Continuous Multi-user Activity Tracking via Room-Scale mmWave Sensing

Argha Sen
IIT Kharagpur, India
arghasen10@gmail.com

Anirban Das*
NIIT University, India
anirbanfuture@gmail.com

Swadhin Pradhan
Cisco Systems, USA
swadhinjeet88@gmail.com

Sandip Chakraborty
IIT Kharagpur, India
sandipc@cse.iitkgp.ac.in

ABSTRACT

Continuous detection of human activities and presence is essential for developing a pervasive interactive smart space. Existing literature lacks robust wireless sensing mechanisms capable of continuously monitoring multiple users' activities without prior knowledge of the environment. Developing such a mechanism requires simultaneous localization and tracking of multiple subjects. In addition, it requires identifying their activities at various scales, some being macro-scale activities like walking, squats, etc., while others are micro-scale activities like typing or sitting, etc. In this paper, we develop a holistic system called *MARS* using a *single* Commercial off-the-shelf (COTS) Millimeter Wave (mmWave) radar, which employs an intelligent model to sense both macro and micro activities. In addition, it uses a dynamic spatial time-sharing approach to sense different subjects simultaneously. A thorough evaluation of *MARS* shows that it can infer activities continuously with an accuracy of $> 93\%$ and an average response time of ≈ 2 sec, with 5 subjects and 19 different activities.

KEYWORDS

mmWave, FMCW Radar, Multi-user Activity Recognition

1 INTRODUCTION

Imagine living in an intuitively interactive space without the need to understand its grammar. One doesn't need to interact in a specific way or use *select* voice commands [57] or always wear something [27]. Sharing this intelligent space with others does not also degrade the individual user experience. Interestingly, this vision of seamless smart spaces is not novel and quite dated [20]. However, we are yet to occupy this kind of space regularly. For this vision to become an everyday reality, we argue that there is a need for multi-user continuous room-scale activity tracking through passive sensing. This paper attempts to create an activity-sensing system that can be used to make indoor living spaces truly intelligent.

However, it is crucial to establish what features make such a passive activity-sensing system desirable for wide-scale deployment. Learning from the decades of research in wireless sensing [8, 9, 25, 34, 62], we argue that: 1) monitoring multiple subjects, 2) monitoring different activity over time (for single subject), 3) multi-activity support including macro-scale (activities involving significant body movements) and micro-scale (activities involving minor body movements) activities, 4) real-time inference of activities, and 5) continuous subject tracking[1] are critical for such

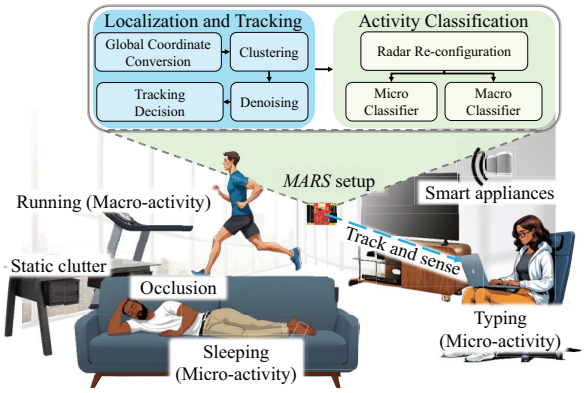


Figure 1: Overview of *MARS* .

a system to be successful and widely adopted. Notably, existing works [10, 12, 23, 34, 50, 52] closest to our vision primarily focus on a subset of the above objectives. For example, [10, 12, 23, 50] supports single-user activity tracking. [34] can track multiple users but only considers short-duration actions. On the other hand, [39] supports a large class of activity recognition; however, for multi-user tracking, the developers must provide feedback to locate and track users using a $1 \times 1 \times 2m^3$ bounding box for isolation.

We examined different modalities for our envisioned system to achieve continuous passive wireless sensing, including Wi-Fi, UWB, acoustic, and mmWave. As our goal is to make the system self-contained, easy to maintain, and readily deployable, we intentionally ignore the multi-modal approaches [25, 44] due to their added complexity and modalities where users need to carry or wear something [35]. Acoustic sensing is compelling because of its cheap hardware but suffers from range, resolution, and the impact it has on users [31]. Wi-Fi sensing has also been deployed extensively in the literature [8, 9, 11, 21, 26]. However, as discussed in [16], Wi-Fi has limited resolution when simultaneously separating activities at different scales (macro and micro) due to its narrow bandwidth. UWB and mmWave are the most compelling technologies that provide high bandwidth and hence, better resolution for sensing a wide range of activities [28, 50, 61]. However, UWB overlaps with the Wi-Fi in the supported frequency range and thus may experience high interference from Wi-Fi deployments; therefore, we consider mmWave to realize the above vision. Recent exploration in the direction of identification [63], position tracking [56], action recognition [28, 50], vital-signs [19, 60], speech sensing [33, 36], etc. justifies the practicality of mmWave sensing for daily activity monitoring. However, we observe that current mmWave sensing literature does not address continuous human activity monitoring

*The author was affiliated to IIT Kharagpur during this project.
[1]We use the term *continuous tracking* to indicate time-shared tracking of multiple subjects, where the system may *intermittently* switch from one subject to another, but the time of switch should be less than the typical activity duration, ensuring every performed activity by the subjects is tracked.

979-8-3503-6201-5/24/$31.00 ©2024 IEEE
DOI 10.1109/IPSN61024.2024.00018

over a longer time or space in an indoor environment, thereby restricting it from being a pervasive practical solution.

Gaps: The majority of previous studies [10, 12, 50, 59] have reported a high level of sensing accuracy as the subject is kept within the main lobe (−15°< radar lobe angle < 15°) of the radar's field-of-view (FoV). In practice, however, the indoor movement of a subject can be completely random, and thus, activities cannot be detected when the subject is outside the radar's FoV. Incorporating multiple mmWave radars to track multiple users within the same room will significantly increase the complexity due to the complex interference patterns from multi-path signals over the same or overlapping frequency bands. Furthermore, there are countless activities a user can engage in, ranging from macro activities involving major body movements (like cleaning the room) to micro activities involving lesser body movements (like typing on the phone). Most of the previous works [10, 50] primarily consider macro activities for a single user, which are easy to detect due to the rich doppler patterns in the reflected mmWave signals. Nevertheless, in reality, a subject can perform both macro and micro activities over time, whereas different subjects can work on different things simultaneously. Also, tracking activities from multiple subjects is challenging as household objects and motion artifacts across subjects can cause noise from static and dynamic multi-path reflections. In a nutshell, in contrast to the existing works, the critical challenges are three-fold that we aim to address in this paper: 1) handling multiple users, 2) addressing multiple activities seamlessly, and 3) making it continuous in real-world settings with COTS mmWave devices.

Motivated by these gaps and empowered by our vision, we first divide the activity grammar into two subsets – (i) *Macro activities* that involve significant body movements (like *changing clothes*) and (ii) *Micro activities* that need minor movements of body parts (like *typing*). Next, to track users' activities seamlessly, we divide the problem into two parts (Figure 1): (i) *localization and tracking* in a way that multiple users' positions can be tracked in every scenario with a single mmWave radar and (ii) continuous opportunistic *activity monitoring* to distinguish both macro and micro activities. The primary challenges involved in the multi-user localization with a single mmWave radar are: (i) scenarios when the subject is present inside the room but not within the FoV of the mmWave sensor, (ii) creation of *zombie* subjects due to multi-path reflections, (iii) associating subjects based on their Radio Frequency (RF) reflections in scenarios when the users cross each other, and (iv) blind-spots during multi-user tracking due to occlusions by other subjects. Additionally, for continuous activity monitoring across multiple subjects, detecting both macro and micro-scale activities simultaneously with the same mmWave radar configuration is not feasible. For example, the radar with a high-doppler resolution can capture better micro movements but adds more noise in capturing macro activities. In contrast, low-doppler resolution can capture macro movements but fails to detect micro activities.

Contributions: To mitigate these challenges, in this work, we propose ***MARS*** , a mmWave-based sensing system: **M**ulti-user **A**ctivity tracking via **R**oom-scale **S**ensing. In summary, we contribute in the following ways:

(1) We build an end-to-end prototype for continuous multi-user activity monitoring using a single mmWave Frequency Modulated Continuous Wave (FMCW) radar using a novel technique that rotates and scans complete 360° opportunistically. The approach develops methods for dealing with zombies, static clutters, and blind spots utilizing a single rotating radar, thereby avoiding the complex interference patterns that arise from multiple radars. Further, *MARS* can track complex multi-user movements (like two users walking in opposite directions) in less than 5 sec latency (median latency ≈ 2 sec) by employing intelligent handling of pointcloud clusters captured from a single radar.
(2) *MARS* employs a novel method of differentiated stacking of the captured range-doppler frames as well as opportunistic switching of radar configurations in order to detect macro and micro activities simultaneously. By doing so, to the best of our knowledge, we design a system that can monitor the *highest* number of human activities in the mmWave domain (1.6× nearest baseline Vid2Doppler [10]). In contrast to the existing works, *MARS* can run on an edge device for real-time monitoring of activities performed by multiple subjects within a room.
(3) We performed a thorough evaluation of *MARS* at diverse setups and have shown its superiority compared to several other baselines. In classifying the macro and micro activities, we can achieve an accuracy of 97% and 93%, respectively, with an average response time of ≈ 2s. We open-source our implementation and sample dataset to reproduce our results: https://github.com/arghasen10/MARS.git.

