mmTrack: Passive Multi-Person Localization Using Commodity Millimeter Wave Radio

Chenshu Wu, Feng Zhang, Beibei Wang, and K. J. Ray Liu

Department of Electrical & Computer Engineering, University of Maryland, College Park, MD, USA

Origin Wireless Inc., Greenbelt, MD, USA

{cswu, fzhang15, bebewang, kjrliu}@umd.edu

Abstract—**Passive human localization and tracking using RF signals have been studied for over a decade. Most of the existing solutions, however, can only track a single moving subject due to the coarse multipath resolvability limited by bandwidth and antenna number. In this paper, we break down the limitations by leveraging the emerging 60GHz millimeter-wave radios. We present mmTrack, the first system that passively localizes and tracks multiple users simultaneously using a single commodity 60GHz radio. The design of mmTrack consists of three key components. First, we significantly improve the spatial resolution, limited by the small aperture of the compact 60GHz array, by performing digital beamforming over all receive antennas. Second, we propose a novel multi-target detection approach that tackles the near-far-effect and measurement noise. Finally, we devise a robust clustering technique to accurately recognize multiple targets and estimate the respective locations, from which their individual trajectories are further derived by a continuous tracking algorithm. We implement mmTrack on a commodity 802.11ad device and evaluate it in indoor environments. Our experiments demonstrate that mmTrack detects and counts multiple users precisely with an error ≤ 1 person for 97.8% of the time and achieves a respective median location error of 9.9 cm and 19.7 cm for dynamic and static targets.**

I. INTRODUCTION

Over the past decade, we have witnessed the emergence of RF-based indoor localization systems that shift from an active device-based theme [10], [34] to a passive device-free manner [19], [29] - localize and track a user without requiring her/him to carry or wear any device. Passive indoor localization can enable a wide range of applications that demand ubiquitous tracking of users' locations while not favoring or allowing a user to carry or wear a device. For example, it allows a smart TV to localize users and intelligently rotate the viewing angle or turn on/off when users are coming/leaving. It would also enable an energy-efficient building to localize and count the occupants and adjust the heating and lighting accordingly. It can be used in a smart home to control the Internet of Things (IoT) devices in response to the users' locations. All of these applications would become more practical and attractive in natural indoor settings when multi-person localization is enabled in ubiquitous contexts without invading user privacy.

Prior works have taken steps towards this goal, and considerable progress has been made, especially for single target tracking [11], [19]. However, most of the existing works fail in the presence of multiple persons, mainly because the reflected signals from all targets and other objects in the environment are superimposed together. Besides, these works mainly focus on the tracking of a moving target and need cumbersome calibration to localize static objects. To achieve multi-person localization, many of past proposals resort to a dense deployment of many devices and/or prior training [3], [30]. WiTrack2.0 [2] eliminates this need by building a specialized hardware with many antennas in a quite large form factor. The latest work [9] localizes up to 3 persons using commodity off-the-shelf (COTS) WiFi, but needs multiple receivers and assumes the knowledge of the number of persons. For all of these systems, the resolvability to disentangle the weak signals reflected off multiple targets is fundamentally limited by the antenna number and the system bandwidth.

In this work, we break down the limitation by leveraging an opportunity in the emerging 60GHz 802.11ad/ay technology (*a.k.a* WiGig), which is already available in commercial routers [14] and will soon be available in smartphones [1]. We present *mmTrack*, the first system that achieves *multi-person localization and tracking using a single commodity 60GHz millimeter wave* (mmWave) device. Different from 2.4GHz/5GHz WiFi, 60GHz radios offer high directionality with large phased arrays in small size and precise time-of-flight measurements thanks to the large bandwidth, circumventing indoor multipaths and underlying competent space and time resolution for accurate localization of multiple persons. For example, the commodity device we use in this work has a 32-element array with 3.52GHz bandwidth centered at 60GHz. By designing mmTrack, we aim to reuse commodity 60GHz networking chipsets as a radar for multi-person localization.

However, enabling multi-person passive localization using 60GHz radio is not an easy task as it sounds like. Multiple challenges need to be addressed. First, despite the many antennas, the spatial resolution is limited due to the small aperture of the antenna array. For example, our experimental device has an array size of 1.8 cm × 1.8 cm, and the on-chip analog conventional beamforming only provides an angular resolution of 15°, which is frequently inadequate to separate nearby targets. To boost the spatial resolution, instead of using the built-in analog beamforming, mmTrack performs digital beamforming on the received Channel Impulse Response (CIR), which achieves a much higher spatial resolution in distinguishing two close angles compared with conventional beamforming. Specifically, we employ a non-parametric beamformer, *i.e.*, the Minimum Variance Distortionless Response (MVDR) beamformer [4], with a direction scan interval of 2° in both azimuth and

elevation dimensions, resulting in a *cell* resolution of 2° (for azimuth) × 2° (for elevation) × 4.26 cm (for range) in space.

Second, due to the high carrier frequency, signals attenuate rapidly over the propagation distance, making it challenging to locate distant targets. The measurement noises further exacerbate the weak reflection signals and complicate the detection in a large coverage. To address this challenge, we devise an adaptive target detection approach based on a novel metric independent of the absolute power to find the ranges of interests, which adapts to the diverse responses of targets at different distances.

Third, given the reflections from multiple users and different parts of their bodies, it is non-trivial to sort the occupied cells for location estimates of individuals, especially when they are close to each other. Additionally, different from device-based localization that can distinguish users by their device identity, it is non-trivial to sort multiple users' locations and recover their respective trajectories. To overcome these challenges, we employ the widely used k-means clustering and augment it with a set of robust techniques. Then we formulate the problem of successive tracking of multiple users as a weighted bipartite graph matching problem, which effectively recovers the trajectories of multiple users even in the presence of location errors, missing detections, and dynamic user numbers.

We have built a prototype of mmTrack using a COTS 60GHz 802.11ad device. The device offers a radar-like mode in addition to its original networking function, similar to the emerging dual-function communication and radar design [35]. In our experiments, we operate the testbed in its radar-like mode: transmit over one antenna and receive from all the 32 elements on the receiver that is co-located with the transmitter. We conduct experiments in indoor space with multiple dynamic/static users. Experimental results show that mmTrack achieves median location accuracy of 9.9 cm and 90%tile accuracy of 22.5 cm for dynamic targets, and median location accuracy of 19.7 cm and 90%tile accuracy of 35.6 cm for static targets. In addition, mmTrack detects and counts multiple users precisely, with no error in user number for 79.0% and an error ≤ 1 for 97.8% of the time, respectively. We envision mmTrack takes an important step towards practical multi-person localization and tracking and sheds light on mmWave sensing via 60GHz radios.

