before practical usage is achieved. GaitSense is designed to address these issues.

3 PRELIMINARY

In this section, we provide preliminary knowledge of gait, and modeling and extraction of BVP, a feature capturing kinetic characteristics of human motion.

3.1 Preliminaries on Gait

Gait means the walking manner of a person. The gait cycle of one leg can be partitioned into two stages, and Fig.3 takes the left leg as an example. The stance phase begins when the heel of the left foot strikes the ground and ends when the right foot toes off the ground. Then the swing phase follows and ends when the heel of the right foot strikes the ground. For different people, the gait cycle will vary as it is determined by the underlying musculoskeletal structure of the person. Such biometric discrepancy contributes to the possibility of utilizing gait patterns to discriminate different persons and achieve human identification [2].

Although walking is a complicated motion as it involves different body parts including both torso and limbs, prior work [36] has proved that human identity extracted from gait is embodied mostly in limbs motions. Hence, we mainly focus on the movement patterns of limbs in our paper to recover human movements for identification.

3.2 Preliminaries on BVP

CSI from Wi-Fi. The physical layer Channel State Information (CSI) describes channel characteristics of the Wi-Fi signal propagation environment. Suppose that there is a moving person in the monitoring area, the received CSI with multipath effects can be modeled as follows:

$$H_m(f,t) = (H_m(f) + \sum_{l=1}^{L_{dynamic}} \alpha_{m,l}(f,t) e^{j2\pi \int_{-\infty}^{t} f_{D,l}(u)du}) \cdot e^{\varepsilon_m(f,t)}, \tag{1}$$

where $H_m(f)$ refers to the static reflection paths. $L_{dynamic}$ is the number of dynamic reflection paths. $\alpha_{m,l}$ is the amplitude attenuation of l^{th} path on m^{th} antenna. $f_{D,l}$ is the Doppler Frequency Offset (DFO) caused by the moving person, corresponding to the path length changing rate of the l^{th} reflection path. Term $e^{\varepsilon_m(f,t)}$ is the phase noise introduced by hardware imperfection on m^{th} antenna, which can be calibrated by conjugate multiplication of CSI from two antennas of the same Wi-Fi NIC [22], and static phase offset can be manually removed according to [33].

BVP From CSI. When a person performs a gait activity, the whole body is exposed to Wi-Fi signals, causing dynamic multipath components in the received CSI signals. Each dynamic reflection path causes different DFO, which are superimposed at the receiver and form the observed DFS profile. DFS can be extracted by applying time-frequency-analysis [31] on the received CSI signals. However, DFS may NOT stay consistent across different walking settings for the same person, as DFO only captures the radial velocity component and is determined by the relative position of the moving target from the perspectives of the Wi-Fi transmitter and receiver.

A recent work [43] proposed to leverage DFS from multiple Wi-Fi links to generate BVP, an ideal motion indicator that is resilient to environmental changes and the location and orientation of the target. BVP describes power distributions over velocity components at each timestamp in the body coordinate system, whose origin is at the location of the target and the positive direction of the x-axis is determined by the orientation of the target. BVP is an N × N matrix, where N is the number of possible values of velocity components decomposed along each axis of the body coordinate.

To recover BVP from DFS, a quantitative model can be established to mathematically depict their relationship. For a single reflection path, we can suppose that the moving target is on an ellipse whose foci are the transmitter and receiver at any instant. The velocity component projected on the norm direction causes changes in reflection path length and induces DFO as the following shows:

$$f(\vec{v}) = a_x v_x + a_y v_y, \tag{2}$$

Here, a_x and a_y are projection coefficients to project v_x and v_y onto norm direction. They are determined by the location of the transmitter, receiver, and target, which can be provided with the help of existing localization systems (please refer to [43] for details).

Searching over the velocity and attenuation of all the potential reflection paths by approximating observed DFS with reconstructed DFS from combined DFO, BVP can be recovered from multiple links' CSI information.

As BVP is theoretically immune to the changes of environment, and location and orientation of the moving target, it is a favorable feature to indicate gait patterns and characterize human identities, especially for arbitrary walking traces. However, BVP is not readily applicable for gait recognition, because it is modeled for in-place human activities (arm gestures) but gait involves

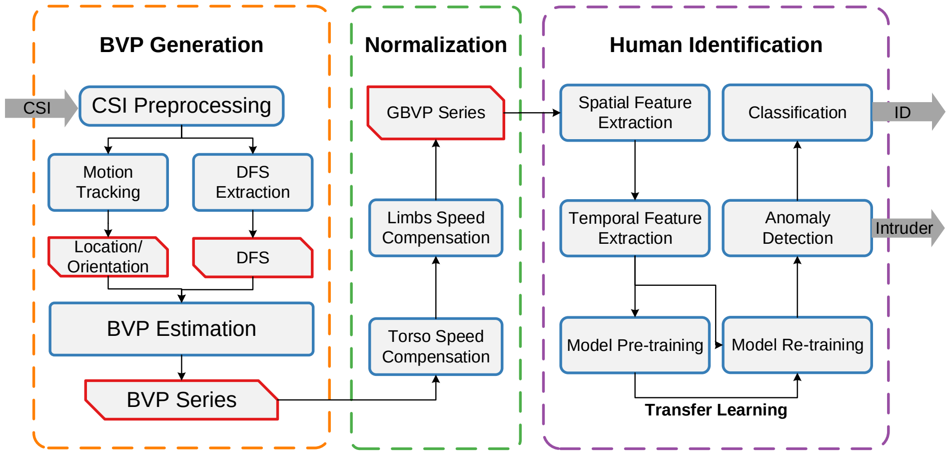


Fig. 4. Work flow of GaitSense.

the collaborative motion of the arm, torso, and legs. Besides, the BVP extraction algorithm is too complex to be applied to real-time services.

4 SYSTEM DESIGN

4.1 Overview

Fig.4 shows the workflow of our system. GaitSense first collects CSI measurements from multiple Wi-Fi links and removes random phase noises caused by hardware imperfection by applying the denoising algorithms in [14, 22, 33]. GaitSense then performs motion tracking and time-frequency analysis, followed by GBVP estimation with a novel and agile algorithm. With GBVP as input, GaitSense implements a deep neural network composed of both convolutional layers and recurrent layers to capture underlying spatial and temporal characteristics of GBVP. GaitSense further utilizes transfer learning to adapt to new users with only a few gait samples collected from the user. Finally, GaitSense performs anomaly detection to identify illegal users from the predefined user set.

In the following sections, we will discuss the detailed design of GaitSense, including GBVP extraction in §4.2, the recognition model and methods for reducing training data in §4.3, as well as illegal user detection algorithms in §4.6.

4.2 GBVP Extraction

Widar3.0 [43] proposed an environment-independent motion feature BVP to portray in-place human activities (gestures), which is resilient to location, orientation, and environmental changes. However, BVP can not be directly used for gait recognition and the reasons are two-fold. 1) BVP assumes that the reflection objects on the human body are at fixed locations during movements. This assumption holds valid for gesture recognition because the motion range for arms is at the decimeter level and can be ignored compared to the relative location between the human torso and Wi-Fi devices. However, for gait recognition, users can walk for several meters and the locations of reflection points will change drastically during walking. Hence, the velocity profile should be reformulated. 2) The complexity of the BVP extraction algorithm is especially high and can hardly be applied in real-time systems. Gait recognition technique is commonly used in security or commercial scenarios and demands an agile recognition system. Therefore, GaitSense designed a rigorous and

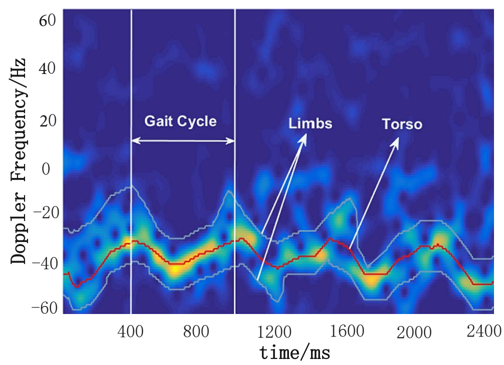


Fig. 5. Torso and limbs motion in DFS profile.

agile feature extraction algorithm to acquire environment-independent and gait-specific feature GBVP (gait-BVP) to tackle these concerns.