2 PRELIMINARIES AND PILOT STUDY

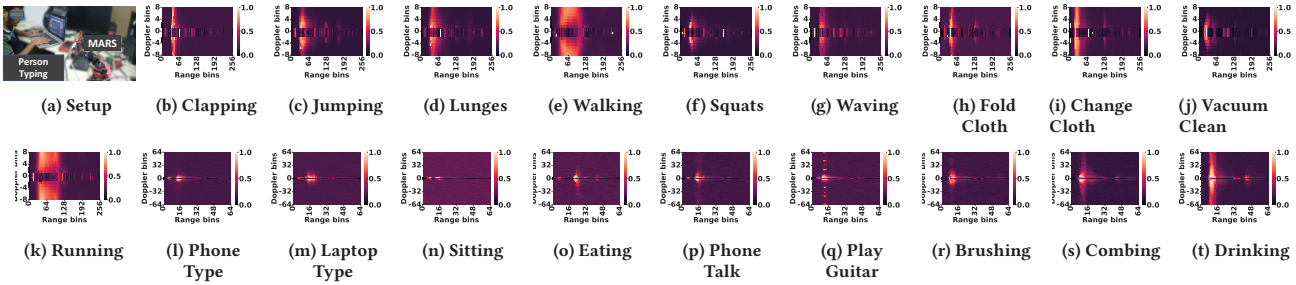
In this section, we empirically illustrate the key foundational underpinnings of *MARS* through pilot studies.

2.1 Preliminaries

The primary working principle of COTS mmWave radars is centered on FMCW [42] that transmits continuous frequency chirps and performs *dechirp* operation by combining the transmitted signal (TX) with the signal reflected (RX) from objects to create an *Intermediate Frequency* (IF) signal. From this IF signal, we extract (1) *Pointcloud*, a discrete set of points representing the detected objects and (2) *Range*, the distance of the detected objects from the radar.

2.1.1 Range estimation. The distance information between the object and the radar can be obtained by measuring the frequency difference between the reflected and transmitted signals [42]. This frequency gap, also known as *beat frequency* (f_b), arises after a Round Trip Time (RTT) of, say τ. If T_C is the transmit time of the mmWave chirp across a bandwidth of B, then the slope of the FMCW chirp can be given as $S = \frac{B}{T_C} = \frac{f_b}{\tau}$. The RTT delay, τ, can be specified as $\tau = \frac{2d}{c}$ where d is the *distance of the detected object* and c is the *speed of light*. Thus, the detected object's distance can be given as, $d = \frac{c}{2}.\frac{T_C}{B}.f_b$. To calculate f_b, a Fast Fourier Transform (FFT), called *range-FFT*, is performed on the IF signal, which produces frequency peaks at locations where the reflecting object is present. Locating these frequency peaks in turn estimates the range.

2.1.2 Velocity estimation. To measure the velocity of a moving target, the radar transmits N number of chirps separated by a transmission time of T_C. If a subject moves with a speed of v, the



(a) Setup (b) Clapping (c) Jumping (d) Lunges (e) Walking (f) Squats (g) Waving (h) Fold Cloth (i) Change Cloth (j) Vacuum Clean

(k) Running (l) Phone Type (m) Laptop Type (n) Sitting (o) Eating (p) Phone Talk (q) Play Guitar (r) Brushing (s) Combing (t) Drinking

Figure 2: Standard deviation (std) in the range-doppler heatmaps captured during the entire activity duration. (a) Sample setup for data collection, (b)-(k): Macro activities with low doppler resolution, (l)-(t): Micro activities with high doppler resolution. Activities having similar body movements have similar patterns, but the difference can be captured in the temporal domain.

phase difference between two successive RX chirps corresponding to the motion, vT_C, can be given as, $\Delta\phi = \frac{4\pi vT_C}{\lambda}$. A second FFT, called *doppler-FFT*, is performed on these phasors to determine the movement or velocity of the object. This information is captured in a 2D matrix called *range-doppler* $\mathbb{D}_{D\times R}$ where D and R correspond to the numbers of *doppler bins* and *range bins*, respectively.

2.1.3 Pointcloud estimation. The pointcloud is estimated through the standard CFAR algorithm [38] that detects peaks of the range-doppler matrix corresponding to the detected objects. More details on the pointcloud estimation can be found in [43]. The pointcloud consists of the coordinates (x_i, y_i, z_i), doppler variation (d_i), and the received power (p_i) of the detected objects. The pointcloud set (S) for N number of detected objects can be given as $S = \bigcup_{i=1}^{N}\{(x_i, y_i, z_i, d_i, p_i)\}$.

2.2 Pilot Study

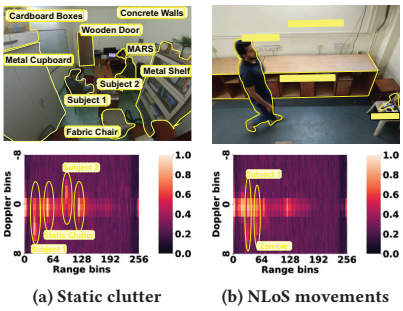
We consider 19 different activity classes from *Activities of Daily Living* (ADLs), *Instrumental Activities of Daily Living* (IADLs) [7], and *daily indoor exercises* – (i) *macro activities* like walking, running, jumping, clapping, lunges, squats, waving, vacuum cleaning, folding clothes, changing clothes, and (ii) *micro activities* like laptop-typing, phone-talking, phone-typing, sitting, playing guitar, eating food, combing hair, brushing teeth, and drinking water. In contrast to the existing literature that primarily uses voxelized pointcloud [14, 50, 55] or 1D doppler [10, 12], in this paper, we explore range-doppler 2D heatmaps for activity classification; the primary motive is to find a parameter that can detect both macro and micro activities simultaneously from different users. For this purpose, we conducted a set of *pilot experiments* to explore to what extent range-doppler information can be used in capturing human activity signatures and how the indoor setting impacts such sensing capability.

2.2.1 Feasibility study for range-doppler. Figure 2 shows the standard deviation in the range-doppler heatmaps captured during the activity. Notably, standard deviation technique removes static powers (-3 to 2 doppler bins) in the heatmap; thus, we see a low-power value in these doppler bins. We observe that each activity has different signatures captured by the range-doppler heatmap. Although the plotted standard deviation looks similar for some activity pairs with similar body movements (like walking/running, jumping/lunges), there are temporal changes in the heatmaps; for example, "running" induces a faster change than "walking". Thus, combining the observations from range, doppler, and time, we can get different signatures. Notably, the macro activities have more robust patterns due to the magnitude of movement involved. Even though the micro activities have relatively weaker signatures, they can be distinctively captured with a *higher doppler resolution* (−64 to +64 doppler bins, in contrast to −8 to +8 doppler bins used for macro activities).

2.2.2 Impact of static clutters. Static clutters are any object (walls, furniture, etc.) that are stationary but can reflect the mmWave signal and therefore, generates unwanted signatures in the range-doppler data. We consider a scenario with two subjects – *Subject 1 and Subject 2*, both sitting inside the room, as shown in Figure 3(a). The room also contains multiple static clutters, such as wooden sheets and walls. From the corresponding range-doppler heatmap, we observe multiple peaks at the range bins corresponding to both the subjects and the static clutters. Indeed, the static clutters produce a higher magnitude along the zero doppler axis, thereby signifying zero or no movement. On the other hand, the dynamic movements of the subjects are positioned across non-zero doppler bins. A major takeaway from the range-doppler heatmap is that static clutters are easily identifiable by their zero-doppler signatures.

2.2.3 The effect of Non Line of Sight (NLoS) movements. To study the NLoS reflections, we first ask a single subject (Subject 1) to stand close to a wall and make some movements (macro-scale) as shown in Figure 3(b). From the corresponding range-doppler heatmap, it can be observed that the subject's movements are captured at two different instances at two different range bins. Of the two visible peaks, the more substantial peak belongs to the actual user's movement, whereas the other instance, also termed as a *zombie subject*, occurs due to the multi-path reflection from the wall.

2.2.4 Impact of radar configurations on determining users' activity. To understand how the radar configuration affects the patterns in the activity signatures, we ask one subject to switch his activity from jumping to two micro activities, namely, sitting in a chair and phone typing, and finally, walking out of the room. The subject is asked to repeat the pattern twice to collect the corresponding range-doppler data under *low and high doppler resolution*. From the



(a) Static clutter (b) NLoS movements

Figure 3: Range-doppler signatures for two experimental scenarios.

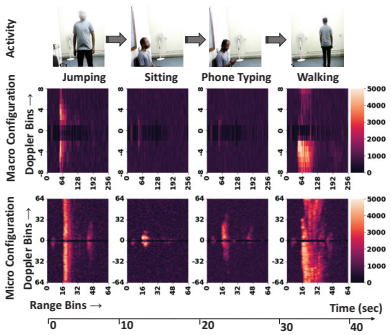
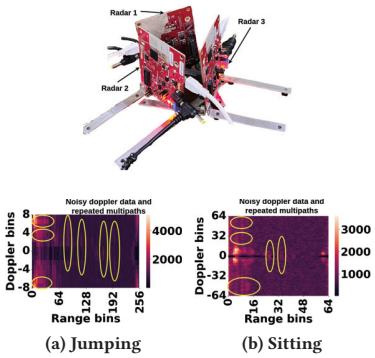


Figure 4: Range-doppler signatures (standard deviation for individual activity windows) over time.



(a) Jumping (b) Sitting

Figure 5: Multi-radar challenges.

std in the heatmap across the entire activity time axis Figure 4, it is evident that low doppler resolution is adequate for capturing macro activities like walking and jumping. Still, typing and sitting does not have any significant signatures. On changing the radar configuration to *high doppler resolution*, we observe that micro activities like typing and sitting have better visibility. However, with this, the macro activities (walking, jumping) generate noisy data due to the higher resolution. Therefore, *different doppler resolutions* is crucial to capture the signatures corresponding to different activities.

2.2.5 Impact of multiple radars. To have entire room coverage, we have taken three radars and kept them in a colocated position with 120°to each other as shown in Figure 5. We observe that incorporating multiple radars within the same room leads to complex interference patterns in the range-doppler heatmaps. The same or overlapping frequency bands lead to interference in the mmWave chirps and also cause more multipath effects. As shown in Figure 5, the range-doppler heatmaps are very noisy and have complex interference patterns which are not easily separable. This indicates that using multiple radars for 360°coverage makes the system complex; therefore, we need some alternate solution.