In summary, our core contributions are as follows:

- As far as we are aware of, mmTrack is the first work that enables multi-person passive localization and tracking by leveraging a COTS 60GHz mmWave radio with a dual-function radar communication system. We promote the spatial resolution by digital beamforming based on MVDR and propose a novel object detection approach that tackles the near-far-effect and measurement noise.
- We devise a robust algorithm based on k-means clustering that can accurately and robustly determine the number of users and estimate their respective locations. We achieve continuous tracking of multiple trajectories by a novel algorithm using weighted bipartite graph matching.
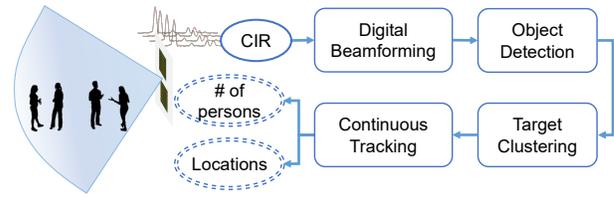
Fig. 1. An overview of mmTrack

- We prototype and evaluate mmTrack, and demonstrate it can localize moving and static users with a median accuracy of 16.24 cm and track the continuous trajectories of multiple users simultaneously.

In the rest of the paper, we first present an overview of mmTrack in §II, followed with object detection in §III and target clustering in §IV. We implement and evaluate mmTrack in §V. We review the literature in §VI and conclude in §VII.

II. MMTRACK OVERVIEW

mmTrack is a wireless system that can accurately localize and track multiple users by using the RF signals reflected off the users' bodies. Different from many existing works, it can localize multiple static users with a single radio.

mmTrack is built upon commodity Qualcomm 60GHz chipsets. 60GHz WiFi technology, with established IEEE 802.11ad/ay standards, is becoming mainstream in wireless devices to enable high rate networking and rich user experience. Different from the 2.4GHz and 5GHz WiFi that typically has only 2 to 3 antennas and 20MHz to 80MHz bandwidths, 60GHz WiFi radios offer large phased antenna arrays in compact forms and large bandwidths centered at a high-frequency band. Specifically, the 60GHz radio we use for mmTrack operates at 60GHz center frequency and 3.52GHz bandwidth. It uses one commercial 802.11ad chipset with two standard 6×6 antenna arrays (each with 32 elements in total, the other four locations are reserved) for Tx and Rx in a small form factor of 1.8 cm×1.8 cm, allowing it to be mounted on a single, compact device. The device is a full-duplex chip towards the emerging joint radar communication systems [35], which allows radar-like operation of simultaneous transmitting and receiving. The Tx sequentially transmits 32 pulses of a known sequence, each of which, after reflection on a target, is received and correlated by one Rx element to estimate CIR. The 60GHz radio provides a range resolution of 4.26 cm, and the on-chip beamforming offers a 3 dB beamwidth of 15°.

However, enabling multi-person localization using 60GHz signals entails several challenges, including limited spatial resolution due to the very small aperture of the compact antenna array, fast attenuation due to the high carrier frequency, measurement noises that not only blur target locations but also produce ghosts, and inter-person blockage. As illustrated in Fig. 1, mmTrack tackles these challenges by four main components: 1) *Digital beamforming* that achieves much narrower beamwidth than the on-chip beamforming; 2) *Object detection* that adaptively detects the presence of objects at various distances; 3) *Target clustering* that identifies the
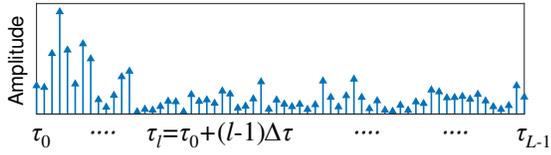
Fig. 2. The measured CIR: τ_0 and $\Delta\tau$ represent the propagation delay of the first received path and the time resolution of the CIR, respectively.

number of targets along with their respective locations; and 4) *Continuous tracking* that resolves individuals' successive trajectories.

III. OBJECT DETECTION VIA DIGITAL BEAMFORMING

A. CIR on 60GHz Radio

mmTrack uses one element of the transmit antenna array to continuously transmit the beacon signals with a constant rate F_s. The CIR measured by the m-th antenna $h_m(\tau)$ can be expressed as

$$h_m(\tau) = \sum_{l=0}^{L-1} a_{m,l}\delta(\tau - \tau_l), \tag{1}$$

where L is the number of the CIR taps, $\delta(\cdot)$ is the Delta function, and $a_{m,l}$ and τ_l denote the complex amplitude and the propagation delay of the l-th tap, respectively. The time resolution of the measured CIR, *e.g.*, $\Delta\tau$ in Fig. 2, is determined by the bandwidth B of the transmitted signal, *i.e.*, $\Delta\tau = 1/B$. At each time slot, mmTrack captures $M \times L$ complex values, *i.e.*, $h_m(\tau_l)$ where $m = 1, \cdots, M$ and $l = 0, \cdots, L-1$, and the location information of multiple persons can be inferred from these measurements. To facilitate the notations in the following, we define the vector $\mathbf{h}(\tau_l) = [h_1(\tau_l), \cdots, h_M(\tau_l)]^T$ to record the complex channel gains of all the receive antennas at a propagation delay τ_l.

The COTS device we use reports CIR with up to 256 taps. An example of a real CIR measurement with the first 64 taps is shown in Fig. 2. Previous work [16], [17] also exploits CIR under standard networking mode. Due to the unsynchronized Tx and Rx, however, either the CIR is not accurate enough for computing the range [17] or it needs tedious efforts to extract the CSI [16]. To our best knowledge, mmTrack is the first to extract precise CIR via radar-like operations implemented on a 60GHz networking device attached with an extra array.

B. Beamforming

For each propagation delay τ_l, assume that there are $N(\tau_l)$ reflected signals impinging on the receive antenna array with different azimuths ϕ and elevations θ, as shown in Fig. 3. To simplify the notations in the following, we omit the dependence on τ_l if not mentioned. Then, the CIR can be formulated as

$$\mathbf{h} = \begin{bmatrix} \mathbf{s}_1(\theta_1, \phi_1), & \cdots, & \mathbf{s}_N(\theta_N, \phi_N) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} n_1 \\ \vdots \\ n_N \end{bmatrix}, \tag{2}$$

where $\mathbf{s}_i(\theta_i, \phi_i)$ denotes the steering vector pointing to (θ_i, ϕ_i) corresponding to the direction of the i-th reflected signal, *i.e.*,
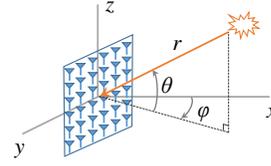
Fig. 3. Coordinate system: θ and ϕ denote the elevation and azimuth respectively, and r denotes the distance between the reflector and the device.

the normalized phase response of the antenna array for a signal coming from the direction (θ_i, ϕ_i), x_i denotes the complex amplitude of that signal and n_i stands for additive thermal noise. The more concise matrix representation can be written accordingly as

$$\mathbf{h} = S\mathbf{x} + \mathbf{n}. \tag{3}$$

The reflected signals from different directions can be distinguished by combing the channel gains of receive antennas linearly with different weights $\mathbf{w}$, which is also known as beamforming techniques.

