Gait Recognition for Co-Existing Multiple People Using Millimeter Wave Sensing

Zhen Meng,[1] Song Fu,[1] Jie Yan,[1] Hongyuan Liang,[1] Anfu Zhou,[1]
Shilin Zhu,[2] Huadong Ma,[1] Jianhua Liu,[3] Ning Yang[3]
[1]Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, BUPT, Beijing, China
[2]University of California San Diego, San Diego, CA, United States
[3]OPPO Co. Ltd., Beijing, China
{zhenmeng, fusong, yanjie, hongyuanliang, zhouanfu, mhd}@bupt.edu.cn,
shz338@eng.ucsd.edu, {liujianhua, yangning}@oppo.com

Abstract

Gait recognition, *i.e.*, recognizing persons from their walking postures, has found versatile applications in security check, health monitoring, and novel human-computer interaction. The millimeter-wave (mmWave) based gait recognition represents the most recent advance. Compared with traditional camera-based solutions, mmWave based gait recognition bears unique advantages of being still effective under non-line-of-sight scenarios, such as in black, weak light, or blockage conditions. Moreover, they are able to accomplish person identification while preserving privacy. Currently, there are only few works in mmWave gait recognition, since no public data set is available. In this paper, we build a first-of-its-kind mmWave gait data set, in which we collect gait of 95 volunteers 'seen' from two mmWave radars in two different scenarios, which together lasts about 30 hours. Using the data set, we propose a novel deep-learning driven mmWave gait recognition method called mmGaitNet, and compare it with five state-of-the-art algorithms. We find that mmGaitNet is able to achieve 90% accuracy for single-person scenarios, 88% accuracy for five co-existing persons, while the existing methods achieve less than 66% accuracy for both scenarios.

Introduction

Gait is an important biological feature of human beings. In recent years, gait recognition has been found versatile applications in security check, health monitoring, and novel human-computer interaction. For instance, the home automation system can automatically adjust the lighting brightness or temperature according to each person's preference once it identifies each person via gait recognition.

The most common gait recognition is based on computer vision, which utilizes camera to capture visual images when a human walks, and then performs the identification (Makihara et al. 2017; Li et al. 2018; Chao et al. 2019; Li, Liu, and Ma 2019). However, the vision solution bears several limitations. *Firstly*, it raises severe privacy threats for capturing the real images of human daily life, especially considering that the camera could be hijacked by malicious users. *Secondly*, cameras are easily affected by lighting conditions. They cannot obtain clear images in dark/weak-light/blockage scenarios, which lead to gait recognition failures.

Recently, researchers propose to utilize wireless signal for gait recognition. In particular, each person's unique walking posture leads to a unique wireless signal variation pattern. By analyzing the pattern, we can identify the person just from the wireless signal, which eliminates the privacy concern and is also independent of the lighting conditions. Moreover, driven by the emerging 5G technologies, gait recognition using the 5G mmWave wireless signal has gained much interest. Compared with the traditional recognition using low-frequency omni-directional Wi-Fi signal (Zou et al. 2018), mmWave gait recognition is promising to achieve higher accuracy, in particular for co-existent multiple persons, because mmWave radios, with 100× bandwidth, can provide much fine-grained spatial resolution.

Despite the potential, however, there is not a mmWave gait data set open to researchers, which hinders the research progress. In this paper, we build and publicly achieve a first-of-its-kind mmWave gait data set, which is collected from 95 volunteers and lasts about 30 hours in total. In particular, we use two mmWave devices in two different scenes as shown in Figure 1. We let each device capture the reflected mmWave signal from walking persons, which forms point clouds. Then we segment the point cloud of multi-people who are walking at the same time to get a single person's gait point cloud data. We propose a novel method to tagging the data: *firstly*, we use clustering algorithm DBscan (Ester et al. 1996) to automatically cluster point clouds without given the number of clustering categories. *Secondly*, we use the Hungarian algorithm (Kuhn 2010) to track the point cloud clusters of one person's routes. *Thirdly*, the routes are matched to corresponding volunteers one by one. We believe that the data set can provide an open base for designing and comparing various mmWave gait recognition methods in a fair way, so as to facilitate further research.

Using the data set, we evaluate five state-of-the-art deep learning based gait recognition algorithms, and have two observations: *(i)* the accuracy of gait recognition decreases with the increase of the number of co-existent walking people. *(ii)* the accuracy of gait recognition increases if us-

ing more mmWave sensing devices hlrespecially when Co-existing multiple people. Motivated by the observations, we propose a new deep learning driven algorithm mmGaitNet. In contrast with previous work mID (Zhao et al. 2019), mmGaitNet designs new neural network model to extract features for each attribute of point cloud, so as to achieve higher recognition accuracy in multi-people co-existing scenarios. The contributions of this paper lie in two aspects:

- We build a first-of-its-kind mmWave gait data set, and use the data set to evaluate the existing deep learning based gait recognition algorithms.
- We design a new mmWave gait method mmGaitNet, which outperforms the existing methods and achieves about 90% accuracy even for multi-person co-existent scenarios.