3 METHODOLOGY

We consider a single mmWave radar to track multiple subjects performing diverse activities over time. *MARS* relies on the following assumptions about device constraints and activity grammars:

(1) We consider that human activities can be classified into *macro* and *micro* based on the amount of body movements involved. *Micro* activities involve low-velocity movements, where the user is primarily static and performs the activity with minor movements in the body parts, like typing on a phone or laptop, brushing, combing, etc. Macro activities involve more extensive body movements like exercises, folding or changing clothes, vacuum cleaning, etc.
(2) At a single instance, the subject performs either a macro or a micro activity, but not both (like typing on the phone while exercising; the macro activity will suppress the micro activity); however, they can switch between macro and micro activities over time.
(3) Each activity is performed for a minimum duration Δ. The value of Δ depends on the hardware setup, particularly the response time of the radar to complete a 360° scan of the entire room. We consider $\Delta \geq 5sec$ in our setup based on the hardware components used to develop the prototype.
(4) The radar needs to be placed at a location such that two subjects should not occlude each other while performing some micro-activities (occlusion can be handled for macro-activities like walking or running, as we discussed in Subsection 3.1.5).
(5) For similar types of micro-activities like *phone typing* and *laptop typing*, the surface area of the devices differentiates the activities; for example, it is assumed that the surface area of a smartphone keyboard will be much smaller than the surface area of a laptop keyboard.

To have the end-to-end user localization and activity monitoring pipeline based on the above assumptions, we divide the problem into two sub-problems as highlighted in Figure 6: (i) subject detection followed by the localization and tracking of the subjects, and (ii) activity classification for individual subjects. Next, we discuss two modules addressing these sub-problems.

3.1 Localization and Tracking

MARS relies on the *pointcloud data* to localize subjects and track their movements. Motivated by the challenges discussed in Subsection 2.2, we perform the following steps.

3.1.1 Isolate subjects from static clutters. In practice, multiple static objects can be present within the FoV of the radar. As we are interested in identifying subjects' movement, the background, corresponding to stationary objects, needs to be removed. For this, we remove the zero-valued doppler bins for segregating the static objects (clutters). With this step, the mmWave radar can generate a pointcloud that does not contain static obstacles to isolate the subjects. Once *MARS* starts receiving the pointcloud, it tracks subjects by converting its pointcloud coordinate to a global coordinate.

3.1.2 Global Coordinate Conversion. When the subject is present within the room but outside the radar's FoV, localization, and activity recognition of the subject is not feasible. As a solution, we

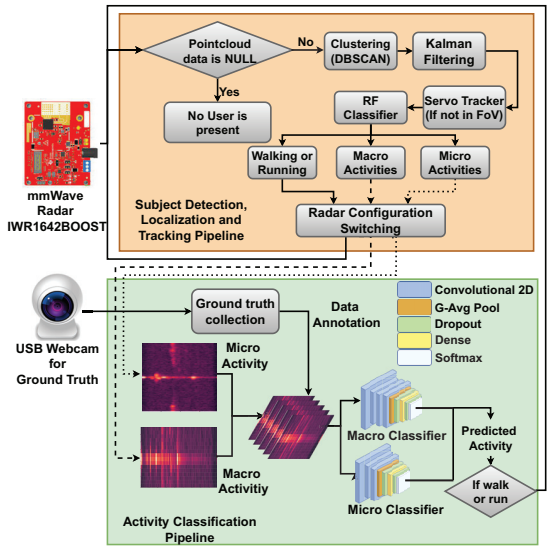


Figure 6: System architecture.

mount the mmWave radar on top of the rotor axis of a servo motor. This enhances the FoV of the radar to 360°. However, rotating the sensor will directly change the reference coordinate system of the estimated pointclouds. Therefore, instead of keeping the local coordinate system w.r.t. the radar, we use a magnetometer to keep a global reference coordinate system. The magnetometer provides the reference azimuthal angle w.r.t. the earth's magnetic pole. Consider a user at $P(x, y)$ in the radar coordinate system. The radar is oriented by an angle of θ w.r.t. the magnetometer. So in the global coordinate system, the angular position of the object is at $(\theta + \phi)$, where $\phi = \tan^{-1}(\frac{y}{x})$. Equation 1 illustrates the transformation of the radar coordinate system to the global coordinate system.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} r\cos(\theta+\phi) \\ r\cos(\theta+\phi) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

With the above transformation matrix, the pointclouds are now referenced w.r.t. the magnetometer and does not suffer from any coordinate shift due to the rotation of the radar.

3.1.3 Tracking multiple subjects. Based on the pointcloud data, we have information about all the subjects; however, we also get noisy pointclouds due to the movements of the subjects. To tackle this, we take the pointcloud data in a queue format and pass this information to Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [24] for clustering. Each cluster is associated with a unique ID to associate subjects with their respective clusters. Now, to detect the presence of a new subject, *MARS* compares the detected pointcloud clusters between two consecutive frames received from the radar. If the Euclidean distance between the centroid of a cluster over a new frame and that over the previous frame is less than ϵ, we keep the respective cluster ID the same as before. In the case of a newly discovered cluster, we assign it a new cluster ID, indicating a new subject. Note, ϵ is a hyperparameter, and in our setup, we keep it as 10cm, signifying the minimum range resolution for a subject.

3.1.4 Tracking movement of individual subjects. After the clustering step in the pipeline, each cluster corresponds to the pointcloud information associated with each subject. However, some of these clusters may correspond to zombie subjects, as discussed earlier. We observe that *pointclouds for zombie clusters have a low reflection power* and thus are generated less frequently when compared to the pointclouds caused by the actual subject. So, we first apply a mode function on the pointcloud queue for each cluster to filter out the pointclouds generated more frequently due to the actual subject's presence. The remaining noisy outliers get removed with this approach. However, due to the uncertain movements of the subjects, two subjects may impede each other while crossing. This may lead to *blind spots* in the pointcloud data.

3.1.5 Handling blind spots during multi-user tracking. To track each subject seamlessly, we apply a Kalman filter [40] on the pointcloud queue for each cluster. The Kalman filtering technique uses the prior knowledge of the state of an object and then *predicts* and *updates* the location and velocity of that object for the next frame. For precise tracking of individual clusters instead of a static Kalman gain, we opted for Recursive Kalman Filter (RKF) to estimate the subjects' motion states. RKF can recursively generate the error covariance matrix and Kalman gain at each stage of the update process. With this step, we can estimate the subjects' state when the actual pointcloud data is unavailable due to occlusion by other subjects' movement or errors in the former pipeline.

3.1.6 Servo-based tracking. Usually, the azimuthal FoV of the radar is 120^o which can localize and track subjects. To enhance this FoV, we rely on servo-based tracking. As soon as we have the final coordinates of the denoised pointclouds, we check if each subject is within the main lobe of the radar, i.e., $\leq \pm 15°$. Otherwise, we rotate the servo towards the subject by an angle of $\tan^{-1}(\frac{y}{x})$, to generate high-fidelity pointclouds and range-doppler heatmaps which are needed for activity classification. We stop the rotation when the subject is within $\pm 15^o$ so that the doppler remains unaffected for the next activity classification task. Gradually with the rotation process, a new pointcloud queue is generated for subjects that were earlier outside the FoV.

3.1.7 Handling Multi-user Localization under Mobility. Occasionally, a user may completely exit the radar FoV due to complex mobility scenarios involving multiple users. For example, when the radar tracks one user, another user may work in the opposite direction. To handle such scenarios, the system maintains these global coordinate clusters generated by the pointcloud queues for each user, and the servo rotation process is scheduled sequentially for each subject during the next activity classification task. If a user completely exits the FoV, whether by moving in the opposite direction of the current tracking FoV or by leaving the indoor space, *MARS* detects such a scenario by observing a reduction in the current cluster count by one. In response, we initiate a full 360°rotation search. This search ensures that we can re-establish the location information of the lost user in complex scenarios. Additionally, it checks if the cluster count reduction is due to a user moving out of the indoor space, and gradually removes the cluster in the next frame. Thus, we continuously track multiple users by associating them with their global coordinate clusters, ensuring that we never

lose sight of them even when the servo rotates to a different FoV and solve the challenge of multi-user tracking under mobility. As explained in Section 3.1.4, it's worth noting that occasionally, a cluster may form due to noisy pointclouds resulting from multipath reflections. Eventually, however, such clusters will be eliminated due to the mode filter applied to each cluster's pointcloud queue.

3.1.8 Monitoring state change of a subject. Once a subject is tracked, *MARS* monitors the possible state changes of that subject by utilizing the pointcloud data. Broadly, it performs a high-level classification to check whether (i) the subject is walking or running inside the room or (ii) the subject is static and performing some macro/micro-activities. For this purpose, we capture the mean, standard deviation, kurtosis, and skewness in the denoised pointcloud queue for each cluster for a time window of 1 seconds. These features are fed to a Random Forest Classifier to predict the subject's activity scale. Based on the prediction, we continue the localization and tracking if the subject is walking or running. Else, *MARS* enables the macro or micro activity classifier opportunistically based on the inference.

3.2 Macro/Micro Activity Monitoring

We keep two different radar configurations to capture the classes of micro and macro activities. For macro activities, *MARS* uses a low doppler resolution of 16 doppler bins (captures major body movements but eliminates the details that may generate noise), while for capturing micro activities, it uses a high doppler resolution of 128 doppler bins (captures minor body movements with finer details). Using range-doppler enables us to easily switch the radar configurations at different resolutions to recognize both macro and micro activities from different users.

