2024 IEEE 30th International Conference on Parallel and Distributed Systems (ICPADS) | 979-8-3315-1596-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICPADS63350.2024.00014

MMTAI: Biometrics-assisted Multi-person Tracking with mmWave Radar

Yande Chen[1], Yuan He[1*] , Yimiao Sun[1], Awais Ahmad Siddiqi[1], Jia Zhang[1], Xiuzhen Guo[2]
[1]School of Software and BNRist, Tsinghua University, Beijing, China
[2]College of Control Science and Engineering, Zhejiang University, Hangzhou, China
Email: cyd22@mails.tsinghua.edu.cn, heyuan@tsinghua.edu.cn,
{sym21, xidq19, j-zhang19}@mails.tsinghua.edu.cn, guoxz@zju.edu.cn

***Abstract*—mmWave-based human tracking is a key enabling technology for smart applications. Most of the existing works on this topic employ the conventional approach of device-free object localization, which treat any person as a general moving target rather than distinguish different persons. As a result, the existing approaches have poor performance in the scenarios of multi-person tracking, especially when there are crossovers among different persons' trajectories. This paper presents MMTAI, a novel approach for multi-person tracking with a mmWave radar. By exploiting mmWave sensing to capture a human's biometric features, MMTAI augments mmWave radar based human tracking with the ability of identifying different persons. Specifically, MMTAI is able to sense persons' scalp responses to the signals and their head-shoulder distances, which are then continuously mapped to their trajectories using a bipartite matching algorithm. We implement MMTAI with a commercial mmWave radar and evaluate its performance under various settings. The results show that in the multi-person tracking scenarios, MMTAI has a median tracking error of 12.33 cm, which is 35.88% lower than that of the state-of-the-art approach.**

***Index Terms*—multi-person tracking, millimeter wave radar, human identification**

I. Introduction

Location information is crucial for many indoor and outdoor applications [1]–[4]. Since most people's activities are indoors, indoor human tracking has many attractive applications such as smart home [5], elderly care [6] and so on [7]–[9]. Moreover, this technique can be extended to multi-person tracking, providing much more practicability. Conventional multi-person tracking approaches like mmTrack [10] and m^{3}Track [11] achieve high accuracy in most cases, yet they are prone to trajectory crossovers which may hinder the operation of such approaches.

The root cause of this problem is the lack of information to identify persons after the crossovers. If the persons are differentiable, they can be re-associated with their past trajectories, and the tracking solution can work smoothly with crossovers. Recent efforts have explored different biometric features to distinguish persons, including persons' gait [12] and skeleton [1]. These methods are effective in differentiating persons, yet they rely on exposure of the lower half of the body to the radar, which is likely to be blocked by indoor blockages in complex indoor scenarios.

*corresponding author

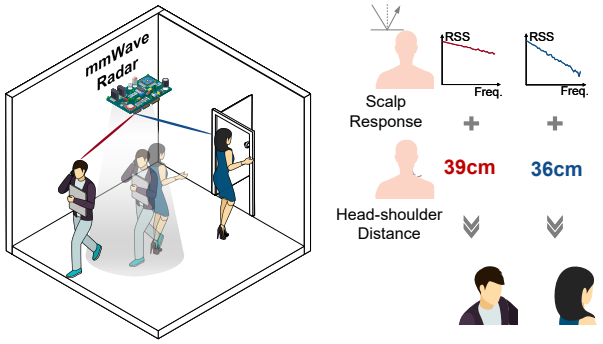

Fig. 1: The application scenario of MMTAI.

Based on the above discussion, one may find that a reliable and accurate approach for indoor multi-person tracking is still missing. Inspired by the research progress on material identification in recent years [13], [14] and the high spatial resolution of mmWave radar, we propose MMTAI[1], a mmWave-based technique which tracks multiple persons more resiliently and accurately by extracting their unique biometric features. The high-level idea of MMTAI is illustrated in Fig. 1. As persons walk in the monitoring area of the mmWave radar, MMTAI extracts the following two distinct biometric features of each person from the reflected signals: the *scalp response* and the *head-shoulder distance*, which help to match the persons before and after the crossover.

However, it's non-trivial to put this high-level idea into practice, where several critical challenges in design should be tackled, which are summarized as follows:

Subtle Biometric Features. Generally, there is only a 2-cm difference among persons' head-shoulder distance, smaller than the typical range resolution of mmWave radar. What's worse, the impact of scalp characteristics on the received signal is much smaller than that of the head-radar distance. To differentiate persons with these biometric features, MMTAI improves the accuracy of biometric features measurement with a novel usage of the Multi-Signal Consolidation technique [15]. Specifically, a group of chirps with different starting

[1]MMTAI is the abbreviation of mmWave-based Tracking Assisted by Identification.

979-8-3315-1596-6/24/$31.00 ©2024 IEEE
DOI 10.1109/ICPADS63350.2024.00014

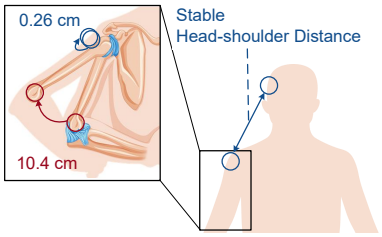

Fig. 2: Illustration of a typical arm movement.

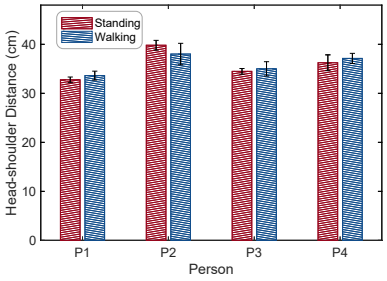

Fig. 3: Estimated head-shoulder distances of four persons.