In this work, a non-parametric beamformer, Minimum Variance Distortionless Response (MVDR) beamformer (*a.k.a* the Capon beamformer) [4], is utilized due to its high spatial resolution compared with the conventional beamformer. This is because MVDR beamformer minimizes the power of interference and noise from other angles while ensures the distortionless response towards the looking direction. Besides, the reason a non-parametric beamformer is preferred is that no prior assumptions about the structure of the reflected signals are required. The MVDR beamformer for direction (θ, ϕ) is defined as

$$\mathbf{w}(\theta, \phi) = \frac{R_{\mathbf{h}}^{-1}\mathbf{s}(\theta, \phi)}{\mathbf{s}^H(\theta, \phi)R_{\mathbf{h}}^{-1}\mathbf{s}(\theta, \phi)}, \tag{4}$$

where $R_{\mathbf{h}}$ denotes the correlation matrix of $\mathbf{h}$, *i.e.*, $R_{\mathbf{h}} = \mathbb{E}[\mathbf{h}\mathbf{h}^H]$. Then, the spatial spectrum obtained by MVDR beamformer for each direction (θ, ϕ) can be written as

$$\begin{aligned} P(\theta, \phi) &= \mathbb{E}[|\mathbf{w}^H(\theta, \phi)\mathbf{h}|^2] \\ &= \frac{1}{\mathbf{s}^H(\theta, \phi)R_{\mathbf{h}}^{-1}\mathbf{s}(\theta, \phi)}. \end{aligned} \tag{5}$$

Equation (5) builds the link between the measured CIR and the distribution of the energy of the reflected signals *w.r.t.* the receive antenna array in the space. Note that in principle, $P(\theta, \phi)$ is independent of the transmit antenna, and in practice, it is observed that different transmit antenna produces similar quality of the estimation. Therefore, mmTrack only selects a single antenna from the transmit antenna array to reduce the complexity of mmTrack.

In practice, $R_{\mathbf{h}}$ at time slot t is estimated by the sample correlation:

$$R_{\mathbf{h}}[t] = \frac{1}{W}\sum_{i=0}^{W-1} \mathbf{h}[t-i]\mathbf{h}^H[t-i], \tag{6}$$

where W denotes the window length. The choice of the window length W depends on the desired tradeoff between the precision and responsiveness of mmTrack.

(a) Power w.r.t range   (b) Number of points w.r.t range
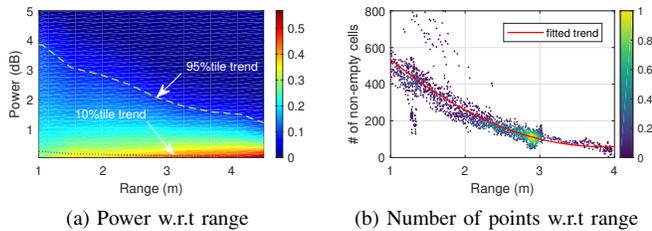
Fig. 4. Measurements on the near-far effect and the effects of weighted resampling. Colors indicate the data density.

In the following, we use $P_t(\theta, \phi, \tau)$ to denote the spatial spectrum at time slot t, propagation delay τ, and direction (θ, ϕ). In practice, θ and ϕ are quantized with a resolution of $2°$ to reduce the computational complexity of computing the spatial spectrum, while theoretically, the sector can be arbitrarily narrow.

C. Target Detection

The primary purpose of traditional radar systems is to detect the locations of the *point* targets far away from the system. Differently, mmTrack aims to detect multiple targets close to the device and extract the 3D information, *i.e.*, the azimuth, elevation, and range of every reflection point of a human body that implies the location and even the body silhouette. There are two significant challenges for multi-target detection:

1) **Noise and interference**: Due to the thermal noise and hardware internal interferences, there exist spurious spikes in the obtained spatial spectrum which may cause false detection of targets;

2) **Near-far-effect**: It is hard to detect a distant target in the presence of a nearby target, mainly due to two reasons: 1) the blockage by the nearby target, and 2) the imbalance between the energy of the EM waves reflected off the nearby target and the distant target.

In the following, we first present measurements on the near-far-effect, and then introduce two key elements in the target detection module of mmTrack that overcome the above challenges: *range finding* and *3D information extraction*.

Measurements on Target Reflections. We empirically analyze the two issues by about 6,000 real measurements of targets present at different locations. Fig. 4a shows the power of the reflected signals with respect to the target ranges, which implies that 1) the received power decreases over distances, and 2) the received power varies at the same distance due to various factors such as their relative locations and orientations to the device, the surface conditions, and noise and interference, etc. Further in Fig. 4b, we depict the varying numbers of reflection points with respect to the ranges. As seen, the number of non-empty directions rapidly drops when the distance increases. This is because the target (human body) becomes relatively "smaller" from the viewpoint of the device at larger distances and thus occupies fewer sectors while fills in more sectors at closer locations. In addition, the direction of some reflecting signals might not be in the field of view of the device or be too weak to be received for distant targets.
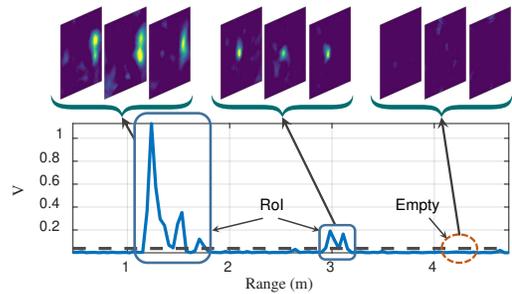
Fig. 5. Illustration of the metric for the detection of the RoI: two subjects stand in front of the device.

We note that the above analysis is environment independent since there are few multipaths for 60GHz signals.

Range Finding. The purpose of range finding is to robustly detect distant targets even in the presence of close-by targets. Compared with the near targets, the energy of the signals reflected off the far targets is usually very weak. To increase the "visibility" of the far targets, for each specific range (or propagation delay), we calculate the variation of the energy distribution of the spatial spectrum $V_t(\tau)$, which is defined as $V_t(\tau) = \mathrm{Var}_\theta[\mathrm{Var}_\phi[P_t(\theta, \phi, \tau)]]$, where $\mathrm{Var}_\theta[\cdot]$ denotes the variance over parameter θ. A large $V_t(\tau)$ implies that the energy distribution of the reflected signals for that range is highly non-uniform, indicating the presence of a target in that specific range, while for the range where no target presents, the energy of the reflected signals is usually small and uniformly distributed, as illustrated in Fig. 5, where two subjects stand still in front of the device. Then, the set of the range of interest (RoI) at time slot t are formed as $\mathrm{RoI}(t) = \{\tau | V_t(\tau) > \eta, \forall \tau\}$, where η is a preset threshold. To accommodate the time-varying interference and noise, we use a multiple of the median value of $V_t(\tau)$ as the threshold for each time slot t, *i.e.*, $\eta(t) = \alpha \mathrm{Med}_\tau[V_t(\tau)]$, where α denotes a constant coefficient, and $\mathrm{Med}_\tau[\cdot]$ denotes the median value over τ. The reason we use the median as the threshold is that, in practice, the targets are usually sparsely distributed in the monitoring area, which results in the sparsity of $V_t(\tau)$ in τ.

