mID: Tracking and Identifying People with Millimeter Wave Radar

Peijun Zhao[†], Chris Xiaoxuan Lu[*,†], Jianan Wang[§], Changhao Chen[†], Wei Wang[†], Niki Trigoni[†], and Andrew Markham[†]

[†]Department of Computer Science, University of Oxford, United Kingdom
[§]DeepMind, London, United Kingdom

***Abstract*—The key to offering personalised services in smart spaces is knowing where a particular person is with a high degree of accuracy. Visual tracking is one such solution, but concerns arise around the potential leakage of raw video information and many people are not comfortable accepting cameras in their homes or workplaces. We propose a human tracking and identification system (mID) based on millimeter wave radar which has a high tracking accuracy, without being visually compromising. Unlike competing techniques based on WiFi Channel State Information (CSI), it is capable of tracking and identifying multiple people simultaneously. Using a low-cost, commercial, off-the-shelf radar, we first obtain sparse point clouds and form temporally associated trajectories. With the aid of a deep recurrent network, we identify individual users. We evaluate and demonstrate our system across a variety of scenarios, showing median position errors of 0.16 m and identification accuracy of 89% for 12 people.**

***Keywords*-millimeter wave radar, tracking, identification**

I. INTRODUCTION

Knowing 'who is where' is a key requirement for emerging applications and services in smart spaces, such as personalized heating and cooling, security management, efficiency monitoring, natural light adjustment, background music selection, etc [1]. For these and other pervasive services to be truly seamless, tracking and identification need to be performed with high accuracy and without active human effort.

Currently, most identification methods in smart spaces are device-based. Via the carried token, such as ID/swipe cards, active badge, smartphone, smartwatch, these methods identify users by the unique identifier of their personal devices. However, an implicit assumption made with these techniques is that the user and their identifying device are inseparable, which is not always the case.

In order to cope with more general scenarios, device-free methods have been proposed and are increasingly being adopted for human identification. Vision based techniques (e.g., cameras) are widely used methods in this category and have good performance when given a clear, frontal view of the face. However, cameras are intrusive and have a low user acceptance in domestic and commercial settings [2]. In contrast, radio frequency based methods are less intrusive and have also been utilized for device-free identification. For instance, it has been found that the variations in ambient WiFi signals can be used to recognize people while they walk [3], [4]. Unfortunately, such methods require a separate transmitter and receiver, and are limited to cases when users walk between the transmitter and receiver. The mmWave radar is a transceiver, so only requires a single device for tracking and identification. More importantly, existing WiFi CSI techniques are incapable of simultaneously tracking and identifying multiple people in the same scene. Point cloud based sensors, such as LiDAR and depth cameras are also able to identify and track people [5]. However, LiDAR is too expensive for home use while depth cameras only have a limited tracking range and accuracy. As an optics based sensor, they have similar user acceptance concerns as with conventional cameras.

In this paper, we introduce mID, a system that identifies people by their unique characteristics as sensed by a millimeter-wave (mmWave) radar. MMWave radar provides highly precise ranging by analyzing the reflection from obstacles in the environment, such as humans. It has a number of interesting properties. For example, a mmWave radar can be concealed behind furniture, as it is able to penetrate thin layers of different kinds of material [6], unlike optics based sensors. This property makes mmWave radar significantly more unobtrusive by concealing itself inside furniture or walls. The unobtrusive nature of mmWave radar means domestic users are more likely to accept it, list like how Amazon Echo has been more widely accepted by users than web camera [7].

Exploiting these characteristics, we developed our gait recognition pipeline based on a commercial-off-the-shelf mmWave Radar. Our device is based on a single chip solution and operates in the $77-81$ GHz band. To the best of our knowledge, this is the first work to use the point cloud generated by a mmWave radar to track and identify people while they are walking.

The contributions of this work are as follows:

- We designed and implemented a human tracking and identification system using mmWave radar, which is capable of providing highly accurate tracking and identification in multi-person scenarios.
- We are the first to identify people from mmWave radar

*Chris Xiaoxuan Lu is the corresponding author.

