mmWave-based Crowd Sensing for Metaverse Applications

Marios Raspopoulos
INSPIRE Research Centre
UCLan Cyprus
Larnaca, Cyprus
0000-0003-1513-6018

Iacovos Ioannou
CYENS Centre of Excellence
University of Cyprus
Nicosia, Cyprus
0000-0002-1562-5543

Louis Nisiotis
INSPIRE Research Centre
UCLan Cyprus
Larnaca, Cyprus
0000-0002-8018-1352

***Abstract*—The growing adoption of the Metaverse raises critical challenges in real-time crowd sensing, where traditional vision-based systems struggle to balance high-resolution monitoring with user privacy. Cameras and other optical sensors, while effective in tracking movement and interactions, inherently capture personally identifiable information, creating ethical and legal concerns. This paper explores the use of millimetre-wave (mmWave) radar technology as a privacy-preserving, high-resolution solution for real-time crowd sensing in Metaverse applications. Recognising the limitations of traditional monitoring methods, such as visual surveillance and mobile-based tracking, this study presents a simulation-based framework for evaluating mmWave-enabled visitor tracking within a museum-style environment. A MATLAB-based simulator models realistic human mobility and sensor data, incorporating error models obtained from an experimental precision analysis of mmWave sensors. A combination of DBSCAN and K-Means clustering is then applied to estimate crowd formations, density, and mobility flows. Results demonstrate the effectiveness of mmWave in identifying dynamic crowd behavior while preserving user anonymity, highlighting its potential for immersive digital twins, XR experiences, and intelligent environment management in Cyber-Physical-Social Systems that underpin the Metaverse.**

***Index Terms*—mmWave Sensing, Metaverse, Clustering, Crowd monitoring**

I. Introduction

Crowd management and control in public and private spaces represents a critical area of interest across fields such as urban planning, emergency response, event organization and coordination, and the development of smart infrastructures. This growing attention is typically driven by several key factors, including the need for efficient navigation in complex environments, the demand for applications that enhance user experience in shared spaces, and rising concerns over public safety and the risks associated with dense crowds. Traditional crowd monitoring systems such as visual surveillance (e.g., CCTV), infrared sensors, and mobile device tracking usually raise privacy concerns, could be sensitive to environmental conditions and typically depend on user participation or device presence.

Crowd control also appears to have an important and distinctive role in the Metaverse concept, which is highly integrated with real-world environments through eXtended Reality technologies (XR - the umbrella term encapsulating augmented, virtual and mixed reality applications, software and hardware technologies). Often regarded as the next significant technological evolution, the Metaverse is conceptualized as a seamless integration of physical and virtual realities within unified digital environments, enabling people to engage, collaborate, and interact within immersive, interconnected spaces with applicability opportunities across multiple domains [1], [2].

Building on this vision, understanding and managing human presence within shared hybrid spaces becomes essential to creating responsive and intelligent Metaverse applications. As users navigate physical spaces augmented by virtual layers, real-time insights into crowd behavior can support adaptive system behavior, safety monitoring, and efficient resource allocation, among other capabilities. Crowd sensing, therefore, becomes a crucial capability for achieving dynamic and intelligent user-centered interactions in Metaverse applications that accurately reflect real-world complexity.

In this context, having knowledge about crowd flows and densities has multiple benefits. Within the Metaverse, crowd sensing can serve as a valuable tool for gathering data related to user behavior and environmental dynamics through capturing information such as users' physical locations and interactions with physical spaces and digital objects to support real-time modelling of Metaverse applications with accurate representations, enhanced interactivity, and adaptive system responses [3]. For example, crowd sensing can offer enhanced spatial awareness in digital twin applications, enhancing realism in XR interfaces; it has the potential to enable or facilitates contextual adaptation in virtual overlays; for instance, in XR-aided retail or museum environments it could enable adaptive content delivery or direct users in less congested areas; it provides a safe and immersive user experience in public XR applications by preventing physical overlaps, collisions or bottlenecks; it can support hybrid event coordination and management in smart crowded environments synchronizing the virtual and physical crowd movement and many other applications. This paper presents an approach that leverages millimeter-wave (mmWave) sensing to non-intrusively localize users within a public environment—specifically, an art

This work was co-funded by the European Union under the programme of social cohesion "THALIA 2021-2027", through the Research and Innovation Foundation (Project: EXCELLENCE/0524/0218).

gallery—and to identify large visitor clusters and characterize their flow. Such information is valuable for Metaverse applications aiming to enhance visitor experiences through the integration of robots, digital twins and/or other technologies.