GBVP formulation. To foumulate GBVP, we first define an operator $\otimes$ as:

$$\mathcal{A} \otimes \mathcal{B} \triangleq \sum_{i=1}^{M} \sum_{j=1}^{N} \mathcal{A}_{(i,j,*)} \cdot \mathcal{B}_{(i,j)}, \tag{3}$$

where $\mathcal{A} \in R^{M\times N\times P}$ and $\mathcal{B} \in R^{M\times N}$. Hence, the operation result $\mathcal{A} \otimes \mathcal{B} \in R^{P}$ is equivalent to multiply each element of $\mathcal{B}$ with corresponding vector in the third dimension of $\mathcal{A}$ and then sum them up. Using the defined operator, we formulate GBVP as follows:

$$[GBVP] = \min_G \sum_{i=1}^{L} |\mathrm{EMD}(D^{(i)}(G), [DFS]^{(i)})| + \eta \|G\|_0, \tag{4}$$

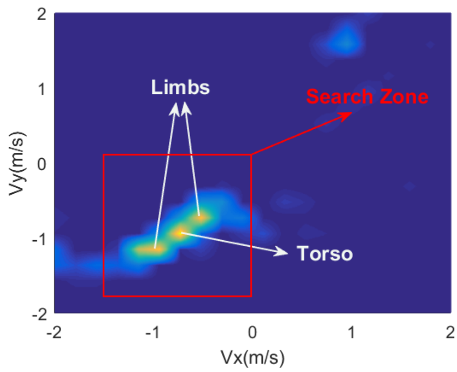
where $G \in R^{N\times N}$ is GBVP. $D^{(i)}(G) \in R^{N\times N\times F}$ is the reconstructed DFS from GBVP for i^{th} Wi-Fi link. $[DFS]^{(i)}$ is the observed DFS on i^{th} link. L is the total number of links. N is the number of possible values of velocity components along x/y axis. F is the number of frequency bins in DFS. $EMD(\cdot,\cdot)$ is the Earth Mover's Distance [23] and $||\cdot||_0$ is the number of non-zero elements in GBVP. η is the sparsity coefficient. Equation.4 is similar to BVP but the following formulation makes GBVP more agile and adaptive to torso movements:

$$D^{(i)}(G) = SUB(A^{(i)}) \otimes G, \tag{5}$$

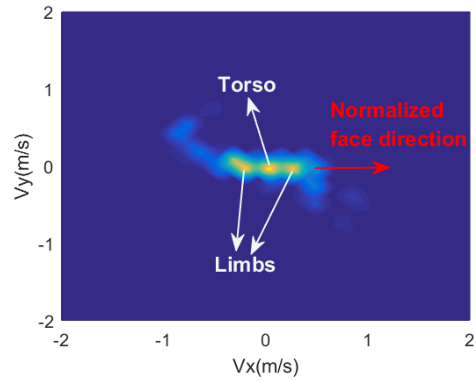
where $A^{(i)} \in R^{(N\times N\times F)}$ is the coefficient matrix to map GBVP into DFS on i^{th} link. The $SUB(\cdot)$ is the operator to cherry-pick the most relevant elements in the coefficient matrix to reduce the searching space of GBVP and eventually reduce algorithm complexity. We will shortly introduce the algorithm on how to do that reduction. Each element in the coefficient matrix can be determined by:

$$A^{(i)}_{(j,k,m)} = \begin{cases} 1, & f_m = f^{(i)}(< v_j, v_k >, t) \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where $< v_j, v_k >$ is the corresponding velocity vector of the $< j, k >^{th}$ element in GBVP matrix. f_m is the m^{th} frequency sampling point in the DFS profile. $f^{(i)}(\cdot, t)$ is a mapping function to convert target velocity into DFS observation for i^{th} link at time t (as illustrated in Equation.2).



(a) Search zone within GBVP.



(b) Normalized GBVP with speed compensation.

Fig. 6. GBVP extraction and normalization.

Different from BVP, the mapping function is dependent on the location of human torso and is time-varying. Theoretically, $f^{(i)}(\cdot, t)$ should be updated whenever the location of user's torso changes (we dismiss the relative location of limbs to torso). In practice, we only update this function when the location changes over a specific threshold (i.e., 0.5 meters) for computational efficiency. In GaitSense, we acquire target user's torso position and orientation by existing Wi-Fi-based passive tracking system [11, 20, 21].

We would like to clarify that even though GBVP is an extended version of BVP, they are different in both model formulation and motion representation. First, the extraction algorithm of BVP is based on the static location of the human torso while GBVP is for moving activities. Second, BVP only captures the motion of human arms, which is applicable in gesture recognition tasks. However, GBVP portrays both the motion of the torso and limbs, which is applicable in full-body activity recognition tasks such as gait recognition or fitting activity recognition. GBVP fills the gap for human activity recognition that is not fully covered by BVP features. We believe, with the collaborative use of BVP and GBVP, human-centered wireless sensing will be pushed into an environment-independent manner.

Accelerating GBVP extraction. For activity recognition tasks with Wi-Fi, only a few major reflection paths are considered and Widar3.0 [43] leveraged this sparsity by adding a regular term in target function to recover BVP. However, this sparsity is not fully exploited and the algorithm is still too cumbersome to be applicable. Our key insight is that the spatial correlation embodied in human motion could potentially be used to reduce the complexity of the GBVP extraction algorithm.

We observe that, during gait activity, the limbs would swing to opposite sides of the torso with limited speeds and the velocity of all the reflection paths would cluster around the velocity of the torso. This observation can be confirmed by Fig.5. This profile is constructed with CSI for 4 continuous cycles from one user. The red curve demonstrates the major energy corresponding to torso motion and the white curves demonstrate the residual energy corresponding to the motion of limbs. It is clearly shown that both sparsity and spatial clustering phenomena exist in the reflected signals. Hence, if we can pinpoint the velocity of the torso on the GBVP matrix, then the whole body GBVP components can be searched within a small area centered on the torso component.

Based on the above observations, GaitSense first identifies the maximum frequency bins in DFS from each Wi-Fi link and formulates the relationship between frequency bins and velocity with

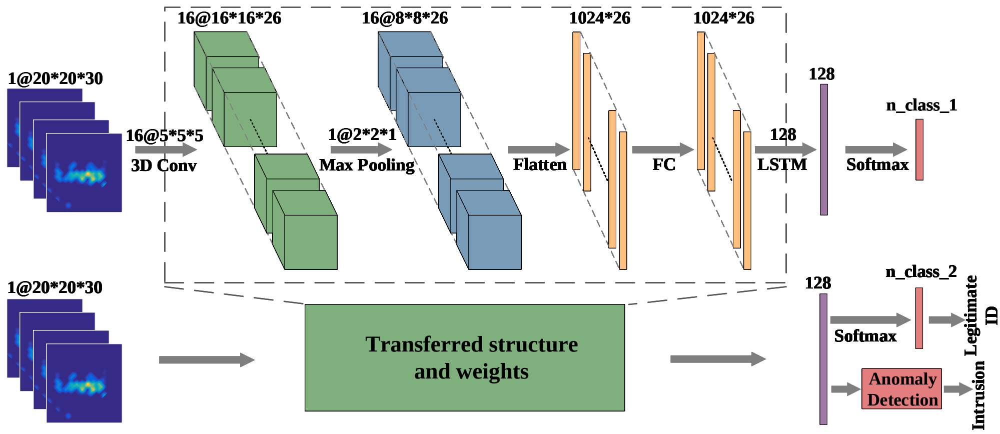


Fig. 7. Gait recognition model.

Equation.2. Solving the equations from multiple links, GaitSense pinpoints the torso velocity on the GBVP matrix. On top of that, GaitSense crops the adjacent elements of the torso component in the coefficient matrix described in Equation.5. The subtracted coefficient matrix is then used for GBVP recovery with Equation.4. After the above process, the GBVP search space is hereby reduced. In our experiments, the crop windows size is empirically selected and a smaller window would result in shorter running time but deteriorative accuracy, vice versa. The torso and limbs components as well as the search zone in GBVP are visualized in Fig.6a.

GBVP Normalization. While GBVP is theoretically only related to the gait of the target, it requires extra normalization to increase the stability as a gait indicator. The reasons are three-fold. First, literature [36] has proven that the torso movement contains little information on gait patterns and needs to be removed. Second, the power of the reflected signals is correlated with torso position relative to transceivers. Third, different walking speeds correspond to different limbs swing speeds, which results in variations of the number of GBVP frames and values floating within each GBVP matrix.