Since not all the RF signals reflected off human body parts can be captured by the receive antenna array, the RoI for a specific target could be discontinuous. Fig. 6 shows an example of the detected RoI using the same data as Fig. 5. Ideally, the RoI for this example should consist of two subsets of continuous ranges corresponding to the two subjects. Due to the non-uniform reflection of the RF signals, however, four subsets are detected. Some points of the targets could thus be lost due to the incomplete RoI detection, *i.e.*, some parts of the target can be missing. To solve this issue, a completion algorithm is applied to the raw RoI, which combines any two neighboring subsets of continuous ranges as long as their minimum distance is smaller than a threshold $\gamma(\tau)$, as shown in Fig. 6. The threshold $\gamma(\tau)$ is dependent on the range of the target since the target closer to the device occupies a larger span of the range. By doing so, the targets are separated in terms of their distances to the device and each subset of the RoI should correspond to at least one target.
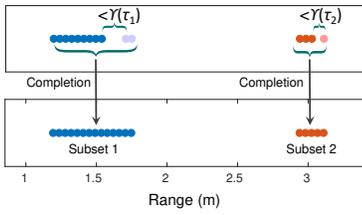
Fig. 6. Illustration of the RoI completion. Missing ranges are included to complete full and continuous RoIs.
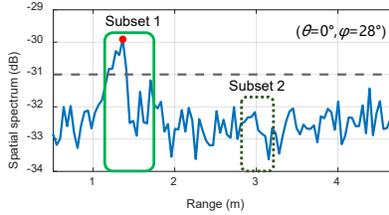

Fig. 7. The spatial spectrum in a specific direction ($\theta = 0°, \phi = 28°$) in which only one subject presents in the range of Subset 1.
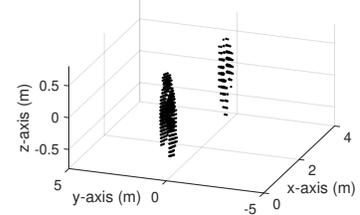

Fig. 8. The 3D information of the two subjects (at (x=1m, y=1m) and (x=3m, y=0m) respectively) extracted from the spatial spectrum.

3D Information Extraction. For each subset of RoI, mmTrack searches all the possible directions within its field of view to detect the presence of targets. Here a specific range along a certain direction is also termed as a *cell* in the space. Specifically, Fig. 7 shows the spatial spectrum *w.r.t.* τ for a certain direction ($\theta = 0°, \phi = 28°$). Based on the noise level of the device, an empirical threshold $\beta = -31$dB is applied to differentiate the signals from the noise, *i.e.*, if the value of the spatial spectrum at some range is larger than β, then a target is likely to occupy that cell, resulting in a detected *point*. To reduce the redundancy of the detected points, for each direction, mmTrack only takes the maximum value as the representative point for each subset of RoI. This is also because that the millimeter-wave cannot penetrate the human body, and most of its energy is reflected and absorbed by the surface of the human body. As illustrated in Fig. 7, mmTrack detects the presence of a target within the Subset 1 of the RoI, denoted as a red dot, while no targets are detected within the Subset 2 for this specific direction. To improve the accuracy of the range estimation for each detected point, a simple fractional 1-D interpolation is applied to range dimension (τ) using two adjacent samples of the peak. Fig. 8 illustrates an example of the object detection results for a two-person case. Note that the 3D information has been transformed from the form (θ, ϕ, r) to 3D locations (x, y, z) in Euclidean space with the device's location as the origin. As seen, thanks to the high spatial resolution, the proposed object detection almost recovers the complete human figures (when they are physically separate). We leave this potential in imaging as a promising future direction and focus on passive localization in this work.

IV. TARGET CLUSTERING AND CONTINUOUS TRACKING

A. Target Clustering

In mmTrack, we mainly employ k-means clustering, one of the most widely used unsupervised clustering techniques, to identify individual targets. The inputs are the outputs of the object detection module, *i.e.*, a set of points $O = \{\mathbf{o_i}, i = 1, \cdots, K\}$ where $\mathbf{o}_i = (x_i, y_i, z_i, w_i)$ denotes a reflection point at location (x_i, y_i, z_i) with reflection amplitude w_i and K is the total number of detect points for the current frame. The goal here is to identify the number of targets and the corresponding locations. However, directly feeding O for clustering does not yield satisfactory results. Therefore, we apply a pipeline of techniques to enhance the target clustering.

Determining Target Number. A key to k-means clustering is to determine the number of clusters k, *i.e.*, the number of targets in mmTrack. We adopt silhouette analysis [20] for this purpose. The silhouette value is a visual measure that characterizes how similar a point is to its own cluster compared to the separation distances to neighboring clusters. Formally, the silhouette value of the ith point defined as

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}, \qquad (7)$$

where a_i is the average distance from the ith point to the other points in the same cluster, and b_i is the minimum average distance from the point to points in a different cluster. In the case of a cluster with size 1, s_i is defined as 0. The measure ranges from -1 to +1, with higher values indicating the point is far away from the neighboring clusters. Then the average of s_i over all data points is used to evaluate how appropriately the data have been clustered. We try k in a proper range and select the value that produces the highest average silhouette value $\bar{s}(k)$.

Considering that the number of targets will not fluctuate within consecutive frames, we apply temporal constraints to further improve the robustness of k selection. Specifically, we attenuate $\bar{s}(k)$ for time moment t if k is not consistent with the majority of the target number within the past several estimates:

$$\bar{s}'(k) = \begin{cases} \kappa \cdot \bar{s}(k), & \text{if } k \neq k' \\ \bar{s}(k), & \text{otherwise} \end{cases} \qquad (8)$$

where κ is a penalty factor smaller than 1 and k' is the majority of the number of clusters over a short window of about one second (10 frames). Then the number of clusters is selected as $k^\star = \arg\max_k \bar{s}'(k)$.

Since the silhouette value is not suitable for a single cluster (*i.e.*, $k = 1$), we build an examiner to determine in prior if there tends to be a single cluster. Our algorithm starts from the global *centroid* of all the points O, and iteratively expands to all neighboring points that satisfy a density larger than p, the minimum number of points within a neighboring circle of radius ϵ. When the procedure terminates, we calculate the area enclosed by the points included in the x-y plane. If the area is sufficiently small, the algorithm will determine the data as one cluster for a single target and will thus skip k-means clustering. In a special case that no points are found, there should be multiple clusters since the data points are separate and thus, there are no dense points around the global centroid. Note that
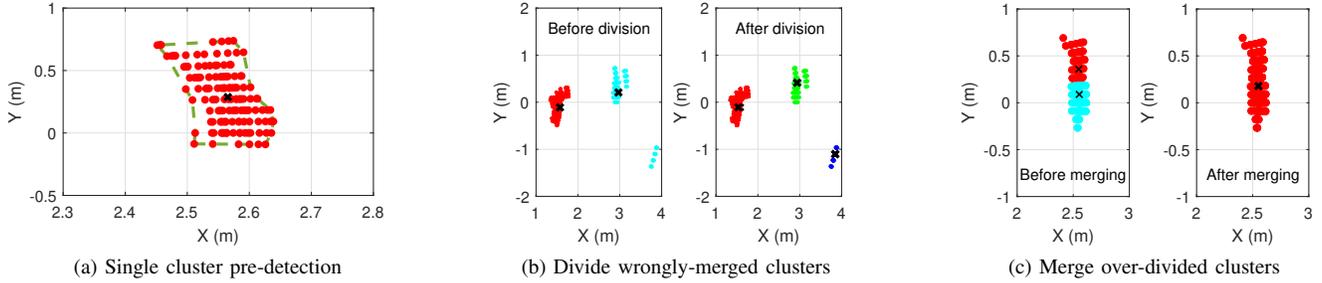
Fig. 9. Robust clustering. mmTrack performs (a) single cluster detection before k-means clustering, and two post validation steps (b) and (c) afterwards.

the algorithm only involves the x and y locations, without the z dimension. Our idea of examining the single target case is inspired by the DBSCAN algorithm [5]. However, we do not directly use DBSCAN for clustering because the data density varies too much over different targets due to the near-far-effect (as shown in Fig. 4b).