The remainder of this paper is organized as follows: in Section II the recent related works in crowd sensing using different technologies are presented, highlighting some of the most promising approaches focusing on works that use mmWave. Section IV describes the methodology and setup used for the mmWave precision analysis experimentation and simulation development, including the clustering techniques. Section V presents the simulation results for a typical museum-style environment, and Section VI, provides a critical discussion and conclusion.

II. Related Work

A crowd is defined as a collective situation where individuals gather at a specific location within a certain time frame and a shared purpose [4]. Different types of crowds exist and vary based on context, and are typically classified according to the nature of the situation, such as casual gatherings, protests, or scheduled events [4]. Common characteristics used to describe crowd scenarios include the participants' intentions, the crowd's dynamics, size and density, as well as its location and the duration of the gathering [5]. A comprehensive review applicable to crowd analysis of public transportation systems can be found in [6]. It is a well-established fact that overcrowding significantly decreases user satisfaction, particularly in public places such as museums, galleries, hospitals, shopping malls, and transportation hubs, as well as the safety risks and operational deficiencies that accompany it. Various technologies have been proposed and used over the years to monitor and control crowds, but many of those come with concerns regarding intrusiveness and privacy implications. For instance, infrared counters and Wi-Fi analytics have been used to regulate visitor flows in museums and art galleries. Another example is the use of Real-time location services (RTLS) and Radiofrequency Identification (RFID), as well as thermal imaging have been used in hospital Emergency Rooms (ER) and waiting areas to optimize patient movement, ensure distancing and facilitate infection control. In retail and transportation applications, crowd-aware systems utilize conventional cameras, complemented by video analytics, as well as Bluetooth or Ultra-Wideband (UWB) tracking. Most of these applications include the identification of people, which is prohibited or strictly governed by regulations like HIPAA and GDPR, or require that the users hold specific devices. The following table summarizes the various technologies available for crowd control, highlighting their advantages and disadvantages. It is for these reasons that millimeter-wave (mmWave) sensing emerges as a promising solution for crowd monitoring.

TABLE I: Comparison of Crowd Sensing Technologies

Technology	Advantages	Limitations
Video Cameras	High resolution, rich visual data	Intrusive, privacy issues, sensitive to lighting conditions
Infrared Sensors	Simple, low-cost, low power consumption	Limited range, unable to distinguish individuals in groups
WiFi/Bluetooth	Utilizes existing personal devices, non-visual	Requires users to carry devices, raises privacy concerns
LiDAR	Accurate spatial and depth sensing, good for open spaces	High cost, affected by reflective or occluded surfaces
mmWave Radar	Privacy-preserving, works in darkness and clutter, tracks multiple targets	Complex signal processing, less intuitive output
Thermal Cameras	Effective in low light, health monitoring use cases	Low spatial resolution, moderate privacy intrusiveness

mmWave is a communication technology that has started being used in modern communication systems such as Wi-Fi (e.g., IEEE802.11ad), while it is planned to be used in 5G (and beyond) communications due to its flexibility to use wider bandwidths and hence its strong potential in achieving much higher data rates and capacity. mmWave systems typically operate in frequencies between 26 and 300 GHz. Their very large availability of bandwidth which leads to fine timing (and hence ranging) resolution, and together with the ease of using phase array antennas (at those frequencies) that enable the estimation of the phase (and thus the angle) could be used for achieving decimeter 3D positioning accuracy or better [7]–[9]. When used in positioning, mmWave is typically referred to as sensing, mostly because of its radar-like operation, which facilitates the localization of passive device-free targets. A mmWave sensor emits high-frequency electromagnetic waves that bounce off surrounding objects and return as echoes whose time and angle of arrival can be measured. This knowledge enables estimation of the position of targets in motion since a fundamental requirement is that these targets generate Doppler shifts. On the other hand, such technology gives rise to various limitations and challenges that are not faced with conventional transmitter-receiver technologies, such as UWB, Bluetooth, and Wi-Fi. These echoes can become mixed together in complicated environments, making it difficult to differentiate specific objects. This becomes especially challenging when using multiple sensors and is particularly critical in applications such as indoor people activity tracking or autonomous vehicle navigation, where the ability to discern and track multiple objects with precision is crucial.

However, the application of many of these technologies raises concerns regarding privacy, environmental dependency, and limited scalability. We highlight some of the most promising approaches, focusing on works that use mmWave.