3.2.1 Segregation of individual subject's activity signatures. As shown in Figure 2, range-doppler is represented as a heatmap image, where the abscissa is the range, the ordinate is the doppler speed containing the power value of subjects' movement. Each subject's activity has its activity signatures in the range-doppler heatmap (See Figure 2). To classify the activity of individual subjects, we first segregate these activity signatures based on the range bins. From the pointcloud data (collected along with the range-doppler), we check if there is a non-zero doppler value in the range profile where the subject is present. If a doppler variation exists, we slice out that Range-doppler heatmap information with padding of ±10 range bins. Additionally, we define another copy of the Range-doppler heatmap for each subject, replacing the remainder with the minimum heatmap value for the subject. In this way, we address the challenge of multiple activity tracking, as each subject has its own activity signatures, and the remaining signatures corresponding to other subjects are suppressed. This modified range-doppler data is fed to the classification model.

3.2.2 Differentiated frame stacking for macro/micro activities classification. These macro or micro activities span over a short period, affecting range-doppler values temporally. We stack 1 sec range-doppler data to capture temporal features, thus achieving a two-dimensional (2D) multichannel array. However, for macro activities, the doppler resolution is low, resulting in a heatmap of size 16×256, while for micro activities, the doppler resolution is high, resulting in a heatmap of size 128×64. This diversity results in different Frames Per Second (FPS) for the range-doppler computation and data transfer. For low-resolution doppler, the FPS is 5, while for the high-resolution doppler, the FPS is 2. Therefore, we stack 5 frames together in the case of the macro activity classifier, while for the micro activity classifier, we stack 2 frames together. This enables us to have the range-doppler for a consistent time period of 1 sec for both scenarios.

Table 1: 2D-CNN architecture (M: macro, μ: micro)

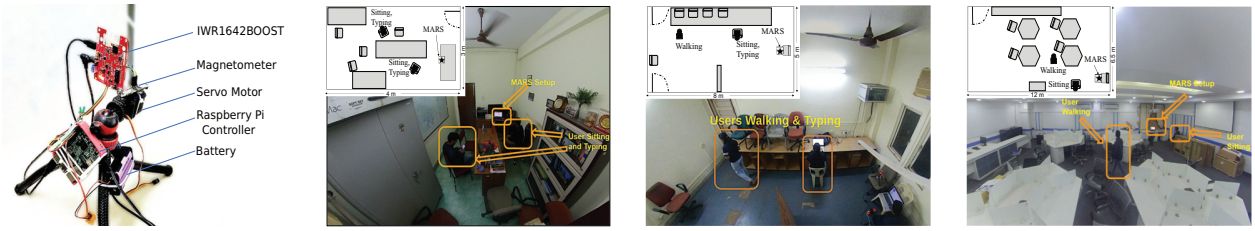
CNN Layer	Parameters							
	Kernel		Stride		Channel		Dropout	
	M	μ	M	μ	M	μ	M	μ
Input Layer	-	-	-	-	5	2	-	-
Conv1	2 x 5	3 x 2	1 x 2	2 x 1	32	32	-	-
Conv2	2 x 3	3 x 3	1 x 2	2 x 2	64	64	-	-
Conv3	2 x 3	3 x 3	1 x 2	2 x 2	96	96	-	-
Conv4	2 x 3	-	1 x 2	-	96	-	-	-
G-avg Pool	-	-	-	-	-	-	-	-
Dropout1	-	-	-	-	-	-	20%	20%
Dense1	-	-	-	-	32	32	-	-
Dropout2	-	-	-	-	-	-	10%	10%
Softmax	-	-	-	-	6	6	-	-

3.2.3 Model Architecture. The 2D range-doppler heatmaps have different spatial patterns for each activity. So, we employ a 2D Convolutional Neural Network architecture (2D-CNN). Convolution 2D operation considers the dependency of neighboring spatial values and the temporal relationship of past t (t = FPS) frames. We use four and three 2D convolutional layers with 'same' padding and Relu activation for the macro and micro activity classifiers. Next, a global average pooling layer is added to extract the average spatial activation across the entire feature map. Finally, we add two successive dropout and dense layers, where the dropout rate is kept as 20% and 10%, respectively. The last layer outputs a joint probability distribution over all possible activities with a softmax activation (detail in Table 1). Although the subject's orientation may not impact the detection of macro activities, the micro activities need precise signatures. As we collect the range-doppler at a higher resolution for micro activity classification, it can sense the movements even when the signal strength is low. As a result, the proposed 2D-CNN model can capture micro activities even when the subject is not directly facing the radar.

3.2.4 Opportunistic Configuration Switching. For each macro and micro activity, *MARS* switches the configuration accordingly (as derived from the step mentioned in Section 3.1.8). Once the activity classification is performed, it checks whether the subjects are still in their activity state. If any subject starts walking or running, the micro and macro classifiers can detect that and switch the configuration back to capture the pointcloud data to reinitiate the *Localization and Tracking* Pipeline. The clustering and denoising filters get restarted to track the subjects' movement.

4 IMPLEMENTATION

As shown in Figure 7a, *MARS* is developed on top of a COTS millimeter wave radar, `IWR1642BOOST` [3]. The system is tested in three different rooms (see Figure 7b, 7c, 7d) – (i) R1, an office cabin of size 4×3 m^2, (ii) R2, a classroom of size 8×5 m^2, and (iii) R3, a laboratory of size 12×6.5 m^2. The ground truth activity of each



(a) (b) (c) (d)

Figure 7: (a) *MARS* hardware setup; and data collection in different rooms: (b) R1, (c) R2, (d) R3.

subject is manually annotated with the help of the video captured using a USB camera. Overall, *MARS* consists of: the front-end radar and the backend processing unit. The radar senses data and generates 2D pointclouds and range-doppler heatmaps. These data entries are transferred via a USB cable with a baud rate of 921600 to the backend `Raspberry Pi-4 Model B` with 1.5GHz Broadcom BCM2711 64bit CPU and 8 GB RAM. We have used `Python 3.9.6`, `TensorFlow v2.10.0`, and `Scikit-learn v1.1.2` for implementing the macro and micro activity classifiers and the opportunistic Random Forest classifier. The models are trained on an iMac-M1 (with 16 GB primary memory running macOS v12.6 with base-kernel version: 21.6.0) and then deployed on the Raspberry Pi-4 for live inference. The training takes 10 minutes for the opportunistic classifier and 20 & 25 minutes for the case of macro and micro classifiers, respectively, with a model size of 7.8 MB, 460 KB, and 334 KB, respectively, for the three cases.

4.1 Hardware setup

4.1.1 Radar Configuration. The IWR1642BOOST radar is configured to use two transmitter and four receiver antennas with frequencies of 77-81 GHz (bandwidth 4 GHz). For the three different use cases, i.e., (i) localization and tracking, (ii) macro activity classification, and (iii) micro activity classification, we have used three different radar configurations (Table 2). For the localization and tracking, we set the frame periodicity as 33.33 millisecond to have 30 FPS to fill the localization queue fast so that clustering and Kalman filter-based tracking can be performed with minimal error. This configuration provides a range resolution of 4.36 cm, with a maximum unambiguous range of 9.02 m. It can measure a maximum radial velocity of 1 m/s, with a doppler resolution of 0.13m/s. The sensor is set to transmit 32 chirps per frame. We use the same radar configuration for the macro activity classification, except we reduce the FPS to 5 to allow the flow of larger range-doppler heatmaps (matrix of size 16×256) via USB. The doppler resolution is kept at 0.01 m/s for the micro-scale activity classification. The size of the range-doppler heatmap is 128×64, supporting 2 FPS frame rate.

4.1.2 Localization and Tracking Setup. To enhance the radar field-of-view to 360°, we have mounted the radar on top of the rotor axis of a `TowerPro MG995 Servo Motor` [2], powered using a 1200mAh Li-ion Rechargeable Battery. This enables the localization and tracking of subjects for the entire indoor space. `GY-273 Compass Magnetometer Sensor` [1] is used to transform the pointcloud coordinates to a global coordinate system.

Table 2: Radar configuration

Parameters	Localization	Macro	Micro
Start Frequency	77 GHz		
End Frequency	81 GHz		
Range Resolution (cm)	4.36		12.5
Maximum Range(m)	9.02		6.4
Maximum Radial Velocity (m/s)	1		0.64
Velocity Resolution (m/s)	0.13		0.01
Azimuthal Resolution (Degree)	14.5°		
Frames per Second	30	5	2
Chirps Per Frame	32		64
ADC Samples per Chirp	256		

4.2 Data Collection Setup

Data collection is carried out for 7 subjects (3 female and 4 male), with ages ranging from 23 to 35, for a total duration of 44 hours across 19 different activity classes (details in 2.2) involving both macro and micro activities. Around 30 hours of the collected dataset are being used as the training set. In total, we have 1584000 samples of the pointcloud dataset, 264000 macro range-doppler samples, and 105600 micro range-doppler samples. In order to train the system for activity classification, the training data is mostly collected in a controlled environment, in which the user is asked to perform certain activities. However, for testing data, other than controlled data collection we also have intentionally kept some scenarios where no subjects are inside the room (around 30 mins of data) and scenarios where subjects can select any task from the activity set and perform in an uncontrolled fashion. Thus we have experimented over different controlled, semi-controlled, and in-the-wild setups, as we also explained later for individual evaluations. To generate the ground truth for localization and tracking pipeline of *MARS* , we manually marked the positions of users' movement in the room's floor map. We asked the users to move in the marked path. We have evaluated the MAE in the marked coordinates and the denoised pointcloud coordinates. Further, we have used *mmWave-Demo-Visualizer* [4] tool, and implemented a patch to extract raw data, containing pointcloud and range-doppler heatmap under different radar configurations. Annotating the video footage captured via another USB camera installed in the room was done with the help of two volunteers.

