GaitCube: Deep Data Cube Learning for Human Recognition with Millimeter-Wave Radio

Muhammed Zahid Ozturk, *Student Member, IEEE,* Chenshu Wu, *Senior Member, IEEE,* Beibei Wang, *Senior Member, IEEE,* and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—Monitoring and identifying gait has recently emerged as a promising solution candidate for unobtrusive human recognition. In order to enable ubiquitous and reliable application, a gait recognition system must be robust to environment changes and easy to use without requiring too much user cooperation and recalibration, while maintaining high accuracy, which is often not satisfied in conventional approaches. In this paper, we present *GaitCube*, a high-accuracy gait recognition system with minimal training requirement using a single commodity millimeter wave (mmWave) radio. To reduce the training overhead, we propose *gait data cube*, a novel 3D joint-feature representation of micro-Doppler and micro-Range signatures over time that can comprehensively embody the physical relevant features of one's gait. With a pipeline of signal processing, *GaitCube* can automatically detect and segment human walking and effectively extract the *gait data cubes*. We implement and evaluate *GaitCube* through experiments conducted at 6 different locations in a typical indoor space with 10 subjects over a month, resulting in >50,000 gait instances. The results show that *GaitCube* achieves an accuracy of 96.1% with a single gait cycle using one receive antenna, and the accuracy increases to 98.3% when combining all the receive antennas. Further, it achieves an average recognition accuracy of 79.1% for testing over different times and unseen locations by using only 2 minutes of training data collected in a single location, enabling a practical and ubiquitous gait-based identification.**

***Index Terms*—Sensor signal processing, gait recognition, mmWave sensing, deep learning.**

I. Introduction

Ubiquitous human recognition acts as an essential element for a variety of applications in smart spaces, such as personalized environmental control, security management, or access control for automatic doors and IoT devices. Mainstream approaches rely on fingerprint identification, face recognition, voice authentication, etc., which usually require the active cooperation of the user within a certain proximity. Radio biometric based on the unique way that a human body alters the multipath radio channel has been proposed [1]–[3], which is, however, very sensitive to environmental changes and thus requires a lot of training/calibration.

Recently, human gait has been proposed as an effective biometric that is useful for more passive person identification, *i.e.*, identification (at a distance) during normal walking without additional cooperation of the user. Ideally, a gait recognition system should satisfy the following conditions in order to enable practical applications:

- **Accurate**: The system should be able to recognize users accurately.
- **Fast registration and response**: The system should require minimal training effort to register a new user, and recognition should be accomplished with short delays. Particularly, a few steps (short period) of walking should be sufficient to achieve accurate and reliable recognition.
- **Environment-independent**: The system should be able to operate at different times and locations without requiring tailored (re)calibration, and should not be affected by changes in lighting, furnishing, and other environmental factors.
- **Contactless**: The system should operate in a contactless manner, without asking the user to carry any device or using any user cooperation.
- **Privacy-preserving**: Even though the system can identify users, it should not reveal sensitive information about the person and surroundings.

In order to realize a gait recognition system with these properties, various modalities have been considered in the literature, such as vision [4], WiFi [5], [6], acoustic sensing [7], wearable sensors [8], and pressure pads [9]. Each of these methods has advantages and drawbacks with respect to the criteria above. For instance, the vision-based systems [4] suffer from environmental changes and impose privacy concerns. Methods using inertial sensors require user cooperation and thus are not so practical. WiFi-based systems provide an attractive solution by exploiting ambient WiFi signals for contactless recognition, but they usually require calibration for each location due to changes in the multipath profile. More importantly, many of the existing systems train and test on the data collected at the same time and/or locations, imposing a high risk of performance loss when generalizing to different locations and times in practical deployments [5], [6], [10].

With the recent proliferation of miniaturized mmWave radar devices, there have been increasing interests in indoor radar sensing applications [11]–[16]. Using an mmWave radar for sensing brings multifold advantages thanks to its shorter wavelength, larger bandwidth, and phased array processing, while still sharing the favorable characteristics of WiFi-based systems (*e.g.*, contactless, privacy-preserving), promising an ideal solution for feasible and practical gait recognition with minimal infrastructure support (*e.g.*, a single mmWave radio). Such an opportunity motivates us to ask the research question

M. Z. Ozturk, C. Wu, B. Wang, and K. J. R. Liu are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, College Park, MD 20742 USA, and also with Origin Wireless Inc., Greenbelt, MD 20770, USA (e-mail: ozturk@umd.edu, cswu@umd.edu, bebewang@umd.edu, kjrliu@umd.edu).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3083934, IEEE Internet of Things Journal

in this work: *Can we build a gait recognition system that can register a user with minimal training effort (e.g., less than one minute) and recognize a person within a single gait cycle (i.e., two steps)?*

We approach this goal through a practical design, *GaitCube*, which uses a single mmWave radar and combines the power of signal processing and deep learning. *GaitCube* introduces a three-dimensional joint-variable representation of micro-Doppler (μD) and micro-range (μR) signatures over time (T), termed as *gait data cube*, to comprehensively embody physical relevant features of one's gait, which is then fed into a neural network for effective learning. With that, *GaitCube* can register a user with minimal walking data and can recognize a user with high accuracy from only a couple of steps. Combining the proposed gait data cube with neural networks also allows *GaitCube*, trained once, to generalize to different locations and time with little performance loss. To deliver a practical system, *GaitCube* further incorporates modules to detect and track human walking automatically and segment gait cycles effectively.

GaitCube addresses multiple challenges to achieve all these properties in one system. First, even with an mmWave radar, it is non-trivial to extract fine-grained μR and μD signatures efficiently and effectively, accounting for the inadequate range resolution, specular reflection, multipath effects, and computational burden. Second, it is usually difficult to build a learning-based gait recognition system that can operate over different locations at different times, as neural networks can easily overfit or learn features related to the environment (but not human gait), especially when the training dataset is limited.

To combat these challenges, *GaitCube* detects and tracks human motion with a simple yet effective algorithm, triggers μD-μR spectrogram estimation only at interested distances and times, and extracts gait information from human walking automatically. To assemble the gait data cubes that maximize gait information, we reshape μD signatures at all the distances the human body spans, which are then aligned in range domain with respect to the human torso, segmented in the time domain with respect to walking cycles, and cropped in the frequency domain. *GaitCube* then identifies environment-independent and physically relevant features from the gait data cubes in tandem with a convolutional neural network (CNN). To boost training and generalize the trained networks, we exploit spatial diversity attributed to multiple receiving antennas, which not only increases the training dataset considerably but also mitigates the specular reflection issue by capturing more spatially independent snapshots of human walk.

We implement *GaitCube* using a commercial mmWave radar and conduct experiments in office space over different locations and times used for training and testing. We recruit 10 volunteers and collect 11 sessions of data, performed at six different locations over a month. During each session, we ask the users to walk for two minutes in each location and collect a total of two hours of walking data, which results in more than 50000 steps. The results demonstrate that *GaitCube* achieves an accuracy of 96.1% using a single gait cycle (two steps) with one receiving antenna, and this accuracy further improves to 98% by aggregating all the antennas or combining multiple walking steps. In the extreme case of using a single session (only two minutes of data) for training and different locations and sessions for testing, *GaitCube* achieves an average accuracy of 79.1%, promising a practical application of gait recognition with minimal training requirement.