A. Crowd Sensing using different Technologies

In the highly cited vision-based work reported in [10], the authors utilize overhead surveillance cameras to detect the motion of users, and then employ a Long Short-Term Memory (LSTM) recurrent neural network model that learns both individual motion patterns and their interactions with nearby individuals. The proposed model incorporates a social pooling layer that helps predict future paths by accounting for human-human interactions in shared crowded spaces. Various other video-surveillance solutions for crowd analysis, including density estimation and behavior recognition using deep learning techniques, are reviewed in [11].

In [12], the authors combine LiDAR scans with camera imagery and utilize filtering and fusion algorithms to control a servo-based swinging platform, enabling accurate real-time tracking of humans in enclosed, crowded areas. Laser-based people tracking is not new, as works exist in the literature since the early 2000s. One such work is the one reported in [13] where a laser-based dense crowd tracking method is proposed that uses a stable feature extraction method based on accumulated distribution of successive laser frames trying to detect paterns of rhythmic leg swings to extract each persons legs followed by a region coherency property to construct an efficient measurement likelihood model and then by using a combination of independent Kalman filter and Rao-Blackwellized Monte Carlo data association filter (RBMC-DAF) detects the presence of people. Although laser-based tracking has proven to be an accurate approach, it is a complex and expensive solution that is also environmentally sensitive (e.g., dark conditions, fog, humidity, and reflecting mirrors).

The work reported in [14] proposes a crowd monitoring system featuring Wi-Fi sensors by means of a small number of automated counting systems and uses data fusion aiming to estimate the pedestrian flows based on real-time Wi-Fi traces at one sensor location, and historic flow rate and Wi-Fi trace information gathered at other sensor locations. Wi-Fi is combined with Bluetooth in [15] to develop an intelligent transportation system that monitors and classifies pedestrian and cyclist activity traffic. Although this work appears solid and interesting, it raises privacy concerns due to the requirement to capture MAC addresses.

Lastly, thermal imaging emerged as another possible solution in crowd analysis, addressing some limitations associated with conventional RGB imaging, such as privacy concerns and environmental challenges like darkness and fog. For instance, the study reported in [16] utilizes a thermal imaging camera (TIC) combined with a deep learning model based on the YOLOv4 algorithm to detect humans during emergency evacuations in low-visibility, smoke-filled fire scenarios. Nevertheless, thermal imaging still has certain limitations related to environmental conditions, particularly ambient heat interference, and does not entirely eliminate privacy concerns.

B. Crowd Sensing using mmWave

Over the last few years, works have started appearing in the literature that make use of the mmWave technology for crowd sensing, leveraging on the benefits that this technology can bring with regard to the concerns regarding privacy, accuracy, responsiveness and environmental sensitivity. These works can be grouped into two categories: (1) crowd density estimation and (2) human activity recognition.

1) Crowd Density Estimation: The authors of [17] propose a radar-based people counting pipeline to detect individuals moving together as a single group, based on the complementary combination of a tracking algorithm and a feature-based classifier that estimates the number of people in each group tracked within the scene of interest. It demonstrates that a combination of features extracted from the range–azimuth domain using a mmWave radar yields robust results. Similarly, the authors of [18] implemented a system using millimeter-wave radar combined with a convolutional neural network (CNN) to detect and classify human behavior.

2) Human Activity Recognition: Human activity recognition, also known as movement pattern analysis, has been shown to be facilitated through mmWave technology. For instance, the work reported in [19] implements a system that utilizes commercial, off-the-shelf radar to obtain sparse point clouds. It then employs a sliding time window to accumulate these clouds, generating a voxelised representation that serves as input to human activity classifiers, which differentiate between five different activities. In [20], the authors introduce "POI-GAN", a novel approach tailored to forecasting pedestrian trajectories in service-centric settings. Pedestrian trajectory data is obtained from the HLK-LD6001A-60G millimeter-wave radar, and proposes a prediction method that leverages the Point of Interest (POI) model and the field of view angle model. A comprehensive survey that discusses the evolving field of human behavior analysis using radar and LiDAR is reported in [21].