4.3 Baselines and Performance Metric

We compare *MARS* activity classifier with three different baselines, (i) **Pointcloud-based: RadHAR** [50], which is based on voxelized 3D pointclouds for classifying six macro activities. For developing the baseline, we have collected the 3D pointclouds

using a TI IWR1443ISK [6], and we train the classifier (as provided in [50]). (ii) **Range-Doppler: Vid2Doppler** [10], which used range-doppler data to classify 12 different activities. With our collected datasets, we transfer-learn the model weights using the open-sourced Vid2Doppler classifier model. (iii) **VGG-16 network** [49] which is pre-trained on the ImageNet [22] dataset, we apply transfer learning to learn new model weights w.r.t. our collected range-doppler matrix. This transfer learning approach helps in reducing the feature extraction part, as all the trained convolutional layers in VGG-16 are used as feature extractors and do not require retraining. The base VGG-16 model has been enhanced with 2D-Global Average Pooling and successive Dropout and Dense layers as done in the 2D-CNN Architecture (see § 3.2.3). (iv) **Pointcloud-based: Pantomime** [39] a combined PointNet++ [41] and LSTM based feature extractor for classifying mid-air gestures. The models are trained with a train-test split of 70%-30% and a validation split of 20% from the training set. In evaluating *MARS* against the baselines, we relied on the **accuracy** metric, which calculates the total number of correct predictions over the total number of predictions made. Additionally, we considered **response time**, measuring the time taken to infer the activity class.

5 EVALUATION

This section provides the detailed performance analysis of *MARS* in comparison with the existing baselines.

5.1 Overall Performance

We consider three scenarios to evaluate the overall performance of *MARS* in comparison to the baselines – (i) single subject, multiple activities over time (*Temporal activity diversity*), (ii) multiple subjects, individual subject performs a single activity over time but different subjects may perform different activities (*Spatial activity diversity*), (iii) multiple subjects, each performing different activities over time (*Spatio-temporal activity diversity*). We performed these experiments in a room *R*2; later, we discuss the impact of the room size with a *leave one out* train-test method.

5.1.1 Impact of different activities over time. Here, we asked the subjects to choose four activities (two macro and two micro) in a logical sequence and perform each for at least 10 sec within a room. For example, a subject may first do some exercise through jumping (macro), then sit (micro), then take their phone and type a message (micro), and finally walk to leave the room (macro). As shown in Figure 8(a), *MARS* takes the least response time in inferring the activities with the highest accuracy in comparison to the baselines. We observe that the response time for the first activity takes $\approx$ 2 sec, which involves the bootstrap time to denoise and cluster the data for localizing the subject. When a configuration switch is necessary (macro to micro or vice versa), the average response time is $\approx$ 3.14 seconds. Without a configuration switch, the average response time is $\approx$ 1.08 seconds. In comparison, the baselines perform worst in the response time due to more extended frame stacking (2 sec and 3 sec, respectively, for RadHAR and Vid2Doppler) and longer classifier inference time ($\approx$ 4 sec for VGG-16 and Pantomime). Longer response times of the baselines directly impact lowering the number of hits in the activity time window, as shown in Figure 8(a) w.r.t. *MARS* , which has a low response time due to smaller frame stacking (1 sec) and reduced inference time ($\approx$ 0.08 sec) with a light-weight model architecture.

5.1.2 Impact of multi-user activities. In the second scenario, we pick four subjects and ask three of them to choose one activity from the set of macro activities and the remaining one to choose one from the set of micro activities. After determining the subjects' location and states, *MARS* configures the low doppler resolution and classifies the macro activities simultaneously, with a response time of $\approx$ 3s at the beginning. However, using 1 sec of frame-stacked data, it can gradually infer the three macro activities simultaneously with a response time of 1.04s. For the subject performing the micro activity, it switches the configuration to high doppler resolution and classifies the same with a response time of 2.08 sec, resulting in an average response time of 1.9 sec with an accuracy of 98% in the entire activity time window of 10 sec (see Figure 8(b)). RadHAR and Vid2Doppler show poor performance as they are built focused on macro activities only and are trained only for single-user activity classification. Pantomime, on the other hand, uses separate pointclouds (within $1\times1\times2$ m^3, as defined in [39, Sec 6.4]) for the activity classification task. These pointclouds easily overlap with each other, especially in scenarios where three subjects are performing macro activities, and thus show lower accuracy. Interestingly, we observe that the average number of correct inference for *MARS* is higher in this case (spatial diversity) compared to the previous one (temporal diversity), as the radar needs less configuration switching.

5.1.3 Impact of different activity over time for multi-user. In the final scenario, we ask four subjects to simultaneously perform four different activities of their choice (with at least one micro activity and one macro activity) in sequence within a room, where they switch the activity approximately every 10 seconds. As shown in Figure 8(c), the average response time of *MARS* in this scenario is $\approx$ 2 sec with 94% average accuracy, in a time window of 10s. Thus, the overall performance of *MARS* demonstrates its potential to be adopted as a real-time system for multi-subject scenarios.

5.2 Performance of Opportunistic Classifier

With the dataset collected across different scenarios, under the localization configuration (as mentioned in Table 2), we first perform a train-test split of 70% : 30%. The Opportunistic Classifier (as discussed in Sec. 3.1.8) is trained with the 70% training dataset with a validation split of 20% from the training set. According to our observations, the pointcloud dataset can accurately classify macro and micro activity sets with 90%, and 99% accuracy, respectively. However, a slight overlap exists (of 10%) between the macro activity class and the *walking* or *running* class, as under both the scenarios, there exists a significant variation in the pointcloud data. Interestingly, this variation is similar during the activity initiation period, and gradually, the variation becomes more separable with time.

5.3 Performance of Activity Classifiers

We next evaluate the performance of the macro and the micro activity classifiers w.r.t. the baseline in terms of the accuracy. As shown in Figure 8(d), the accuracy for *MARS* is 97% in the case of macro activities and 93% in the case of micro activities. The lower accuracy for RadHAR is primarily because it relies on the pointcloud dataset

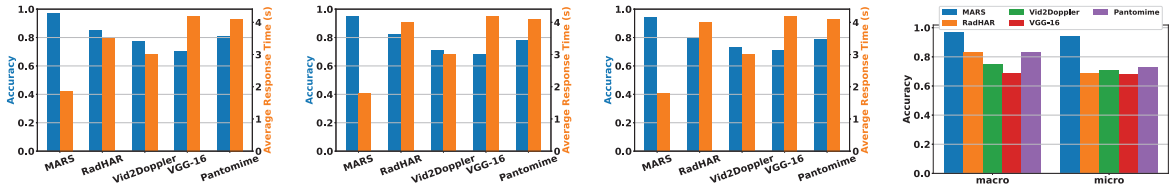


Figure 8: Accuracy and average response time while predicting (a) different activities over time (single-user), (b) multi-user, (c) different activities over time for multi-user, (d) overall accuracy of *MARS* .

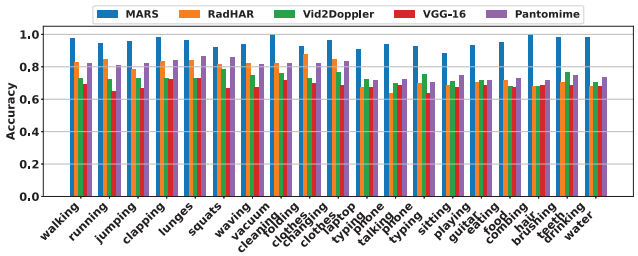


Figure 9: Accuracy across different activities.

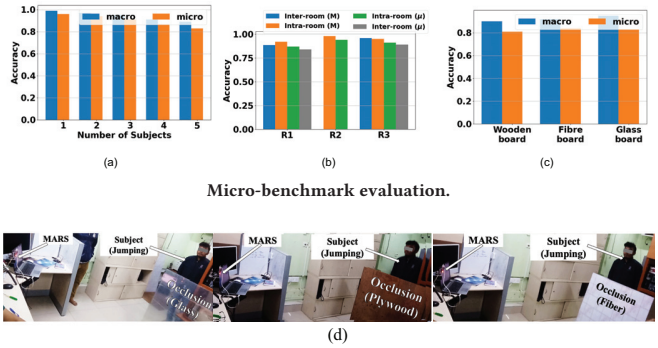


Figure 10: Impact on (a) # of subjects, (b) different rooms (M: macro, μ: micro), (c) different blockages and (d) Scenarios with blockages: glass, plywood, and fiber respectively.

for the voxel formation and generates sparse pointclouds in case of micro activities. Lower accuracy for Pantomime is due to its sparse pointclouds, which can easily overlap with other subjects, and its radial velocity resolution of 0.87 m/s (see [39, Sec 5.1]) which is almost 9× our velocity resolution for macro configuration (0.13m/s). For Vid2Doppler, the poor accuracy is primarily because it takes only 32 doppler bins, which are unsuitable for micro activity monitoring, and the model feature extraction part is pre-trained on macro activity datasets. As the body movements in the case of macro activities are significant, thus the classifier can segregate individual classes with an excellent accuracy (close to ≈ 0.97). In the case of the micro activities, the body movements are less significant, but with the proposed classification pipeline with a higher doppler resolution, we can achieve an accuracy of 0.93. Among the micro activities, laptop typing, eating food, and playing guitar involve higher body movements, and thus for these particular activities, we observe higher accuracy (see Figure 9). Activities such as sitting, typing, and talking on a phone are carried out while subjects sit on a chair. Thus, the doppler shift for these activities is very low. When the subject talks on a phone, the overlap with the *sitting* class is more significant (≈ 10%). In Figure 9, we show activity wise accuracy of *MARS* w.r.t. the baselines. Although *MARS* supports higher number of activities compared to the baselines (19 in *MARS* versus 5 and 12, respectively for [50] and [10]), the classification accuracy of the baselines significantly drops in comparison to *MARS* .

5.4 Results: Micro Benchmarks

5.4.1 Impact of number of subjects. In Figure 10(a), we show the variation in the accuracy for both the macro and micro activity classifiers with the different number of subjects present inside the room. With an increase in the number of subjects up to 5, we observe a direct impact on the accuracy for both macro and micro classifiers, but by only ≈ 10%. We compared *MARS* with the baselines and have reported in Table 3.