Thereafter, GaitSense first compensates the torso speed by applying translation and rotation on GBVP. The translation displacement is $\|\vec{v}_{torso}\|_2$ and the rotation angle is $\angle\vec{v}_{torso} - \angle\vec{v}_{ref}$ where $\vec{v}_{ref}$ is the manually selected reference orientation. This transformation procedure is similar to moving the target human to a treadmill, on which the target performs fixed-speed walking.

GaitSense then normalizes the sum of all elements in each GBVP to 1. It is based on the observation that the absolute reflection power contains environment information while the relative power distribution over physical velocities doesn't. Lastly, GaitSense scales each single GBVP with a scaling ratio $\frac{v_{ob}}{v_{tg}}$, where v_{ob} is the observed walking speed and v_{tg} is the target walking speed. GaitSense then resamples GBVP series over time with a resampling ratio of $\frac{v_{tg}}{v_{ob}}$, which is based on the hypothesis that the total displacement of the limbs relative to the torso is analogous across different walking speeds.

After normalization, only human identity information is retained while gait-irrelevant factors are removed. The normalized GBVP is visualized in Fig.6b, where torso speed is compensated and walking direction is normalized to a fixed direction.

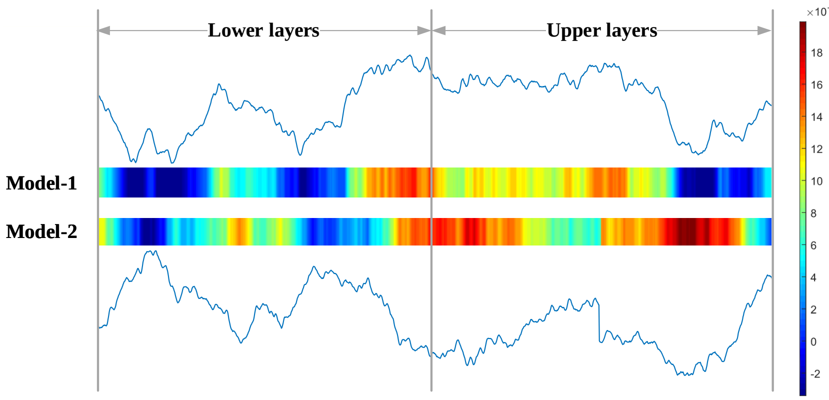


Fig. 8. Network weights trained from two datasets (lower layers share commonalities).

4.3 Recognition Mechanism

Fundamental model. Generally speaking, each single GBVP frame captures limbs' velocity distribution relative to the torso, and the GBVP series exhibit how the distribution varies over time. As shown in the upper half of Fig.7, we adopt a deep neural network (DNN) to best depict the characteristics of GBVP.

The input of the fundamental DNN model is of size 20×20×30, as velocity is quantized into 20 bins along the axis of the body-coordinate system, and the GBVP series is adjusted to 30 snapshots after normalization. GaitSense first applies 3D CNN onto the GBVP series for spatial feature compression and time redundancy reduction. Convolution operations along with the time domain also alleviate single GBVP estimation error. 16 convolutional filters of size 5×5×5 output 16 3D matrices of size 16×16×26. Then a max-pooling layer is applied to down-sample feature maps to the size of 8×8×26. By flattening the feature maps except for the time dimension, we obtain a vector series of size 1,024×26. And a fully connected (FC) layer is appended.

Recurrent layers are also incorporated into the model to model the temporal dynamics of the vector series. Considering the long-term characteristics of GBVP as a gait cycle that always lasts for a duration of more than one second [17], regular RNN suffers from the vanishing gradient problem [24], which hinders them from being used for long-term information extraction. Thus, instead of regular RNN, we adopt a better variation of RNNs: Long Short Term Networks (LSTM) [25]. The output of LSTM is then encoded with the softmax layer to do multiclass classification.

Transfer learning for reducing training efforts. Despite the fact that the structure of the fundamental model is not that sophisticated, the DNN model still demands enormous training data to converge. And when a new user is added to the human identification system, he/she must perform massive gait activities. Evaluation results in §5 show that there will be a rapid reduction in recognition accuracy even if the amount of training data decreases slightly.

Our solution was inspired by the observation that neural networks trained on similar datasets often share commonalities, i.e., the model trained on similar datasets undergoes analogous convergence procedure to some extent [27, 35]. This characteristic is exploited in a well-known research realm called *Transfer Learning*.

To testify the validity of Transfer Learning in gait recognition, we tune the fundamental model from scratch on two independent datasets separately, each of which is composed of GBVP series from two different users. We visualize and compare the network weights from the two converged models.

As can be seen from Fig.8, the lower layers of the neural network have analogous distributions of weights while the upper layers vary a lot. This phenomenon paves the way to transfer the information learned from different datasets and alleviate data collection efforts.

To leverage this generalizability between datasets, GaitSense first trains a model on the pre-collected large-scale dataset, which consists of GBVP from *n_class_1* persons. Then GaitSense replaces the softmax layer in the fundamental model with a different shape of *n_class_2* and initiates it randomly. The remaining weights of the model are initialized with the weights copied from the pre-trained model. The lower half of Fig.7 shows how transfer learning is applied in GaitSense. We will demonstrate that starting from the transferred structure and weights, our model can converge on the new dataset with significantly fewer data instances in §5.

4.4 Data Augmentation

By transferring the DNN model to new user sets, the scale of dataset required for model retraining is significantly reduced, relieving the need for extra data collection and model adaptation efforts when applying the system to new users. However, given that the neural network is heavily parameterized, it still requires a considerable amount of training data collected from the users. To further boost the agility of GaitSense, we propose a series of data augmentation algorithms to expand the scale of the collected dataset virtually.

As is discussed in §4.2, the GBVP feature is a 3-dimensional matrix. The first two dimensions represent the velocity along the horizontal plane and the third dimension represents time. This physically plausible trait of GBVP allows us to generate novel data samples by directly deforming the original samples. Specifically, we design three methods to generate synthetic data to augment the training dataset.

- ***Smoothing to remove temporal mutations.*** Each frame of the GBVP series represents the distribution of signal power over different velocities. This distribution should transform continuously over time because the movements of the human body are successive. However, based on our observation, the extracted GBVP series have mutations over time, making it challenging to capture the consistent features of gait movements. The mutations are due to the following reasons. First, radio frequency (RF) signals have a specular reflection effect when reflected from the human body [1, 42], which makes some of the reflected signals not received by the Wi-Fi receiver. Consequently, some of the elements in GBVP frames are missing and this phenomenon is varying over time. Second, the GBVP extraction algorithm is based on the compressed sensing technique and is prone to local optimums. This will break the continuity structure of the GBVP series over time. To remedy this, we smooth the GBVP series over time with a sliding window to produce a sanitized version of the sample. This smoothing process alleviates the temporal mutations and better captures the transformations of the human body during walking.
- ***Randomizing to introduce diversity.*** The previous method augments data by removing temporal mutations. In this entry, we adopt the randomization technique to introduce diversity into the GBVP series. Specifically, we randomly generate sparse matrices that follow a normal distribution and add them to each frame of GBVP. The maximum value of non-zero elements is set to half of the maximum value of GBVP and the sparsity is set to 20% empirically. Note that each frame of GBVP is randomized with the same random distribution, which is equivalent to randomizing over both spatial and temporal dimensions.
- ***Averaging over adjacent gait cycles to enhance walking periodicity.*** Human gait motion has periodical patterns and each cycle embodies the complete gait feature for a user. We will demonstrate in §5 that a single gait cycle suffices for gait identification. However, during

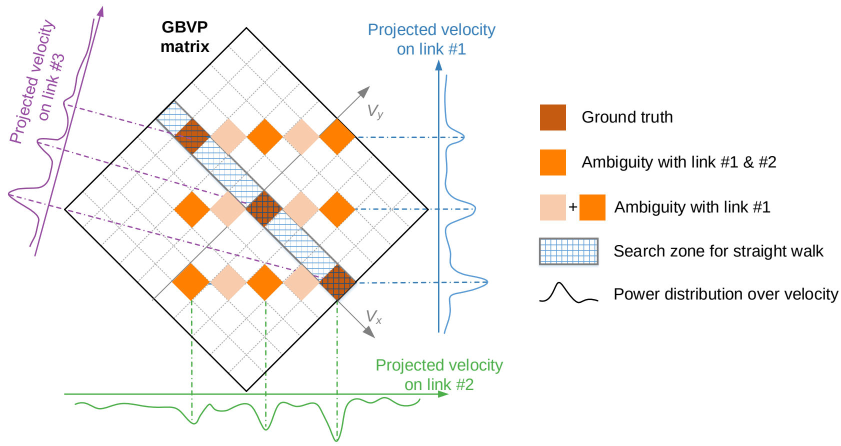


Fig. 9. Ambiguity exists in GBVP estimation due to the constrained field of vision for each Wi-Fi link. Ambiguity can be removed with a single link when the user walk on a linear track.

human walking, several factors affect the stability of gait patterns across different cycles. First, the changes in surrounding environments would affect the walking manners. For example, the body would push more weights on the outer legs when turning at the corner, making the gait pattern distorted. Second, the relative locations of the Wi-Fi devices to the human body are continuously changing when the users walk towards different locations. This phenomenon induces perturbations in the received signals and hence affects the GBVP periodical patterns. To remedy this, we split GBVP series into different cycles and the adjacent cycles are averaged to preserve the periodicity. This creates an extra number of synthetic samples for model training.