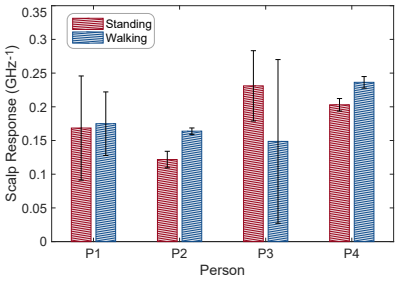

Fig. 4: Estimated scalp responses of four persons.

frequencies are applied, and the change in received signal strength (RSS) and phase with the starting frequency implies the two biometric features respectively.

Dynamic Interference on Measured Biometric Features. The position, posture and mutual interference of persons are unpredictable and affecting the biometric features. For instance, the measurement of the scalp response is interfered with multipath signals which vary across different positions. Fortunately, the effect is negligible during a short time. Therefore, MMTAI exploits the bipartite graph to match persons between adjacent time instants, and then re-identifies persons after a short-time crossover based on the biometric features.

In this paper, we make the following contributions:

- We propose MMTAI, a novel technique to achieve resilient and accurate indoor multi-person tracking. At the core of MMTAI is the exploration of persons' unique biometric features, the scalp response and the head-shoulder distance, to re-identify persons after crossovers or other forms of trajectory breaks.
- MMTAI has a tailored design to deal with the challenges in biometrics-assisted multi-person tracking, including utilizing the Multi-Signal Consolidation technique to distinguish subtle biometric features and employing the bipartite graph to reconstruct the trajectories.
- We implement MMTAI using a commercial mmWave radar and evaluate its performance in various settings. The results show that in the multi-person tracking scenarios, MMTAI has a median tracking error of 12.33 cm, which is 35.88% lower than that of the state-of-the-art approach.

The rest of this paper is organized as follows: Sec. II introduces the biometric features we use and unfolds their feasibility. Then, Sec. III elaborates on the detailed design of MMTAI. The implementation and evaluation results of MMTAI are presented in Sec. IV. We discuss practical issues in Sec. V and summarize related works in Sec. VI. This work is concluded in Sec. VII.

II. Primer of Biometric Features Used in MMTAI

This section introduces biometric features utilized in MMTAI and their feasibility.

A. Principles of Biometric Feature Selection

The biometric feature are designed to distinguish different persons in the application scenario of MMTAI. Based on the application, a biometric feature should have:

1) Accessibility. The body parts regarding the biometric feature should be visible to the radar.

2) Diversity among people. Great diversity of the biometric feature among people makes it easy to distinguish them.

3) Short-time invariance. During the moving period of multiple persons, the biometric feature and its measured value should be stable for each person.

Although the commonly used biometric features including the skeleton, gait and vital signs are highly differentiating and invariant, they are inaccessible in some scenarios. To maximize the accessibility and maintain acceptable diversity and invariance, MMTAI employs the head-shoulder distance and the scalp response. In the following section, we assess their feasibility in greater detail.

B. The Head-Shoulder Distance

The head-shoulder distance is highly accessible to the radar, diverse among people and nearly invariant with their activities.

There is likely to be a line-of-sight (LoS) path between the head/shoulders and the radar, especially when the radar is deployed on the ceiling. Besides, the head-shoulder distance typically spans from 30 cm to 43 cm across individuals with a standard variance of 2 cm. Moreover, the head-shoulder distance serves as a relatively stable metric. Although the head movement is highly flexible, it follows the rules below in most cases: 1) the neck only tilts front and back, not left and right; 2) both side of the neck does not stretch or compress significantly; 3) the head rotates freely. Under these rules, the absolute position of the reflector on the head is nearly constant unless the person is right under the radar. Similarly, the reflection point on the shoulder exhibits minimal movement in absolute coordinates as shown in Fig. 2.

Therefore, the height of the head, the height of the shoulder, the shoulder-shoulder distance and the head-shoulder distance are all short-time invariant features. As discussed later in Sec. III-C, the height of each reflector is more error-prone than the distance between two reflectors; besides, one of the shoulders

may be invisible to the radar. Consequently, we take the head-shoulder distance as one of the biometric features in MMTAI.

Leveraging the mm-level ranging technique discussed in Sec. III-B, it becomes feasible to distinguish between individuals based on their unique head-shoulder distances. Experimental results in Fig. 3 also validate this feature.

C. The Scalp Response

The head-shoulder distance is highly accessible to the radar, diverse among people and nearly invariant with their activities. The reflectivity of the human scalp is approximately linearly associated with the frequency of the mmWave signal, and the linear coefficient (named the scalp response) is a distinct feature among people as shown in Fig. 4. Although the hairstyle and clothing of the scalp affect the scalp response, those factors can be assumed invariant within a short period.

The reflectivity $|H_r(f)|$ is determined by the thickness and tissue of human skin. The skin can be simplified as a single-layer structure, as the major portion of the reflected signal originates from the epidermis, the top layer of human skin [16], [17]. Furthermore, the reflectivity can be derived using Fresnel's Equation:

$$|H_r(f)| = |\frac{\sqrt{\varepsilon_a f + \varepsilon_b} - 1}{\sqrt{\varepsilon_a f + \varepsilon_b} + 1}| \,, \tag{1}$$

where $\varepsilon_a f + \varepsilon_b$ representing the complex dielectric permittivity of the epidermis at frequency f.The scalp response can be defined as the RSS ratio of signals at different frequencies.

III. Design

A. Overview

MMTAI comprises three sequential modules: human detection, biometric feature measurement, and multi-person tracking, as depicted in Fig. 5.