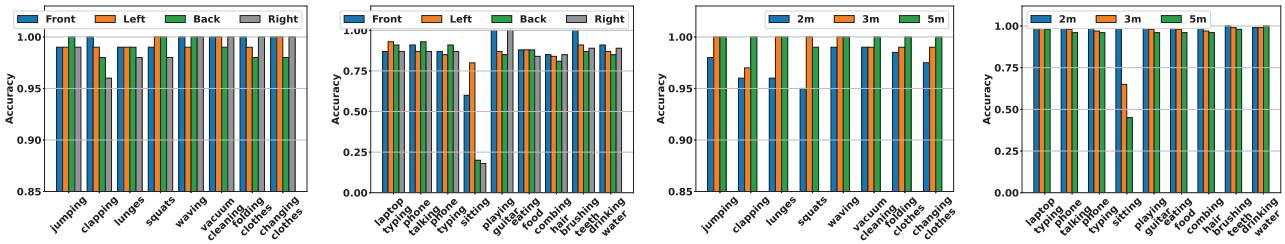
Table 3: Accuracy under different number of subjects

# of subjects	1	2	3	4	5
MARS	0.99	0.97	0.95	0.91	0.89
RadHAR	0.88	0.84	0.82	0.77	0.7
Vid2Doppler	0.78	0.76	0.61	0.55	0.2
VGG-16	0.69	0.7	0.66	0.51	0.52
Pantomime	0.87	0.83	0.78	0.76	0.72

Table 4: Accuracy under different occlusions

Occlusions	MARS	RadHAR	Vid2Doppler	VGG-16	Pantomime
Wooden Board	0.86	0.77	0.68	0.69	0.78
Fibre	0.92	0.81	0.71	0.73	0.83
Glass	0.89	0.75	0.69	0.71	0.81

5.4.2 Impact of room structure. We studied *MARS* extensively in the three rooms R1, R2, R3 as discussed in Sec. 4. We consider two cases – (i) *Inter-room Training*, where we train *MARS* over the data collected at R2 and test over R1 and R3, and (ii) *Intra-room Training*,



(a) Macro classifier (w.r.t. orientation) (b) Micro Classifier (w.r.t. orientation) (c) Macro classifier (w.r.t. distance) (d) Micro Classifier (w.r.t. distance)

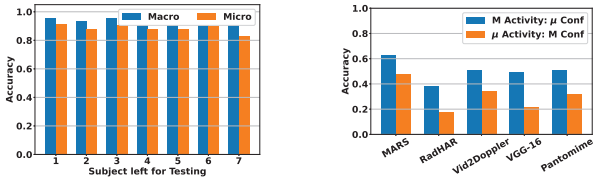
Figure 11: Accuracy at different orientations and distances.

where we train and test the models using the data collected over the same room. As shown in the Figure 10(b), the accuracy for all the rooms is > 90% for both the classifiers. Interestingly, for R1, the multipath-reflection effect is more significant due to the smaller room size, so the accuracy is lower; however, the impact of intra-room and inter-room is not very significant as *MARS* learns the features related to the doppler patterns of the subject's activity rather than room-specific features. Inter-room training and testing (leave one out) also indicate that the model does not **overfit**.

5.4.3 Impact of blockages. We have tested *MARS* with different blockages which acts as a source of occlusion for the mmWave signal, such as (i) Wooden board, (ii) Fibreboard, (iii) Glass board, etc (shown in Figure 10(d). However, mmWave at higher frequency shows higher penetration loss, although it can penetrate materials like plywood, glass, and fiber. Thus for macro activities, at least, we can achieve an accuracy > 90% (see Figure 10(c)). However, the accuracy for micro activities is lower, as small phase variations are attenuated more quickly than macro phase variations. We have compared *MARS* with the baselines and have reported in Table 4.

5.4.4 Impact of subject orientation. We test *MARS* under different body orientations of the subject, i.e., (i) front, (ii) left, (iii) right, (iv) back, w.r.t. the setup. As observed in Figure 11(a) under different orientations, the macro activity classifier can recognize the activity classes with an accuracy of ≈ 0.97. But in the case of the micro activity classifier (shown in Figure 11(b)), the accuracy is lower, especially for the case of phone talking. During talking on phone, the subject's body orientation, such as back or right (while holding the phone in left hand) significantly impacts the activity classification due to the occlusion of small-scale body movements. Nevertheless, it is comforting to see that for other micro-activities, the accuracy is always > 0.80, even when the subject is at a complete opposite orientation from the radar.

5.4.5 Impact of distance. Figure 11(c) shows the variation in the accuracy for the macro activity classifier under different distances from the subject. The classification is reported for up to a distance of 5m. The accuracy is ≈ 0.97 as observed. However, Figure 11(d) indicates that the accuracy for micro activity classifier sometime drops (up to 10%) with the increase in the distance. Due to signal attenuation, the doppler shift for micro activities sometimes becomes



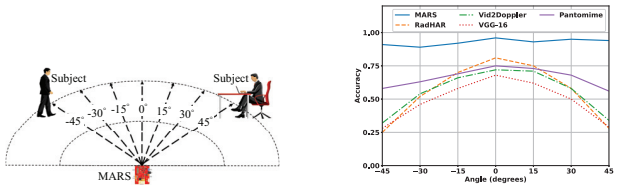
(a) (b)

Figure 12: Accuracy for (a) Leave one subject out cross validation; (b) for Macro (M) activities under micro (μ) configuration and μ activity under Macro configuration.

undetectable with increasing distance. Interestingly, macro activities demonstrate improved performance at greater distances (>3m, as depicted in Figure 11c). This advantage arises from the radar's conical field of view (FoV) adeptly capturing the expansive body movements associated with macro activities at extended distances. This phenomenon is unique to macro activities, since increasing distance substantially decreases accuracy for micro activities.

5.4.6 Impact of Leave-one-subject-out. We evaluated *MARS* on leave one subject out cross validation scenarios. As shown in Figure 12a, the accuracy for macro activities is > 90%, and that of micro activities is > 88%, which validates the robustness of *MARS* . This ensures the model is not overfitting the dataset.

5.4.7 Cross Configuration Evaluation. We evaluated *MARS* when subjects are performing macro activities while the sensor is in micro configuration and vice versa. The motivation behind doing so is empirically validating the importance of radar reconfiguration on activity classification. As shown in Figure 12b, the classification accuracy drops significantly to 66% for *MARS* . The reason behind such a poor classification for all the methods is twofold: (i) Under micro configuration, the doppler changes are noisier and thus confused with other macro activities, and (ii) under macro configuration, small changes due to micro activities cannot get captured properly. Macro activities under micro doppler resolution show higher accuracy as higher doppler resolution can classify disjoint activity classes such as changing clothes and jumping easily but due to rapid changes in the doppler resolution it gets overlapping classification on similar activities like jumping, lunges, squats, etc.



(a) (b)

Figure 13: (a) Subjects' angular position from the radar, (b) accuracy under different articulation angle.

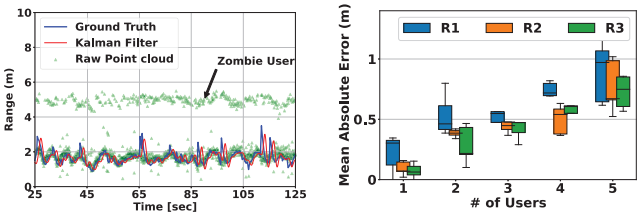
5.4.8 Impact of Articulation angle. We conducted a thorough evaluation of *MARS* across a range of articulation angles, varying from -45° to $+45^\circ$ in 15° increments. Subjects were instructed to perform various activities within these angular regions, and the experiments were conducted in two distinct sessions: **Session 1:** We utilized *MARS* , which can dynamically rotate until the subject is within the FoV before predicting the activity. **Session 2:** Data was collected using a fixed radar setup, similar to the approach detailed in [39, Sec 6.5] for evaluating the baselines. The results, illustrated in Figure 13b, demonstrate the superior accuracy of *MARS* compared to the baseline methods. Notably, the baseline accuracies were highest within the $\pm 15^\circ$ range, aligning with the FoV. However, accuracy gradually decreased with increasing angle, as the radar's sensing capabilities became limited in capturing the full range of the participant's arm movements. Even at 0°, the lower accuracy of Pantomime and RadHAR can be attributed to the inherent noise and potential corruption of pointcloud data during micro activities.

5.5 Localization and Tracking Performance

Figure 14(a) shows a snapshot of the subject tracking pipeline to the ground truth distance. In this selected experiment, we observe that the raw pointcloud data for the subject cluster contains the signature of a zombie subject arising due to multi-path reflections (present at a distance of $\approx$ 5m). With the proposed localization approach, this zombie subject's pointcloud gets suppressed. In Figure 14(b), we show the mean absolute error (MAE) in subject localization in three different rooms R1, R2, and R3, w.r.t. the ground truth, under different numbers of subjects present inside the room who are walking simultaneously within the FoV of the radar at an average speed of 0.7-1.1 m/s. Although the MAE in the localization is $<$ 60 cm with three simultaneous subjects, the MAE gradually increases with increasing the number of subjects walking simultaneously. Since R1 is smaller, the MAE is higher because the pointcloud data is noisier with more multi-path reflections. However, it is comforting to see that even with the five users walking randomly at a normal to moderately high speed, the MAE is within $\approx 1m$.

5.6 Response Time for Multi-User Localization

With the current setup, we can detect multi-user activities with a faster response time compared to the baselines. However, in some specific corner cases, we have a lag in the response time. Here, we will discuss these scenarios: (i) **Scenario 1:** Normally, the subject localization task takes an average response time of $\approx$ 2s; however,

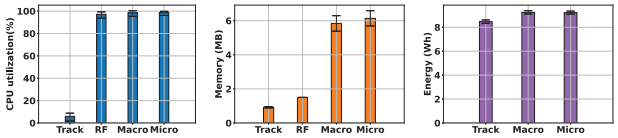


(a) (b)

Figure 14: (a) Qualitative analysis of Kalman tracking, (b) MAE with different numbers of walking subjects.