4.5 Extending to a Single Wi-Fi Link

In GaitSense, we adopt six Wi-Fi links (including one transmitter and six receivers) to capture human walking. This setting enables orientation-agnostic recognition. However, in real-world scenarios, the users may walk on fixed tracks with specific directions (e.g., corridor, aisle, etc), under which circumstances, a single link is sufficient for GBVP estimation. Hence, it is attractive to extend GaitSense to work with a single Wi-Fi link. In this section, we first discuss the ambiguity that exists in the GBVP estimation algorithm and present the reasons why multiple links are required to solve this problem. Then, we discuss the method to extend the algorithm to a single link when users walk on straight tracks, relieving the need for heavy hardware deployment.

In GaitSense, we quantize GBVP as a discrete matrix with dimension as $N \times N$, where N is the number of possible values of velocity components decomposed along the X-axis and Y-axis. The value of each element represents the power of the reflected signal from all the objects with the corresponding velocity. For each Wi-Fi link, if we draw an ellipse across the reflection object with the transmitter and receiver as foci, only the velocity along the norm direction of the ellipse can induce DFO. In other words, DFS profiles only reflect the projected velocity along the norm direction. The relationship between projected velocity and GBVP is depicted in Fig.9. To estimate the GBVP matrix, we have to reversely perform this projection procedure. Suppose three reflection objects are observed and their ground truth velocities are marked with brown color in Fig.9. Each link can only capture the projected velocity along its norm direction, which creates ambiguity in

GBVP estimation. Specifically, with only the information from link #1, all the elements along the projection direction (marked with pink and orange colors) are potentially the target ones. To remedy this, we have to add more links. With link #1 and #2, ambiguity still exists in the elements marked with an orange color and three links are sufficient in this case. However, when the projection of ground truth elements overlaps on link #3, more links are demanded to resolve this ambiguity.

The above discussion presents the most general model for GBVP estimation where users may walk on arbitrary tracks. However, in realistic scenarios, we can leverage the prior information of walking tracks to relieve the need for more Wi-Fi links. As is discussed in §3, human limbs swing forward and backward during walking, which means most of the body parts move along the same direction (the facing direction). This phenomenon is also observed from our measured GBVP features (Fig.6). When users walk along a linear track, the dominant elements in the GBVP matrix will align on a line. For example, in Fig.9, a user walks along X-axis and the three ground truth elements only occur in the search zone marked with a dashed grid. In this case, we only need link #1 to recover the location of the three elements without ambiguity. Generally speaking, when the user walks on a linear track, we only need one link that is not perpendicular to the walking direction to recover the GBVP matrix.

In practice, if GaitSense is deployed to perform recognition on a linear and fixed track, we place one pair of Wi-Fi devices with their line-of-sight (LOS) perpendicular to the walking track. The sub-area in Equation.5 is cropped by the direction of the track:

$$D^{(i)}(G) = SUB(A^{(i)} * M^{(\theta)}) \otimes G, \tag{7}$$

where θ is the angle of the walking direction, $M^{(\theta)}$ is the mask matrix formulated as:

$$M^{(\theta)}_{i,j} = \begin{cases} 1, & |angle(< v_i, v_j >) - \theta| < \epsilon \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

where $< v_i, v_j >$ is the corresponding velocity vector of the $< i, j >^{th}$ element in GBVP matrix. ϵ is the threshold (set to 0.1 rad). By applying the mask matrix, the non-zero elements in the GBVP matrix are constrained to be within a linear zone and ambiguity is circumvented. When GaitSense is deployed for arbitrary tracks, we still need at least three Wi-Fi links to estimate GBVP for recognition.

4.6 Anomaly Detection

Anomaly detection is another important problem for human identification as illegal users should be discriminated against the legal ones even if no data instance of intruders is collected in the training dataset.

We split a dataset of 6 persons into 5 legitimate users and 1 illegal user, and train the classifier to favor 5 legitimate users. Thereafter, we feed all of the 6 users' data into the model and extract the 128 dimension vectors of LSTM output as features for anomaly detection. We then apply the t-SNE algorithm to visualize those features in Fig.10.

As can be seen, the feature points from legitimate user 1-5 are tightly clustered, while those from illegal user 6 (purple points) spread across the feature space randomly. The key insight is that these outliers mostly occur at the edge space of *each* cluster. We describe our anomaly detection algorithm in Algorithm.1.

Algorithm.1 can be divided into three steps. First, during the training time, GaitSense calculates the mean average neighbour distance for each legitimate class and treats them as the *class density*. Second, during the testing time, GaitSense identifies the K-nearest neighbours of the test sample, and regards the most prevalent class in its neighbours as its *potential class*. At last, for the test

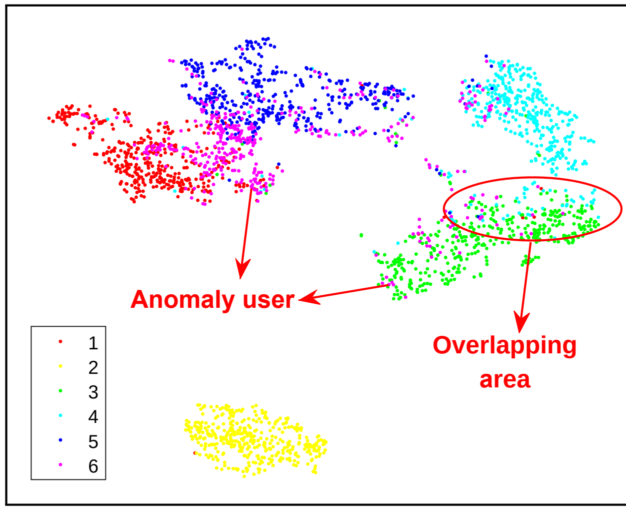


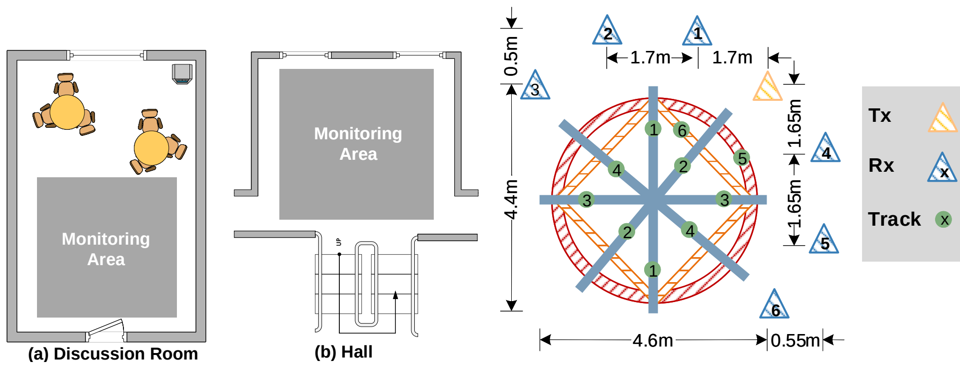
Fig. 10. t-SNE visualization of legitimate and illegal user distribution (illegal user marked as 6).