Robust Clustering. The k-means algorithm, however, might result in error clusters. There are two types of typical errors: 1) the points of two targets are falsely merged as one cluster, or 2) the points of one single target are wrongly divided into multiple clusters. To improve the robustness, we additionally perform post validation techniques by leveraging the geometric properties of the detected target points.

We first compute the *medoid* of each cluster by k-means. Then we calculate two distances in the 2D plane: the *intra-cluster medoid distance*, *i.e.*, the distance of every point within the same cluster to its medoid; and *inter-cluster medoid distance*, *i.e.*, the distance between the medoids of two clusters. To correct the first type of error, we examine the number of points whose intraDis exceeds a pre-defined maximum value representing the maximum of typical human size in the top view. If the selected points form a considerable portion of the cluster, they are then sorted out as a new cluster. Fig. 9b illustrates an example of the partition operation. Note that it is critical to use medoid here, rather than the more commonly used centroid. The centroid, as the geometric center, could be largely biased by the wrongly included points, and thus, its distance to all points would be averagely smaller. Differently, the medoid will be always constrained within the cluster, rendering a larger distance to the points in other cluster.

In the second case, usually, the resulted clusters are extremely close to each other since the points are from the same target. We therefore examine the inter-cluster medoid distance and merge those close clusters as one, as shown in Fig. 9c.

Although a few hyper-parameters are used in the above robust clustering, we note that these parameters are environment independent and generally applicable since they are mostly related to the physical properties of human bodies.

Target Location Estimation. When we obtain the clustering results, a target's location is intuitively estimated as the geometric medoid of all the points belonging to the corresponding cluster. Since we have 3D coordinates for each data point, the target's height information is also available as the maximum height of all points within the cluster.

B. Continuous Tracking and Counting

So far, we have achieved passive localization of multiple persons. Given a snapshot at a specific time, mmTrack can output the number of persons and their respective location estimates. However, to continuously track the trajectories of multiple targets still entails challenges. First, due to the LOS limitation of the 60GHz signals, one person could be completely blocked by the other during free movements, making the hidden person invisible to the device. Miss detection might also happen even when there is a clear LOS view; Second, there are false alarms due to measurement noise; Third, the clustering algorithms may output a wrong number of targets and thus erroneous location estimates. Below, we overcome the challenges by solving the continuous tracking problem via weighted bipartite graph matching.

Assume that there are q trajectories, each of which the last location appears at (x_i, y_i) at time $t-1$ (denoted as $u_i = (t-1, x_i, y_i), i = 1, 2, \cdots, q$). Denote the latest estimates of k targets at time t as $v_j = (t, x_j, y_j), j = 1, 2, \cdots, k$. The goal of the continuous tracking problem is to associate each of the latest estimates to an existing trajectory or a newly created one. A straightforward solution is to splice each new estimate to the closest trajectory. Such a greedy solution, however, could lead to frequent mismatches. Instead, we attempt to find the optimal association by minimizing the sum of the distances for all the k targets.

We model this task as a weighted bipartite graph matching problem. As shown in Fig. 10(a), we maintain the last location of every target that already present as a vertex in the left subset U of the bipartite graph. Here we only consider targets who recently appear. In other words, targets who have disappeared for a certain time are considered leaving, and the corresponding trajectories are simply terminated. The other subset V on the right side contains the latest k location estimates as vertexes. Then we add an undirected edge (u, v) for every pair of vertexes $u \in U$ and $v \in V$. Each edge gets a weight defined as the Euclidean distance $d_{u,v}$ between the locations of u and v. Then by finding a perfect matching of the constructed graph, in which every vertex in each subset is incident to exactly one edge of the matching, we will obtain an optimal solution that associates the newly arrived estimates to existing trajectories. Finding a perfect matching for a weighted bipartite graph is a classical problem in graph theory and can be solved by the KM (Kuhn and Munkres) algorithm [15].
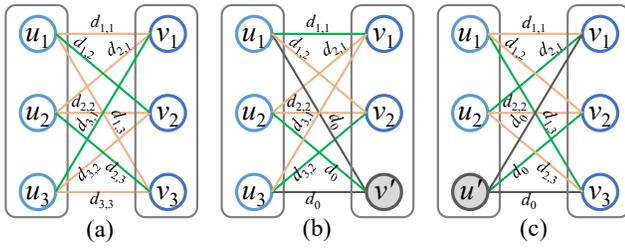
Fig. 10. Weighted bipartite graph matching. (a) Normal case. (b) Missing existing target. (c) Newly appeared target.

Fig. 11. Experimental scenario and device setup.

A perfect matching is only possible for a bipartite graph with an equal number of vertexes in each partition. In practice, however, targets may come and go, making the vertex number in each subset varying over time. To overcome this issue and ensure a perfect matching exists, we employ virtual vertexes for the smaller subset. Specifically, in case an existing target leaves or is missing (Fig. 10b, where there are fewer vertexes in the latest estimates), we add fake vertexes to V as inexistent targets. Similarly, in Fig. 10c, if a new target presents and there are fewer vertexes in U, we also insert a virtual vertex. For both cases, we assign an identical weight of d_0 to all edges associating with the virtual vertexes. By doing so, we form a bipartite graph that has a definite perfect matching. For the case in Fig. 10b, the matching corresponding to the virtual vertexes are simply ignored. While for Fig. 10c, a new trajectory is initiated for every virtual vertex, and the corresponding location in the latest estimates becomes the starting location of the trajectory.

The proposed algorithm does not assume a fixed number of targets. Neither does it need prior knowledge of how many persons there are. By successively appending the upcoming estimates to existing or newly created trajectories, we are able to track the continuous locations of individual targets and certainly count the number of persons. Finally, we perform spline smoothing [6] on the trajectories for post-processing.

V. EVALUATION

A. Methodology

We prototype mmTrack and conduct real-world experiments using Qualcomm 802.11ad chipset. The platform, including the chips and the software tool to export and collect CIR, is provided by Qualcomm. As shown in Fig. 11, the chipset is equipped with two antenna arrays, both of which have 32 antennas arranged in a 6×6 topology. We configure one Tx antenna to transmit pulses of known sequence and receive signals from all of the 32 Rx elements. The experimental area is roughly a sector of 110° with a radius of about 5 m, which we believe is sufficient to cover typical rooms.