III. Why mmWave-based crowd Sensing

Having reviewed the literature and considering the specific requirements that crowd sensing imposes regarding privacy, environmental robustness, and accuracy, it appears that mmWave emerges as a promising solution. Unlike technologies such as cameras, Wi-Fi/Bluetooth, which could lead to identification of users, mmWave is considered to be positioned at the lower end of the intrusiveness spectrum because it does not collect visual or user-specific data, making it suitable for sensitive areas (e.g., hospitals, schools) and compliant with strict privacy regulations. Considering crowd sensing applications, mmWave holds distinct advantages over other traditional approaches: (1) offers enhanced spatial resolution as it could potentially distinguish better between individuals spaced closely together, which is a challenge for infrared or thermal sensors, (2) offers environmental and visual robustness as it can operate effectively in a wide range of environmental conditions including darkness, smoke, and fog, (3) it is non-Intrusive and privacy-Compliant since mmWave radars do not capture identifiable information, (4) has high refresh rate enabling real-time tracking since mmWave sensors can detect movement and update positions in real time, essential for dynamic crowd scenarios, (5) has fairly low infrastructure requirements given that a single mmWave unit can cover a relatively wide area, reducing the number of sensors required compared to camera-based systems and (6) has low Power and compact design as mmWave modules are energy-efficient and can be embedded into battery-powered IoT devices. In fact, mmWave communication can be applied to enable and support Metaverse applications [22], [23], as it allows high-speed data transfer [24], making it particularly well-suited for the intensive bandwidth and low-latency demands of immersive Metaverse experiences [22].

A. Potential Application Examples in Metaverse Systems

In Metaverse applications, particularly those involving crowd sensing and visualization, mmWave technology offers the opportunity to capture and transmit fine-grained spatiotemporal data in real-time, enabling dynamic and responsive representations of user behavior within physical settings and in virtual environments. An example of a Metaverse application where this is highly relevant is the Intelligent Reality Virtual Museum prototype, developed within the context of a Cyber-Physical-Social Metaverse system under development at [Research group and relevant citations omitted for peer review process]. This system combines advanced AI and Large Language Models (LLMs) with Digital Twins, Robotics, and gaming technologies to deliver immersive cultural heritage experiences that seamlessly blend the real and digital worlds. It serves as an example of a CPSS Metaverse application, demonstrating the convergence of multiple emerging technologies to transform how cultural heritage is experienced, preserved, and interacted with in the Metaverse. In such an example system, mmWave technology can be used to support non-intrusive tracking of visitor movement and density within the digital twins of physical spaces, helping to establish and maintain real-time synchronization between the physical and virtual environments. This capability enables the generation of heatmaps to visualize visitor flow and highlight areas of interest within a physical space, which can be used to adapt the virtual experience dynamically. Moreover, the system can leverage this data to dispatch a robotic guide to specific locations within the physical or virtual museum, offering contextual assistance or information to visitors based on their position and behavior.

Another potential application example is a project developed by (Citation and group information removed for peer review), focusing on indoor air quality monitoring by non-experts. The project utilizes Virtual Reality and digital twins to enhance spatial understanding and support the analysis of indoor environmental conditions. In this project, ecological sensors were deployed within a mixed-use industrial office to capture historical data on air quality metrics. The system infrastructure facilitates continuous data collection and storage, allowing users to visualize the physical space through 2D and VR-based digital twin interfaces. This supports data visualization, spatial awareness, user immersion, and intuitive data exploration, enabling non-expert users to perform trend analysis and make informed decisions related to air quality and building performance. In such types of systems, mmWave technology can be used to generate real-time heatmaps to visualize human presence and movement patterns within the indoor environment. When coupled with sensor readings, this can help identify correlations between human activity levels and variations in air quality, helping to identify usage-based inefficiencies or ventilation issues to further enhance decision-making and support more targeted interventions for improved building management strategies.

IV. Methodology

Considering the potential of mmWave technology for practical applications within Metaverse environments, particularly for crowd sensing and spatial analytics, and to facilitate the investigation and development of crowd estimation algorithms, a demonstration scenario simulating a physical museum/gallery-style environment is employed. We developed a MATLAB-based simulator that models visitor behavior in these environments and generates synthetic mmWave sensor data corresponding to pre-defined sensor positions. This data is generated pseudorandomly for each simulated visitor position based on an error model derived from a set of mmWave measurements conducted to characterize and analyze the precision of a typical mmWave sensor in terms of range, azimuth, elevation, and the number and spread of echoes detected when tracking human targets moving within its field of view.