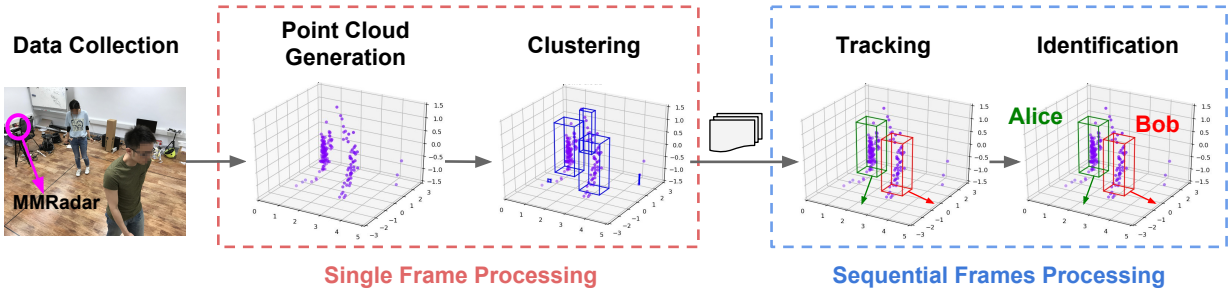


Figure 1: System Overview

point clouds, using deep recurrent neural networks.

- We evaluated the tracking and identification ability of our system, demonstrating median tracking accuracy of $0.16m$ and identification accuracy of 89% for 12 people.

The roadmap of this paper is: the basic principles behind mmWave radar are introduced in Section 2; Section 3 describes mID, which is our proposed mmWave gait identification system; Section 4 introduces the implementation details of mID and the setup of our testbed; Section 5 evaluates the system and Section 6 discusses the limitations of mID. Related work is given in Section 7 and Section 8 concludes the paper.

II. Introduction to mmWave Radar Background

MMWave radar is based on the principle of frequency modulated continuous wave (FMCW) radar. FMCW radar has the ability to simultaneously measure the range and relative radial speed of the target. Detailed principles of FMCW radar are briefly introduced below[1].

A. Range Fourier Transform (range-FFT)

FMCW radar uses a linear 'chirp' or swept frequency transmission. The chirp is characterized by a start frequency f_c, bandwidth B and duration T_c. When receiving the reflected signal, the radar front-end computes the frequency difference between the transmitter and the receiver with a mixer, which produces an Intermediate Frequency (IF) signal, from which the distance between the object and the radar can be calculated as:

$$d = \frac{f_{IF} c}{2S} \tag{1}$$

where c represents the light speed $3 \times 10^8 m/s$, f_{IF} is the frequency of the IF signal, and S is the frequency slope of the chirp, which is calculated by B/T_c. To detect objects at different ranges, we perform an FFT on the IF signal, and each peak represents an obstacle at a corresponding distance. This is called 'range-FFT'.

[1]Please refer to https://training.ti.com/mmwave-training-series for more information.

B. Doppler Fourier Transform (Doppler-FFT)

A small change in the distance of the object leads to a large shift in the IF signal phase, so we can obtain the relative velocity of the detected object by transmitting two chirps with an interval of T_c and measuring the phase difference ω

$$v = \frac{\lambda \omega}{4\pi T_c} \tag{2}$$

where λ is the wave length.

Using this technique, objects moving a different velocities at the same distance can be distinguished from one another.

C. Angle Estimation

Transmitters emit chirps with the same initial phase. With simultaneous sampling from multiple receiver antennas, we can estimate the Angle of Arrival (AoA), due to slight differences in phase of the received signals. For two antennas, the AoA can be calculated with

$$\theta = sin^{-1}(\frac{\lambda \omega}{2\pi d}) \tag{3}$$

The final AoA can be calculated with the average result from different receiver pairs. The estimation is most accurate at $\theta = 0$ and decreases with $|\theta|$.

III. System Design

mID is a tracking and identification system that exploits the unique properties of millimeter wave radar. It operates by transmitting an RF signal and recording its reflections off objects. By analyzing the point cloud generated, it then infers the people's trajectories and identifies them from a database of known users. The mID system consists of four modules that operate in a pipelined fashion, as shown in Fig. 1:

1) *Point Cloud Generation*. In this module, a FMCW radar transmits millimeter waves and records the reflections from the scene. It then computes sparse point clouds and removes those points corresponding to static objects (i.e., points that appeared in the previous frame).