Human Detection Module identifies all persons from the raw mmWave signal. After removing background reflections, the reflectors are localized accurately with the proposed ranging technique and grouped by their corresponding person.

Biometric Feature Measurement Module provides estimations for both the head-shoulder distance and the scalp response of each person. After detecting the head and shoulders, each person's head-shoulder distance and scalp response are estimated by analyzing the RSS and received phase.

Multi-person Tracking Module matches the persons at the current time instant with those at the previous time instant. By continuously running this module, a trajectory of each person can be reconstructed. Furthermore, this module accounts for potential trajectory breaks caused by joining, leaving, blockage, and crossover with human re-identification mechanics.

B. Human Detection

To obtain the 3D position of each reflector on each human body, the human detection module involves the stages below.

Signal Preprocessing. This stage removes the background reflections and decomposes the raw I/Q signal regarding the range, azimuth angle and elevation angle. The composition

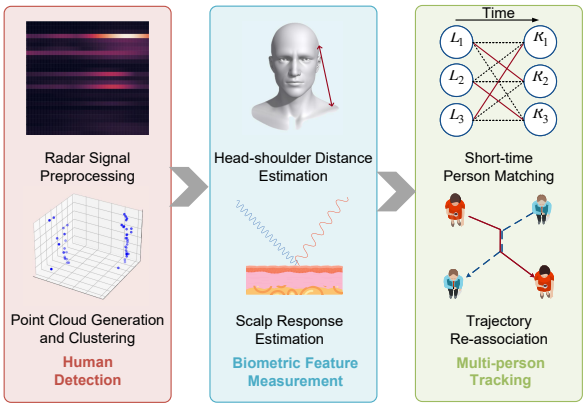

Fig. 5: Overview of MMTAI.

of all background reflections can be estimated by averaging the raw signal over the measuring period and then removed by subtracting the average from the raw signal. Afterward, the reflected I/Q signal components from different ranges are separated with a range-FFT operation. The i-th range bin contains the reflected signals $\left[\frac{(i-1/2)c}{2B}, \frac{(i+1/2)c}{2B}\right]$ away from the radar, where c denotes the speed of light and B denotes the bandwidth of the FMCW signal. Subsequently, the robust Capon beamforming technique is applied to estimate the distribution of the received signal strength and phase over the azimuth and elevation angles for each range bin.

Point Cloud Generation. This stage identifies the reflectors and estimates their 3D coordinates. The head and shoulders among the reflectors will be further detected and utilized to calculate the head-shoulder distance. The reflectors can be detected by finding the local maxima of the RSS distributions above. The azimuth-elevation heatmap of each range bin is processed independently. Thus, the head and the shoulders can be identified as distinct reflectors in most cases. Besides, the special cases are handled later in Sec. II. Furthermore, a fine-grained method is designed to localize the reflectors in the 3D space. Although the coordinates can be calculated from the range bin index and the azimuth and elevation angles, the localization error can hardly satisfy the requirements to distinguish persons with biometric features.

The error mainly comes from the ranging error (up to $\frac{1}{2}\frac{c}{2B}$). To reduce the error, MMTAI proposes a novel ranging technique combining the analysis of mmWave propagation and our specially designed signals, and mitigates the impact of angle errors in Sec. III-C.

Inspired by the Multi-Signal Consolidation technique in [15], we design a chirp group to measure the distance from the radar to the same body part. All chirps in the group have identical bandwidths, and each chirp has a different starting frequency.

After the round-trip propagation between the radar and the body part, the received signal phase φ is determined by the

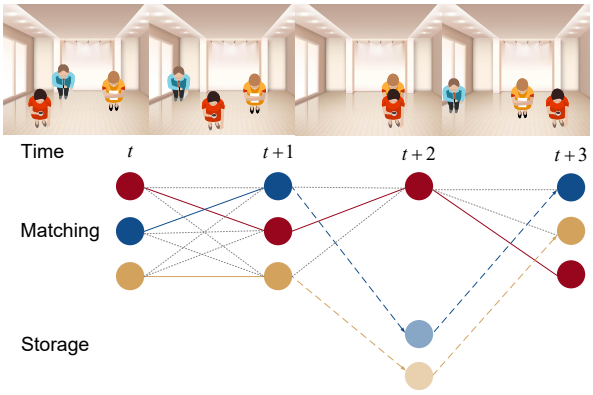

Fig. 6: Illustration of the short-time person matching algorithm with re-association.

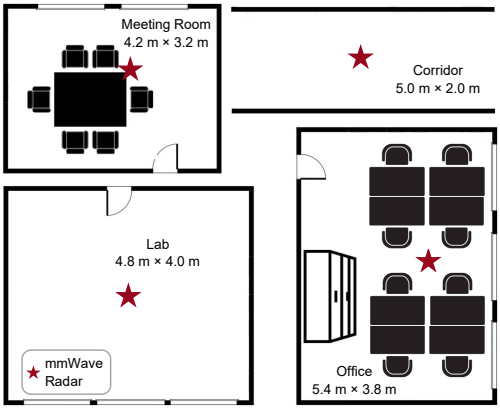

Fig. 7: Environmental settings.

starting frequency f_0 of the FMCW signal and the distance d:

$$\varphi(f_0, d) = \frac{1}{B}\int_{f_0}^{f_0+B} \frac{4\pi f d}{c} \mathrm{d}f \tag{2}$$

Within the duration of a chirp (60 μs), the change of the distance from the radar to the body is typically no more than 300 μm. Therefore, we assume the distance d as constant, and the phase linearly changes with the starting frequency:

$$\frac{\partial \varphi}{\partial f_0} = \frac{1}{B}\left(\frac{4\pi(f_0+B)d}{c} - \frac{4\pi f_0 d}{c}\right) = \frac{4\pi d}{c} \tag{3}$$

Furthermore, the radar-reflector distance can be calculated:

$$d = \frac{c\Delta\varphi}{4\pi\Delta f_0} \tag{4}$$

Different from the Step-Frequency Continuous Wave in [18], this technique breaks the range resolution limit with properties of both FMCW and step-frequency properties.