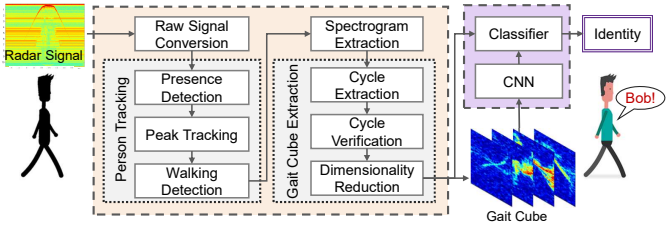


Fig. 1: Overview of *GaitCube*

Our core contributions are the following:

- We present *GaitCube*, a human recognition system that learns from gait data cubes, a 3D joint-feature representation of micro-Doppler, micro-range, and time for radar signals, allowing it to achieve remarkable accuracy with a single gait cycle while needing minimal data for training.
- We propose a signal processing pipeline to form the gait cubes that can detect and segment human walking automatically, and extract the μR-μD-T gait cubes effectively, and a deep learning model for classification.
- We implement and experiment *GaitCube* on a commercial mmWave radar and validate its performance with 10 people over different environments/locations and time.

The remainder of the paper is as follows. Section II gives an overview of the system. Section III explains the details related to radar-cube processing module, and Section IV elaborates on the learning based classification module. Section VII discusses related works and Section VIII concludes the paper.

II. System Overview

GaitCube inputs raw radar data and outputs identification results at time instances that correspond to gait cycles. To register a user, the system needs to collect a minimal amount of walking data (*e.g.*, two minutes) from the user. Then one can be recognized by *GaitCube* when she/he walks normally in front of the radar.

As shown in Fig. 1, *GaitCube* consists of two main modules, gait cube extraction and classification. The gait cube extraction module inputs raw radar data and extracts the gait data cubes. It first tracks a person's walking trace by three submodules of presence detection, peak tracking, and walking detection. Then it extracts the spectrogram around the person and therefore constructs the Doppler (or speed) dimension of the gait cubes. Further, it segments the data in the time domain, with respect to the extracted gait cycles each with a single step, and removes unstable walking data by gait cycle validation. Consecutive valid steps are aligned together to construct the μR-μD-T gait cubes. The resulted *gait cube* represents the reshaped μD and μR signatures at different distances from the transmitter, which are aligned in range domain with respect to the human torso, segmented in the time domain with respect to walking cycles (steps), and cropped in

the frequency domain to include maximum gait content, while minimizing dimensionality.

The output gait cubes are fed into the classification module that outputs the user identity. The classification module includes a CNN to extract useful features from gait cubes, and a fully connected layer that allows us to concatenate some other hand-crafted features to augment the output representation.

III. Gait Cube Processing

In this section, we first review the basics of the radar we use in this work, and then present the pipeline of signal processing to form the gait cubes.

A. FMCW Radar Basics and Preprocessing

Our system relies on a frequency-modulated continuous-wave (FMCW) radar, which transmits a signal with linearly increasing frequency, and the distance from an object is measured by calculating the frequency shift between the transmitted and the received signal. This single transmission is called a chirp, and the range resolution is determined by the bandwidth. When there is no multipath and a single rigid object is placed at a distance R, the received signal $y(t)$is given as [17]:

$$y(t) = A_{rx} \exp\left(j\left(2\pi\left(f_0(t-\tau) + \tfrac{\beta}{2}(t-\tau)^2\right) + \theta_{rx}\right)\right), \tag{1}$$

for $t \in [0, T)$, where τ represents the time duration of an electromagnetic wave from object to radar, β is the frequency slope of the linear chirp, θ_{rx} is the phase offset at the receiver, and A_{rx} is the amplitude of the returned signal. A single chirp is sent and received repeatedly, and we denote the received signal for chirp k as $y_k(t) \triangleq y(t - kT)$, for $k = \lfloor t/T \rfloor$ where T is the chirp duration. For simplicity, we assume t to be discrete as the reported signal is sampled and digitized, and drop the subscript k. A_{rx} is given as [18]:

$$A_{rx} = \frac{G_{ant}\lambda\sqrt{P\sigma}}{4\pi^{1.5}R^2\sqrt{L}}, \tag{2}$$

where G_{ant} represents antenna gain, λ is the wavelength, σ is the target cross-radar section, and L represents other losses. This equation is given for rigid objects, under no multipath assumption, and does not necessarily apply to indoor environments. When there is multipath, (1) is modified as:

$$y(t) = \sum_{k=1}^{N} h_k A_{rx} \exp\left(j\left(2\pi(f_0(t-\tau_k) + \tfrac{\beta}{2}(t-\tau_k)^2) + \theta_{rx}\right)\right), \tag{3}$$

where h_k denotes the scaling of the returned signal for time instance τ_k, arising from multipath. We first note that, due to directionality and reflection characteristics of mmWave signals, the channel response usually have a strong dominant part, stemming from line-of-sight (LOS) propagation, along with additional paths having lesser energy. This information is used to reject the multipath effect to some extent, and will be utilized in Section III-B for tracking. As the frequency shift is equivalent to time difference, this information can be converted to range information by Fourier transform of $y(t)$, *a.k.a* Range-FFT. Time-range signal, *a.k.a* Channel Impulse Response (CIR), is sampled at certain time indices, and denoted as:

$$Y(r,k) = \sum_{n=0}^{N-1} y_k(n) \exp\left(-j\tfrac{2\pi rn}{N}\right), \tag{4}$$

where N is the number of FFT points. Range-Doppler spectrogram is calculated by applying FFT on the time domain signal for a frame over long time. We define N_f as the number of samples per frame, N_{ov} as the amount of overlap between frames, and $n_k(i)$ as the long time indices for frame k where $n_k(i) \triangleq (k-1)(N_f - N_{ov}) + i$ for $i \in \{1, ..., N_f - 1\}$. Consequently, range-Doppler spectrogram is defined as

$$G(f,r,k) = \left| \sum_{i=0}^{N_f-1} W(i) Y(r, n_k(i)) \exp\left(j\tfrac{2\pi iTf}{N_f}\right) \right|^2, \tag{5}$$

where $W(i)$ represents a finite length windowing function used to fine tune the resolution between time and frequency domains. In *GaitCube*, instead of calculating the spectrogram for all range bins (r), we first locate the human body, and only calculate the spectrogram of nearby distances that contain body motions to reduce computational complexity.This nearby range-Doppler spectrograms will be used to construct *gait data cubes*, an integral part of the *GaitCube*.

B. Human Body Tracking

We first present how to track the users, whose locations will be used to construct the range dimension of gait cubes.

The received time-range signal (CIR) in (4) can be decomposed as:

$$Y(r,n) = Y_b(r,n) + Y_d(r,n) + \varepsilon(r,n), \tag{6}$$

where Y_b represents the background reflection from surrounding objects, Y_d stands for the reflection from a moving subject, and ε denotes the additive noise. To track a moving body, we extract the variance for each range bin as:

$$\mathbf{Z}_Y(r,k) = \frac{1}{N_f} \sum_{i=1}^{N_f} (Y(r, n_k(i)) - \bar{Y}(r, n_k))^2, \tag{7}$$