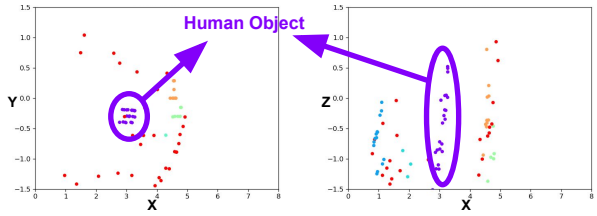


Figure 2: 2D Projections of a 3D Point Cloud (see Sec. III-A). Purple points corresponding to a human object are compact in the horizontal x-y plane but dispersed in the vertical z-axis.

2) *Point Cloud Clustering*. In this module, potential human objects are detected by merging individual points into clusters.
3) *Tracking*. In this module, mID associates the same human object in consecutive frames and uses a multiple object tracking algorithm to maintain trajectories of different people.
4) *Identification*. In this module, a recurrent neural network is used to recognize user identities from the sequential data of each user.

We refer the reader to a video based demonstration of mID for simultaneously tracking and identifying two people[2].

In the following subsections, we describe each of these components in detail.

A. Point Cloud Clustering

The generated sparse point clouds are dispersed and not informative enough to detect distinct objects. Moreover, although static objects are discarded through clutter removal, the remaining points are not necessarily all reflected by moving people. As shown in Fig. 1, this noise can be significant and lead to confusion with points from nearby people. To determine which points in the scene are caused by reflections from people, mID first merges points into clusters using DBScan, a density-aware clustering method that separates cloud points based on the distance in the 3D space. A major advantage is that it does not require the number of clusters to be specified *a priori*, as in our case people walk in and fade out of the monitored scene at random. Additionally, DBScan can automatically mark outliers to cope with noise.

However, in a real-world measurement study, we observed that points of the same person are coherent in the horizontal (x-y) plane, but more scattered and difficult to merge along the vertical (z) axis. Fig. 2 illustrates an example. We hence modify the Euclidean distance to place less weight on the contribution from the vertical z-axis in clustering:

$$D(p^i, p^j) = (p_x^i - p_x^j)^2 + (p_y^i - p_y^j)^2 + \alpha * (p_z^i - p_z^j)^2 \quad (4)$$

[2]https://www.youtube.com/watch?v=3m84xZo6E_A

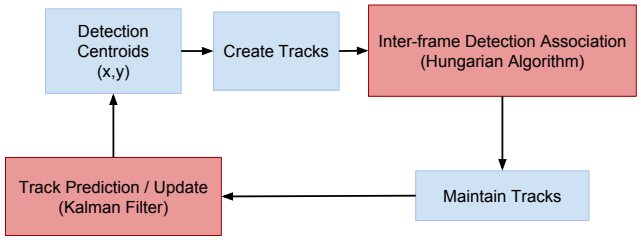


Figure 3: Workflow of moving object tracking in Sec. III-B.

where p^i and p^j are two different points and the parameter α regulates the contribution of vertical distance.

B. Moving Object Tracking

To capture continuous individual point clouds to track and identify a person, we require an effective temporal association of detections as well as correction and prediction of sensor noise. Fig. 3 illustrates the flow of our multi-object tracker. We essentially create and maintain tracks for object detections from each frame. A new track is created for each object detection which is either from the first incoming frame or one which cannot be associated with an existing track. Inter-frame object association is based on the Hungarian Algorithm. If a track object is undetected for D continuous frames, we mark the track as inactive and exclude it from successive associations. Finally, we apply a Kalman Filter to predict and correct tracks. These two components are discussed in more detail below.

1) Detection and Association: We use the Hungarian Algorithm which is an effective combinatorial optimization algorithm. Our objective is to create an association between each object detection and maintained track objects so that the combined distance loss is minimized. Here we are facing a many-to-many assignment problem where the cost matrix can be non-square because the number of active tracks K_1 and the number of object detections at the current timestamp K_2 can be different. Given K as the greater of K_1 and K_2, we essentially augment the true cost matrix with dummy entries to construct a $K \times K$ matrix M where $M_{i,j}$ represents the distance of centers between track object i and object detection j in current frame. If $M_{i,j}$ exceeds step size threshold θ we set the cost to be a large number L to avoid association given the intuition that j should be a joining person. If a detection is mapped to an augmented dimension, or if it is mapped to a correspondence with cost L, we ignore such mappings and create a new track for this detection. Similarly, if a track object is mapped to an augmented dimension or a correspondence with cost L, we treat the track object as undetected. This method enables us to successfully maintain tracks of detections.