Table 5: Avg. Response time for multi-user localization when simultaneous tracking of multiple users is not possible.

Scenario	Avg. Response time (s)	MAE (m)
Scenario 1: Users walking in opp. directions	4.2	0.12
Scenario 2: User walks while *MARS* classifies activities	3.22	0.31



(a) CPU (b) Memory (c) Energy

Figure 15: Resource consumption of *MARS* .

in cases where users move in opposite directions, the servo tracker is only able to track one user, resulting in the other user entering a blind zone. However, as soon as the pointcloud queue count decreases by one, we resume a 360°rotation search. This allows us to determine the subjects' locations with an additional response time due to servo rotation and refilling the pointcloud queue ($\approx$ 4.2s shown in Table 5). This, however, does not compromise the localization accuracy, as the average MAE remains close to $\approx$ 0.12 m, on per with single user localization MAE. (ii) **Scenario 2:** When the radar is focused on monitoring a specific user's macro/micro activities, other users in the queue may start moving, causing the queue to lose track of them. However, once the system returns to the localization and tracking pipeline, we are once again able to deduce the subject's new location with some additional response time of $\approx$ 3.22s (including activity inference time of $\approx$ 0.08s, radar configuration switching time of $\approx$ 1.14s and localization response time of $\approx$ 2s) with an MAE of 0.31 m (shown in Table 5).

5.7 Resource and Energy Benchmarks

We measure the resource and energy consumption of the back-end processing unit, i.e., RPi-4, under different scenarios. As observed in Figure 15(a), 15(b), the CPU and the memory utilization in case of the localization and the tracking pipeline is low when the subject is not present inside the room. As the subject enters the room, the Opportunistic classifier gets initiated. As a result of feature computation and pipeline initiation, we observe significantly higher CPU and memory utilization. After that, when the activity classification

Table 6: Comparison of the state-of-the-art systems

Title	Macro Activities	Micro Activities	Continuous Monitoring	Real-time Inference	Multi-Person Monitoring
IMU2Doppler [12]	✓	×	✓	×	×
Mobi2Sense [61]	✓	✓	×	✓	✓
RF-Action [34]	✓	✓	×	✓	✓
RF-Net [23]	✓	×	×	×	×
RF-Pose [62]	✓	×	✓	×	✓
Vid2Doppler [10]	✓	×	✓	✓	×
RadHAR [50]	✓	×	✓	×	×
m-activity [55]	✓	×	×	×	×
RF-Diary [25]	✓	×	✓	×	×
Jiang et. al. [28]	✓	×	×	×	×
Cominelli et. al. [21]	✓	×	×	×	×
MARS	✓	✓	✓	✓	✓

pipeline gets initiated, we observe that memory utilization increases significantly due to large-scale feature computation and loading of the trained model in the memory. Finally, using a Monsoon Power Monitor [5], we measure the overall energy consumption of the RPi under the three scenarios – (i) localization and tracking, (ii) macro activity classification, (iii) micro activity classification. As observed in Figure 15(c), the energy consumption is comparatively higher in the case of macro and the micro activity classification because of the higher CPU and RAM utilization.

6 RELATED WORK

Some studies have used wearables for active sensing techniques, mostly to detect human activity [30, 45]. Even though such methods are useful, they are not seamless and pervasive enough. Our next discussion explores alternative passive sensing methods, mainly those based on acoustics and radio frequency.

Acoustic-based: Passive acoustic sensing [32, 53, 54] involves the creation of audio chirps that are reflected away from nearby surfaces and detected by a *microphone*. In this direction, both macro [13], as well as micro activities [32], were studied. Acoustic-based sensing involves several frequencies, including ultra-sounds [28]. Due to the reliance on audio frequency, acoustic-based approaches are susceptible to environmental noise, interference, and microphone orientation. In addition, multiple individuals affect the acoustic signature in an unpredictable manner [54].

RF-based: Radio-frequency (RF) in the form of WiFi [23], RFIDs [29], UWB radars [61] has been studied for capturing human dynamics. Many have explored WiFi Channel State Information (CSI) [21, 21, 51, 52]; as the phase and amplitude of radio waves are impacted by the movements of objects in their path [17]. Both macro and micro-level [11] activities have been studied in this direction. Alternatively, UWB is suitable for penetrating walls and capturing movements [19]. Regarding WiFi, CSI extraction from signals is usually a complex process. Other environmental dynamics, such as door movement, furniture movement, and electromagnetic interference, can also affect the signal. More specifically, the environments with different sizes and layouts have different multipath effects on the received WiFi signals [15]. Also, WiFi modality suffers from poor range resolution [18]. Some works [25, 34] rely on FMCW techniques with specialized hardware aiming for a relatively higher depth resolution [25]. However, this specialized hardware is usually expensive compared to COTS hardware [16]. [34] can track multiple users, mainly for short-duration actions like hand-shaking and falling. Also A COTS FMCW mmWave sensor such as IWR1642 [3] demonstrates a better range resolution as compared to the specialized device used in [25] (~3.75cm vs. ~8cm) [25, 42]. Compared to WiFi CSI, UWB has a higher achievable resolution [16]; however, it has a well-known spectrum coexistence issue. RFIDs also have a limited range of around 5 meters. Instead, mmWave-based sensors can detect small movements at a finer level. Due to its shorter wavelength, mmWave can create stronger reflections even from smaller objects [28]. The works [12, 37, 48, 50, 55, 58] employing this modality rely on emitted mmWave signals in the form of chirps and exploit the received signal reflected by the surroundings to capture activity signatures. COTS mmWave radars often use FMCW chirps for this purpose. Features, such as pointclouds [39, 50, 55], range-doppler [10, 12, 46, 47], etc., have been proven to be effective in movement detection. Previous works such as Vid2Doppler [10] and RadHAR [50] focus on single-user macro activity tracking. MultiTrack [52] and Pantomime [39] provide long-range localization and multi-user tracking but does not ensure continuous monitoring.

In contrast to previous works, this work uses mmWave sensing to continuously track activities since it is minimally intrusive on privacy and captures micro-movements. The single modality is sufficient for continuous activity monitoring of multiple individuals. Our approach also detects the most number of activities (both macro and micro simultaneously) in the mmWave domain with a dynamic environment when multiple users are present. Table 6 highlights the advantages of *MARS* compared to the state-of-the-art contributions in the relevant domain.

7 CONCLUSION

We need simple yet effective ways for humans to interact with our smart spaces. Existing ideas, however, use techniques that are both invasive and difficult to integrate. The key insight prompted us to design and develop *MARS* , a lightweight yet highly effective mmWave-based continuous activity monitoring system. Through experiments, *MARS* proves its effectiveness of single subject tracking with a mean absolute error of just 45cm despite supporting global coordinates. After that, it demonstrates field-deployable accuracy of 97% and 93%, respectively, for multiple macro and micro-scale activities. Based on the results, we are confident that *MARS* will seamlessly adapt to human activities in all situations encountered in real-world scenarios. The existing form of *MARS* is incapable of capturing micro activities beyond five meters due to signal attenuation with an increasing range, which is a fundamental challenge in mmWave. Instead of limiting our evaluation to simplistic functional accuracy, we evaluated the performance of *MARS* based on different counts, orientations, distances, and even energy consumption footprints, comparing it to the state-of-the-art baselines that demonstrate its superiority.

REFERENCES

[1] 2011. Robodo SEN40 GY-273 HMC5883L Module Triple Axis Compass Magnetometer Sensor Module for Arduino. https://www.amazon.in/Robodo-Electronics-SEN40-HMC5883L-Magnetometer/dp/B0787LH8XR. [Accessed April 2, 2024].

[2] 2013. Tower Pro MG995 Servo Motor. https://www.towerpro.com.tw/product/mg995/. [Accessed April 2, 2024].

[3] 2018. IWR1642BOOST. https://www.ti.com/tool/IWR1642BOOST. Accessed April 2, 2024.

[4] 2018. mmWave Demo Visualizer — dev.ti.com. https://dev.ti.com/gallery/view/mmwave/mmWave_Demo_Visualizer/. [Accessed April 2, 2024].

[5] 2020. High Voltage Power Monitor | Monsoon Solutions . https://www.msoon.com/high-voltage-power-monitor. [Accessed April 2, 2024].

[6] 2020. IWR1443BOOST. https://www.ti.com/tool/IWR1443BOOST. Accessed April 2, 2024.

[7] 2022. ADLs-IADLs. https://betterhealthwhileaging.net/what-are-adls-and-iadls/. [Accessed April 2, 2024].

[8] Fadel Adib and Dina Katabi. 2013. See through walls with WiFi!. In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*. 75–86.

[9] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart homes that monitor breathing and heart rate. In *ACM CHI*. 837–846.

[10] Karan Ahuja, Yue Jiang, Mayank Goel, and Chris Harrison. 2021. Vid2Doppler: Synthesizing Doppler radar data from videos for training privacy-preserving activity recognition. In *ACM CHI*. 1–10.

[11] Kamran Ali, Alex X Liu, Wei Wang, and Muhammad Shahzad. 2015. Keystroke recognition using wifi signals. In *ACM MobiCom*. 90–102.

[12] Sejal Bhalla, Mayank Goel, and Rushil Khurana. 2021. IMU2Doppler: Cross-Modal Domain Adaptation for Doppler-based Activity Recognition Using IMU Data. *ACM IMWUT* 5, 4 (2021), 1–20.

[13] Gaddi Blumrosen, Ben Fishman, and Yossi Yovel. 2014. Noncontact wideband sonar for human activity detection and classification. *IEEE Sensors Journal* 14, 11 (2014), 4043–4054.

[14] Pingping Cai and Sanjib Sur. 2023. MilliPCD: Beyond Traditional Vision Indoor Point Cloud Generation via Handheld Millimeter-Wave Devices. *ACM IMWUT* 6, 4 (2023), 1–24.

[15] Chen Chen, Gang Zhou, and Youfang Lin. 2023. Cross-Domain WiFi Sensing with Channel State Information: A Survey. *Comput. Surveys* 55, 11 (2023), 1–37.