Related Works

Gait Recognition

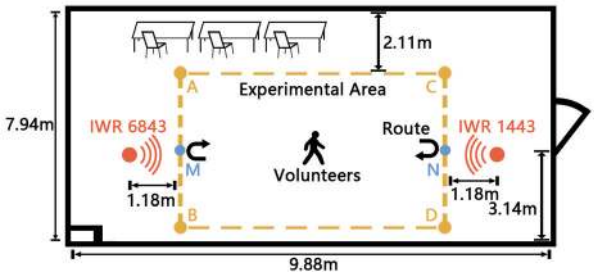
Gait recognition has versatile applications in security check, health monitoring, and novel human-computer interaction. People try to solve the problem with many different methods, such as the methods based on computer vision or alternatives based on various wireless perception. In particular, traditional computer-vision-based gait recognition methods (Makihara et al. 2017; Li et al. 2018; Chao et al. 2019; Li, Liu, and Ma 2019) perform very well in terms of accuracy, but bears several limitations: firstly, camera invades people's privacy for capturing the real image of life, which will cause leakage of personal information, particularly considering that cameras may be attacked and hijacked. Secondly, cameras are easily affected by lighting condition. they cannot obtain a valid image in the dark environment. To solve the problems mentioned above, researchers attempt to utilize wireless signal to capture the gait data of people. In those wireless sensing works, most methods are based on Channel State Information (CSI), such as WiFiU (Wang, Liu, and Shahzad 2016), wiwho (Zeng, Pathak, and Mohapatra 2016) and AutoID (Zou et al. 2018). However, WiFi signals are difficult to be segmented to isolate the impact of each person, so they are unable to identify multiple people at the same time.

mmWave and Wireless Sensing

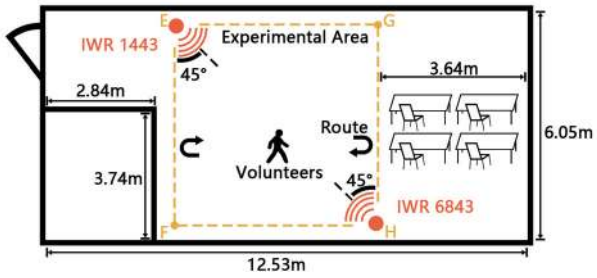
With the rise of 5G, mmWave sensing is expected to play a more important role in gait recognition. The estimated human pose is utilized to do human identifying. Zhao *et al.* demonstrates the feasibility of human tracking and identifying capacity (Zhao et al. 2019) with commercial, off-the-shelf mmWave radars. However, it only can identify individual users.

In addition, mmWave radios are becoming versatile for other sensing applications. Zhou *et al.* proposes autonomous environment mapping (Zhou et al. 2019b) using commodity mmWave network device and robot navigation (Zhou et al. 2019a) in dynamic environment. Yang *et al.* demonstrates vital sign monitoring (Yang et al. 2016) with commercial, off-the-shelf mmWave radars. Wang *et al.* demonstrates gesture recognition (Wang et al. 2016) with Soli (Lien et al. 2016).

In a boarder field, wireless sensing via ubiquitous WiFi/RFID/4G/mmWave signals is becoming another sensing venue besides the conventional vision based venue, which bears the unique characteristic of 'see-through' (Adib and Katabi 2013). Wireless sensing is firstly utilized for indoor localization (Yang, Zhou, and Liu 2013), human posture/gesture estimation (Pu et al. 2013), material identification (Wang et al. 2017) and even for the sub-mm-level vibration recovery (Wei et al. 2015). We note that mmWave sensing operating at high frequency band represents the frontier of wireless sensing, which provides high spatial resolution and enables more robust sensing.



(a) scene1



(b) scene2

Figure 1: The experimental scenes. Scene1 simulates the application scenario of volunteers walking in the corridor facing the devices. In this scenario, the effective horizontal detection angle of the devices is $\pm 60^{\circ}$. Scene2 simulates the application scenario of volunteers walking in a living room. In this scenario, the effective horizontal detection angle of the devices is $\pm 45^{\circ}$.