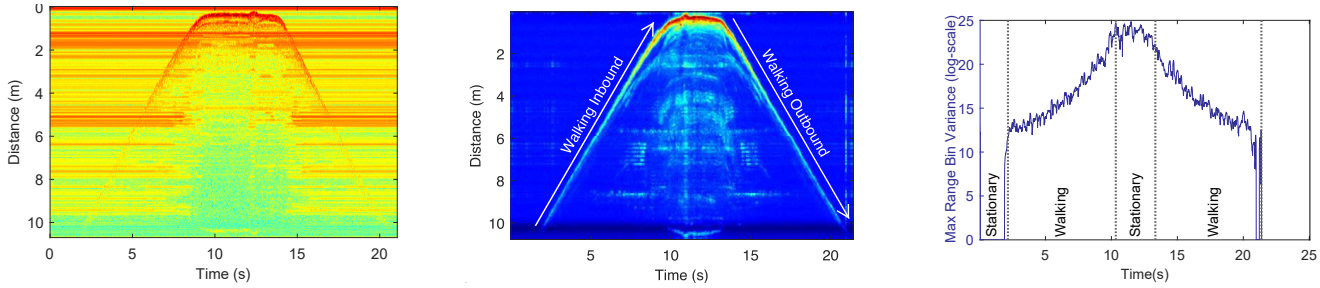
where $\mathbf{Z_Y}(r,k)$ denotes the variance of Y at range r and time-frame k, and $\bar{Y}(r,k) \triangleq \frac{1}{N_f}\sum_i Y(r, n_k(i))$, denotes the average value of the CIR for frame k. Assuming the noise to be uncorrelated with motion, and assuming the motion and noise to be zero mean, the following relation can be established:

$$\mathbf{Z}_Y(r,k) \approx \frac{1}{N_f} \sum_{i=1}^{N_f} |(Y_d(r, n_k(i)))|^2 + \sigma_\varepsilon^2, \tag{8}$$

which suggests that, dominant motion could be detected by maximizing $\mathbf{Z}_Y(r,k)$ over range dimension.

An exemplary raw data frame can be seen in Fig. 2a, whereas the corresponding variance-time plot can be seen in Fig. 2b. As can be seen, the background effects are reduced significantly in the time-variance plot. Fig. 2c further illustrates the maximum-variance trace, which demonstrates that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3083934, IEEE Internet of Things Journal



(a) Time-Range plot on Raw Data (b) Time-Variance plot as described in (7) (c) Peak tracking and decision output

Fig. 2: Person Tracking

the variances at range bins with the presence of a person are several orders of magnitude higher than background variance.

Extracting this maximum-variance trace, however, does not allow one to immediately detect and track motion, as there are impulse noises, discontinuities, and multipath effects in the environment. Thus we apply several additional steps to the person tracking submodule to detect human presence and motion and extract the walking trace.

Presence Detection: *GaitCube* detects the presence of a subject by thresholding the amplitude of a smoothed maximum-variance trace. *GaitCube* smooths maximum-variance trace with a median filter before thresholding. In addition, short presence intervals are filtered by combining with longer intervals to reduce false alarms. Even with a stationary person or motion perpendicular to the radar axis, presence detection robustly detects a person due to minute body motions caused by breathing and heart pulses.

Peak Tracking: For time instances with presence, *GaitCube* extracts the subject trace in range-time domain. An arbitrary trace on $\mathbf{Z}_Y(r,k)$ is defined as:

$$\mathbf{p} = \{(p(i), i)\}_{i=1}^{L}, \tag{9}$$

where $p(i)$ denotes the distance (range bin) of the person from the radar at timestep i, and L is the total length of the trace. Note that $p(i)$ only includes the range information, and the system cannot determine exact location of the person. Our person trace requirement has two objectives, extracting high variance bins (dominant motion), while preserving a smooth trace. First, we define $E(\mathbf{p}) \triangleq \sum_{i=1}^{L} \mathbf{Z}_Y(p(i), i)$ as the total energy of the trace, and $C(\mathbf{p}) \triangleq \sum_{i=1}^{L-1} P(p(i+1), p(i))$ as the cost function for overall smoothness of the trace, where the cost function P controls the change between two consecutive indices of the trace. For our application, we define P as:

$$P(m,n) = \begin{cases} \epsilon(|m-n|) & |m-n| < T_{th} \\ \infty & \text{otherwise} \end{cases}, \tag{10}$$

where we limit the displacement between consecutive frames by forcing $|p(i+1) - p(i)| < T_{th}$. Maximum deviation threshold T_{th} could be selected based on maximum speed of an object, and ϵ controls the cost for deviation from a direct path. Based on these cost functions, we define the person trace as:

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} E(\mathbf{p}) + C(\mathbf{p}), \tag{11}$$

which is solved by dynamic programming in *GaitCube*. $\mathbf{p}^*$ is the output of the peak tracking module, and used by walking detection to segment the data with respect to walking time instances. Recall in Fig. 2b, some of the distant paths are affected by the nearby human body, because of the blockage and multipath, yet these are easily avoided by our peak tracking algorithm, as it limits maximum deviation, and calculates the location of a user for a longer duration, instead of an independent location estimation for each radar sample.

Walking Detection: Lastly, *GaitCube* utilizes a basic walking detection module to extract time indices with inbound or outbound walking with respect to the radar in order to reduce signal processing overhead. The walking detection algorithm is applied to the speed estimates to extract stationary and non-stationary periods.

We estimate the approximate body speed as:

$$v(i) = \begin{cases} \frac{1}{D}\left(\mathbf{p}^*(i) - \mathbf{p}^*(i-D)\right) & i > D \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

where D is simply the time offset for calculating the speed of body. Choosing D greater than 1 helps to reduce variance of the speed estimation, stemming from limited range resolution. Using the speed estimate, *GaitCube* extracts the walking segments with the following decision rule:

$$m(t) = \begin{cases} 1, & |v(t)| > v_{walk} \\ 0, & |v(t)| < v_{stat} \text{ or } t = 0 \\ m(t-1) & \text{otherwise} \end{cases}, \tag{13}$$

where v_{walk} is the speed threshold for detecting walking, and v_{stat} is the stationary threshold for the magnitude of speed. Note $v(t)$ could be positive or negative since walking inbound and outbound introduces positive and negative Doppler shifts, respectively. This detector allows the amplitude of the input signal to swing between (v_{stat}, v_{walk}) without changing previous decisions. As the speed $v(i)$ is defined with respect to the change of distance between the radar and the device, *GaitCube* can remove most of the walking that does not traverse multiple range bins by selecting a high amplitude for v_{walk}.

C. Spectrogram Extraction

We extract the spectrograms at and around the person trace by using short-time Fourier transform (STFT), as given in (5). Since only the range values nearby the human torso are of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3083934, IEEE Internet of Things Journal

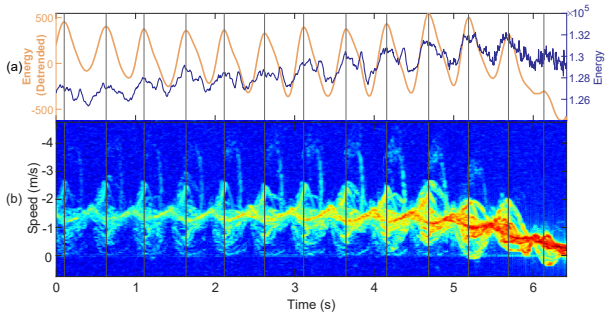


Fig. 3: Cycle extraction from (a) energy of the log-spectrum, and (b) 2D spectrogram with corresponding cycles

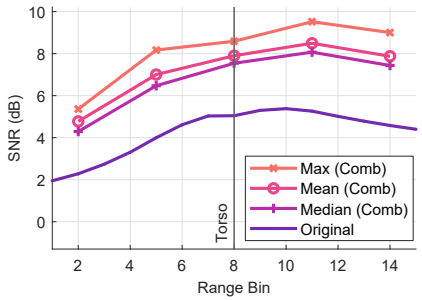


Fig. 4: Comparison of output SNR with different downsampling methods, along with the original signal

interest for calculation, we denote this aligned spectrogram cube as G_{align} and is given by:

$$G_{align}(f,r,n) = \{G(f, p(n) - N_{body} + r, n)\}, \quad 1 \leq r \leq 2N_{body}+1, \tag{14}$$

where N_{body} denotes the number of leading and trailing range bins, centered at the range of human torso. Consequently, $G_{align} \in \mathbb{R}^{N_{frame}} \times \mathbb{R}^{2N_{body}+1} \times \mathbb{R}^{N_{time}}$. Spectrogram-range cube reveals Doppler frequency shifts caused by motion of multiple limbs, referred as μ-Doppler phenomenon [19]. Different limbs, such as arms, legs and feet move at different speeds during walk, and the received spectrograms are superposition of all these effects, which are expected to reserve more information than a single snapshot at the human torso.