Algorithm 1 Anomaly detection algorithm

Input: $\Theta[i], (i = 1, 2, \ldots, M)$, the legitimate users' feature data of all M classes.
$\Phi[i], (i = 1, 2, \ldots, m)$, the legality unknown users' feature data to be classified.
K, the nearest neighbour parameter.
ω, the threshold parameter.
Output: $\phi[i], (i = 1, 2, \ldots, m)$: the legality of $\Phi[i], (i = 1, 2, \ldots, m)$

Calculate neighbour density for each legitimate class:
for $i = 1$ to M **do**
  For each sample θ_j in $\Theta[i]$, calculate its mean distance to K nearest neighbours, denoted as d_j
  $Density[i] = average(d)$
end for
Diagnose legality of each sample in Φ:
for $i = 1$ to m **do**
  Identify K nearest neighbours of $\Phi[i]$ in Θ
  Regard the most prevalent class τ in its neighbours as the potential class
  Calculate mean distance to its K nearest neighbours in $\Theta[\tau]$, denoted as λ
  if $\lambda > \omega \cdot Density[\tau]$ **then**
    $\phi[i] = Illegal$
  else
    $\phi[i] = Legal$
  end if
end for
return ϕ

sample, GaitSense computes its average distance to the nearest K samples in that *potential class*. If it is larger than the weighted *class density*, the test sample is considered illegal, vice versa.



(a) Floor plan of rooms.
(b) Tx/Rx positions and walking tracks.

Fig. 11. Experimental settings.

5 EVALUATION

This section presents the experimental settings and the detailed performance of GaitSense.

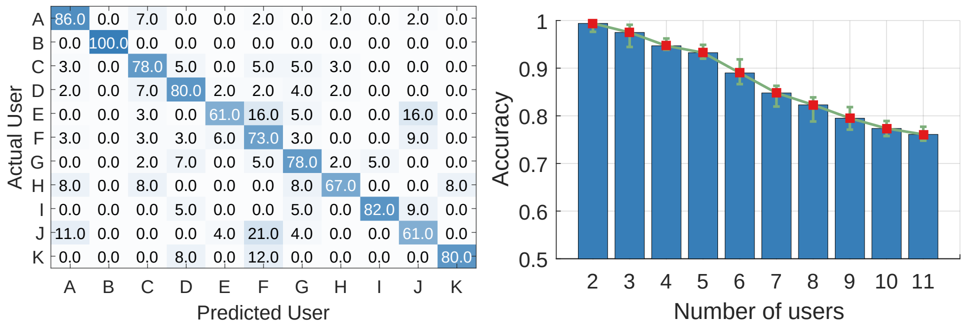
5.1 Experimental Methodology

Implementation. GaitSense consists of one transmitter and six receivers. Each of the transceivers is a mini-computer physically sized 170mm × 170mm × 50mm and equipped with an Intel 5300 Wi-Fi card. The operating system is Ubuntu 10.04 with Linux CSITool [6] installed to log CSI measurements. The Wi-Fi cards work at channel 165 with a center frequency of 5.825 GHz. CSI measurements are logged at a rate of 1,000 Hz and the CSI processing and feature extracting algorithms are implemented in MATLAB, while the learning model is implemented in Keras with TensorFlow as backend. Table.2 presents the size of the implemented DNN model.

Table 2. DNN Model Parameters

Layer	3D Conv	Max pooling	FC-1	LSTM	FC-2	Total
Params.	2,016	0	65,600	74,496	387	142,499

Evaluation setup. We conducted experiments under two different indoor environments with different layouts: an empty discussion room with desks and chairs and a wide hall beside stairways. Fig.11a demonstrates the floor plan of monitoring area. Benefiting from GBVP's robustness to device deployment, we empirically setup a monitoring area of 4.6m × 4.4m square, which is sufficient to collect at least 5 steps of gait data. The setup of transceivers and walking tracks is demonstrated in Fig.11b. The 4.6m × 4.4m square is a typical setting to perform gait recognition in indoor scenarios especially in the office or household environments. However, people can also custom their device deployment settings according to their needs, and three requirements should be met. First, the receivers should be within 6 meters of the transmitter. This is because the signal amplitude decays during propagation and we observed a tumble in system performance when the distance exceeds 6 meters. Second, no more than three devices should be in a straight line. This rule guarantees that the ambiguities in the GBVP estimation procedure are removed to the utmost extent. Third, the LOS path for each transmitter-receiver pair should not be blocked by objects and the users. This is because the signal reflection model only holds valid when the LOS path exists.



(a) Confusion matrix.

(b) Accuracy for user diversity.

Fig. 12. Overall accuracy.

GaitSense is designed to handle various walking tracks, but it's hard to traverse the endless instances of tracks. Our evaluation results have shown that 2 steps are sufficient to portray the identity of users and a walking orientation error within 50 degrees exerts minor influence on recognition accuracy. Hence, an arbitrary walking track can be segmented into sub-tracks, each can be quantized into one of the 8 directions separating 45 degrees apart. Hence, we designed four linear tracks, including two perpendicular lines to both axes and two diagonal lines shown in Fig.11b. Each of these tracks has enough length for five steps and users can walk on both ways. To evaluate the system performance on non-straight tracks, we also design non-linear tracks with circle and rhombus shapes, which are labeled #5 and #6 in Fig.11b. For the linear tracks, we manually annotate the human locations mainly for benchmark analysis. For track #5 and #6, we adopt Wi-Fi-based human tracking systems to provide the human locations for case study in §5.6.

We recruited 11 volunteers (4 females and 7 males) to participate in our experiments, covering the height from 155cm to 186cm and weight from 44kg to 75kg, and age from 20 to 28 years old. These volunteers were asked to walk normally with different speeds on those tracks and each data sample contains five steps. Specifically, 10 of the volunteers were asked to walk in each direction of four tracks 50 times in Hall, contributing to 10 users $\times$ 4 tracks $\times$ 2 directions $\times$ 50 instances for datasets. 3 of the volunteers in the Discussion Room contributed to 3 users $\times$ 4 tracks $\times$ 2 directions $\times$ 25 instances in the dataset. All experiments were approved by our IRB. The datasets are released and incorporated into our previously published Widar3.0 datasets at the website[1].

5.2 Overall Performance

All the data collected from 11 volunteers in Hall were mixed and randomly split into two datasets, discarding the walking tracks and speed. We employed standard 10-fold cross-validation to evaluate the accuracy. Fig.12a shows the confusion matrix for 11 users.

The overall identification accuracy for all of the 11 users is 76.2%. From the confusion matrix, most users experience an accuracy of over 75% except for user E, H, and J, which may be attributed to their walking manner of putting hands in pockets, leading to infrequent motions of arms and induce less significant features in GBVP. User J and F are likely to be confused, which may be caused by their similarities in body shape.

Fig.12b further shows the identification accuracy for different user numbers. Basically, the identification accuracy declines with more users involved, which is intuitive because more categories

[1]http://tns.thss.tsinghua.edu.cn/widar3.0/index.html

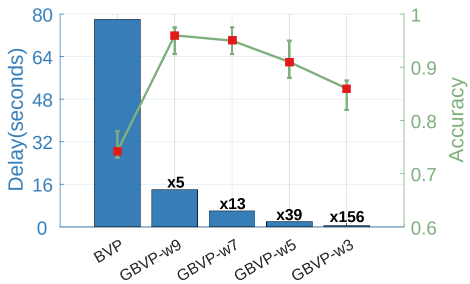


Fig. 13. GBVP extraction delay and accuracy (system delay is reduced up to 156 times with smaller window size).

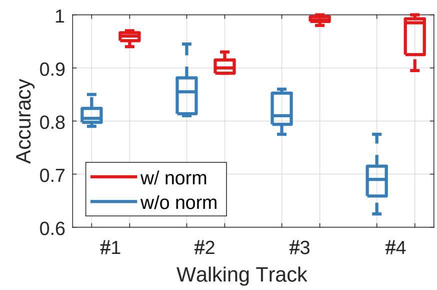


Fig. 14. Cross tracks performance.

would lead to more crowded feature clusters in feature space. The accuracy for two users is above 99%, meaning that the extracted gait features are distinct for identification. Notably, GaitSense holds its accuracy above 93.2% for about 5 users, demonstrating its potential for smart home applications where there are only a few users in indoor scenarios.

5.3 Generalizability Evaluation

Now we evaluate the performance of proposed generalization mechanisms, including GBVP extraction accelerating algorithm, GBVP normalization algorithm and transfer learning framework.