Data Collection Methodology

We collect gait data of volunteers with two commercial, off-the-shelf mmWave radars in two scenes as shown in Figure 1. Scene1, as shown in Figure 1(a), imitates a corridor where the open free for walking is a long rectangle. When a person is close to the device, the device cannot scan the whole body of the person because the effective vertical monitoring angle of the device is less than $\pm 20^{\circ}$. Hence, the location of the two devices is 1 meters away from the point M and N. The height of the devices is 1m. The angle between the

first device (*i.e.*, IWR6843) and line-segments AB is $0°$, and the angle between the second device (*i.e.*, IWR1443) and the line-segments CD is $0°$ too. Scene2, as shown in Figure 1(b), imitates an office or homeroom, where the open space is square. The two mmWave sensing devices (*i.e.*, IWR1443 and IWR6843) are placed at the location of point E and H, with a height of 1m. The angle between IWR6843 and line-segments EF and the angle between IWR1443 and line-segments GH are both $45°$.

mmWave Sensing Device

We use IWR1443[1] and IWR6843[2] as the mmWave sensing devices. The two devices are essentially FMCW radars equipped with multiple antennas (*i.e.*, an antenna array). The FMCW radars periodically transmit a signal in sinusoidal waveform whose frequency increases linearly with time. This type of signal is also named a chirp and can measure the range as well as velocity and angle of sensing targets (Richards et al. 2010). In the experiments, the two devices are configured to use all their three transmitter antennas and four receiver antennas to generate the 3D point cloud data. The devices both output a frame of 3D point cloud, in every 0.1s. The detailed configuration parameters of the devices are as follows:

The configuration of IWR6843. The device transmits 32 chirps per frame. The start frequency of the chirp is set to 60.25GHz. The bandwidth B is set to 3.75GHz. The Chirp Cycle Time Tc is set to 169.33µs. The Idle Time is set to 93µs. The ADC Valid Start Time is set to 7µs. The Ramp End Time is set to 83.33µs. The Frequency Slope is set to be 45GHz/ms. With such a configuration, IWR6843 has a range resolution of 4.4cm and a maximum unambiguous range of 8m. In terms of velocity, it can measure a maximum radial velocity of 2.35m/s, with a resolution of 0.15m/s.

The configuration of IWR1443. The device is set to transmit 16 chirps per frame. The start frequency of the chirp is set to 77GHz. The bandwidth B is 4GHz. The Chirp Cycle Time Tc is set to 131.14µs. The Idle Time is set to 81µs. The ADC Valid Start Time is set to 7µs. The Ramp End Time is set to 57.14µs. The Frequency Slope is set to be 70GHz/ms. With such a configuration, IWR1443 has a range resolution of 4.4cm and a maximum unambiguous range of 8m. In terms of velocity, it can measure a maximum radial velocity of 2.35m/s, with a resolution of 0.3m/s.

Time Synchronization

In order to use the data collected by two devices at the same time, we use timestamps to mark the collected data. To ensure computer time consistency, we run time synchronization Network Time Protocol (NTP) on two computers, while one computer acts as the NTP server and the other as the client. We use the client computer to synchronize with the time of the server. To reduce synchronization time, we connect the two computers directly using a cable. Using above method, the time difference between client and server is less than 5ms. The walking speed of persons is commonly less than 2m/s. Person can be deemed as quasi-stationary at the 5-ms interval.

[1]http://www.ti.com/tool/IWR1443BOOST

[2]http://www.ti.com/tool/IWR6843ISK

Table 1: The number of volunteers in the data set. For example, The number 25 in row 3vol-sim and scene2, fixed route column means that there are 25 volunteers take part in an experiment, in which a volunteer needs to walk with two other volunteers on fixed route in scene2.

Scene	scene1		scene2	
Route	Fixed	Free	Fixed	Free
1vol-sim	20	20	30	25
2vol-sim	20	20	30	30
3vol-sim	15	10	**25**	25
4vol-sim	15	10	15	15
5vol-sim	10	10	10	10

Data Set

We collect a total of 30 hours of 3D point cloud data from 95 volunteers[3]. The data set contains two types of walking trajectories: fixed route and free route where there are up to 5 volunteers walking at the same time. Fixed route means that volunteers walk from one side to the other along a straight line, and free route means that volunteers walk casually on any route in the specified area as shown in Figure 1.

Data Composition

We deliberately ask concurrent people to walk simultaneously all the time during the gait collecting time. We make such setting because it represents the most challenging case. For fixed route, volunteers walk back and forth on the fixed route 25 times; for free route, volunteers walk freely about 10 minutes. The number of volunteers in the data set is shown in Table 1. For example, the number 25 in row 3vol-sim and scene2, fixed route column means there are 25 volunteers take part in this experiment, and each volunteer walks along with two other volunteers on fixed route in scene2.