The relationship between frequency shift and speed is given as:

$$\Delta f = \frac{2v}{\lambda}, \tag{15}$$

where v, Δf and λ denotes speed of the object, frequency shift, and wavelength, respectively. As noted by earlier research, human motion induces both μD and μR signatures [20], which could be captured by a high-bandwidth (fine-range-resolution) radar and μD signatures only cannot capture the rich spatial information.

Fig. 3(b) depicts a single slices of $G_{align}(f,r,k)$ at $r = 9$, where $N_{body} = 8$, $k \in \{0, ..., 1567\}$ (corresponds to $[0, 6.4]$s), and $f \in [-385, 2493]$Hz.

D. Gait Cycle Estimation

Next, G_{align} is fed to the gait cycle estimation submodule, in order to segment with respect to the walking cycles.

Cycle Extraction: In this subsection, we discuss how to extract gait cycles from the estimated gait cubes. As observed in [20], the energy of spectrogram images (in our case, G_{align}) corresponding to the human torso could be used for extracting walking cycles. At the same time, as given in (2), the energy of the reflected signal also depends on the range and multipath profile, and we cannot use it directly to estimate gait cycles. In addition, FFT-based or autocorrelation-based methods are not preferable, as they only provide *average* period of a walk.

By using the detrended log-energy of the spectrum, *GaitCube* can successfully extract individual steps and gait cycles. In order to remove erroneous peaks, a criterion based on minimum distance between peaks is used, and *GaitCube* retrieves the peaks with higher amplitudes. A comparison of the raw energy and the detrended log-energy with a smoothing filter applied can be found in Fig. 3.

Cycle Verification: In order to use these gait cubes as representative samples of human gait, *GaitCube* extracts valid walking steps and gait cycles (*i.e.*, cycles with full walking speed) out of all periods. Unstable walking periods, *e.g.*, initial and final steps of a walking segment, need to be removed. To do so, *GaitCube* utilizes two criteria:

- Walking distance and duration extracted from the variance-range plot: We observe that acceleration and deceleration steps traverse shorter distances, and can be distinguished by thresholding on distance and duration.
- Variations in the torso speed: After thresholding, *GaitCube* extracts the median speed of each step and gait cycle, and removes those cycles with a speed variation exceeding 25% of median speed.

The first criterion validates a full gait cycle, and the second one removes acceleration and deceleration cycles, while leaving room for speed differences between left and right steps. These procedures ensure extraction of physically relevant and useful features, which will be used by the classifier algorithm. In contrast to the purely deep learning based approaches, *GaitCube* reduces computational complexity, and requires a very short window (*e.g.* 0.5 s) to capture useful gait information. Consequently, *GaitCube* enables rapid detection and classification.

Dimensionality Reduction Lastly, we reduce the dimensions of the gait cube before feeding it into the CNN. For the range domain, we investigate downsampling G_{align}. To achieve optimal combination, we investigate various combining methods such as maximum, mean, and median, as the signal is correlated between consecutive range bins. The same experiment is repeated 24 times, and the output SNRs are provided in Fig. 4. As seen, extracting maximum effectively smooths person trace and results in higher SNRs in our setting, and is used in *GaitCube*.

For frequency domain, *GaitCube* removes frequency bands with very little speed information. Walking inbound and outbound introduces positive and negative Doppler shifts, respectively, and the other half spectrum (frequency bins) does not have any useful signal content. *GaitCube* extracts frequencies that correspond to walking speeds of $[v_{min}, v_{max}]$.

Lastly, *GaitCube* reduces dimensionality in the time domain by resizing the data cube to a fixed size. As each cube has

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3083934, IEEE Internet of Things Journal

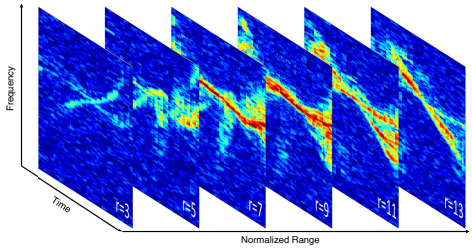


Fig. 5: A Gait Cube shown at various range bins separately

TABLE I: Neural Network Layers

Layer Name	Filter (Kernel)	Output Size
Conv. Layer 1	$(5,5)$, stride:$(2,2)$	$(12,112,48)$
Pooling Layer 1	$(2,2)$, stride:$(2,2)$	$(12,56,24)$
Conv. Layer 2	$(5,3)$, stride:$(3,1)$	$(24,18,22)$
Batch Norm.		
Pooling Layer 2	$(2,2)$, stride:$(2,2)$	$(48,9,11)$
Conv. Layer 3	$(3,3)$, stride:$(2,2)$	$(48,4,5)$
Batch Norm. + Flatten		$(1,960)$
Fully Conn. (FC)		$(1,120)$
FC with additional features	$(1,126)$	$(1,60)$
Softmax		$(1,10)$

time-series information that relates to the same stages of a gait cycle, resizing preserves the *signature*. After all these methods, an example *gait cube* can be seen in Fig. 5, where we have provided various range bins of gait cube prior to downsampling.

Within the current pipeline, we note that two procedures may cause loss of useful information related to human gait. Extracting *aligned* gait cubes removes range information, whereas resizing in the time domain throws away valuable speed information. To alleviate the effects of these, *GaitCube* extracts several additional features from the gait cubes. Specifically, *GaitCube* extracts trace length (*i.e.* stride length) and cycle duration, which are readily available, as well as the mean and the variance of the speed, which can be calculated easily. These features provide additional information to the classifier, as they relate to the mechanics of the human gait.

IV. Deep Learning Model for Classification

In this section, we provide the design of the learning network. In specific, *GaitCube* uses a basic CNN coupled with several convolutional layers, followed by batch normalization and pooling layers with ReLU nonlinearity. In order to mitigate the quadratic power decay on the amplitude of the returned signal, as given by (2), we scale gait cubes by extracting log-magnitudes, and further apply normalization. CNN is used for feature extraction from gait cubes, and its output is combined with the aforementioned additional features by concatenation and fed through a few fully connected layers, followed by a softmax layer to extract output probabilities of each class, whose details can be seen in Table I. We further apply dropout to the outputs of the fully connected layers with probability $p = 0.5$, as this is shown to reduce overfitting [21].

In order to use the CNN model with the outputs of the previous module, we resize gait cubes in the time domain to have a fixed size. The typical human walking speed is around 1 m/s (3.6 km/h), with two steps per second. Therefore, the input size to the CNN is set to be $(5, 227, 99)$, with dimensions (range $\times$ frequency $\times$ time), where $n = 99$ corresponds to 506 ms chosen with respect to the empirical walking speed.

The imbalance between the right and left steps can also provide useful information for identifying people [22]. To exploit this phenomenon, we merge two consecutive gait cubes prior to feeding into the CNN. As merging two cubes with respect to the time dimension increases the size of gait cubes, we resize these *merged* gait cubes to the CNN input size. As *GaitCube* cannot distinguish the left and the right foot, the classifier is trained with both sequences, using overlapping full gait cycles, whereas further evaluation on using a single step is also done.