A preliminary experiment shows the feasibility of the ranging technique. The median ranging error of a static object within 20-100 cm away is 0.20 cm. Each chirp group includes 42 chirps with linearly increasing starting frequencies from 60 GHz to 61.348 MHz. The precision of the received signal phase is $\frac{2\pi}{N_{FFT}}$ that can reach 0.010 rad with $N_{FFT} = 640$ and the random error of them is 0.003 rad. The combination of these data reveals the potential of mm-level ranging.

Clustering. From the point cloud, persons can be detected. A person corresponds to a group of dense points much closer to in-group points than outliers. We feed the coordinates in all frames during a period to the clustering algorithm DBSCAN to group points by their corresponding person.

C. Biometric Features Measurement

Detecting the Head and Shoulders. To detect the head and shoulders from all the reflectors on each human body, we design an approach based on the spatial constraints.

In the assumed body poses (i.e. standing or sitting) of MM-TAI's application scenario, the head and shoulders are among the highest skeletal joints. After removing limb reflections by applying velocity thresholding, the points with high height in the point cloud are likely to be the head or shoulders.

Besides, the head-shoulder distances and the shoulder-shoulder distance are constrained by the skeleton structure and should be within a certain range. Therefore, for each triplet of points in the point cloud, a score can be calculated regarding the pair-wise distance and the point-wise height, and the triplet with the best score is chosen as the head and shoulders. Furthermore, the point with the nearest-to-middle projection on its opposite edge is regarded as the head.

Specially, if only two points are left after velocity thresholding, the higher one of them is regarded as the head, and the other is viewed as the shoulder.

Head-shoulder Distance Estimation. Although the head-shoulder distance can be directly calculated from the coordinates, the limited angle resolution of the mmWave radar makes the error overwhelm the diversity among persons.

To solve this problem, we notice that the radar-point range of the coordinates are much more accurate than the azimuth and elevation angles. Based on this, we propose a guess-and-check technique utilizing accurate ranging results.

We guess the 3D coordinates H_G, S_G of the head and shoulder and calculate the head-radar-shoulder angle $\theta = \angle H_G R S_G$. After that, another estimation can be obtained by:

$$|HS|_G = \sqrt{|RH|^2 + |RS|^2 - 2|RH| \cdot |RS| \cos\theta} \tag{5}$$

The guessed coordinates are scored as $(|HS|_G - |H_G S_G|)^2$, and the coordinates with the smallest score are chosen for the head-shoulder distance estimation.

Scalp Response Estimation. The change of dielectric permittivity with frequency characterizes the head skin. The RSS of a chirp starting at f_0 and ending at $f_0 + B_1$ is:

$$RSS(f_0) = \frac{1}{B_1}\int_{f_0}^{f_0+B_1} H_r(f) H_p(f) \mathrm{d}f, \tag{6}$$

where $H_p(f) \propto 1/(f^2 d^4)$ is the Rayleigh free-space attenuation and $|H_r(f)|$ is the skin reflectivity at frequency f.

The scalp response can be defined and approximated as

$$SK(f) = \frac{\mathrm{d}RSS(f_0)}{RSS(f_0)\mathrm{d}f_0} = \frac{\varepsilon_a}{2\left(\sqrt{\varepsilon_b}(1+\varepsilon_a B/2)+1\right)} \quad (7)$$

D. Multi-person Tracking

Short-time Person Matching. In MMTAI, a continuous tracking scheme is implemented, wherein a short-time person matching is performed between persons at each time instant and those from the previous time instant. The matching problem is formulated as a minimum-cost perfect bipartite matching problem, where the persons at the previous and current time instant are modeled as left-nodes and right-nodes.

Every two nodes on different sides share an edge with a cost indicating their dissimilarity. The cost is a weighted mean of the persons' two biometric features, velocity and location.

The weights are applied to the normalized feature values and adaptive to the Euclidean distance. When the distance is greater than a threshold T_d determined by the radar's range/angle resolution and the maximum velocity of humans, the nodes are unlikely to belong to the same person, and the weight of distance is W_{far}. Otherwise, the weights are automatically tuned to maximize the median relative difference between the optimal and suboptimal matches in a training trajectory. The parameters are: $W_{hs} = 0.8$, $W_{scalp} = 0.2$, $W_v = 0.1$, $W_{near} = 1$, $W_{far} = 5$, $T_d = 20\ cm$.

After modeling, the Kuhn-Munkres algorithm is applied to solve the minimum-cost perfect matching for each time instant. Concatenating the matching results continuously depicts the trajectory of each person. The solid lines in Fig. 6 depict the minimum-cost perfect matchings and the trajectories.

Trajectory Re-association. Persons can join and leave the area, crossover with each other or move behind an obstacle, causing fluctuation in the number of detected persons and imbalance of the bipartite graph. Consequently, some nodes cannot be matched and their corresponding trajectories may break. Conventional bipartite matching algorithms typically disregard these unmatched nodes, resulting in lost tracking targets and performance degradation.