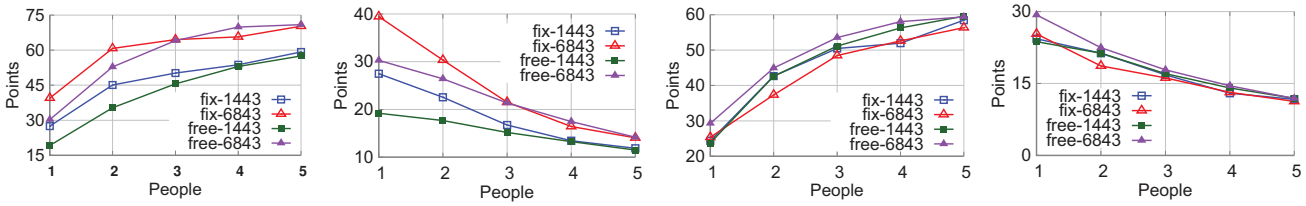
Volunteers

We collect mmGait from 95 recruited volunteers, 45 male and 50 female volunteers. The age of the volunteers is between 19 and 27. More than half of the volunteers are between 20 and 21 years old. The height of the volunteers is between 150cm and 185 cm. More than two-thirds of the volunteers are between 160 cm and 180 cm in height. The weight of the volunteers is between 41kg and 115kg. More than half of the volunteers are between 50kg and 65kg in weight.

Data Characteristics.

The data collected by our sensing devices is quite sparse, as shown in Figure 2. The number of points belong to a volunteer in the point cloud is very small. Note that a human

[3]https://github.com/mmGait/people-gait.



(a) The number of points for a group of volunteers in scene1

(b) The number of points for a volunteers of multi-data in scene1

(c) The number of points for a group of volunteers in scene2

(d) The number of points for a volunteers in scene2

Figure 2: The number of mmWave reflection points. In scene1 (a) and scene2 (c), as the number of volunteers walking increases simultaneously, the number of points in the point cloud increases. However, the increase in the number of points becomes smaller. In scene1 (b) and scene2 (d), as the number of walking volunteers increases simultaneously, the number of points for a volunteer in the point cloud decreases.

body acts as a reflector rather than a scatter. At any point in time, our device can capture only a subset of the radio frequency reflections off the human body. The data in Figure 2 demonstrates that: as the number of simultaneous walking volunteers increases, the number of points in the point cloud increases slowly. On the other hand, the number of points belong to a volunteer in the point cloud decreases. The reason lies in that the output capacity of the device is limited. The 3D point cloud data generated by IWR1443 contains up to 64 points. The 3D point cloud data generated by IWR6843 contains up to 100 points. As the number of points in the point cloud increases, the devices could only output the points with higher confidence. What's more, as the number of points in the point cloud increases, the probability of a volunteer occluded by other volunteers increases. For example, when five people are walking, the devices could only detect three people most of the time. Using two devices working at the same time can effectively increase the density of point cloud and reduce the occlusion of volunteers.

Data Annotation

Data Processing

The devices have removed many noise points which are reflected by static objects utilizing static clutter removal algorithm CFAR (Richards et al. 2010). However, there are still many noise points by high-order reflection between walking people and static objects, such as the situation shown in Figure 3. These noise points usually have larger distance to the receiver, compared with the points reflected directly from walking people. In mmGait, we adopt DBSCAN clustering algorithm to remove the noise points in the point cloud, besides to segment the point cloud formed by multiple people. An example is given in Figure 4.

For the multi-people gait recognition, we use the DBSCAN clustering algorithm to divide the points in a frame into different groups. Each group on behalf of a single person. One advantage of DBSCAN is that we do not need to preset the number of clusters. In our data set, the closest distance between two side-by-side people is about 0.3m which is a normal social distance. In this case, the detection and identification accuracy is about 86.5%. The impact factors of the accuracy are the distance between people and the severity of their mutual covering. The accuracy will drop if the distance becomes smaller or one person is blocked by another because the point clouds of two or more people may merge together and become hard to separate. Figure 4 shows the results of clustering. Then we use a matching algorithm for tracking each person's gait.

We track the clustered point cloud using Hungarian algorithm to get the continuous gait data of volunteers. Hungarian algorithm matches the clusters in current frame to those clusters that appear in the previous frames. The weight matrix of Hungarian algorithm includes the location of clustering categories generated in the last 10 frames, which helps alleviate the clustering interruption caused by the sparseness of point cloud.

Data Merge

A single device is unable to meet the requirements of the experiment. Therefore, we decide to utilize two devices to collect gait data concurrently, which can greatly increase the number of points in point clouds, and also reduce the mutual covering of volunteers. For example, when we use a single device to monitor two volunteers walking at the same time, the volunteers may cover each other. If we use two devices, they can capture full point cloud of both volunteers without blockage. The data collected by the two devices are in different rectangular coordinate systems, because the positions of devices is different. In order to use the data collected by the two devices, we convert the data into the same rectangular coordinate system as follows.