We use cross entropy-loss, combined with L_2 regularization on the weights to train our classification module. The cost function is given as:

$$\mathcal{L} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{N_c} z_{n,c}\log \hat{z}_{n,c} + \alpha\|w\|^2, \qquad (16)$$

where N_c is number of identities (*i.e.* classes), N is the total number of samples, $z_{n,c}$ is the one-hot encoding for sample n, $\hat{z}_{n,c}$ are the outputs of the network at softmax layer, α is the regularization parameter, and w are weights of the network. We train our network with stochastic gradient descent, with a batch size of 80, and a learning rate of 10^{-2} and reduce the learning rate gradually. The maximum number of epochs is set to 50. We implement deep learning module of *GaitCube* in PyTorch, with NVIDIA CUDA toolkit. Using NVIDIA RTX 2080S GPU with Intel Core i7-9700 CPU, the training procedure takes less than 15 minutes, whereas testing a single sample takes 0.1 ms on average.

TABLE II: mmWave Radar Parameters

Parameter	Value
Carrier Frequency (f_c)	77 GHz
Effective Bandwidth	3.52 GHz
Chirp Duration	19.1 μs
Chirp Frequency per Tx Antenna	6250 Hz
Chirp Slope	209 MHz/μs
ADC Sampling Frequency	15 Msps
Range Resolution	4.19 cm
Maximum Range	10.91 m

V. Implementation and Evaluation

A. Implementation

We implement *GaitCube* using a commodity off-the-shelf (COTS) mmWave radar, IWR1443BOOST. The radar operates at 77GHz and has 3 transmitter and 4 receiver antennas. We configure the device to use two transmitter antennas in time-domain multiplexing mode due to the hardware limitations on the sampling rate and capture the received signal on all antennas simultaneously. Placement of the transmitter antennas enables us to exploit virtual antenna array concept by creating 8 virtual receiver antennas, and therefore obtain *less-dependent* measurements of the physical environment, and capture specular reflection from different limbs. We provide the radar parameters in Table II.

GaitCube uses the parameters given in Table III for gait-cube extraction. In our experiments, the observed maximum speed on gait-cubes is around 4 m/s, corresponding to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3083934, IEEE Internet of Things Journal

TABLE III: Preprocessing Parameters

Parameter	Notation	Value
Window length	N_f	512 ($\approx$ 81 ms)
Window type		Hamming
Overlap amount	N_{ov}	480 ($\approx 77ms$)
Speed resolution		2.38 cm/s
Time step for speed calculation	D	256 ms
Walking detection thresholds	(v_{walk}, v_{stat})	$(0.7, 0.3)$ m/s
Maximum trace deviation	T_{th}	1 sample (10 m/s)
Range for gait cubes	$\pm N_{body}$	$\pm$33.52 cm
Speed range (walking inbound)	$[v_{min}, v_{max}]$	$[-0.75, 4.75]$ m/s
Speed range (walking outbound)	$[v_{min}, v_{max}]$	$[-4.75, 0.75]$ m/s

Fig. 6: Experimental area and data collection setups

foot motion. On the other hand, some volunteers swing their arms significantly in the opposite direction while walking, which requires the inclusion of negative speeds for walking inbound (and vice versa for outbound). Hence, we select our speed range as $[-0.75, 4.75]m/s$ for inbound walking, and $[-4.75, 0.75]$ for outbound walking.

To select proper range values for *GaitCube*, we examine the SNR of the received signals at different ranges and show the results in Fig. 4. As seen, the very first range bins have very little signal content, while the trailing (behind the torso) range bins still have significant signal content, which stems from the multipath effect. Hence, by experimentation and visual inspection of gait cubes, we limit the range dimension of gait cubes into $(-33.52, +33.52)$ cm, which corresponds to ± 8 range bins. After extraction of the gait cubes with ± 8 ranges, *GaitCube* downsamples in the range domain with a factor of three by extracting maximum, as discussed in Section III.

The current implementation exploits receiver diversity by treating each antenna as spatially independent in order to boost dataset size for training and thus reduce data collection effort for each user. In our experiments, we also evaluate *GaitCube* by combining the receiver predictions together via majority voting, as will be discussed in Section V-C.

B. Data Collection

For experimentation, we recruited 10 users (4 female, 6 male), where each user walks in six different setups at two to four days, with at least a week in between consecutive data collection. The overall data collection takes more than 1 month, with a total of 11 sessions, as some users were

TABLE IV: Experimental Setting

Environment	Description (width x distance)
Setup 1	1.3×5 m corridor, $4.8 \times 6m$ open space
Setup 2	6×15 m open space, with two columns
Setup 3	5×3 m open space, $2 \times 8m$ corridor
Setup 4	2.3×7 m wide corridor, $8 \times 5m$ open space
Setup 5	6.5×2.8 m open space, $1.5 \times 10m$ corridor
Setup 6	1.5×12 m corridor

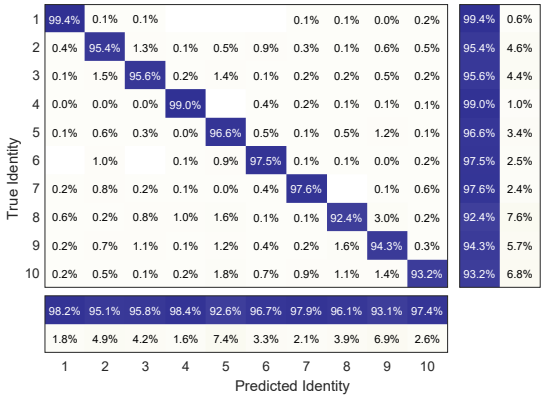


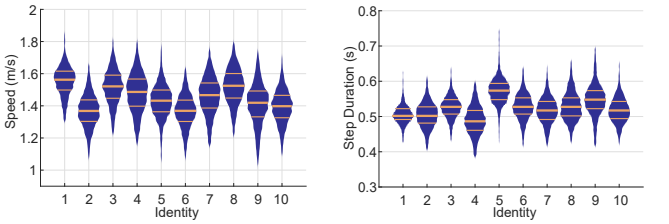
Fig. 7: Confusion matrix for person identification.

available on different days. Ages of the users vary from 23 to 59, with weights varying from 50kg to 77kg and heights varying from 160cm to 174cm. During data collection sessions, users wear arbitrary clothes and accessories, and we do not ask for any specific requirement.

The experimental setup for data collection is given in Fig. 6. In each setup, we place the radar on a cart, with an approximate height of 1m. We ask users to walk 5 times in each way in all setups, where the path length is approximately 10-12 meters. Our algorithm segments walking instances and directions automatically, and the total duration of the walk per user per setup is around 2 minutes, depending on walking speed and distance. In order to test the generalization of our method to multiple environments with different multipath signatures, we select different sections of a large office (>1000 m^2). We provide detailed descriptions of the setups in Table IV. As shown, these setups have open spaces, corridors with varying widths and lengths, and different combinations of both environments. This enables a more thorough analysis of the system.

The data collection procedure results in a total of about 52000 non-overlapping gait cubes from 10 people, at six setups. The number of samples in each setup varies less than $\pm 3\%$. The data distribution between walking inbound and outbound are quite uniform, at 51.1% and 48.9%, respectively. In order to illustrate the challenges with this time and location varying data collection procedure, we provide a violin plot showing the distribution of walking speeds and step durations of different users in Fig. 8a. As shown, our users have diverse walking speeds and gait cycle durations, even the speeds of the same user vary over time. Although this information is not sufficient for accurate classification, it improves the performance of *GaitCube* slightly, as will be explained in the next section.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3083934, IEEE Internet of Things Journal



(a) Speed Distribution (b) Gait Cycle Duration Distribution

Fig. 8: Distribution of walking data, where horizontal lines represent 25, 50 and 75th percentile, respectively

C. Performance

In this section, we evaluate the performance of *GaitCube* under different setups. Unless stated otherwise, we implement 6-fold cross-validation on our dataset, with separation based on data collection setup to ensure independence among realizations, and show its performance over different times and locations. In addition, we report the testing accuracies with interpreting each receiver's decision independently, unless stated otherwise.