A. Precision Analysis

The precision analysis was conducted to characterize the ranging and angular capabilities of the mmWave sensors used in the simulator and extract the range, azimuth and elevation errors at different distances and orientations of the target. In this experimental analysis, a Texas Instruments IWR1843BOOST sensor was placed at a fixed location and range measurements were collected on a straight line (1 to 8 m) every 1 m with a 1.7 m-tall man standing away from it, making minor movements with his hands to generate Doppler shifts and enable the detection. To characterize the ability of the sensors to conduct range measurements at different angles, the orientation of the sensor was systematically varied from 0 to 60 degrees in the azimuth plane and from 0 to 45 degrees in the elevation plane (15^o step in both cases). At every point, the sensor was set to receive measurements for 5 seconds while the target was moving, and after excluding outliers (using DBSCAN with low ε), the mean values of the range, elevation and azimuth readings are recorded. Using these three values, we estimate the 3D position of the target and we compare it to the true position to calculate the 3D Euclidean error.

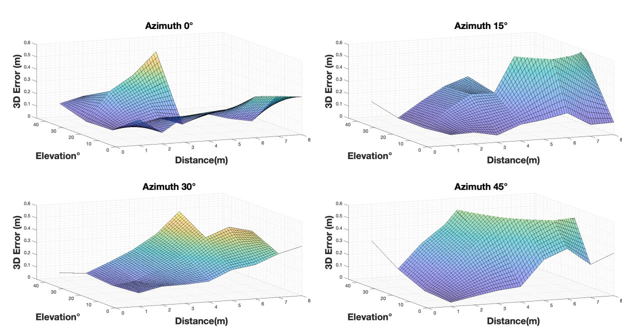


Fig. 1: Precision Analysis Results

Figure 1 shows the 3D Euclidean error achieved under different orientations of the sensor with respect to the target and at different distances from the sensor. The evaluation revealed a varying level of accuracy contingent on azimuth and elevation angles up to a maximum distance of $8m$. Beyond 8 m the sensor was not able to detect the target. When the sensor was aligned at 0 degrees azimuth, it demonstrated exceptional 3D accuracy, which was slowly degrading as the distance and elevation from the target were increased. As it was expected, at large elevation angles (e.g. 45 degrees) and at long distances, the sensor was failing to provide a measurement. This is indicated by the gaps in the surface plots in the figure. The contour plots at the bottom of each surface plot indicate the range of Distance/Elevation values for which the error is below an intuitively-selected accepted 3D positioning error ($0.5m$). As expected, the positioning accuracy appears to deteriorate more rapidly as the azimuth angle increases, limiting the usability and reliability of the sensors at very low elevation and azimuth angles. In a scenario where multiple sensors are used, one could use the range, azimuth and elevation measurements as a measure of the reliability of the single-anchor position estimation and either use or discard the particular anchor from the entire positioning algorithm.

The output of this precision analysis is a data structure that contains the range, azimuth and elevation errors of the measurements at different ranges and orientations, and this is fed as input to the mmWave Data Generator (see section IV-B)

B. Mobility and mmWave Simulator

As noted above, a custom-developed MATLAB-based simulator was developed to model visitor 2D or 3D movements in art-related bounded environments and generate data from mmWave sensors defined in such environments. The simulator is based on the assumption that exhibits in museums and art galleries are typically arranged to guide visitors through a comfortable route and that they are reasonably spaced apart from each other. It assumes that visitors form clusters in front (in the case of paintings) or around the exhibits (for artefacts or sculptures), and that they move from one exhibit to another either as groups or individually. It is assumed that atop each exhibit, there is at least 1 mmWave sensor that generates mmWave data for each user roaming in front of it.

Given these requirements, the simulator accommodates both clustered and unclustered users (defined by an unclustered ratio), each exhibiting distinct mobility behaviours. Users are assigned specific mobility types—restricted within clusters, free movement, or movement toward other clusters. The simulator enforces constraints such as minimum distances between clusters and users, and allows for customisable parameters including the total number of users, cluster configurations, and simulation duration. This flexibility also makes it applicable for studies in wireless communication networks, smart city planning, and IoT environments.

To offer flexibility in supporting various user behaviours and spatial constraints, enabling detailed analysis of mobility patterns and cluster dynamics, the input to the simulator is:

- **use_3D**: Boolean flag for 3D or 2D simulation
- **bounding_box**: 2D or 3D environment dimensions of the entire space, e.g., $[X_{\min}, X_{\max}; Y_{\min}, Y_{\max}; Z_{\min}, Z_{\max}]$
- **sim_time**: Total simulation time (seconds)
- **tick**: Time step (seconds)
- **NU**: Total number of users (fixed integer)
- **define_clusters_explicitly**: Boolean flag for explicit definition or random generation of clusters
- **NC_min, NC_max**: Minimum and maximum number of clusters (integers)
- **min_users_per_cluster, max_users_per_cluster**: Minimum and maximum number of users allowed per cluster
- **cluster_radius_min, cluster_radius_max**: Minimum and maximum radius for each cluster
- **min_cluster_distance**: Minimum allowed distance between cluster centers
- **min_user_distance**: Minimum allowed distance between any two users
- **pedestrian_velocity**: User movement speed (m/s)
- **unclustered_ratio**: Fraction of users not assigned to any cluster (0 to 1).
- **mobility_distribution**: Defines the mobility behaviour of users: 0 = restricted within a cluster, 1 = to move arround freely, 2 = move between clusters. These are defined as percentages (e.g. 80% restricted, 5% free and 15%: move between clusters)