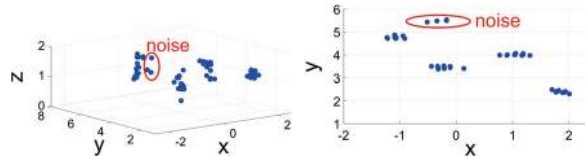
Coordinate transformation. We convert the point clouds data into the same coordinate system by rotation and translation of the coordinate system. Firstly, we rotate the coordinate system of the two devices clockwise to make the two coordinate systems in the same direction. Secondly, we translate the coordinate system of IWR6843 consistent with IWR1443. The translation formula is as follows:

$$x' = xcos(\theta) - ysin(\theta)$$
$$y' = xcos(\theta) + ysin(\theta)$$

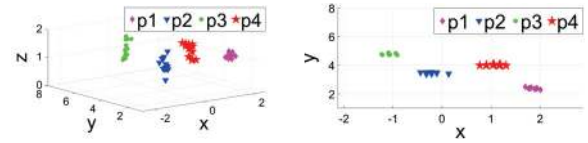
where θ is the rotation angle of a coordinate system, *(x, y)* is the coordinate of one point in the original coordinate system, *(x', y')* is the coordinates of the point in new coordi-

(a) Signal at t (b) Signal at $t+1$ (c) Signal after CFAR

Figure 3: The noise source. Signal B is a signal multi-reflected by the static objects. Signal A is a signal multi-reflected by walking people and static objects. Signal B will be undone after static clutter removal algorithm (CFAR), because it is the same at time t and $t+1$. Signal A has changed due to the walking people and cannot be undone.



(a) original 3D point cloud of 4 subjects (b) original 2D point cloud of 4 subjects



(c) 3D point cloud of 4 subjects after clustering (d) 2D point cloud of 4 subjects after clustering

Figure 4: The 2D and 3D point clouds of 4 people walking at the same time.

nates after transformation.

Merging process. We merge the point clouds collected by the two devices according to the timestamps of point cloud. Firstly, we give each point cloud a new attribute called the device name for recording the device ID that collects the point clouds. Secondly, we combine the coordinate-converted point clouds collected by the two devices into the same file. Thirdly, In particular, we sort all point cloud according to the collection time of the point clouds. We merge the point clouds from two devices whose time difference is less than a specified threshold. In our experiments, we set the threshold to be 50ms. We find that the mean value of the time difference of the merged point clouds is 24ms.

mmGaitNet

Neural Network Structure

The network directly consumes the point cloud takes into account the unique properties of point cloud that the number of points in point cloud is very small. If the point cloud is mapped to a picture, there will generate a lot of redundant data and the network consumption time will grow too high. Points of 3D point cloud have five properties, range, mirror speed, horizontal angle, pitching angle, and signal noise

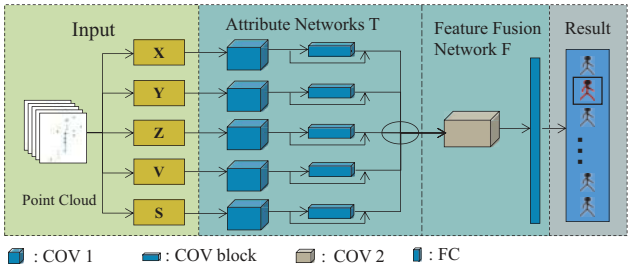


Figure 5: Overview of mmGaitNet. COV 1 means a layer of 7×7 spatio-temporal convolutions kernel with 2×2 strides. COV block means layer1 of ResNet18. COV 2 means a layer of 3×3 spatio-temporal convolutions kernel with 1×1 strides. FC: a fully connected layer.

ratio. According to the distance, horizontal angle, pitching angle of the points, we can derive the three-dimensional coordinates of points. Correspondingly, the network consists of five identical attribute networks $T = \{T_1, T_2, T_3, T_4, T_5\}$ and a fusion network G as shown in Figure 5. The inputs are point clouds' five attributes X, Y, Z, V, and S, where X, Y, Z denote the spatial location, V denotes the radial speed and S denotes the signal strength of the points. These attribute values represent different characteristics of a point. In order to extract time and space information at the same time, the input of each attribute network is a $p \times t$ matrix, where p denotes the number of points in point cloud, and t denotes time. The attribute features of point cloud extracted by attribute network are not comprehensive enough to identify people. We need feature fusion network to fuse the features extracted by each attribute network and distill the overall characteristics of the point cloud to identify people. The final fully connected layer F output the class score which is:

$$scr = F(Cat(G(X), G(Y), G(Z), G(V), G(S))) \quad (1)$$

where Cat(.) mean concatenate. The loss function of the network is:

$$\begin{aligned} L(scr, l) &= -\log\left(\frac{\exp\left(scr\left[l\right]\right)}{\sum_j \exp\left(scr\left[l\right]\right)}\right) \\ &= -scr\left[l\right] + \log\left(\sum_j \exp\left(scr\left[l\right]\right)\right) \end{aligned} \quad (2)$$

The final classification result is the jth person marked as shown in Figure 5, where $j = max\left(scr\right)$.

Implementation and Training

Each attribute network takes 30 frames (3 seconds) of the point cloud with a size of 128 points as input. In practice, the practical number of points in a cloud always less than 128. To solve the issue, we multi-copied the points a point cloud as padding points of the point cloud if the point's number of the real point cloud is less than 128.