To solve this problem, MMTAI manages a storage for the unmatched nodes. It stores essential information to re-identify persons, including locations and biometric features. The mechanics of the storage is two-fold: for unmatched left-side nodes, their locations and biometric features are saved in case it is blocked instead of leaving; for unmatched right-side nodes, their locations and biometric features are looked up in the storage to determine whether it is a newly-joined node or an old one resumed from crossover or blockage.

IV. Evaluation

A. Evaluation Setup

Testbed. We have successfully developed a prototype of MMTAI leveraging a single commercial off-the-shelf (COTS) mmWave radar TI IWR6843AOPEVM Rev.G. The chirp group configuration is identical to that in Sec. III-B.

Environmental Settings. We select four typical indoor scenarios for evaluation, across different sizes and complexity: 1) A lab (4.8×4.2 m^2) with little multipath; 2) A meeting room (4.2×3.2 m^2) with moderate multipath; 3) A corridor (5.0×2.0 m^2) with rich multipath; 4) An office (5.4×3.8 m^2) with rich multipath and a 1×0.5×1.5 m^3 cabinet in the center. Fig. 7 shows the settings in the lab, where the red star stands for the mmWave radar (on the ceiling) and the persons can walk anywhere on the floor. The ground truth trajectory is obtained from the monitoring video, and a reference grid is drawn on the floor to help localize the person in any frame.

Data Collection. The persons naturally perform various daily activities in the monitoring area, including walking, looking around, looking down, telephoning, running and sitting. To obtain ground truth, we pinpoint each person's heels in the video at keyframes. Subsequently, we interpolate their positions in other frames using linear interpolation. MMTAI localizes all persons every 10 ms, and the Euclidean distances between the estimated positions and the ground truth are calculated. All the distances are aggregated to analyze the performance.

Subjects. The experiments involve 3 male subjects: S1 (190cm, 68kg), S2 (188cm, 79kg), S3 (180cm, 65kg). All the experiments are IRB-approved, and all data are anonymized.

B. Overall Performance

We first evaluate the tracking accuracy of MMTAI and compare its performance with that of the state-of-the-art method, mmTrack [10]. Intuitively, Fig. 8 illustrates estimated trajectories of MMTAI in the two-person scenario, where both two persons can be tracked with a low error, and the estimated trajectories are well coordinated with the ground truth (GT).

Moving on to Fig. 9, we present the results of both methods with varying numbers of persons under tracking. In the 1-person and 2-person scenarios, MMTAI exhibits impressive median errors of 9.56 cm and 11.11 cm, respectively, surpassing mmTrack, which demonstrates median errors of 14.68 cm and 19.43 cm for the same scenarios. As the complexity increases with 3 persons under tracking, both methods experience a noticeable rise in median errors. However, MMTAI maintains a superior performance compared to mmTrack, recording median errors of 20.87 cm and 29.44 cm, respectively. The challenges faced, such as severe multipath and more frequent crossovers, contribute to the elevation in errors, but MMTAI remains resilient. In all three scenarios, MMTAI achieves a remarkable median error of 12.33 cm, signifying a notable 35.88% reduction when compared to mmTrack, which shows a median error of 19.23 cm.

C. Impacting Factors

We analyze the impact of four different factors on the performance of MMTAI, that is, environment, trajectory-breaking behaviors, activities during tracking and the location of persons relative to the radar.

Environments. We deploy MMTAI in the four selected environments. In each environment, we conduct three groups of experiments to track 1, 2 or 3 persons (1P/2P/3P). The results are shown in Fig. 11. Due to the large vacancy of the laboratory, the median errors in this scenario are the lowest,

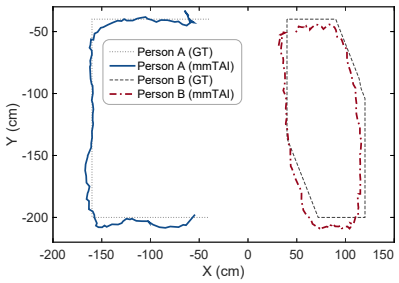

Fig. 8: Estimated trajectories of two persons.

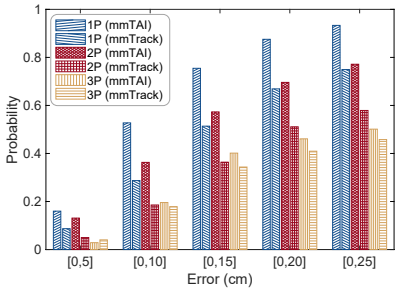

Fig. 9: Overall performance.

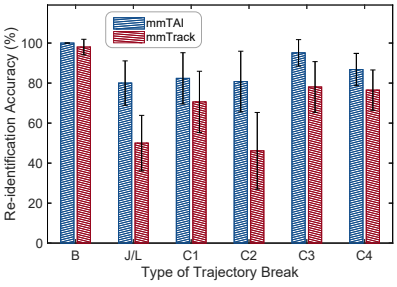

Fig. 10: Impact of trajectory breaks.

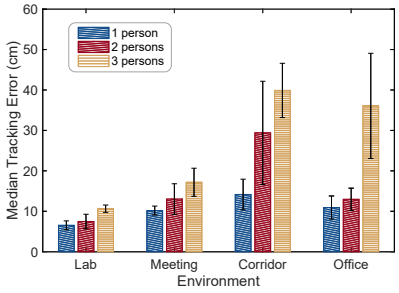

Fig. 11: Impact of different environments on multi-person tracking.

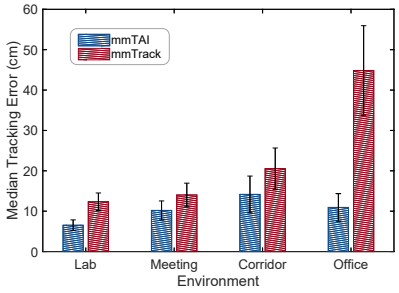

Fig. 12: Comparison of performance in different environments.