Accelerating performance. The process of GBVP extraction is accelerated by the novel acceleration algorithm. To validate the efficiency and effectiveness of GBVP over BVP, we collect 400 samples of CSI from three volunteers and each corresponds to 4 steps. We then extract BVP as well as GBVP with different accelerating windows sizes (as described in §4.2). The system is running on a server with 32 cores of Intel Xeon CPU E5-2620 v4 @ 2.10GHz and Matlab2016b installed. The system running delay and recognition accuracy are demonstrated in Fig.13. Note that the time consumption of the DNN model is within 6 ms for a gesture that lasts 2 seconds, which is too trivial to be considered. The end-to-end delay of GaitSense is approximate to that of the GBVP estimation algorithm. As can be seen, even with 2 seconds of CSI as input, the BVP extraction would last for unbearably 78 seconds. However, with our proposed accelerating algorithm, the feature extraction speed can be accelerated 156 times while maintaining the recognition accuracy to some extent. The GBVP extraction delay is 0.5 seconds with a window size of 5x5 and the accuracy holds above 83%, which enables the system to respond in realtime. We believe the proposed accelerating algorithm would push the BVP to a broader application prospect on other motion recognition scenarios.

Normalization performance. For walking track independence, we randomly select two users' data from all the datasets, using one track for test and the remaining three tracks to train our model, ignoring their walking speed. As is shown in Fig.14, the overall accuracy with normalized GBVP is significantly above that without normalization. The second track benefits least from normalization because we selected its direction as the reference direction and the GBVP from other tracks are rotated to match this orientation. For walking speed independence, we classify all the collected data into 6 categories, each with quantized speed. We then select one category for test and the others to train the model. Fig.15 demonstrates the remarkable improvements with normalization. The fifth category benefits the least from normalization because we selected its speed as the reference speed and GBVP with other speeds are normalized to this case.

Transfer learning and data augmentation performance. Transfer learning is applied to re-train the model on a new dataset from an existing pre-trained model, which needs fewer data

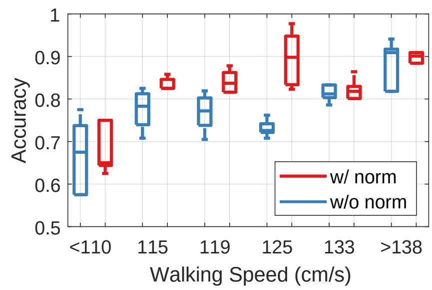


Fig. 15. Cross speeds performance.

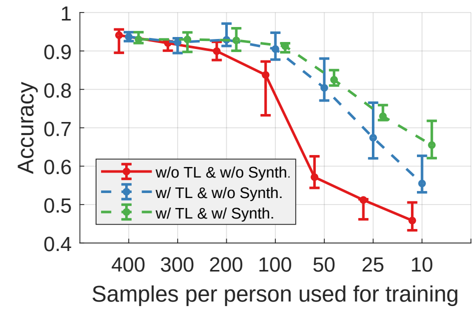


Fig. 16. Impact of training samples.

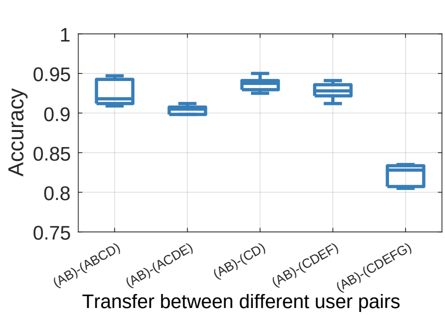


Fig. 17. Cross user pairs performance.

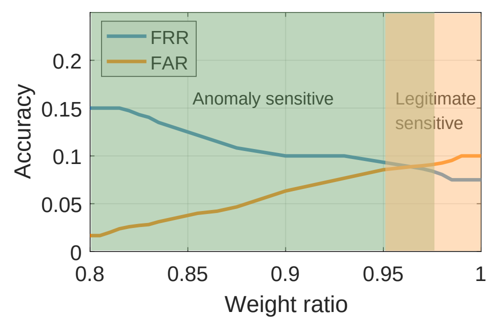


Fig. 18. Impact of weight ratio.

samples for the model to converge. Data augmentation techniques are used to generate synthetic datasets from the authentic data and use both of the datasets for model training. To verify the effectiveness of these methods, we randomly selected four ([A,B,C,D]) users' data with a total of 4 × 400 samples containing all the tracks and speeds. These samples are feed into a pre-trained model that has already been tuned on two ([A,E]) user's data. By gradually decreasing the number of data involved in re-training, results can be found in Fig.16. We also evaluate the system performance when the augmented data samples are used for model retraining. In Fig.16, the **TL** represents **T**ransfer **L**earning and **Synth** represents Synthetic dataset. With transferred model structure and weights, only 50 data samples from each user are needed to keep the recognition accuracy above 80%. When the synthetic dataset is used, the accuracy increases by 3%-5%. But when training the model without transfer learning and without data augmentation, accuracy can hardly exceed above 60% with such few data. We also evaluated transformations between different user pairs with limited data samples. Fig.17 exhibits consistency in generalizability between different user set pairs. The slight degradation of accuracy for user pair (A,B)-(C,D,E,F,G) may attribute to user G's great similarities in gait patterns with the other users.

5.4 Anomaly Detection Evaluation

This section presents the performance of the anomaly detection algorithm. Before we give the detailed evaluation results, several definitions should be clarified. Fail Alarm Rate (FAR) defines the ratio of illegal users are classified as legitimate users. False Rejection Rate (FRR) defines the ratio of legitimate users classified as intruders. Looking back into Algorithm.1, the weight ratio poses a significant impact on FRR and FAR.

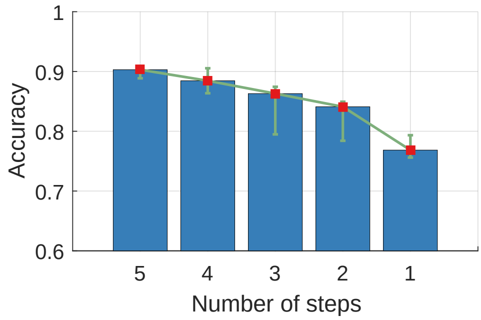


Fig. 19. Impact of gait instances.

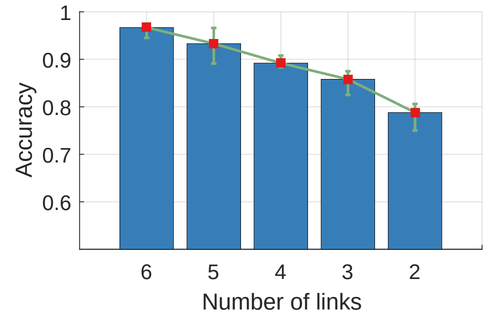


Fig. 20. Impact of Wi-Fi links number.

We randomly select 5 users to construct the legitimate set and another user as an intruder. By gradually changing the weight ratio parameter, results can be found in Fig.18. Intuitively, a small weight ratio would produce a narrow confidential space for legitimate users, which may lead to a large FRR. While a large weight ratio would produce a loose confidential space, which may falsely embrace illegal users, causing higher FAR. The green part in Fig.18 represents a high risk for fail alarm and the orange part represents a high risk for false rejection. To seek a balance between FRR and FAR, we empirically set the weight ratio to 0.96, which produces 9% for both FAR and FRR. Note that, due to the environment-independent feature of GBVP, the weight ratio only needs to be tuned once and can generalize to other scenarios.

5.5 Parameter Study

This section presents the impact of parameters to the performance of GaitSense.

Impact of gait instances. To evaluate the impact of gait instances on identification accuracy, we randomly select four users and let them perform 5 steps of walk. We then manually split steps into different numbers by DFS peaks and valleys. Results can be found in Fig.19. The performance falls slightly from 5 steps to 2 steps but tumbles to below 78% with a single step. The reason is that a full gait cycle contains 2 consecutive steps, each of which is insufficient for the representation of identity. Meanwhile, gait is a periodic motion and the repetition of gait cycles introduces trivial extra information for gait characteristics. Result from temporal memorability of LSTM in our model, GaitSense is capable of retaining distinctions from single gait cycles. Hence, we claim that 2 consecutive steps are sufficient for human identification. In practice, we suggest using four steps for more robust performance.

Impact of link numbers. In the formulation of extracting GBVP for gait, we adopted 6 Wi-Fi links, which potentially contain redundancy. In this section, we evaluate the impact of link numbers on system performance. We randomly selected four users for classification and randomly prune partial links. Results can be found in Fig.20. The accuracy gradually slides when involved Wi-Fi links reduce. This is because fewer links capture fewer reflection paths from the human body, and theoretically at least 3 links are necessary to recover valid GBVP. With only 2 links, accuracy drops to below 80% and is hardly beyond research usage.