Overall performance: To evaluate the overall accuracy, we use a full gait cycle for training and testing and illustrate the confusion matrix in Fig. 7. As seen, *GaitCube* achieves a mean accuracy of 96.1%, with recall and precision rates higher than 92.6% and 92.4% respectively for all users.

Impact of the distance from radar: We investigate the effect of the distance from the radar by extracting accuracy for different range values, and observe a lower accuracy at short and long distances as shown in Fig. 9. As the distance from the radar increases, the received signal has a lower SNR, which reduces the performance. On the other hand, when the object is very close to the radar, the gait cube gets distorted because of the radiation pattern from a point source, and the performance of the system decreases slightly. As stated before, we treat the receiving antennas independently, and even with an equal gain combining scheme with 8 antennas, we should obtain an SNR improvement of 9dB. This scheme can possibly improve the identification accuracy at longer distances significantly.

Impact of gait cycle duration: As *GaitCube* resizes gait cubes during preprocessing, we also evaluate the performance based on the duration of the gait cycle. As can be seen from Fig. 10, most of the gait cycles have a duration in between $(0.4, 0.6)$s and the performance in these regions is the highest. Our gait-cube extraction sometimes falsely outputs long half gait cycles (up to 1.1s), which is tolerated by the classification module to some extent. Longer gait cubes decrease the overall performance of *GaitCube*, yet the number of those samples is at least an order of magnitude lower, and does not contribute to the overall error significantly.

Impact of walking direction: We train our algorithm in both walking directions, as mentioned before. As shown in Fig. 11, our system does not favor one particular direction, and inbound and outbound accuracies are within 1% difference. We also try training in one direction and testing in another, but do not obtain meaningful results, even with reflecting gait cubes with respect to frequency and range. This is because changing walking direction affects the reflection surface, and the range-Doppler signature is distorted significantly. Thus, in practice, in order to recognize a user regardless one is coming or leaving, we need to train in both directions.

Impact of the number of steps: We also evaluate the performance with a varying number of steps involved in one sample. To that end, we train the system with half-cycles (single step) and combine the scores of consecutive samples to construct a decision rule. We average softmax scores (*i.e.* probabilities) over several steps and extract the decision based on the combined score. Performance improvements with respect to the number of steps can be seen in Fig. 12. *GaitCube* achieves an accuracy of 98.3% with three steps, which slightly decreases to 97.3% using a single cycle of two steps, and maintains 94.5% accuracy even using a single step. The remarkable performance allows *GaitCube* to recognize a user with minimal delay, *e.g.*, one single step.

Impact of the training data size: To further investigate low training data requirement, we vary the number of instances per setup from 1 to 5 for training and show the accuracy in the testing setup. As seen from Fig. 13, even with using only one single round-trip walk at each environment for training ($\approx$20s), *GaitCube* achieves a minimum accuracy of 84.8% and a mean accuracy of 89.6%. In this evaluation, we do not test the model against the remaining data in the setup trained, as this would artificially improve the accuracy.

Impacts of gait cubes: To better understand the performance gains from different components, we also conduct a basic ablation study. We investigate the improvement stemming from additional features and/or 3D data cubes. Our results are presented in Fig. 14, where we investigate using the proposed gait cube (FBS), using front or behind slices only (FS and BS), or having a single center slice (SS) of the cube. We also compare the performance with and without the additional hand-crafted features ('+' in the method name denotes using those features). As seen, using full gait cubes plus additional features (FBS+) provides better performance than the other methods, where BS+ and SS+ perform the worst. This is caused by the multipath effect, as the trailing range bins are affected more than leading range bins (Fig. 4), whereas capturing all of the information seems more useful. Nevertheless, we note that even using a single slice, an accuracy of over 90% can be achieved thanks to the many building blocks of *GaitCube*. On the other hand, although the hand-crafted features provide an accuracy gain of 0.6%, *GaitCube* learns the gait signatures well, as the performance gain is marginal, and high accuracy is preserved even with long duration cycles, as discussed before.

Combining receivers for testing: In practice, the system only needs to make a single decision at a time, allowing to fuse all the receiving antennas for better performance. Here we experiment, using a single step, with a simple majority voting scheme, and combine softmax scores of each receiving antenna for overall output. As shown in Fig. 15, the performance increases by 1.3% on average and by 3.1% in Setup 1. We plan to explore different combining schemes in the future.

Rapid training and generalization: We further investigate the question: Can we train *GaitCube* with just two minutes

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3083934, IEEE Internet of Things Journal

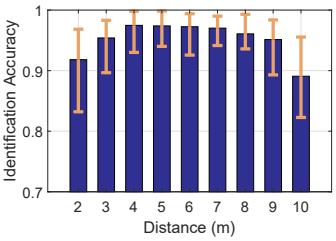


Fig. 9: Accuracy w.r.t. distance from the radar

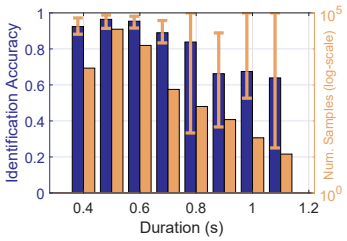


Fig. 10: Accuracy w.r.t duration of half cycles (*e.g.* a step)

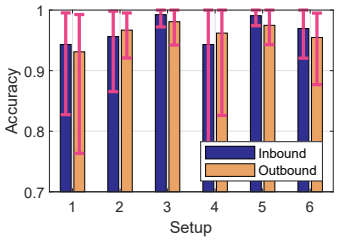


Fig. 11: Accuracy w.r.t different walking directions

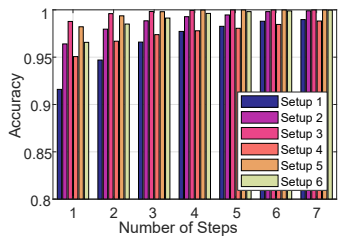


Fig. 12: Accuracy w.r.t number of steps per sample for testing

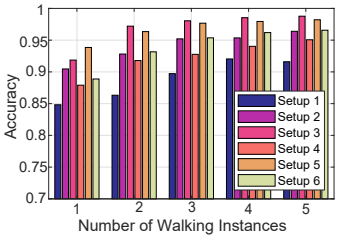


Fig. 13: Accuracy w.r.t number of instances used for training

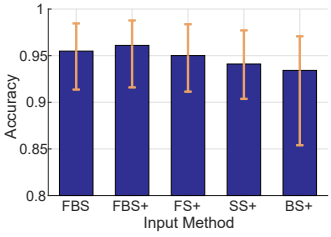


Fig. 14: Different methods of building gait cubes

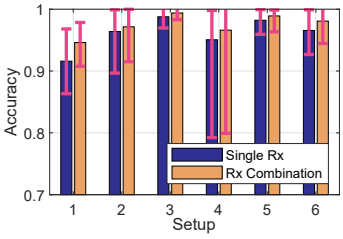


Fig. 15: Accuracy gain by receiver combination for testing

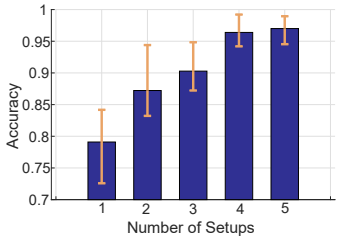


Fig. 16: Accuracy w.r.t number of setups used for training

of walking data at one location and generalize it over time and locations? This use case is particularly challenging, as it neither allows the use of multiple locations nor different time instances for training data. We study this problem with a more practical setting and implement receiving antenna combining. To that end, we investigate the performance with respect to the number of setups used for training, as in Fig. 16. We provide the mean accuracy obtained by training at varying numbers of locations, as well as the error bars representing the deviation from the mean value for different testing setups. Our system achieves 79.1% mean accuracy with one-shot learning, and using two setups further boosts the accuracy to 87.2%. In these two cases, the minimum accuracy is 72.5% and 82.7%, which could be acceptable for a practical application. By these results, we conclude that *GaitCube* can learn *generalized gait signatures by less than 2 minutes of data at one location*.