The output of this mobility model consists of user trajectories arranged in a 3D array $[NU \times D \times T]$, representing the positions of each user over time, which serves as input for the mmWave measurement generator. A typical snapshot of this mobility model is shown in Figure 2.

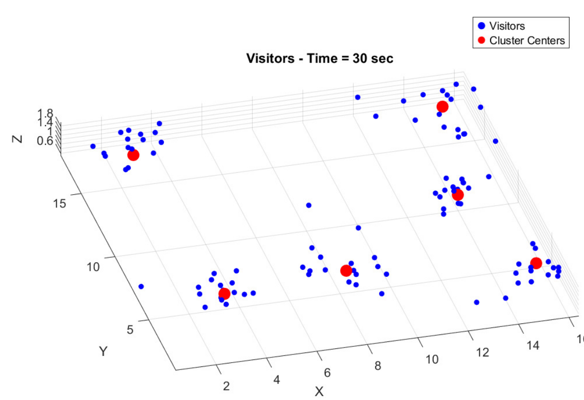


Fig. 2: Typical Mobility Model Output for a preset set of visitor clusters

For each user position, a cloud of measurement points is generated around the true user location, corresponding to reflections from the user's body. The radius (σ) of this cloud is determined based on the user's cross-sectional area; our precision analysis has indicated a typical value of sigma for the human body of 0.5 m. For each reflection point within this cloud, the true range, azimuth, and elevation relative to each mmWave sensor in the environment are calculated as follows:

$$\begin{aligned} range_{\text{true}} &= \|\mathbf{p}_{\text{sensor}} - \mathbf{p}_{\text{target}}\| \\ azimuth_{\text{true}} &= \arctan 2(y_{\text{rel}}, x_{\text{rel}}) \\ elevation_{\text{true}} &= \arcsin\left(\frac{z_{\text{rel}}}{range_{\text{true}}}\right) \end{aligned}$$

where $(x_{\text{rel}}, y_{\text{rel}}, z_{\text{rel}})$ is the relative position vector from the sensor to the reflection point. Subsequently, measurement errors corresponding to these true measurements ($range_{\text{err}}$, $azimuth_{\text{err}}$, $elevation_{\text{err}}$) are obtained using the previously conducted precision analysis, employing interpolation when necessary. Finally, the simulated measurements are generated by adding normally distributed noise with zero mean and standard deviation equal to these extracted errors to the actual (true) measurement values, as follows:

$$\begin{aligned} range_{\text{meas}} &= range_{\text{true}} + \mathcal{N}(0, range_{\text{err}}) \\ azimuth_{\text{meas}} &= azimuth_{\text{true}} + \mathcal{N}(0, azimuth_{\text{err}}) \\ elevation_{\text{meas}} &= elevation_{\text{true}} + \mathcal{N}(0, elevation_{\text{err}}) \end{aligned}$$

Given the simulated measurements, the spherical coordinates are calculated.

$$\begin{aligned} r &= range_{\text{meas}} \\ \theta &= azimuth_{\text{meas}} \quad \text{(in radians)} \\ \phi &= elevation_{\text{meas}} \quad \text{(in radians)} \end{aligned}$$

The corresponding Cartesian coordinates relative to the sensor are computed as:

$$\begin{aligned} x_{\text{rel}} &= r \cdot \cos(\phi) \cdot \cos(\theta) \\ y_{\text{rel}} &= r \cdot \cos(\phi) \cdot \sin(\theta) \\ z_{\text{rel}} &= r \cdot \sin(\phi) \end{aligned}$$

It is important to note that these Cartesian coordinates correspond to the body frame coordinate system of each sensor. To properly determine the coordinates of the point of reflection within the room's coordinate plane, it was essential to align the sensor's coordinate system (body frame coordinate system) with that of the room (Local Coordinate System). Achieving this alignment involves a series of calculations that account for the sensor's yaw, pitch, and roll. These adjustments were critical in ensuring that the sensor's data corresponds accurately to the room's coordinate plane, allowing for reliable 3D positioning. Assuming that the anchor is first rotated by an angle ψ around the z-axis (yaw), then by an angle θ around y-axis (pitch) and finally by an angle ϕ around the x-axis (roll) the 3 ×3 rotation matrix $\mathbf{R} = R_z \cdot R_y \cdot R_x$.