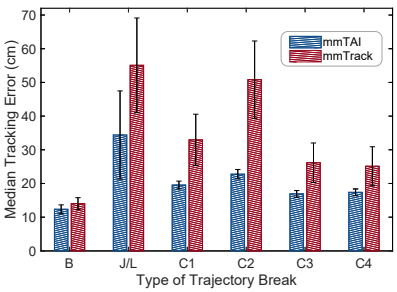

Fig. 13: Comparison of Re-identification accuracy.

which are 6.52 cm, 7.47 cm and 10.64 cm. By contrast, the median errors in the 3-person setting rise drastically to 39.89 cm and 30.07 cm respectively in the corridor and office because the small area strengthens crossover and blockages.

We also compare the performance of MMTAI and mmTrack in the environments as shown in Fig. 12. In the lab, meeting room and corridor, both MMTAI and mmTrack achieve a median tracking error within 20 cm. By contrast, the median error of mmTrack is 44.85 cm in the office, which is 75.63% greater than that of MMTAI (i.e., 10.93 cm). The main cause is the severe blockage between the radar and persons.

Trajectory-breaking Behaviors. During the tracking process, various trajectory-breaking behaviors may occur, such as crossovers, joining or leaving the monitoring area, or being obstructed by indoor blockages. While these situations often lead other tracking methods to encounter failures, MMTAI exhibits much better robustness and resilience.

Fig. 10 shows the performance of MMTAI and mmTrack when facing dense trajectory-breaking behaviors. Both mmTrack and MMTAI perform well when the person passes by blocked areas frequently. Nevertheless, when the person frequently joins/leaves the area or crossovers with others, we can see better error resilience in MMTAI than mmTrack. The median errors of MMTAI are 31.26 cm and 19.70 cm, lower than those of mmTrack (i.e., 57.38 cm and 33.33 cm).

Furthermore, we validate the trajectory re-association ability of MMTAI after six trajectory-breaking behaviors, including blockage (B), joining/leaving the area (J/L) and crossovers (C1-C4 in Fig. 14). The results are shown in Fig. 13.

The tracking performance is positively related to the re-association accuracy. Moreover, the re-association accuracy is determined by the duration and strength of interference brought by the behavior. Specifically, the duration of B, C1-C4 and J/L increase progressively, and their corresponding re-association accuracy decrease progressively; C2 is harder to handle than C1 due to the location prediction based on the velocity before crossover, while C3 and C4 are better resolved because persons tend to remain a gap between them during a longer crossover period, yielding weaker mutual interference.

Activities of the Persons. We conduct experiments where the persons are asked to perform four activities: looking around, looking down, telephoning and running.

As shown in Fig. 15, in the meeting room, if the person walks, looks around or looks down, the median tracking error is 10.17 cm, 11.21 cm and 14.19 cm, respectively. In contrast, when the person is telephoning, the error increases to 18.78 cm. The raise of the arm may mislead the detection algorithm to treat the hand as the head and hinder the measurement of head-shoulder distance. What's worse, if the person runs, his/her body will wag from side to side, causing jitter in the point cloud. Thus, the estimated trajectories will be more noisy and the error increases to 21.98 cm.

Location of the Persons. Finally, we evaluate the impact of the person's location on the tracking error. A person walks with a range of 0.7-1.5 m, 1.5-2.5 m or 2.5-3.5 m from the radar and an azimuth angle of 0-20°, 20-40° or 40-60°. Fig. 16 shows no significant difference in the error across each location, indicating the robustness of MMTAI.

Time Consumption. On an Intel i7-12700H CPU with 20 threads, the radar signal preprocessing costs 120 ms for each

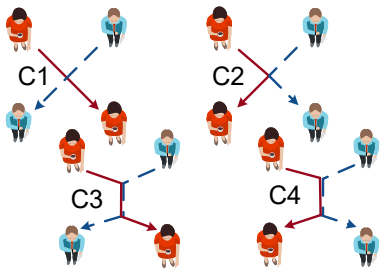

Fig. 14: Types of trajectory crossover.

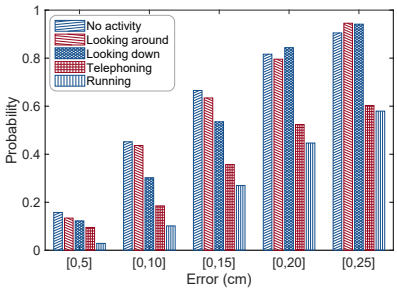

Fig. 15: Impact of different environments.

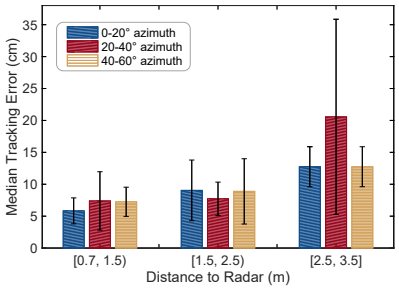

Fig. 16: Impact of the person's location.

second of raw data, and the following steps of MMTAI cost 15 ms. The overall time cost supports real-time application. When the person count N increases, the overall time consumption increases slightly. The preprocessing time remains constant as the size of the raw data depends only on the radar configuration; the biometric feature measurement requires minimal computation and can be ignored; the multi-person tracking algorithm has a complexity of $O(N^3)$ that guarantees the real-time workflow of MMTAI for up to 11 persons.

V. Discussion

We complete the discussion of MMTAI with some practical issues of applicability and general use.