[16] Zhe Chen, Chao Cai, Tianyue Zheng, Jun Luo, Jie Xiong, and Xin Wang. 2021. Rf-based human activity recognition using signal adapted convolutional neural network. *IEEE TMC* 22, 1 (2021), 487–499.

[17] Zhenghua Chen, Le Zhang, Chaoyang Jiang, Zhiguang Cao, and Wei Cui. 2018. WiFi CSI based passive human activity recognition using attention based BLSTM. *IEEE TMC* 18, 11 (2018), 2714–2724.

[18] Zhe Chen, Tianyue Zheng, Chao Cai, Yue Gao, Pengfei Hu, and Jun Luo. 2023. Wider is Better? Contact-free Vibration Sensing via Different COTS-RF Technologies. *Proc. of the 42nd IEEE INFOCOM* (2023).

[19] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. 2021. MoVi-Fi: Motion-robust vital signs waveform recovery via deep interpreted RF sensing. In *ACM MobiCom*. ACM, 392–405.

[20] Adrian Clark, Andreas Dünser, Mark Billinghurst, Thammathip Piumsomboon, and David Altimira. 2011. Seamless interaction in space. In *ACM OzCHI*. 88–97.

[21] Marco Cominelli, Francesco Gringoli, and Francesco Restuccia. 2023. Exposing the CSI: A Systematic Investigation of CSI-based Wi-Fi Sensing Capabilities and Limitations. In *IEEE PerCom*. 81–90.

[22] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *IEEE CVPR*. Ieee, 248–255.

[23] Shuya Ding, Zhe Chen, Tianyue Zheng, and Jun Luo. 2020. RF-net: A unified meta-learning framework for RF-enabled one-shot human activity recognition. In *ACM SenSys*. 517–530.

[24] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *KDD*, Vol. 96. 226–231.

[25] Lijie Fan, Tianhong Li, Yuan Yuan, and Dina Katabi. 2020. In-home daily-life captioning using radio signals. In *ECCV*. Springer, 105–123.

[26] Xiaonan Guo, Bo Liu, Cong Shi, Hongbo Liu, Yingying Chen, and Mooi Choo Chuah. [n. d.]. WiFi-enabled smart human dynamics monitoring. In *ACM SenSys*.

[27] HexiWear. [n. d.]. HexiWear. https://www.mikroe.com/hexiwear

[28] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. 2018. Towards environment independent device free human activity recognition. In *ACM MobiCom*. ACM, 289–304.

[29] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. 2014. Bringing gesture recognition to all devices. In *USENIX NSDI*. 303–316.

[30] Isah A Lawal and Sophia Bano. 2019. Deep human activity recognition using wearable sensors. In *ACM PETRA*. 45–48.

[31] Dong Li, Shirui Cao, Sunghoon Ivan Lee, and Jie Xiong. 2022. Experience: practical problems for acoustic sensing. In *ACM MobiCom*. 381–390.

[32] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2022. LASense: Pushing the Limits of Fine-grained Activity Sensing Using Acoustic Signals. *ACM IMWUT* 6, 1 (2022), 1–27.

[33] Huining Li, Chenhan Xu, Aditya Singh Rathore, Zhengxiong Li, Hanbin Zhang, Chen Song, Kun Wang, Lu Su, Feng Lin, Kui Ren, et al. 2020. VocalPrint: exploring a resilient and secure voice authentication via mmWave biometric interrogation. In *ACM SenSys*. ACM, 312–325.

[34] Tianhong Li, Lijie Fan, Mingmin Zhao, Yingcheng Liu, and Dina Katabi. 2019. Making the invisible visible: Action recognition through walls and occlusions. In *IEEE/CVF ICCV*. 872–881.

[35] Chen Liu, Jie Xiong, Lin Cai, Lin Feng, Xiaojiang Chen, and Dingyi Fang. 2019. Beyond respiration: Contactless sleep sound-activity recognition using RF signals. *ACM IMWUT* 3, 3 (2019), 1–22.

[36] Tiantian Liu, Ming Gao, Feng Lin, Chao Wang, Zhongjie Ba, Jinsong Han, Wenyao Xu, and Kui Ren. 2021. Wavoice: A noise-resistant multi-modal speech recognition system fusing mmwave and audio signals. In *ACM SenSys*. ACM, 97–110.

[37] Chris Xiaoxuan Lu, Stefano Rosa, Peijun Zhao, Bing Wang, Changhao Chen, John A Stankovic, Niki Trigoni, and Andrew Markham. 2020. See through smoke: robust indoor mapping with low-cost mmwave radar. In *ACM MobiSys*. 14–27.

[38] Ramon Nitzberg. 1972. Constant-false-alarm-rate signal processors for several types of interference. *IEEE TAES* (1972), 27–34.

[39] Sameera Palipana, Dariush Salami, Luis A Leiva, and Stephan Sigg. 2021. Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds. *ACM IMWUT* 5, 1 (2021), 1–27.

[40] Jacopo Pegoraro and Michele Rossi. 2021. Real-time people tracking and identification from sparse mm-wave radar point-clouds. *IEEE Access* 9 (2021).

[41] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems* 30 (2017).

[42] Sandeep Rao. 2017. Introduction to mmWave sensing: FMCW radars. *Texas Instruments (TI) mmWave Training Series* (2017), 1–11.

[43] S Rao. 2020. Introduction to mmwave radar sensing: Fmcw radars. *Texas Instruments* (2020), 1–70.

[44] Reuters. 2022. Tesla will remove more vehicle sensors amid Autopilot scrutiny. https://auto.economictimes.indiatimes.com/news/passenger-vehicle/cars/tesla-will-remove-more-vehicle-sensors-amid-autopilot-scrutiny/94654643 [Accessed April 2, 2024].

[45] Muhammad Moid Sandhu, Sara Khalifa, Kai Geissdoerfer, Raja Jurdak, and Marius Portmann. 2021. SolAR: Energy positive human activity recognition using solar cells. In *IEEE PerCom*. IEEE, 1–10.

[46] Argha Sen, Anirban Das, Prasenjit Karmakar, and Sandip Chakraborty. 2023. mmAssist: Passive Monitoring of Driver's Attentiveness Using mmWave Sensors. In *COMSNETS*. IEEE, 545–553.

[47] Argha Sen, Avijit Mandal, Prasenjit Karmakar, Anirban Das, and Sandip Chakraborty. 2023. mmDrive: mmWave Sensing for Live Monitoring and On-Device Inference of Dangerous Driving. In *PerCom*. IEEE, 2–11.

[48] Xian Shuai, Yulin Shen, Yi Tang, Shuyao Shi, Luping Ji, and Guoliang Xing. 2021. millieye: A lightweight mmwave radar and camera fusion system for robust object detection. In *IoTDI*. 145–157.

[49] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[50] Akash Deep Singh, Sandeep Singh Sandha, Luis Garcia, and Mani Srivastava. 2019. Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar. In *ACM mmNets*. ACM, 51–56.

[51] Elahe Soltanaghaei, Rahul Anand Sharma, Zehao Wang, Adarsh Chittilappilly, Anh Luong, Eric Giler, Katie Hall, Steve Elias, and Anthony Rowe. 2020. Robust and practical WiFi human sensing using on-device learning with a domain adaptive model. In *ACM BuildSys*. 150–159.

[52] Sheng Tan, Linghan Zhang, Zi Wang, and Jie Yang. 2019. MultiTrack: Multi-user tracking and activity recognition using commodity WiFi. In *ACM CHI*. 1–12.

[53] Lei Wang, Tao Gu, Wei Li, Haipeng Dai, Yong Zhang, Dongxiao Yu, Chenren Xu, and Daqing Zhang. 2023. DF-Sense: Multi-user Acoustic Sensing for Heartbeat Monitoring with Dualforming. In *ACM MobiSys*. 1–13.

[54] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *ACM IMWUT* 1, 4 (2018), 1–20.

[55] Yuheng Wang, Haipeng Liu, Kening Cui, Anfu Zhou, Wensheng Li, and Huadong Ma. 2021. m-activity: Accurate and real-time human activity recognition via millimeter wave radar. In *IEEE ICASSP*. IEEE, 8298–8302.

[56] Teng Wei and Xinyu Zhang. 2015. mtrack: High-precision passive tracking using millimeter wave radios. In *ACM MobiCom*. ACM, 117–129.

[57] Wikipedia. [n. d.]. Amazon Echo. https://en.wikipedia.org/wiki/Amazon_Echo

[58] Jiahong Xie, Hao Kong, Jiadi Yu, Yingying Chen, Linghe Kong, Yanmin Zhu, and Feilong Tang. 2023. mm3DFace: Nonintrusive 3D Facial Reconstruction Leveraging mmWave Signals. In *ACM MobiSys*. 462–474.

[59] Chengxi Yu, Zhezhuang Xu, Kun Yan, Ying-Ren Chien, Shih-Hau Fang, and Hsiao-Chun Wu. 2022. Noninvasive human activity recognition using millimeter-wave radar. *IEEE Systems Journal* (2022).

[60] Bo Zhang, Boyu Jiang, Rong Zheng, Xiaoping Zhang, Jun Li, and Qiang Xu. 2023. Pi-ViMo: Physiology-inspired Robust Vital Sign Monitoring using mmWave Radars. *ACM TIOT* 4, 2 (2023), 1–27.

[61] Fusang Zhang, Jie Xiong, Zhaoxin Chang, Junqi Ma, and Daqing Zhang. 2022. Mobi2Sense: empowering wireless sensing with mobility. In *ACM MobiCom*.

[62] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *IEEE CVPR*. 7356–7365.

[63] Peijun Zhao, Chris Xiaoxuan Lu, Jianan Wang, Changhao Chen, Wei Wang, Niki Trigoni, and Andrew Markham. 2019. mid: Tracking and identifying people with millimeter wave radar. In *DCOSS*. IEEE, 33–40.