Impact of orientation error. In the GBVP normalization process, we rotate GBVP to identify with the reference orientation, which demands a precise estimation of walking direction. However, orientation extracted from state-of-the-art motion tracking techniques contains prominent errors. To evaluate the impact of orientation error on human identification accuracy, we generated training and testing sets by manually providing trace orientation, and added controllable orientation disturbance

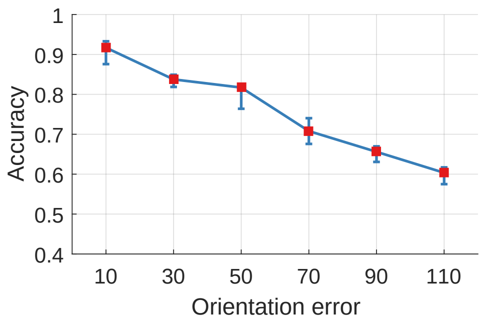


Fig. 21. Impact of orientation error.

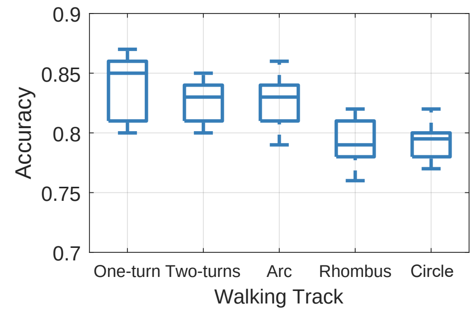


Fig. 22. Performance on non-linear tracks.

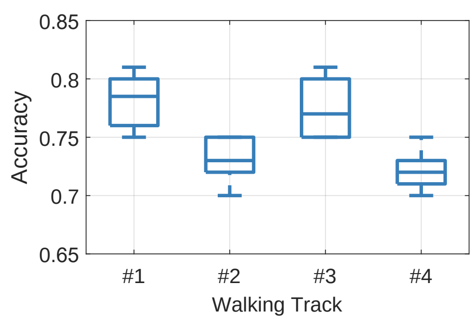


Fig. 23. Performance with a single link.

to the test set. Results can be found in Fig.21. As the illustration showed, an orientation error within 50 degrees doesn't noticeably deteriorate accuracy, while an orientation error above 50 degrees witnesses an unacceptable dilution in identification accuracy. Hence, the four tracks with eight orientations designed in our evaluation implementation are sufficient to represent more complicated walking traces.

5.6 Case Study

This section presents the system performance for some practical cases.

Performance on non-linear tracks. In the above experiments, the users are required to walk on linear tracks and the human locations are annotated manually. However, this setting is not practical for real-world deployment, where users may walk on arbitrary tracks with various shapes. To test the system performance on non-linear tracks, we perform this case study evaluation. In this case study experiment, three users' data are from tracks #1~#4, and two users' data are from all the tracks marked in Fig.11b. For the circle-shaped track #5, we split it into two semicircles and define the shape as the *"Arc"*. For the rhombus-shaped track #6, we split it into four lines and combines them into the *"One-turn"* and *"Two-turns"* shapes. All the non-linear tracks start from the point closest to the X-axis. To acquire the locations of users during walking, we adopt the system proposed in Widar [20], which uses similar device deployment settings with ours. The tracking results and the GBVP extraction results are synchronized according to the timestamps logged by CSITool. As is shown in Fig.22, the average recognition accuracy for track *"One-turn"* is 85% and it decreases to 83% for tracks *"Two-turns"* and *"Arc"*. The accuracy eventually decreases to 79% for tracks *"Rhombus"* and *"Circle"*. The results reveal a decrease in performance when users walk for longer distances. This is because the tracking system is based on velocity estimation and is prone to cumulative errors. The localization results will deviate from ground truth when the users walk further and break the structure of GBVP features. We believe this limitation can be mitigated by using more accurate tracking systems [11, 21] and we leave it for future work. However, according to our evaluation on the impact of gait instances in §5.5, gait can be recognized with a small number of steps benefiting from the periodicity property of human walking. Hence, the impact of the walking track and its length are trivial to the performance of GaitSense.

Performance with a single Wi-Fi link. GaitSense extracts GBVP feature from CSI signals, which is mathematically an underdetermined procedure and requires CSI readings from at least three Wi-Fi links to mitigate the ambiguity. However, as is discussed in §4.5, when the users walk on linear tracks, this problem becomes easily solvable with a single Wi-Fi link. This setting is common in real-world scenarios, where users may walk along corridors or aisles. With fewer links used, the system becomes more practical for deployment. In this experiment, we evaluate the performance of GaitSense when only a single link is used. Specifically, five participated users walk on tracks #1∼#4, and only the transmitter (Tx) and one receiver (Rx) are used accordingly. For tracks #1 and #2, Tx and Rx-3 are used and for tracks #3 and #4, Tx and Rx-6 are used. We adopt this setting because only the movements perpendicular to the links can induce DFO. The evaluation results are shown in Fig.23. As is shown, the accuracies are around 78% for tracks #1 and #3, which reveals that a single Wi-Fi link can also produce satisfying recognition accuracy. The performance tumbles to around 72.5% for tracks #2 and #4, which is because these tracks are not fully perpendicular to the links and have some leakage in the information. This experiment demonstrates that GaitSense is capable of using one single Wi-Fi link for user recognition, as long as the walking tracks are fixed and perpendicular to the link. However, when users are allowed to walk on arbitrary tracks, at least three links are required to perform recognition.

6 RELATED WORK

Our work is focused on human identification based on gait patterns, which can be roughly categorized into vision-based, sensor-based, and radio-based works.

Sensor-based gait identification. Sensor-based approaches fall into two categories: floor sensors and wearable sensors. In the former type, geophone [18] or pressure sensors [3] are installed under/on the floor to detect slight floor vibrations or measure foot shape, orientation, and contact force distribution when human feet strike the ground, and structural vibration patterns are extracted to depict human identification. These techniques are susceptible to sensor position, floor structure, and walking speed. In the latter type, accelerometers [39] or rotation sensors [16] are explored to record accelerations and rotation variance during walking. However, these techniques are prone to sensor noise and non-gait motions. While GaitSense extracts gait features that are immune to walking speed, track variance, and imposes no intrusion on target users.

Vision-based gait identification. Vision-based approaches leverage cameras to generate some series of human silhouettes from the video frame and then apply DNN-based classifier on the sequence of human silhouettes [4, 28, 41]. These methods require a continuous stream of camera data or LOS but are prone to ambient light conditions. They are also criticized for privacy concerns. While GaitSense only extract feature from Wi-Fi signals, which is unrelated to light conditions and is able to work under dark environments. GaitSense doesn't capture target images and can better preserve privacy.

Radio-based gait identification. Radio-based gait identification methods fall into two categories: using specialized devices and using Wi-Fi devices. In the former type, Frequency-Modulated Continuous-Wave (FMCW) radar [12] is applied to analyze DFS induced by human motion. These techniques are less appealing due to their high costs of dedicated infrastructures. In the last type, WiWho [37] and WiFiU [30] leverage CSI collected from commercial Wi-Fi devices to depict human identification, but require the user to walk on a predefined track. A recently published work Gait-Way [32] proposes to extract gait features from the speed variations, which imposes no limitations on the walking manners. However, due to its cumbersome recognition pipeline, a large number of training samples are demanded to adapt the system to novel users. In comparison, GaitSense designs several methods including transfer learning and data augmentation to relieve the needs for massive training data, requiring only dozens of walking instances to tune the system.

Recently, a pioneering work XModal-ID [13] proposes to combine vision technology with wireless signals for human identification. The key idea underpinning this work is to capture human mesh models with cameras and simulate wireless signal that can be reflected from the model. By comparing the synthesis signals with the observed signals, XModal-ID can perform human identification with only one gait sample and the image of the person, relieving the need for a large-scale training dataset. Our work, however, exploits another way for human identification with deep learning-assisted technology, and both works set new stages for the wireless sensing field.

7 CONCLUSION

In this paper, we present GaitSense, a Wi-Fi-based person identification framework that is robust to walking trajectory with few training efforts. GaitSense first proposes an enhanced gait-specific feature, which is environment, trajectory, and speed independent theoretically. GaitSense then reduces the training efforts for new users by transfer learning technique. At last, GaitSense leverages a novel anomaly detection algorithm to detect illegal users. Experimental results show that GaitSense achieves an accuracy of 93.2% for 5 users for arbitrary tracks and the performance retains above 81.6% even when training data decreases to 20% of that needed by the original system.