2) Track prediction and correction : We use a Kalman Filter to correct for sensor noise and to offer predictive guidance in scenarios where tracked objects are undetected

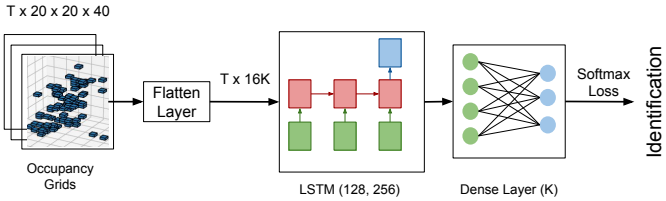
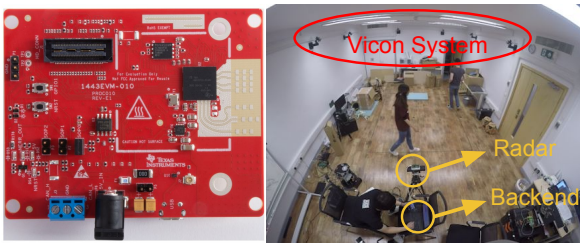


Figure 4: Classification Network Structure (see Sec. III-C). T represents the number of data frames used for identification and K represents the number of people to distinguish. Numbers in bracket represent layer sizes.

due to occlusion or temporary loss from the sensing region. For each track we maintain a state which consists of location and velocity along the x and y axes. For each track the initial state consists of the first detection location and velocity. At each successive time step, the Kalman Filter updates the current state variables with transition matrix along with corresponding uncertainties. Based on the current position and velocity, it estimates a new position/velocity as well as new covariance. The Kalman Filter produces estimates that tend to be more accurate than those based on a single sensor measurement, especially in our case with occasional undetected track objects.

C. User Identification

After the points corresponding to human objects are determined, we can use tracklets to recognize their identities. Specifically, from each frame in the trajectory, we use a fixed-size bounding box to enclose the points of potential human objects, and voxelize it to form an occupancy grid. Note that the occupancy grids inherently encapsulate body shape information. For instance, tall people tend to have higher center of mass. By feeding the sequential occupancy grids to a classifier, the ID of a tracklet is recognized based on both movement characteristics, i.e., gait, and body shape information. The tracklet used in mID is segmented with a sliding window method. A window contains consecutive occupancy grids for 2 seconds, with a 75% overlapping ratio with the previous window. Extracting useful features directly from the occupancy grids is difficult, as most feature engineering methods are not effective for point cloud classification tasks [8]. The Long-short Term Memory (LSTM) network is an established recurrent neural network architecture suited for sequential data classification which is able to learn the features automatically through network training. We therefore propose using it as the identity classifier in mID. The 3-D data is first flattened and then each frame is converted into a feature vector. This is then passed into a bi-directional LSTM network followed by a dense layer. Lastly, a softmax layer is used to output the final classification result (see Fig. 4).



IWR boost sensor Experiment Setting

Figure 5: Experiment Settings. Vicon system is a high-precision tracking system used for acquiring ground truth trajectories.

IV. IMPLEMENTATION

A. Testbed Setup

mID is developed on top of a commercial, off-the-shelf millimeter wave radar, IWR1443boost[3]. The system was tested in a room with a Vicon optical tracking system which is able to provide ground-truth position of each marker with an error within $1cm$. Ground truth identities of the training and testing samples are manually labeled. The system consists of two parts, the radar and the backend. The radar senses data and generate 3D point cloud, which is transferred to the backend computer for further processing, as shown in Fig. 5. We implemented the deep neural network classifier with the Keras library and a Tensorflow backend.

B. mID Configuration

1) Sensor Setup: The IWR1443Boost sensor was configured to use all three transmitter antennas and all four receiver antennas in order to generate 3D point cloud data. Start f_c and end frequencies were set to $77GHz$ and $81GHz$ respectively, so the bandwidth B was $4GHz$. The Chirp Cycle Time T_c was set to $162.14\mu s$ and the Frequency Slope S was set to be $70GHz/ms$. With such a configuration, mID has a range resolution of $4.4cm$ and maximum unambiguous range of $5m$. In terms of velocity, it can measure a maximum radial velocity of $2m/s$, with the resolution of $0.26m/s$. The sensor was set to transmit 128 chirps per frame and the number of frames per second was 33.