Unknown Person Rejection: In a different scenario, we change our focus from *identification* and investigate the *verification* problem. We randomly select 5 users and train a classifier based on these users only. We test the detection module with the complete dataset, which includes 5 additional people, who have not been trained for the system. We apply a decision threshold on the probability scores to label samples as the known or unknown class. We find the optimal threshold by minimizing the error on ROC. The results indicate that, without any prior training, *GaitCube* can reject an unregistered user with 86.2% precision and 83.2% recall rates. A more robust method would include training with many different users that include an *unknown* class, which is adopted by some wireless sensing work, as in [5]. As main use-case of *GaitCube* is *identification*, we leave further investigation of the *verification* problem to a future work.

VI. Discussions

Multi-Person Support: Our current implementation only supports a single person tracking and identification at a time. This would be already useful for many applications that usually authorize a single person at a time, *e.g.*, an entrance system or an IoT customization scheme. If multiple users are present at different distances from the radar, a simple extension of looking at different ranges separately would enable multi-user support for *GaitCube*. For more complicated scenarios, we plan to investigate the radar device's potential for multi-target detection [23], [24].

Walking Direction: Our system requires a user to walk towards or away from the device to ensure the successful construction of gait cubes. Although this somewhat imposes some limitations, *GaitCube* can recognize a user with good accuracy even with only one step. Therefore, we believe *GaitCube* is a practical system even with this limitation. Some other works [24], [25] relax the constraints by relying on point clouds and collecting data with different walking directions or with multiple devices. These approaches, however, have much more complicated and extensive training procedures, and may not be environment independent.

Larger Dataset: Many works in the literature build a dataset of a dozen of users [23], [25]–[27], while some of them may have more users (yet usually with one single location at a single time for both training and testing) [5], [10], [24], [28]. Due to the budget constraints and the COVID-19 pandemic, we limit our dataset to 10 users as well. But encouraged by the remarkable performance, we plan to validate *GaitCube* on a larger dataset later.

VII. Related Work

Wireless sensing has been an emerging field [31]–[35] in recent years, with applications to vital signs monitoring (*e.g.* breathing, heart-rate) [11], [12], indoor localization [36], [37], gesture recognition [38], as well as gait recognition. Among different wireless sensing paradigms, WiFi and radar-based systems are of interest for our gait identification system design.

WiFi-based: WiFi based solutions usually extract features from Channel State Information (CSI) [5], [6], [10], [22], [39]–

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3083934, IEEE Internet of Things Journal

TABLE V: Comparison of *GaitCube* with the state-of-the-art-methods

Method	Input	Tracking	Accuracy (People)	Location	Resp. Time	Acc. in Diff. Env	Training Data	Range	Area	Num. Parameters
Vandersmissen *et al.* [29]	Spectrum (2D)	None	84.9% (5)	Indoor, same loc.	3s	78.6%(5)	20 mins	$9.5m$	Open spaces	32k+
Radar-ID [27]	Spectrum (2D)	None	85.6% (10)	Outdoor, same loc.	2s	N/A	7 mins	$10m$	Outdoor	61M+
mID [25]	Point Clouds	Kalman Filter	89.0% (12)	Indoor, same loc.	2s	N/A	9 mins	$5m$	Open space	8.5M+
mmGaitNet [24][1]	Point Clouds	Clustering-based	90.0% (10)	Indoor, same loc.	3s	86.0% (10)	16 mins	$8m$	Open space	N/A
Pegoraro *et al.* [30]	Spectrum (2D)	Kalman Filter	98.27% (4)	Indoor, same loc.	2s	N/A	20 mins	$10m$	Open space	75k+
MU-ID [23]	Range-Time-Velocity	Clustering-based	97.7% (10)	Indoor, same loc	1 step $\sim 0.5s$	78.0% (5)	1 min[2]	8m	Open space, corridor	4M+
GaitCube	Range-Spectrum(3D)	Peak Tracking	96.1% (10)	Indoor, new loc.	1 step $\sim 0.5s$	79.1% (10)	2 mins	11m	Open spaces, corridors	140k

[1] Uses multiple radar devices for recognition, and can separate multiple users.
[2] Operates on single direction data (*i.e.* walking towards the device)

[42]. Either hand crafted features from the WiFi CSI [5], [10], [39], [40] are extracted, or automatic, learning based systems [41]–[43] have been proposed. One major disadvantage is the fact that WiFi-based systems are difficult to generalize well to new locations due to the poor resolution of multipath indoors. Most of the early works use a single location at a single time and shuffles training/testing data randomly, rendering uncertain performance at another location/time [5], [6]. As validated by [43], the performance of several classical works using WiFi, such as [5], [6] for identification, decreases significantly when tested at a new location. Another method is extracting physical features from WiFi CSI, such as speed information and building a classifier using related features, as done in [22]. Although this approach is shown to be less environment-dependent, it cannot capture speeds of different limbs and the authors report an accuracy of 70% for 10 users. In addition, recent works [43]–[45] resort to deep learning to enable cross-domain recognition, but may still require (unlabeled) data collection from a new environment to enable operation at a different location.

Radar-based: Compared to the WiFi devices, radars offer much higher resolution and thus enable extraction of the fine-grained motion signatures for classification. Early works employ customized radars for characterizing human gait [46], activity recognition [47], extracting motion of different limbs based on their range-Doppler signatures [20], and reconstructing human posture [48]. Recently, compact radars have become commercially available. By using Doppler signatures reported by COTS devices, different methods have been proposed for gait-based identification, as in [26], [27], [29], [30]. In addition, point cloud based methods are also proposed [24], [25], which extract the point cloud data directly from the radar device and apply deep learning methods to process and identify human gait. Using point clouds are prone to errors reported by the radar device during the point cloud generation especially for rich multipath environments, as noted in [25], limited by range (*e.g.* [25] operates at 5 meters), used in open spaces only, as in [24], [25] and require computationally intensive methods to process the point cloud data. Similar to WiFi-based works, many of the existing works use data collection done at once and do not verify their performance in more practical settings [25], [27], [30]. The latest work [23] uses a 2D range-time-(max-velocity) plot and focuses on human foot for identification, which potentially misses rich information of human gait. Although the authors report an accuracy of 97% for 10 people tested at the same location as training, their performance decreases to 78% for a dataset of 5 people when tested at a different location. The authors in [28] study the problem of (re)identification over a long period of time and different locations, but they use a custom-built radar with favorable settings. We summarize some of the related work in Table V, and note that the existing methods require much longer response times, due to using point clouds, or fixed window spectrograms, are not tested in new locations and/or affected by location changes severely, require much longer data collection sessions, and have more limited ranges. Different from the previous works, *GaitCube* shows its high performance generalized to different locations and time, and more importantly, can register and recognize a user rapidly with minimal data required, rendering it a more practical gait recognition system.