The attribute network uses 1 layer of convolution. It is a 7×7 spatio-temporal convolution with 2×2 strides. We use batch normalization followed by ReLU activation functions

Table 2: Accuracy of each person using IWR6843. ($x$vol-sim 10(y)) means the case of x volunteers walking at the same time. 10(y) means there are 10 or y volunteers in this experiment.

People \ Method	PL	P-L	DR	Ours
1vol-sim 10(20)	36%	49%	60%	**90%(80%)**
2vol-sim 10(20)	20%	43%	51%	**86%(72%)**
3vol-sim 10(15)	28%	46%	44%	**86%(76%)**
4vol-sim 10(15)	7%	11%	21%	**57%(62%)**
5vol-sim 10	50%	15%	49%	**58%**

after the layer. We use the 3×3 max pooling with 2×2 strides. Next we use a residual blocks of ResNet18 (He et al. 2016) with 2×2 strides and 3×3 spatio-temporal convolutions. We connect the feature values of the channel modules together. The first layer is 3×3 spatio-temporal convolutions with 1×1 strides. We use batch normalization followed by the ReLU activation functions after the layer. We use the Average pooling with 2×2 strides. We use a layer of fully connected layer with 320 input to obtain classification scores. The batch size is 32. The initial value of learning rate lr is 0.01. For each 8 epoch we set $lr = lr \times 0.1$. The optimization function of the network is Adam. We implement our network in PyTorch.

Tasks and Benchmarks

We benchmark in different contexts of the data set, which contains fixed routes, free routes and different scenes.

Data Preparation. When multiple volunteers walk on a fixed route at the same time, the routes of volunteers who are walking together are inconsistent. This is reflected in the difference in the mean value of the x distribution. In order to eliminate the influence of the route on gait recognition, we normalize the point cloud of each volunteer. So that the average value of each volunteer's point cloud on x is 0. mmGait is split into training and testing set with the ratio 80% : 20%. We process the training set and the testing set into two data formats to meet the requirements of the visual data based neural network method and the neural network method which directly consume point cloud. The vision based approach requires the transformation of point clouds into 3D voxel meshes. We map each point cloud to a matrix with $20 \times 20 \times 40$. The length of the unit grid is 0.05cm. We intercept 30 consecutive point clouds in the gait sequence of human walking as a sample which is fed to the network. The point cloud based approach requires us to process point cloud into point cloud with fixed points. We extend the point cloud's points to a fixed points by copying the point cloud itself. We select 30 consecutive frames in the gait sequence as input to the point cloud based neural network.

Benchmark Algorithms

Five existing point cloud classification methods PointNet (Qi et al. 2017a), PointNet++(Qi et al. 2017b), DGCNN(Wang et al. 2018), mID (Zhao et al. 2019) and ResNet is utilized to evaluate on our data set. We identify the identity of volunteers based on the point cloud sequence generated by volunteers walking for a period of time. However, PointNet, PointNet++ and DGCNN are designed to classify static objects, so we use them to extract the feature of point cloud and then utilize LSTM to extract the time information. In the following, we introduce the five algorithms in brief.

Table 3: Accuracy of each person using two sensors.

People \ Method	PL	P-L	DR	Ours
1vol-sim 10(20)	36%	52%	33%	**86%(80%)**
2vol-sim 10(20)	32%	54%	42%	**88%(84%)**
3vol-sim 10(15)	40%	66%	37%	**90%(88%)**
4vol-sim 10(15)	28%	52%	27%	**85%(80%)**
5vol-sim 10	47%	52%	46%	**88%**

PointNet + LSTM (PL). PointNet is an architecture that is designed for 3D point cloud classification. It uses an asymmetric function to deal with the unordered points, which makes the results invariant to the permutation of the input points. PointNet uses a transformation network T-Net. T-Net carries out the arbitrary transformation of point cloud data or feature. In method PL, firstly, PointNet is used to extract features from point cloud. Secondly, the feature vector is fed to LSTM to extract time characteristics. Thirdly, the feature passes a fully connected layer to get the final classification result.

PointNet (no T-net) + LSTM (P-L). In the method P-L, we remove the T-net network from PL. The reason lies in that T-net destroys the consistency of the continuous gait point cloud sequence.

DR. In the method DR, we remove the last two residual blocks of ResNet18. It feeds the five attributes of point cloud as five channels.

PointNet++ + LSTM (P+L). To address the lack of local feature extraction and processing in PointNet, PointNet++ proposes a sample layer and grouping layer to take neighbor points into consideration. By extracting features from both raw points and grouping layers, PointNet++ can obtain abundant multi-scale features. In the method P+L, LSTM is used in the same way as PL.