2) Classifier Training: Through multiple trials we worked out a set of parameters that have the best performance. Each frame of the input data was first flattened to a vector of dimension 16000. A bi-directional LSTM with size 256 and 128 hidden units was used. We set the dropout ratio to 0.5 and used the Adam optimizer. We used a balanced dataset where training/test sample ratio was set to 11:1. To decrease overfitting, we further augmented the training data to 8 times the original size, by shifting the data in X and Y axis respectively for 1 voxel, and rotating each frame by 90°, 180° and 270°. The model was trained for 30 epochs.

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

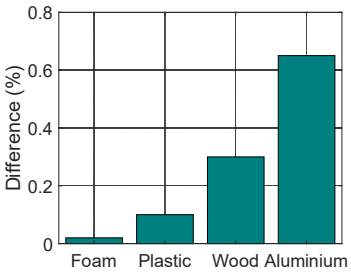


Figure 6: Impact of different materials on point cloud density of mmWave radar. The differences in point cloud density are all below 1% for all cases.

3) Parameters of DBScan Algorithm: DBScan has two parameters, namely Eps which indicates the maximum distance of two points in the same cluster and MinPts which indicates the minimum point number in a cluster to cope with noise points. In practice, we choose 0.05 as Eps and 20 as MinPts. α was set to 0.25 in the customized distance function.

V. Evaluation

A. Sensitivity Analysis

1) Non-Line-of-Sight Conditions: In the first experiment, we study the robustness of mmWave radar under occluded conditions. This property is important because optical imaging based tracking and identification methods, such as RGB and depth cameras, cannot cope with obstructions. We evaluate the robustness of millimeter wave radar with four types of obstructions: foam, plastic wood and aluminium. We used a sheet of each material in turn with a thickness of approximately $3mm$ and a size of 10^5 mm^2. The obstacles are placed $1cm$ away from the sensor so that the signals cannot be transmitted in a line-of-sight condition. We let a user walk back and forth in front of the millimeter wave radar while collecting sensor readings. mID uses the generated 3D point cloud for tracking and identification, so we compare the percentage of change in point cloud density for mID. As can be seen in Fig. 6, mmWave is very robust against non-line-of-sight interference, with less than 1% change in point-cloud density. Robustness to thin obstructions could enable mID to work under furniture or concealed within a picture frame, etc., making it less intrusive.

2) Impact of weighting the vertical axis in DBScan: As introduced in Section 3, it is important to define the weighting parameter to make the DBScan algorithm work better. Therefore, we need to find a suitable value of α to obtain a good clustering result. In practice, we found that $\alpha = 0.25$ results in good clustering performance, as shown in Fig. 7. In contrast, when $\alpha = 0$, (points are effectively projected onto the $x-y$ plane), outliers are merged into the cluster. When $\alpha = 1$, (standard Euclidean distance) points corresponding to a person are split into two clusters.

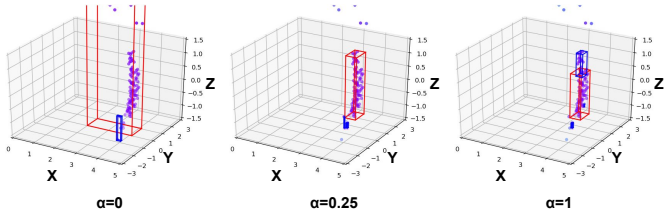
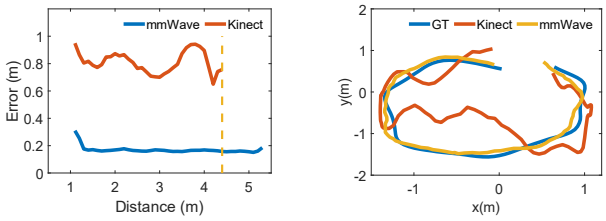


Figure 7: Clustering results with different α. A small α leads to loose clusters containing many noisy points. A large α splits a human object into two clusters. Setting $\alpha = 0.25$ gives the best empirical performance, as shown in the middle one.