Ghost Targets Elimination. MMTAI exploits the reflected signal from the person's head and shoulder for tracking and identification targets. However, ghost targets may exist because of the inevitable indoor multipath effect. MMTAI mitigates this adverse impact by filtering the propagation distance and strength of the reflected signals. In our future work, we may borrow the idea of [19], [20] and [21] to eliminate ghost targets by modeling the process of signal reflection.

Impact of Hair and Clothes. The reflected signals from the hair/clothes and those from human skin are superposed and beyond the range and angle resolution of mmWave radar. However, the reflections of hair [16] and clothes [17] only have a slight impact on the biometric features. Besides, since they are less likely to be changed during a multi-person tracking task, they can hardly degrade the accuracy of MMTAI.

Large-scale Multi-person Tracking. The FoV of the mmWave radar used in MMTAI is 120°, which is configured by the manufacturer and sufficient to monitor most rooms of typical size. If a large indoor coverage is required, MMTAI can be extended in two aspects: deploying the radar at the corner of the ceiling to enlarge the monitoring area, or deploying multiple radars like the vision-based tracking systems.

VI. Related Work

A. Device-free Human Tracking

Device-free human tracking estimates the position of human users without dedicated devices to them, providing much more convenience and scalability. Vision-, RFID-, WiFi- and mmWave-based methods have been widely explored for device-free human tracking.

Vision-based methods track persons in video frames with handcraft features like color, scale and velocity [22]. Based on these features, deep neural networks (DNNs) are usually exploited to track [23], [24].

To compensate for their disability in dark or privacy-concerned places, RFID-based and WiFi-based methods are developed. By analyzing the time of flight (ToF) and channel state information (CSI) of reflected signals, [25]–[27] realize tracking of a single person. Further studies simultaneously track multiple persons by exploring and utilizing more information: mD-Track [28] integrates the ToF, AoA, angle of departure (AoD) and Doppler frequency shift (DFS) to track multiple persons; WiPolar [29] leverages the polarization angle to pair up the AoAs from the same person at two receivers and pinpoints multiple persons.

With higher spatial resolution, mmWave-based tracking schemes provide relatively lower tracking error [10], [11], [30], [31]. mmTrack [10] analyzes the point cloud that consists the position of all reflection points in the field of view: it clusters the reflection points on the same person into one target, and continuously associates targets by location and velocity. m^{3}Track [11] extracts spatial and temporal features from the range-Doppler spectrum of all reflections, feed them into a DNN to reconstruct the 3D posture of multiple persons and track the 3D postures with an extended Kalman filter (EKF). Chen et al. [32] builds a reflection map based on the locations of humans and static reflectors to eliminate multipath and shadow ghosts. These works are limited in identifying persons and suffer from trajectory breaks, which can be tackled by MMTAI.

B. Human Identification

Identification requirements in authentication and security check have explored a variety of human features. MU-ID [33] extracts the motion features of the lower limb from the mmWave range-doppler spectrum and trains a convolutional neural network (CNN) to distinguish persons by gait; Wi-PIGR [34] also applies CNN to extract the gait from WiFi CSIs; AmbiEar [35] extracts the voice features; RF-Identity [36] deploys multiple RFID tags at different heights on a door

frame and integrates their CSIs to identify persons by body shape; DeepBreath [37] can extract respiratory frequencies of two co-located persons with FMCW signals and distinguish them; MSense [38] further supports vital sign extraction during the movement of persons. They explore a wide range of effective biometric features, and MMTAI proposes two novel alternatives.

C. Identification-assisted Human Tracking

As a pioneering work, mmSense [39] characterizes the body shape of multiple persons while tracking. mID [40] considers the gait and trains a long-short time memory (LSTM) network to distinguish persons by both gait and body shape. PALMAR [41] extracts similar features with a CNN and proposes the crossover path disambiguation algorithm to solve the crossover case. [19] can suppress false targets with the body shape feature and resolve multiple forms of trajectory break with trajectory similarity of the same person. Compared with these works, MMTAI can reduce the impact of indoor blockage and perform multi-person tracking more resiliently.

VII. Conclusion

In this paper, we propose MMTAI, a biometrics-assisted multi-person tracking technique. At its core, MMTAI integrates the ability to identify persons to multi-person tracking by discerning distinct persons' biometric features, specifically the scalp response to signals and the head-shoulder distance. This discriminative ability is facilitated by harnessing the fine-grained information encoded in the phase of the reflected mmWave signals and by utilizing a bipartite graph to continuously map the persons' biometric traits with their corresponding trajectories. Extensive experimental results show that in the multi-person tracking scenarios, MMTAI has a median tracking error of 12.33 cm, which is 35.88% lower than that of the state-of-the-art approach. We envision MMTAI as a promising step towards reliable location-based services in smart applications.

Acknowledgement

This work is supported in part by the National Natural Science Foundation of China (No. 62425207 and No. U21B2007).