$$\mathbf{R} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}$$

With reference to Figure 3 and considering that body-frame measurement from a sensor positioned at $A = [x_a y_a z_a]$ is $P_{rel} = [x_{rel} y_{rel} z_{rel}]$ then the local coordinates $P = [xyz]$ of the target can be calculated using:

$$\text{P} = [R \cdot P_{rel}^T]^T + A$$

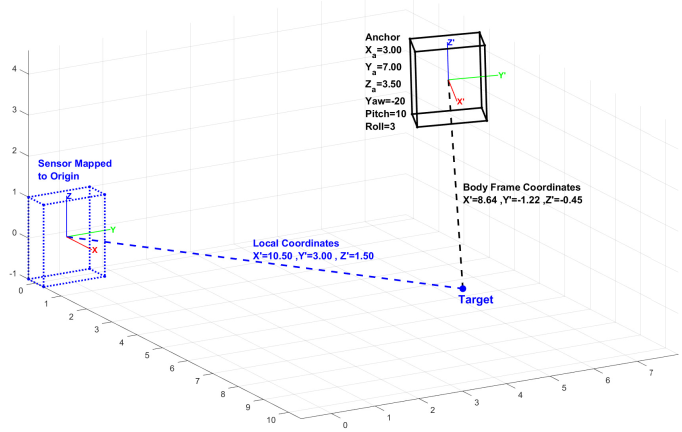


Fig. 3: Body Frame to Local Coordinates Conversion

The generator iterates through all users in the environment and generates a combined set of reflection points for each user, as well as for all sensors within the environment, per time instance. This will then form the input to the clustering algorithm, which will identify and group/cluster the visitors.

In addition to the mobility information, the generator takes as input a structure containing all the sensors in the environment, which are defined in terms of their Cartesian coordinates in the global coordinate system and their orientation (yaw, pitch, roll), the precision analysis data, the maximum number of scatter measurements per user per sensor and the sigma (σ) of the measurement reflection points as defined earlier. A typical output of the mmWave Generator for the users shown in Figure 2 is shown in Figure 4.

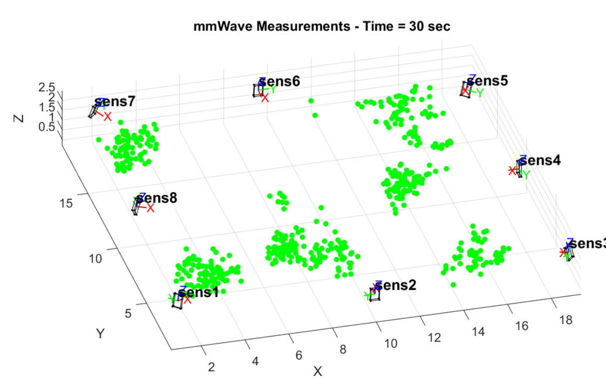


Fig. 4: Typical mmWave Generator Output for the users shown in Figure 2 and 8 mmWave sensors

C. Clustering

In data analysis, clustering algorithms play a pivotal role in interpreting complex datasets. For this investigation, we utilise a combination of DBSCAN and K-means algorithms to identify clusters of users, estimate their centres and radii.

1) K-Means Clustering: K-Means is a centroid-based clustering algorithm that partitions the data into K distinct, non-overlapping clusters. The algorithm iteratively adjusts the centroids to reduce the total variance within each cluster [25]. It requires the number of clusters (K) to be specified in advance, which is a key parameter that directly influences the clustering outcome, as it determines the granularity of the clustering. A value that is too low may merge distinct groups into a single cluster. At the same time, a value that is too high may lead to overfitting, identifying clusters within what is essentially noise, splitting cohesive clusters into multiple, smaller ones.

To overcome the limitation of having to specifically define the value of K we deploy the silhouette method to determine its optimal value. For each data point i, the silhouette score $s(i)$ is defined as:$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}}$, where: $a(i)$ is the average distance between point i and all other points in the same cluster (intra-cluster distance), and $b(i)$ is the minimum average distance between point i and all points in any other cluster (nearest-cluster distance). The silhouette score ranges between: $s(i) \approx 1$: point is appropriately clustered, $s(i) \approx 0$: point lies between two clusters, and $s(i) < 0$: point may have been assigned to the wrong cluster. The average silhouette score $\bar{s}$ across all data points provides a measure of the overall clustering quality. The optimal number of clusters K^* is selected as: $K^* = \arg\max_K \ \bar{s}$.