We carry out experiments on one floor of a typical office building, which is furnished with desks, chairs, computers, and TVs. We recruit 5 volunteers aging from 22 to 57 for testing. Our evaluation includes two scenarios: a static case where users are standing or sitting still at specific locations and a dynamic case where users are walking simultaneously. For each case, we test different locations and traces for different numbers of users. The location error is calculated with respect to the ground truth traces marked in prior. Since we cannot obtain the true user labels of each estimate, we project the estimate to the closest true target to compute location errors.
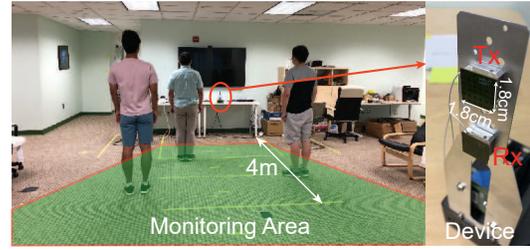
B. Performance

Overall accuracy. Fig. 12a illustrates the overall tracking accuracy of mmTrack in dynamic and static scenarios, respectively. For each scenario, different traces from one or multiple users are integrated to derive the result. As seen, in contrast to many of the existing works that can only track moving persons, mmTrack achieves high accuracy in both static and dynamic scenarios, with a respective median accuracy of 9.9 cm and 19.7 cm, and a respective 90% accuracy of 22.5 cm and 35.6 cm. The accuracy for dynamic targets is higher than static users because dynamic users enjoy time diversity, which is in favor of a better estimation of the correlation matrix $R_{\mathbf{h}}$ and thus improves localization.

The location error is calculated on a 2D X-Y coordinate base. We also evaluate the accuracy in height estimation. As shown in Fig. 12b, mmTrack estimates the target heights with a median accuracy of 11.49 cm and a 90% accuracy of 22.82 cm. In our evaluation, we calculate the maximum height of a target from the ground.

People counting. We evaluate the accuracy of people counting to examine how accurately mmTrack can detect the number of targets. The results are obtained based on over 18K snapshots of dynamic and static targets. As shown in Fig. 12c, mmTrack estimates the user number accurately when there are fewer than three users. However, when more people present simultaneously, the accuracy may decrease due to inter-person blockage and co-located targets. Overall, mmTrack detects the number of targets without error and with error ≤ 1 for 79.0% and 97.8% of the time respectively. And a target can always be detected by mmTrack as long as he/she is sufficiently separate from and not blocked by other users. As shown in Fig. 15g, mmTrack counts and locates 5 users at different locations accurately. The results, however, show that mmTrack barely overestimates but frequently underestimates the number of users. The main reasons are two-fold: mutual shadowing among users and the dominant specular reflection of mmWave, both causing missing body parts or even missing targets.

Multi-person accuracy. We now evaluate the performance of multi-person localization concerning the number of users. As demonstrated in Fig. 13a and Fig. 13b, mmTrack achieves consistently high location accuracy in both dynamic and static

(a) Accuracy of location   (b) Accuracy of height estimation   (c) Accuracy of counting
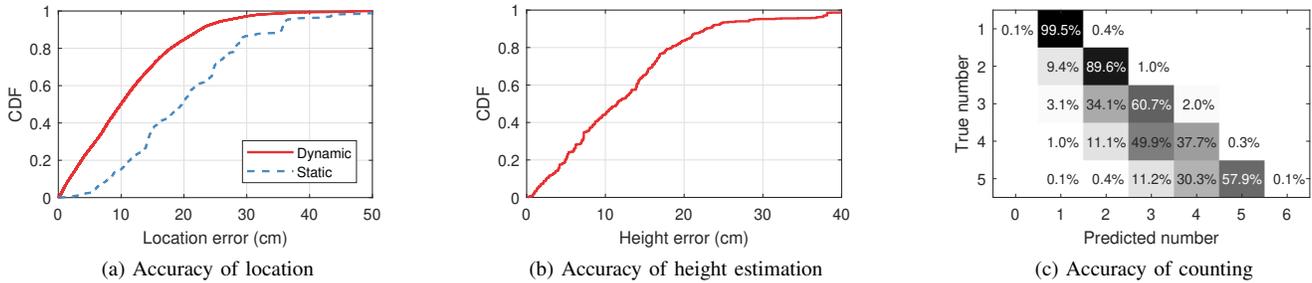
Fig. 12. Overall performance. mmTrack estimates (a) the locations, (b) the heights, and (c) the number of users presented in the area of interests.


(a) Impact of user number (dynamic)   (b) Impact of user number (static)   (c) Impact of range
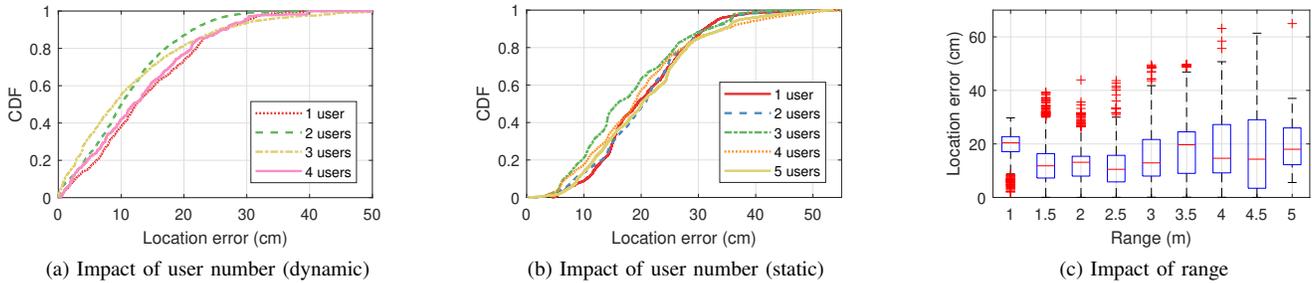
Fig. 13. Multi-person performance. mmTrack achieves similar performance in both (a) dynamic and (b) static scenarios for different number of users. (c) The accuracy also retains over larger ranges.

(a) Location accuracy comparison   (b) Counting without robust clustering
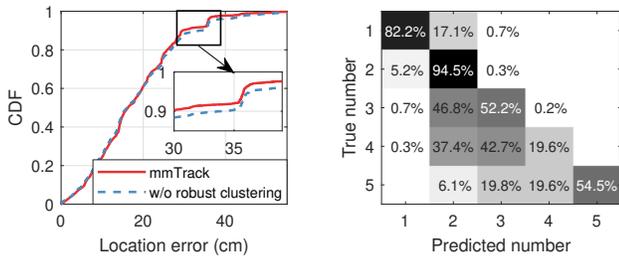
Fig. 14. Benefits of robust clustering, which improves (b) the target detection accuracy significantly, yet only yields marginal gains in (a) location accuracy.

cases for multiple targets. Recalling Fig. 12c, we conclude that more users will lead to more miss detection, but do not affect much the location accuracy. Once a target is detected, mmTrack can estimate the location at remarkable accuracy.

Impact of range. The performance of mmTrack may vary over the distance from the target to the device. As depicted in Fig. 13c, larger errors occur at larger distances, although the median accuracy does not degrade, rendering increased variance of errors. For example, the 90% error touches about 60 cm at the range of 4 m, which is twice lower at 1 m. This is as expected since the reflection signals become weaker and reflection body parts become smaller at larger ranges.