References

[1] Y. Ren *et al.*, "Person re-identification using Wi-Fi signals," in *Proceedings of ACM MobiCom*, 2022, pp. 829–831.
[2] Y. He *et al.*, "Acoustic localization system for precise drone landing," *IEEE TMC*, 2024.
[3] Y. Sun *et al.*, "Indoor drone localization and tracking based on acoustic inertial measurement," *IEEE TMC*, 2024.
[4] W. Wang *et al.*, "MicNest: Long-range instant acoustic localization of drones in precise landing," in *Proceedings of ACM SenSys*, 2022.
[5] T. Wang and D. J. Cook, "Multi-person activity recognition in continuously monitored smart homes," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 2, pp. 1130–1141, 2022.
[6] S. Chen *et al.*, "Afall: Wi-Fi-based device-free fall detection system using spatial angle of arrival," *IEEE TMC*, 2023.
[7] J. Zhang *et al.*, "A survey of mmWave-based human sensing: technology, platforms and applications," *IEEE COMST*, 2023.
[8] Y. Chen *et al.*, "ELASE: Enabling real-time elastic sensing resource scheduling in 5G vRAN," in *Proceedings of IEEE/ACM IWQoS*, 2024.
[9] W. Wang *et al.*, "Meta-Speaker: Acoustic source projection by exploiting air nonlinearity," in *Proceedings of ACM MobiCom*, 2023.
[10] C. Wu *et al.*, "mmTrack: Passive multi-person localization using commodity millimeter wave radio," in *Proceedings of IEEE INFOCOM*, 2020.
[11] H. Kong *et al.*, "m3Track: mmWave-based multi-user 3D posture tracking," in *Proceedings of ACM MobiSys*, 2022.
[12] Y. Zhang *et al.*, "GaitSense: towards ubiquitous gait-based human identification with Wi-Fi," *ACM TOSN*, 2021.
[13] Y. Chen *et al.*, "Wireless sensing for material identification: A survey," *IEEE COMST*, 2024.
[14] C. Wu *et al.*, "mSense: towards mobile material sensing with a single millimeter-wave radio," in *Proceedings of ACM IMWUT*, 2020.
[15] C. Jiang *et al.*, "mmVib: micrometer-level vibration measurement with mmwave radar," in *Proceedings of ACM MobiCom*, 2020.
[16] K. Y. You and Y. Then, "Electrostatic and dielectric measurements for hair building fibers from DC to microwave frequencies," *IAES IJECE*, 2015.
[17] N. Pałka *et al.*, *Transmission and Reflection Characteristics of Textiles in the Terahertz Range*, 2021, pp. 131–144.
[18] K. Iizuka and A. Freundorfer, "Detection of nonmetallic buried objects by a step frequency radar," *Proceedings of the IEEE*, 1983.
[19] M. Jiang *et al.*, "A robust target tracking method for crowded indoor environments using mmWave radar," *Remote Sensing*, 2023.
[20] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave Micro-Doppler signatures," *IEEE TGRS*, 2021.
[21] J. Guo *et al.*, "Dancing waltz with ghosts: Measuring sub-mm-level 2D rotor orbit with a single MmWave radar," in *Proceedings of IEEE/ACM IPSN*, 2021.
[22] J. Zhou and J. Hoang, "Real time robust human detection and tracking system," in *Proceedings of IEEE CVPR*, 2005.
[23] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proceedings of IEEE CVPR*, 2016.
[24] B. Li *et al.*, "SiamRPN++: evolution of siamese visual tracking with very deep networks," in *Proceedings of IEEE CVPR*, 2019.
[25] J. Wang *et al.*, "D-Watch: Embracing "bad" multipaths for device-free localization with COTS RFID devices," *IEEE/ACM TON*, 2017.
[26] M. Kotaru *et al.*, "SpotFi: decimeter level localization using WiFi," *Proceedings of ACM SIGCOMM*, 2015.
[27] Y. Sun *et al.*, "BIFROST: Reinventing WiFi signals based on dispersion effect for accurate indoor localization," in *Proceedings of ACM SenSys*, 2023.
[28] Y. Xie *et al.*, "mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking," in *Proceedings of ACM MobiCom*, 2019.
[29] R. H. Venkatnarayan, M. Shahzad, S. Yun, C. Vlachou, and K.-H. Kim, "Leveraging polarization of WiFi signals to simultaneously track multiple people," in *Proceedings of ACM IMWUT*, 2020.
[30] Y. He *et al.*, "Detection and identification of non-cooperative UAV using a COTS mmWave radar," *ACM TOSN*, 2023.
[31] J. Zhang *et al.*, "mmHawkeye: Passive UAV detection with a COTS mmWave radar," in *Proceedings of IEEE SECON*, 2023.
[32] W. Chen *et al.*, "Environment-aware multi-person tracking in indoor environments with mmWave radars," in *Proceedings of ACM IMWUT*, 2023.
[33] X. Yang *et al.*, "MU-ID: multi-user identification through gaits using millimeter wave radios," in *Proceedings of IEEE INFOCOM*, 2020.
[34] L. Zhang, C. Wang, and D. Zhang, "Wi-PIGR: path independent gait recognition with commodity Wi-Fi," *IEEE TMC*, 2022.
[35] J. Zhang *et al.*, "AmbiEar: mmWave based voice recognition in NLoS scenarios," in *Proceedings of ACM IMWUT*, 2022.
[36] C. Feng *et al.*, "RF-Identity: non-intrusive person identification based on commodity RFID devices," in *Proceedings of ACM IMWUT*, 2021.
[37] S. Yue *et al.*, "Extracting multi-person respiration from entangled RF signals," in *Proceedings of ACM IMWUT*, 2018.
[38] Z. Chang *et al.*, "MSense: Boosting wireless sensing capability under motion interference," in *Proceedings of ACM MobiCom*, 2024.
[39] T. Gu *et al.*, "mmSense: multi-person detection and identification via mmWave sensing," in *Proceedings of ACM mmNets*, 2019.
[40] P. Zhao *et al.*, "mID: tracking and identifying people with millimeter wave radar," in *Proceedings of IEEE DCOSS*, 2019.
[41] M. A. Ul Alam, M. M. Rahman, and J. Q. Widberg, "PALMAR: towards adaptive multi-inhabitant activity recognition in point-cloud technology," in *Proceedings of IEEE INFOCOM*, 2021.