DGCNN + LSTM (DGL). DGCNN is another method which consumes the point cloud directly. DGCNN proposes EdgeConv, which takes k adjacent points as graph structure to extract local features. In the method DGLSTM, LSTM is used in the same way as PL.

mID. The method utilizes bi-direction LSTM network (BiLSTM) to classify point cloud sequences. Firstly, mID maps the point cloud of the point cloud sequences to 3D voxel grid to option the body shape of volunteer. Secondly, the 3D voxel grid is flattened and converted into feature vector. Thirdly, the feature vector is fed to BiLSTM followed by a fully connected layer to get the final classification result. BiLSTM has 128 hidden units and 256 layers.

Table 4: Accuracy of each person under different attributes.

Method	noX	noY	noZ	noV	noS	Ours
Accuracy	77%	83%	83%	82%	86%	**90%**

Accuracy of Benchmark Algorithms

The gait recognition in the scene of multiple people walking at the same time is a key issue in the field of health supervision and behavior analysis in real life. Exploring the number of people walking at the same time is critical to gait recognition. Gait recognition on a fixed route is a fundamental task of gait recognition, which minimizes the impact of other factors on the results. We benchmark the data collected at the situation multiple people walking on fixed route in scene1. We first evaluate five algorithms with accuracy as the evaluation metric.

The results of several methods with better performance are shown in Table 2 and Table 3. Table 2 and Table 3 report the gait recognition performance on data collected by IWR6843 and both IWR6843 and IWR1443 separately. The experiments show two observations. *Firstly*, with the increase of the number of people who are walking at the same time, the accuracy of gait recognition declines. *Secondly*, with the increase of the number of devices, the accuracy of gait recognition increases especially when Co-existing multiple people. A further investigation on the experimental results shows the following findings. *(i)* T-Net is unsuitable for the task since P-L performs better than PL. *(ii)* Each attribute of a point cloud has its own characteristics since our method performs better than DR.

The accuracy of other evaluation methods (P+l, DGL, and mID) on our data set is only 15% on average. PointNet++ carries out down-sample to extract local features. After down-sampling, the number of points is too sparse to extract local features. The experiment shows that our method achieves the optimal performance on mmGait.

In reality, it is also possible for people to walk on any path, thus free route walking should be considered. However, free walking involves changing walking perspective at any time, which brings extra challenges to gait recognition. In our experiment, the recognition accuracy of our method over free-route scenarios is 45%, while DR's accuracy is only 33%, PL and P-L are even worse (less than 20%), possibly due to lack of ability to distill useful classification features in such setting. We leave the problem for future exploration.

Efficiency of mmGaitNet

The running speed of our method is very fast. Our method can return a gait recognition result within 1.5ms when it is running on GPU. The reason lies in that our method consumes point cloud unlike mID. mID map point cloud data to 3D space and generate a lot of extra spatial data. Hence, mID is very slow, which needs 500ms to return a gait recognition result when it is running on GPU.

Robustness of mmGaitNet

We conduct experiments in two different scenes, and the results show that the change of scenes has no effect on recognition accuracy. In particular, we evaluate our method on the data which is collected when two people are walking at the same time in two scenarios. In scene1, our method achieves an accuracy rate of 86%. In scene2, our method achieves an accuracy rate of 93%. The experiments validate that mmGaitNet's performance is resilient to environmental heterogeneity. The reason lies in that the mmWave sensor and our signal processing algorithm is able to remove static reflection points from the different furniture in different environments.

Table 5: Accuracy of each person under different input format of point cloud.

Method	DR	N32	N311	Ours
Accuracy	37%	80%	88%	**90%**

Effect of Input

In order to explore the properties of point cloud data, we design some contrast experiments. The results are shown in Table 4, which demonstrates that each attribute plays an important role in the task. In Table 5, **N32** feeds X, Y, Z as three channels into one attribute network and feed radial velocity and signal noise ratio as two channels into another attribute network. **N311** feeds X, Y, Z coordinate value into one attribute network and feeds radial velocity and signal noise ratio into another two attribute networks respectively. Table 5 illustrates that the similarity between coordinates, radial velocity, and signal noise ratio is smaller than the similarity between coordinates. What's more, the attributes are independent. They represent the different characteristics of the point cloud.

Conclusion

In this work, we build a first-of-its-kind mmWave gait data set. We evaluate multiple baseline gait recognition methods using the data set and propose a new mmWave gait recognition method mmGait. Compared with existing methods, mmGait is able to achieve much higher recognition accuracy even under multiple person co-existent scenarios. We plan to make in-depth study to improve mmWave gait recognition under more dynamic scenarios.

Acknowledgments

We would like to thank all volunteers for their involvement in gait collection. We appreciate the insightful feedback from the anonymous reviewers who helped improve this work. The work is supported in part by NSFC (61720106007,61772084, 61832010), the Funds for Creative Research Groups of China under Grant No. 61921003, the 111 Project (B18008), the Fundamental Research Funds for the Central Universities under Grant 2019XD-A13 and the OPPO research funding.