VIII. Conclusion

In this paper, we propose *GaitCube*, a practical gait recognition system that requires minimal data for training (*e.g.*, less than two minutes of data) and testing (*e.g.*, a single gait cycle) while maintaining high performance over different locations and times. *GaitCube* learns from a distinct gait cube processing with a neural network. Experiments with 10 users at different locations and times show that *GaitCube* achieves an accuracy of 96.1% with a single walking cycle which further improves to 98.8% with two cycles, and the accuracy maintains 79.1% even with only two minutes of training data at a single location, promising it a practical solution for real applications.

References

[1] Q. Xu, Y. Chen, B. Wang, and K. J. R. Liu, "Radio biometrics: Human recognition through a wall," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1141–1155, 2017.
[2] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging wifi-enabled iot," in *Proceedings of ACM MobiHoc*, 2017, pp. 1–10.
[3] S. D. Regani, Q. Xu, B. Wang, M. Wu, and K. J. R. Liu, "Driver authentication for smart car using wireless sensing," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2235–2246, 2020.

[4] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa, "Identification of humans using gait," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1163–1173, 2004.

[5] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of ACM UbiComp*, 2016, pp. 363–373.

[6] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: Wifi-based person identification in smart spaces," in *Proceedings of ACM/IEEE IPSN*, 2016, pp. 1–12.

[7] Kaustubh Kalgaonkar and Bhiksha Raj, "Acoustic doppler sonar for gait recogination," in *Proceedings of IEEE AVSS*, 2007, pp. 27–32.

[8] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S.-M. Makela, and H. A. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *Proceedings of IEEE ICASSP*, 2005.

[9] R. J. Orr and G. D. Abowd, "The smart floor: A mechanism for natural user identification and tracking," in *Proceedings of ACM CHI*, 2000, pp. 275–276.

[10] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *Proceedings of IEEE DCOSS*, 2016.

[11] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proceedings of ACM CHI*, 2015, pp. 837–846.

[12] F. Wang, F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "Vimo: Vital sign monitoring using commodity millimeter wave radio," in *Proceedings of IEEE ICASSP*, 2020, pp. 8304–8308.

[13] S. D. Regani, C. Wu, F. Zhang, B. Wang, M. Wu, and K. J. R. Liu, "Handwriting tracking using 60 ghz mmwave radar," in *Proceedings of IEEE WF-IoT*, 2020.

[14] C. Wu, F. Zhang, B. Wang, and K. J. R. Liu, "mmtrack: Passive multi-person localization using commodity millimeter wave radio," in *Proceedings of IEEE INFOCOM*, 2020.

[15] ——, "msense: Towards mobile material sensing with a single millimeter-wave radio," in *Proceedings of ACM IMWUT/UbiComp*, 2020.

[16] F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "mmeye: Super-resolution millimeter wave imaging," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6995–7008, 2021.

[17] A. G. Stove, "Linear fmcw radar techniques," *IEE Proceedings F - Radar and Signal Processing*, vol. 139, no. 5, pp. 343–350, 1992.

[18] M. I. Skolnik *et al.*, *Introduction to radar systems*. McGraw-hill New York, 1980, vol. 3.

[19] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 2–21, 2006.

[20] O. R. Fogle and B. D. Rigling, "Micro-range/micro-doppler decomposition of human radar signatures," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3058–3072, 2012.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[22] C. Wu, F. Zhang, Y. Hu, and K. J. R. Liu, "Gaitway: Monitoring and recognizing gait speed through the walls," *IEEE Transactions on Mobile Computing*, vol. 20, no. 6, pp. 2186–2199, 2021.

[23] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "Mu-id: Multi-user identification through gaits using millimeter wave radios," in *Proceedings of IEEE INFOCOM*, 2020, pp. 2589–2598.

[24] Z. Meng, S. Fu, J. Yan, H. Liang, A. Zhou, S. Zhu, H. Ma, J. Liu, and N. Yang, "Gait recognition for co-existing multiple people using millimeter wave sensing," in *Proceedings of AAAI*, vol. 34, no. 01, 2020, pp. 849–856.

[25] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "mid: Tracking and identifying people with millimeter wave radar," in *Proceedings of IEEE DCOSS*, 2019, pp. 33–40.

[26] Y. Yang, C. Hou, Y. Lang, G. Yue, Y. He, and W. Xiang, "Person identification using micro-doppler signatures of human motions and uwb radar," *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 5, pp. 366–368, 2019.

[27] P. Cao, W. Xia, M. Ye, J. Zhang, and J. Zhou, "Radar-id: Hhuman identification based on radar micro-doppler signatures using deep convolutional neural networks," *IET Radar, Sonar Navigation*, vol. 12, no. 7, pp. 729–734, 2018.

[28] L. Fan, T. Li, R. Fang, R. Hristov, Y. Yuan, and D. Katabi, "Learning longterm representations for person re-identification using radio signals," in *Proceedings of IEEE/CVF CVPR*, June 2020.

[29] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. D. Neve, and T. Dhaene, "Indoor person identification using a low-power fmcw radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, pp. 3941–3952, 2018.

[30] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2020.

[31] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, 2007.

[32] B. Wang, Q. Xu, C. Chen, F. Zhang, and K. J. R. Liu, "The promise of radio analytics: A future paradigm of wireless positioning, tracking, and sensing," *IEEE Signal Processing Magazine*, vol. 35, no. 3, pp. 59–80, 2018.

[33] K. J. R. Liu and B. Wang, *Wireless AI: Wireless Sensing, Positioning, IoT, and Communications*. Cambridge University Press, 2019.

[34] J. Liu, G. Teng, and F. Hong, "Human activity sensing with wireless signals: A survey," *Sensors*, vol. 20, no. 4, p. 1210, 2020.

[35] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.

[36] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "Pilot: Passive device-free indoor localization using channel state information," in *Proceedings of IEEE ICDCS*, 2013, pp. 236–245.

[37] C. Chen, Y. Chen, Y. Han, H.-Q. Lai, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on wifi platforms: A frequency hopping approach," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 111–121, 2016.

[38] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *Proceedings of IEEE INFOCOM*, 2015, pp. 1472–1480.

[39] F. Hong, X. Wang, Y. Yang, Y. Zong, Y. Zhang, and Z. Guo, "Wfid: Passive device-free human identification using wifi signal," in *Proceedings of EAI MobiQuitous*, 2016, p. 47–56.

[40] Y. Li and T. Zhu, "Gait-based wi-fi signatures for privacy-preserving," in *Proceedings of ACM ASIA CCS*, 2016, pp. 571–582.

[41] C. Lin, J. Hu, Y. Sun, F. Ma, L. Wang, and G. Wu, "Wiau: An accurate device-free authentication system with resnet," in *Proceedings of IEEE SECON*, 2018, pp. 1–9.

[42] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. J. Spanos, "Wifi-based human identification via convex tensor shapelet learning," in *Proceedings of AAAI*, 2018.

[43] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "Crosssense: Towards cross-site and large-scale wifi sensing," in *Proceedings of ACM MobiCom*, 2018, p. 305–320.

[44] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *Proceedings of ACM MobiSys*, 2019, pp. 313–325.

[45] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of ACM MobiCom*, 2018, pp. 289–304.

[46] P. Van Dorp and F. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar & Navigation*, vol. 2, no. 2, pp. 135–145, 2008.

[47] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, 2009.

[48] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Trans. Graph.*, vol. 34, no. 6, Oct. 2015.