(a) Tracking Error (b) Inferred trajectory

Figure 8: Comparison of tracking performance between mmWave radar and Kinect v2. (a) Kinect v2 can only track people within $4.5m$. (b) Inferred trajectory. (GT: Ground Truth by Vicon)

B. User Tracking

To evaluate the tracking accuracy of mmWave radar, we compared it to a Kinect v2, an RGB-D camera which is widely used in homes for gaming. We installed the mmWave radar and Kinect v2 co-located in a room. The Vicon system was set to track a marker placed on the top of a hat worn by the participant. Time was synchronized through a NTP server and the coordinate transformation matrices of the systems were calibrated. A state-of-the-art open source tracking algorithm[4] is used for Kinect v2 people tracking. The tracking error and a comparison of the trajectories is shown in Fig. 8, which shows the median tracking error of mID is $\sim 0.16m$ whereas the Kinect is $0.9m$. Besides a significantly smaller error compared to Kinect v2, the tracking range of mID is also larger than Kinect v2. mID can track people at a distance of more than $5.5m$, whereas Kinect has a tracking limitation of $4.5m$. This could be further extended, at the cost of a reduction in identification accuracy. This shows that the radar based technique is able to achieve highly accurate tracking over a larger tracking area, making it suitable for increasing levels of automation.

[4]based on https://github.com/mcgi5sr2/kinect2_tracker

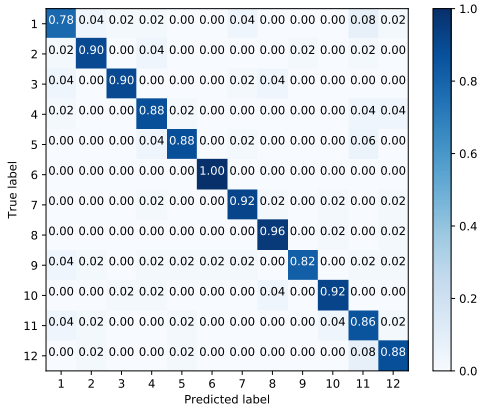


Figure 9: Confusion Matrix of 12 Users.

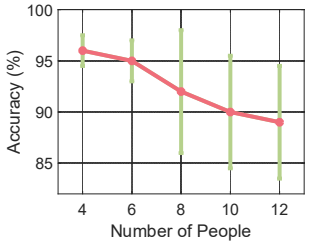


Figure 10: Impact of number of people. mID works reliably with different numbers of participants and performs best when the group size is less than 6, e.g., in home scenarios.

C. Identification Evaluation

We now examine mID with an identification task of 12 participants [5]. The ages of the participants range from 22 to 35, and 3 of the participants are female. The heights of the participants range from 155cm to 188cm, and the weights of the participants range from 55kg to 80kg. The body shape of each participant varies which mimics domestic settings. To mimic real-world complexity, we asked participants to walk on a random trajectory in our testbed for 10 minutes. Note that a number of WiFi CSI identification papers require the user to walk along a predefined trajectory [9].

The point cloud generated by mmWave radar makes it impossible to analyse people's gait with traditional vision based gait recognition methods [10], [11], because it is too sparse to recognize different body parts of the subject. Deep Neural Networks have the ability to automatically extract relevant features of the data while training the model, but it is not straightforward to tell which neural network architecture best suits the problem. We evaluated the performances of neural networks of different architectures and sizes on our dataset by performing an ablation study, which will be discussed in turn below.

1) Identification Performance: Overall, mID is able to reach an accuracy of 89% for 12 people. The confusion matrix is shown in Fig. 9. Note that this performance is non-trivial considering that the original point clouds are very sparse, and demonstrates the utility of using deep-networks for feature extraction.

2) Impact of number of people: In this experiment, we further explore identification accuracy with varying group sizes. Intuitively, a smaller group size should make the problem easier. Fig. 10 shows the performance trend of mID when varying the number of participants from 4 to 12 with a step of 2. As we can see, mID is able to cope with various scenarios and works extremely well in the cases with ≤ 6 people with 95% accuracy. In contrast, WiWho [9], based on WiFi CSI, achieves an identification accuracy of 80% with 6 people. Note that, the number of people in domestic settings is generally less than 6, and we envision that mID could greatly benefit many applications in smart homes of the future.