ACKNOWLEDGMENTS

This work is supported in part by the National Key Research Plan under grant No. 2018AAA0101200, the NSFC under grant 61832010.

REFERENCES

[1] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)* 34, 6 (November 2015), 219.
[2] Asif Ali Banka, Dr Ajaz, and Hussain Mir. 2010. Human Identification Based On Gait. In *Proceedings of ICBT*.
[3] Young-One Cho, Byung-Chul So, and Jin-Woo Jung. 2012. User Recognition Using Sequential Footprints Under Shoes Based on Mat-Type Floor Pressure Sensor. *Advanced Science Letters* 9 (04 2012), 591–596.
[4] Naresh Cuntoor, Amit Kale, and Rama Chellappa. 2003. Combining multiple evidences for gait recognition. In *Proceedings of ICASSP*.
[5] Dexter D. Geralde, Milton M. Manaloto, Dionn Elijah D. Loresca, Jacqueline D. Reynoso, Eden T. Gabion, and Gabriel Rodnei M. Geslani. 2017. Microcontroller-based room access control system with professor attendance monitoring using fingerprint biometrics technology with backup keypad access system. In *Proceedings of HNICEM*.
[6] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool Release: Gathering 802.11N Traces with Channel State Information. *SIGCOMM Comput. Commun. Rev.* 41, 1 (jan 2011), 53–53.
[7] Jutta Hämmerle-Uhl, Karl Raab, and Andreas Uhl. 2011. Robust Watermarking in Iris Recognition: Application Scenarios and Impact on Recognition Performance. *SIGAPP Appl. Comput. Rev.* 11, 3 (aug 2011), 6–18.
[8] Nazia Hossain and Mahmuda Nazin. 2018. EmoVoice: Finding My Mood from My Voice Signal. In *Proceedings of ACM IMWUT*.
[9] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, Wenyao Xu, and Lu Su. 2018. Towards Environment Independent Device Free Human Activity Recognition. In *Proceedings of ACM MobiCom*.
[10] Ju Han and B. Bhanu. 2004. Statistical feature fusion for gait-based human recognition. In *Proceedings of IEEE CVPR*.

[11] Chitra R. Karanam, Belal Korany, and Yasamin Mostofi. 2018. Magnitude-Based Angle-of-Arrival Estimation, Localization, and Target Tracking. In *Proceedings of IPSN*.

[12] Youngwook Kim and Hao Ling. 2009. Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine. *IEEE Transactions on Geoscience and Remote Sensing* 47, 5 (2009), 1328–1337.

[13] Belal Korany, Chitra R. Karanam, Hong Cai, and Yasamin Mostofi. 2019. XModal-ID: Using WiFi for Through-Wall Person Identification from Candidate Video Footage. In *Proceedings of MobiCom*.

[14] Xiang Li, Daqing Zhang, Qin Lv, Jie Xiong, Shengjie Li, Yue Zhang, and Hong Mei. 2017. IndoTrack: Device-Free Indoor Human Tracking with Commodity Wi-Fi. *Proceedings of ACM IMWUT* 1, 3 (sep 2017), 72:1–72:22.

[15] YuanYuan Li and Lynne E Parker. 2008. Intruder detection using a wireless sensor network with an intelligent mobile robot response. In *IEEE SoutheastCon 2008*.

[16] Soumik Mondal, Anup Nandy, Pavan Chakraborty, and G Nandi. 2012. Gait Based Personal Identification System Using Rotation Sensor. *Journal of Emerging Trends in Computing and Information Sciences* 3 (03 2012).

[17] MP MURRAY, AB DROUGHT, and RC KORY. 1964. WALKING PATTERNS OF NORMAL MEN. 46 (March 1964).

[18] Shijia Pan, Ningning Wang, Yuqiu Qian, Irem Velibeyoglu, Hae Young Noh, and Pei Zhang. 2015. Indoor Person Identification Through Footstep Induced Structural Vibration. In *Proceedings of HotMobile*.

[19] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. 2018. Acousticcardiogram: Monitoring Heartbeats using Acoustic Signals on Smart Devices. In *Proceedings of IEEE INFOCOM*.

[20] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. 2017. Widar: Decimeter-Level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi. In *Proceedings of ACM MobiHoc*.

[21] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. 2018. Widar2.0: Passive Human Tracking with a Single Wi-Fi Link. In *Proceedings of ACM MobiSys*.

[22] Kun Qian, Chenshu Wu, Zimu Zhou, Yue Zheng, Zheng Yang, and Yunhao Liu. 2017. Inferring Motion Direction Using Commodity Wi-Fi for Interactive Exergames. In *Proceedings of ACM CHI*.

[23] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 2000. The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision* 40 (November 2000), 99–121.

[24] Hasim Sak, Andrew W. Senior, and Françoise Beaufays. 2014. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. *CoRR* abs/1402.1128 (2014).

[25] Alex Sherstinsky. 2018. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *CoRR* abs/1808.03314 (2018).

[26] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. 2017. Smart User Authentication Through Actuation of Daily Activities Leveraging WiFi-enabled IoT. In *Proceedings of ACM MobiHoc*.

[27] Matthew E Taylor and Peter Stone. 2009. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research* 10, Jul (2009), 1633–1685.

[28] T. Teixeira, D. Jung, G. Dublon, and A. Savvides. 2009. PEM-ID: Identifying people by gait-matching using cameras and wearable accelerometers. In *Proceedings of ACM/IEEE ICDSC*.

[29] Hao Wang, Daqing Zhang, Junyi Ma, Yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. 2016. Human Respiration Detection with Commodity Wifi Devices: Do User Location and Body Orientation Matter?. In *Proceedings of ACM IMWUT*.

[30] Wei Wang, Alex X. Liu, and Muhammad Shahzad. 2016. Gait Recognition Using Wifi Signals. In *Proceedings of ACM IMWUT*.

[31] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and Modeling of WiFi Signal Based Human Activity Recognition. In *Proceedings of ACM MobiCom*.

[32] Chenshu Wu, Feng Zhang, Yuqian Hu, and K. J. Ray Liu. 2020. GaitWay: Monitoring and Recognizing Gait Speed Through the Walls. *IEEE Transactions on Mobile Computing* (2020), 1–1.

[33] Jie Xiong and Kyle Jamieson. 2013. ArrayTrack: A Fine-Grained Indoor Location System. In *Presented of NSDI*.

[34] Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor Localization via Channel Response. *ACM Comput. Surv.* (November 2013), 25:1–25:32.

[35] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks? *CoRR* abs/1411.1792 (2014).

[36] VM Zatsiorky, SL Werner, and MA Kaimin. 1994. Basic kinematics of walking. Step length and step frequency. A review. *The Journal of sports medicine and physical fitness* 34, 2 (June 1994).

[37] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. 2016. WiWho: WiFi-Based Person Identification in Smart Spaces. In *Proceedings of ACM/IEEE IPSN*.

[38] Jin Zhang, Bo Wei, Wen Hu, and Salil S. Kanhere. 2016. WiFi-ID: Human Identification Using WiFi Signal. In *Proceedings of DCOSS*.

[39] Yuting Zhang, Gang Pan, Kui Jia, Minlong Lu, Yueming Wang, and Zhaohui Wu. 2015. Accelerometer-Based Gait Recognition by Sparse Representation of Signature Points With Clusters. *IEEE Transactions on Cybernetics* 45 (2015),

1864–1875.

[40] Yi Zhang, Yue Zheng, Guidong Zhang, Kun Qian, Chen Qian, and Zheng Yang. 2020. GaitID: Robust Wi-Fi Based Gait Recognition. In *Proceedings of WASA*.

[41] Guoying Zhao, Guoyi Liu, Hua Li, and M. Pietikainen. 2006. 3D gait recognition using multiple cameras. In *Proceedings of FGR*.

[42] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-Based 3D Skeletons. In *Proceedings of ACM SIGCOMM*.

[43] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2019. Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi. In *Proceedings of ACM MobiSys*.

[44] Han Zou, Yuxun Zhou, Jianfei Yang, Weixi Gu, Lihua Xie, and Costas J. Spanos. 2018. WiFi-Based Human Identification via Convex Tensor Shapelet Learning. In *Proceedings of AAAI*.

[45] Han Zou, Yuxun Zhou, Jianfei Yang, and Costas J. Spanos. 2018. Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT. *Energy and Buildings* 174 (2018), 309 – 322.