Impact of robust clustering. We examine the gains of the proposed techniques for robust clustering. As depicted in Fig. 14b, without robust clustering, the target counting performance deteriorates considerably. These counting errors, however, do not necessarily lead to location errors, as shown in Fig. 14a. The reason lies in the use of medoid for target location estimation. The medoid of a mistaken cluster still belongs to

one of the true targets and thus endures small location error.

Performance of continuous tracking. We have evaluated the accuracy of localization and tracking. Now we illustrate some examples of successive tracking trace in Fig. 15. As seen in Fig. 15a and 15b, mmTrack recovers the moving trajectories precisely for a single user, regardless of the trace shapes or locations. When there are two or more users, body blockage may happen. As portrayed in Fig. 15c to 15f, except for those blocked moments, the traces of different users are well maintained and recognized. However, successive miss detection may disconnect a specific user's moving trace as segments, as shown in Fig. 15c and 15f.

VI. RELATED WORKS

Device-free Tracking. Wireless signals have been widely studied for device-based [10], [26], [27], [34] and device-free [3], [11], [13], [23], [30], [33] indoor localization and sensing. Past device-free work deploy sensor networks and measure RSS changes for localization [3], [30] and may involve training [21], [22]. Other than RSS, the finer-grained Channel State Information (CSI) is explored to achieve better accuracy [11], [18], [19], [28]. To locate a user, these works either attempt to extract channel parameters such as Angle of Arrival [11], Doppler frequency shifts [18], and/or Time of Flight [8], [19], [29] or employ CSI measurements as location signatures [28]. While high precision could be achieved, they are mostly limited to a single moving target. More antennas [29] or more receivers [9], [24] are required to achieve localization of more than one targets. In one word, these approaches using 2.4GHz/5GHz WiFi have difficulties resolving multipaths and further addressing multiple targets due to the fundamental limits in bandwidth and antenna number.
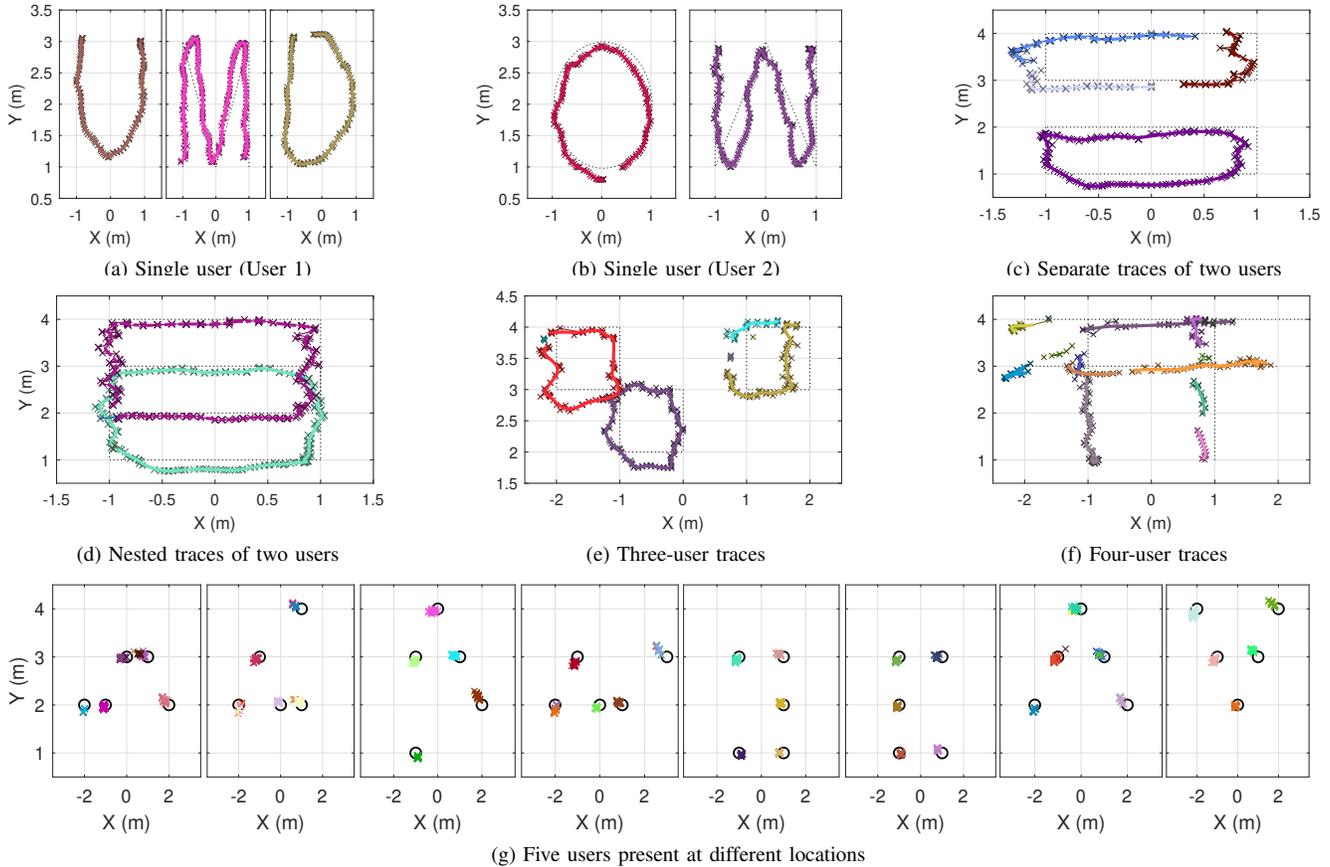
Fig. 15. Example tracking results. Every color indicates a successive trace. Raw estimates are marked with 'x' and the smoothed traces are portrayed on top of them. Ground truths are marked in dashed lines. (a)(b) Traces in character shape by a single user. (c) When two users walk side by side (keeping 2 m separation), the farther target is not detected when blocked by the other user, making his trace disconnected. The locations of either (d) two users or (e) three users are tracked accurately and continuously in the absence of blockage. (f) Missing detection happens more frequently when there are four moving users. (g) In the case of five users, we mainly evaluate localization in static scenarios since a user could be frequently blocked by another if all of them are moving.

The emerging 60GHz radios bring new opportunities and have been utilized for human tracking [14], [17], [25], [36], sensing [32] and imaging [37]. With larger bandwidths, 60GHz radios provide finer time-of-flight measurements, which is leveraged in [25] for high-precision passive tracking of hand motion. However, existing works do not address the accurate localization of multiple human bodies. Moreover, the community has not utilized the spatial resolution provided by a large antenna array, yet instead relies on horn antennas to control beamforming mechanically [25], [32]. To our best knowledge, mmTrack is the first implementation and evaluation of passive multi-person localization and tracking on commodity 60GHz mmWave chips. It is also the first to exploit the joint radar communication system using commodity 802.11ad chipsets.