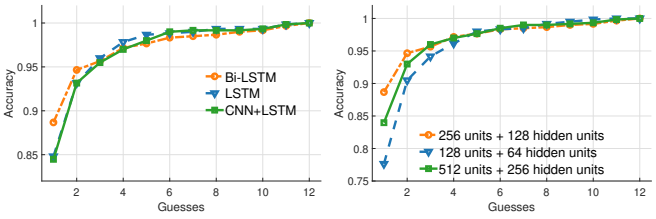
3) Neural Network Architecture Comparison: To explore the best artificial neural network architecture for the identification problem, we compare 3 different architectures that are based on LSTM. We use variants of LSTM as it has been shown to be performant for end-to-end learning of time sequence data. Each model is trained under the same settings: 30 epochs and 0.5 dropout ratio. The LSTM layer used in all the 3 models has the same size of 256 and 128 hidden units. The CNN used in the CNN+LSTM model has two convolution layers, with a max pooling layer after each convolution layer. The CNN is time distributed, which means the data of each frame is first sent into a two-layer 3D CNN for feature extraction, then the sequence data was sent into LSTM for classification.

The accumulative identification accuracy of the different architectures is shown in Fig. 11b. The accuracy differences between architectures become negligible after 3 guesses. However, Bi-directional LSTM converges more quickly and significantly outperforms the other two architectures within fewer guesses. This is presumably because Bi-directional LSTM is able to model the rich temporal correlations in a long sequence of frames from both ends. In contrast, a standard LSTM is essentially a feed-forward network that is difficult to encode the information from the beginning of a long sequence [12]. Such information loss degrades the identity inference performance.

4) Impact of Network Size: Beyond architecture, it is also important to consider the impact of the network size. If the network is too small, it may lack the representation ability, while networks that are too large can suffer from over-fitting problems. We evaluated the Bi-directional LSTM network with 3 different sizes and the comparison of the classification performance is shown in Fig. 11a.

The Bi-directional LSTM with size of 256 and 128 hidden units significantly outperforms the same network

[5]The study has received ethical approval SSD/CUREC1A CS_C1A_18_024



(a) Different Network Architectures

(b) Different Network Sizes

Figure 11: Impact of different network choices.

architecture with a larger or smaller size, especially in the first two guesses. This suggests that a Bi-directional LSTM with size of 256 and 128 hidden units is a good fit for this classification problem.

VI. Discussion and Limitations

Despite the fact that this novel system works well for tracking and identification. We elaborate the limitations.

1) *Large Number of Users.* We have demonstrated the reliable performance of mID when working with a relatively small group of users (12). However, tracking and identifying a large number of people remains an open problem for two reasons. First, the point cloud generated by mmWave is sparse and sometimes the sparsity could significantly disturb human detection and tracking. Secondly, body shape and human gaits used in this work are weak biometrics, and could become ambiguous with the increasing number of users.
2) *Monitoring Range.* In our experiment setting, we set the maximum unambiguous range of mID to $5m$, which is roughly $3/4$ the size of monitored room (see Sec. IV). In principle, the range of the mmWave radar can be as large as 30 m, but this increased range comes at the cost of reduced spatial precision and worse signal-to-noise ratio. If the subject is too far away from the sensor, it is very hard to detect and distinguish them from background noise.
3) *Flat and Planar Surfaces.* As we found in the evaluation, the reflection profile of mmWave can be affected by flat and planar surfaces, such as windows. As a result, noisy 'mirror' human objects appear occasionally when encountering these surfaces. Our experimental site is mainly with walls, which are typically not strong reflectors due to their dielectric properties. However, it is worthwhile to consider the impact of disturbing surfaces in real world deployment.

VII. Related Work

A. Device-Free Gait Recognition

There has been a lot of works on device-free gait recognition. Middleton et al. built a floor sensor based gait recognition system and achieved 80% accuracy over 15 people [13], at high deployment cost. Vision-based methods are one of the most established techniques. Some use silhouette analysis approaches like [11], others use model-based methods like in [14]. Besides monocular vision, multiple cameras, stereo cameras and depth cameras are also utilized in gait recognition tasks [15]. However, as long as video data is used, there will be a risk that users' privacy would be compromised if data leakage occurs.

It has been shown that the Channel State Information (CSI) of WiFi signals, for both $2.4GHz$ and $5GHz$ bands, captures human gait information to a certain extent [4], [9]. These gait recognition systems are easy to deploy as WiFi devices are common in daily lives. However, such methods are generally scene-dependent and cannot handle environmental dynamics very well. Furthermore, these methods struggles to cope with identifying multiple people in the same scene.

A full comparison of different device-free identification methods are provided in Table. I.