References

Adib, F., and Katabi, D. 2013. See through walls with wifi! In *ACM SIGCOMM 2013 Conference, SIGCOMM'13, Hong Kong, China, August 12-16, 2013*, 75–86.

Chao, H.; He, Y.; Zhang, J.; and Feng, J. 2019. Gaitset: Regarding gait as a set for cross-view gait recognition. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019.*, 8126–8133.

Ester, M.; Kriegel, H.; Sander, J.; and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA*, 226–231.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. 770–778.

Kuhn, H. W. 2010. The hungarian method for the assignment problem. In *50 Years of Integer Programming 1958-2008 - From the Early Years to the State-of-the-Art*. 29–47.

Li, S.; Liu, W.; Ma, H.; and Zhu, S. 2018. Beyond view transformation: Cycle-consistent global and partial perception gan for view-invariant gait recognition. In *2018 IEEE International Conference on Multimedia and Expo, ICME 2018, San Diego, CA, USA, July 23-27, 2018*, 1–6.

Li, S.; Liu, W.; and Ma, H. 2019. Attentive spatial-temporal summary networks for feature learning in irregular gait recognition. *IEEE Trans. Multimedia* 21(9):2361–2375.

Lien, J.; Gillian, N.; Karagozler, M. E.; Amihood, P.; Schwesig, C.; Olson, E.; Raja, H.; and Poupyrev, I. 2016. Soli: ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graph.* 35(4):142:1–142:19.

Makihara, Y.; Suzuki, A.; Muramatsu, D.; Li, X.; and Yagi, Y. 2017. Joint intensity and spatial metric learning for robust gait recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 6786–6796.

Pu, Q.; Gupta, S.; Gollakota, S.; and Patel, S. 2013. Whole-home gesture recognition using wireless signals. In *The 19th Annual International Conference on Mobile Computing and Networking, MobiCom'13, Miami, FL, USA, September 30 - October 04, 2013*, 27–38.

Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 77–85.

Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, 5099–5108.

Richards, M. A.; Scheer, J.; Holm, W. A.; and Melvin, W. L. 2010. *Principles of modern radar*. Citeseer.

Wang, S.; Song, J.; Lien, J.; Poupyrev, I.; and Hilliges, O. 2016. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST 2016, Tokyo, Japan, October 16-19, 2016*, 851–860.

Wang, J.; Xiong, J.; Chen, X.; Jiang, H.; Balan, R. K.; and Fang, D. 2017. Tagscan: Simultaneous target imaging and material identification with commodity RFID devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, MobiCom 2017, Snowbird, UT, USA, October 16 - 20, 2017*, 288–300.

Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2018. Dynamic graph CNN for learning on point clouds. *CoRR* abs/1801.07829.

Wang, W.; Liu, A. X.; and Shahzad, M. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 363–373.

Wei, T.; Wang, S.; Zhou, A.; and Zhang, X. 2015. Acoustic eavesdropping through wireless vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, MobiCom 2015, Paris, France, September 7-11, 2015*, 130–141.

Yang, Z.; Pathak, P. H.; Zeng, Y.; Liran, X.; and Mohapatra, P. 2016. Monitoring vital signs using millimeter wave. In *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '16, 211–220. New York, NY, USA: ACM.

Yang, Z.; Zhou, Z.; and Liu, Y. 2013. From RSSI to CSI: indoor localization via channel response. *ACM Comput. Surv.* 46(2):25:1–25:32.

Zeng, Y.; Pathak, P. H.; and Mohapatra, P. 2016. Wiwho: wifi-based person identification in smart spaces. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*, 4. IEEE Press.

Zhao, P.; Lu, C. X.; Wang, J.; Chen, C.; Wang, W.; Trigoni, N.; and Markham, A. 2019. mid: Tracking and identifying people with millimeter wave radar. In *15th International Conference on Distributed Computing in Sensor Systems, DCOSS 2019, Santorini, Greece, May 29-31, 2019*, 33–40.

Zhou, A.; Xu, S.; Wang, S.; Huang, J.; Yang, S.; Wei, T.; Zhang, X.; and Ma, H. 2019a. Robot navigation in radio beam space: Leveraging robotic intelligence for seamless mmwave network coverage. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing, Mobihoc 2019, Catania, Italy, July 2-5, 2019*, 161–170.

Zhou, A.; Yang, S.; Yang, Y.; Fan, Y.; and Ma, H. 2019b. Autonomous environment mapping using commodity millimeter-wave network device. In *2019 IEEE Conference on Computer Communications, INFOCOM 2019, Paris, France, April 29 - May 2, 2019*, 1126–1134.

Zou, H.; Zhou, Y.; Yang, J.; Gu, W.; Xie, L.; and Spanos, C. J. 2018. Wifi-based human identification via convex tensor shapelet learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*.