Radar systems. mmTrack operates the COTS 802.11ad networking device in a radar-like mode, yet differs from traditional radar systems that mostly operate in outdoor space for far-field targets and do not address accurate locations of indoor objects [7], [31]. WiTrack2.0 [2] employs an FMCW Radar to achieve localization of up to 5 static users by relying on their breathing motion. Soli [12] is a small mmWave radar that can sense various hand gestures. Different from these works,

mmTrack is implemented by reusing commodity networking chipsets instead of using dedicated radar hardware, a trend recently emerged as the dual-function radar communication design [35]. Furthermore, mmTrack focuses on accurate localization and continuous tracking of multiple humans.

VII. CONCLUSION

This paper presents mmTrack, the first multi-person localization and tracking system using a single 802.11ad device. We achieve high spatial resolution by performing digital beamforming on the antenna array of the 60GHz chipsets. Then we devise a multi-object detection algorithm and a robust clustering technique to locate and count multiple users. mmTrack also recovers the moving trajectories by a continuous tracking algorithm. Experiments on COTS 802.11ad radios demonstrate that mmTrack achieves a median location error of 9.9 cm and 19.7 cm for dynamic and static targets, respectively, with a people counting error ≤ 1 for 97.8% of the time.

ACKNOWLEGEMENT

We would like to thank Qualcomm for providing the 60GHz testbed on which our experiments were carried out. We are also grateful to the Origin Wireless members for their support.

REFERENCES

[1] Qualcomm 60GHz WiGig/WiFi 802.11ad Chipset World's First Smartphone Edition Complete Teardown Report. https://www.researchandmarkets.com/research/3klmm4, 2018.

[2] F. Adib, Z. Kabelac, and D. Katabi. Multi-person localization via rf body reflections. In *Proceedings of USENIX NSDI*, 2015.

[3] M. Bocca, O. Kaltiokallio, N. Patwari, and S. Venkatasubramanian. Multiple target tracking with rf sensor networks. *IEEE Transactions on Mobile Computing*, 13(8):1787–1800, 2013.

[4] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969.

[5] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of ACM KDD*, volume 96, pages 226–231, 1996.

[6] D. Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational statistics & data analysis*, 54(4):1167–1178, 2010.

[7] Y. Jia, L. Kong, X. Yang, and K. Wang. Through-wall-radar localization for stationary human based on life-sign detection. In *Proceedings of IEEE Radar Conference*, 2013.

[8] K. Joshi, D. Bharadia, M. Kotaru, and S. Katti. Wideo: Fine-grained device-free motion tracing using rf backscatter. In *Proceedings of USENIX NSDI*, 2015.

[9] C. R. Karanam, B. Korany, and Y. Mostofi. Tracking from one side-multi-person passive tracking with wifi magnitude measurements. In *Proceedings of ACM/IEEE IPSN*, pages 181–192. IEEE, 2019.

[10] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti. Spotfi: Decimeter level localization using wifi. In *Proceedings of ACM SIGCOMM*, 2015.

[11] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei. Dynamic-music: accurate device-free indoor localization. In *Proceedings of ACM UbiComp*, 2016.

[12] 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*, 35(4):142, 2016.

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

[14] A. Loch, H. Assasa, J. Palacios, J. Widmer, H. Suys, and B. Debaillie. Zero overhead device tracking in 60 ghz wireless networks using multi-lobe beam patterns. In *Proceedings of ACM CoNext*, pages 224–237. ACM, 2017.

[15] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics*, 5(1):32–38, 1957.

[16] J. Palacios, D. Steinmetzer, A. Loch, M. Hollick, and J. Widmer. Adaptive codebook optimization for beam training on off-the-shelf ieee 802.11 ad devices. In *Proceedings of ACM MobiCom*, pages 241–255. ACM, 2018.

[17] I. Pefkianakis and K.-H. Kim. Accurate 3d localization for 60 ghz networks. In *Proceedings of ACM SenSys*, pages 120–131. ACM, 2018.

[18] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson. Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi. In *Proceedings of ACM MobiCom*, 2018.

[19] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu. Widar2. 0: Passive human tracking with a single wi-fi link. In *Proceedings of ACM MobiSys*, 2018.

[20] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[21] A. Saeed, A. E. Kosba, and M. Youssef. Ichnaea: A low-overhead robust wlan device-free passive localization system. *IEEE Journal of selected topics in signal processing*, 8(1):5–15, 2013.

[22] M. Seifeldin, A. Saeed, A. E. Kosba, A. El-Keyi, and M. Youssef. Nuzzer: A large-scale device-free passive localization system for wireless environments. *IEEE Transactions on Mobile Computing*, 12(7):1321–1334, 2012.

[23] B. Wang, Q. Xu, C. Chen, F. Zhang, and K. R. Liu. The promise of radio analytics: a future paradigm of wireless positioning, tracking, and sensing. *IEEE Signal Processing Magazine*, 35(3):59–80, 2018.

[24] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie. Lifs: low human-effort, device-free localization with fine-grained subcarrier information. In *Proceedings of ACM MobiCom*, 2016.

[25] T. Wei and X. Zhang. mtrack: High-precision passive tracking using millimeter wave radios. In *Proceedings of ACM MobiCom*. ACM, 2015.

[26] C. Wu, Z. Yang, and Y. Liu. Smartphones based crowdsourcing for indoor localization. *IEEE Transactions on Mobile Computing*, 14(2):444–457, Feb 2015.

[27] C. Wu, F. Zhang, Y. Fan, and K. J. R. Liu. RF-based inertial measurement. In *Proceedings of ACM SIGCOMM*, August 19-24 2019.

[28] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni. Pilot: Passive device-free indoor localization using channel state information. In *Proceedings of IEEE ICDCS*, 2013.

[29] Y. Xie, J. Xiong, M. Li, and K. Jamieson. md-track: Leveraging multi-dimensionality for passive indoor wi-fi tracking. In *Proceedings of ACM MobiCom*, 2019.

[30] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An. Scpl: Indoor device-free multi-subject counting and localization using radio signal strength. In *Proceedings of ACM/IEEE IPSN*, 2013.

[31] Y. Xu, S. Wu, C. Chen, J. Chen, and G. Fang. A novel method for automatic detection of trapped victims by ultrawideband radar. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8):3132–3142, 2012.

[32] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra. Monitoring vital signs using millimeter wave. In *Proceedings of ACM MobiHoc*. ACM, 2016.

[33] M. Youssef, M. Mah, and A. Agrawala. Challenges: device-free passive localization for wireless environments. In *Proceedings of ACM MobiCom*, 2007.

[34] F. Zhang, C. Chen, B. Wang, H. Lai, Y. Han, and K. J. R. Liu. Wiball: A time-reversal focusing ball method for decimeter-accuracy indoor tracking. *IEEE Internet of Things Journal*, 5(5):4031–4041, Oct 2018.

[35] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang. Radar and communication coexistence: An overview: A review of recent methods. *IEEE Signal Processing Magazine*, 36(5):85–99, 2019.

[36] Y. Zhu, Y. Yao, B. Y. Zhao, and H. Zheng. Object recognition and navigation using a single networking device. In *Proceedings of ACM MobiSys*, 2017.

[37] Y. Zhu, Y. Zhu, B. Y. Zhao, and H. Zheng. Reusing 60ghz radios for mobile radar imaging. In *Proceedings of ACM MobiCom*, pages 103–116. ACM, 2015.