B. Other RF-based Human-centric Applications

RF sensing has been widely used in human-centric applications, such as fall detection [16], occupancy counting [17], breathing monitoring [18] etc. Google has recently published a touchless gesture interface based on mmWave radar, which recognizes hand gesture in high fidelity [19]. Zhao et al. have recently proposed a human pose reconstruction approach with a customized RF radar. By using the pose extracted from the collocated cameras as supervision signals, their reconstruction network could learn to estimate human skeleton, in both 2D and 3D scenes [20], [21].

VIII. Conclusion

In this paper, we propose mID, a highly accurate tracking and identification system for smart spaces based on millimeter wave radar. With the aid of a commercial-off-the-shelf millimeter wave FMCW radar, we first obtain sparse point clouds. Then, we extract the point clouds representing human objects and associate them to their historical trajectories. Based on the tracklets, a recurrent neural network is used to recognize their identities. Extensive experimental results show that mID achieves an overall recognition accuracy of 89%, and with 12 people in identification with approximately $0.16m$ positioning error. We also demonstrate the ability to simultaneously track and identify multiple people. We envision mID as a promising step towards smart home human identification and tracking, for the sparse point clouds adopted in mID are not themselves as privacy sensitive as vision based techniques, and mID can be concealed inside furniture or walls, which can be highly unobtrusive and gain acceptance by smart home users.

Table I: Comparison of different identification methods and their relative merits.

	Identification Accuracy	Multiple People	Tracking	Environment Independent	Privacy Concerns	Ease of Deployment
Floor Sensor	Moderate	Yes	Yes	No	None	Very Difficult
RGB Camera	Very high	Yes	Yes	Yes	High	Easy
Depth Camera	High	Yes	Yes	Yes	Medium	Easy
WiFi CSI	Moderate	No	Yes	No	Low	Difficult
mID	Moderate	Yes	Yes	Yes	Low	Moderate

REFERENCES

[1] C. X. Lu, H. Wen, S. Wang, A. Markham, and N. Trigoni, "SCAN: learning speaker identity From noisy sensor data," in *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*. ACM, 2017, pp. 67–78.

[2] R. Beringer, A. Sixsmith, M. Campo, J. Brown, and R. McCloskey, "The acceptance of ambient assisted living: Developing an alternate methodology to this limited research lens," in *International Conference on Smart Homes and Health Telematics*. Springer, 2011, pp. 161–167.

[3] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 363–373.

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

[5] C. Benedek, B. Nagy, B. Gálai, and Z. Jankó, "Lidar-based gait analysis in people tracking and 4d visualization," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*, 2015.

[6] D. D. Ferris and N. C. Currie, "Microwave and millimeter-wave systems for wall penetration," in *Targets and Backgrounds: Characterization and Representation IV*, vol. 3375. International Society for Optics and Photonics, 1998, pp. 269–280.

[7] P. Kowalczuk, "Consumer acceptance of smart speakers: a mixed methods approach," *Journal of Research in Interactive Marketing*, vol. 12, no. 4, pp. 418–431, 2018.

[8] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, vol. 1, no. 2, p. 4, 2017.

[9] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: wifi-based person identification in smart spaces," in *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*, 2016.

[10] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002.* IEEE, 2002, pp. 155–162.

[11] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003.

[12] C. X. Lu, B. Du, H. Wen, S. Wang, A. Markham, I. Martinovic, Y. Shen, and N. Trigoni, "Snoopy: Sniffing your smartwatch passwords via deep sequence learning," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, p. 152, 2018.

[13] L. Middleton, A. A. Buss, A. Bazin, and M. S. Nixon, "A floor sensor system for gait recognition," in *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*, 2005.

[14] F. Tafazzoli and R. Safabakhsh, "Model-based human gait recognition using leg and arm movements," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 8, pp. 1237 – 1246, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0952197610001417

[15] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien, "Gait recognition with kinect," in *1st international workshop on kinect in pervasive computing*. New Castle, UK, 2012, pp. P1–P4.

[16] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71–80, 2016.

[17] X. Lu, H. Wen, H. Zou, H. Jiang, L. Xie, and N. Trigoni, "Robust occupancy inference with commodity wifi," in *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2016, pp. 1–8.

[18] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 2015, pp. 837–846.

[19] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, p. 142, 2016.

[20] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7356–7365.

[21] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "Rf-based 3d skeletons," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. ACM, 2018, pp. 267–281.