2) DBSCAN Clustering: DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm. It groups up points that are closely packed together, marking as outliers any points that lie alone in low-density regions. This method is effective for data with clusters of similar density [26], and its algorithm effectiveness relies heavily on two parameters: epsilon (ε) and minimum cluster size ($MinPts$). (ε) sets the neighbourhood radius around each centroid, dictating what is considered close enough to form a cluster, while $MinPts$ defines the minimum number of points a cluster needs, distinguishing core points (points in dense areas) from noise (points in sparse areas). Balancing ε and $MinPts$ is essential to ensure the correctness of the result. Capitalising on the ability of this method to remove outliers, we use it as a pre-clustering step, prior to the K-means algorithm, to clean the data from any points that most likely are not part of clusters. A low value of ε is used in this case.

V. Results

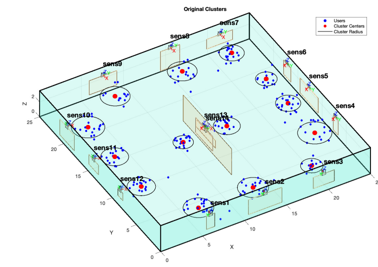
In this section, we present the results of a simulation conducted in a typical art gallery environment. The simulated space is a $25 \times 25\,\mathrm{m}$ square area featuring 14 paintings displayed along the surrounding walls and on a central wall located in the middle of the gallery, as illustrated in Figure 5a. A mmWave sensor is deployed $20\,\mathrm{cm}$ above each painting, oriented to face the space in front of the painting and tilted downward by 15°.

The scenario under investigation involves 250 visitors moving throughout the gallery, forming clusters approximately $2.5\,\mathrm{m}$ in front of each painting. The number of individuals per cluster is randomised between 5 and 20, with each person spaced at least $30\,\mathrm{cm}$ apart. The maximum diameter of each cluster is assumed to be $1\,\mathrm{m}$ greater than the width of the corresponding painting. The simulation also assumes that 5% of visitors remain unclustered and may appear anywhere within the gallery space, while the rest are restricted to stay within their cluster. Figure 5a presents a snapshot of this scenario, showing users forming clusters with centres depicted as red spheres and their corresponding diameters indicated by black circles around each centre.

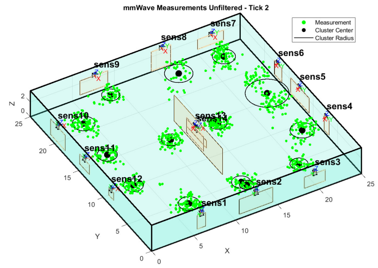
To generate the mmWave data, the simulation assumes that each user produces between 5 and 10 measurements (echoes) originating from different locations on their body, with an average spatial spread of 0.5 m. Figure 5b presents the generated sensor data along with the results of the K-means clustering algorithm, where the maximum number of clusters K was set to 20. The black spheres represent the estimated cluster centres, while the black circles around them illustrate the approximate diameter of each cluster. Although 14 clusters were identified (most of them correctly positioned), it appears that users located in front of sensors 5 and 6 were grouped into a single cluster, whereas two clusters were detected in front of sensor 2, where only one was expected. To address this issue, a filtering step was introduced using the DBSCAN algorithm to remove outliers ($\varepsilon = 1$, $MinPts = 8$) before running the K-means algorithm. The new results are shown in Figure 5c.

The first scenario is rather unrealistic, as it is unlikely that all gallery visitors remain stationary and clustered in front of the paintings without moving around. To simulate a more dynamic and realistic environment, and to evaluate the applicability and effectiveness of our approach, the simulator was run again with increased user variability. Specifically, 15% of users were initialised as unclustered, 10% were allowed to move freely within the gallery, 30% moved between clusters, and the remaining users stayed within their assigned clusters. All visitors were assumed to walk at a speed of 0.5 m/s. The simulation was run for 1 minute, resulting in the user distribution shown in Figure 5d. Due to the increased spread of users across the gallery space, the resulting mmWave measurements were not easily clustered using the same parameters and algorithms as before. To mitigate this, the sensors were reconfigured to discard measurements with azimuth angles beyond $\pm 45^\circ$, elevation angles beyond $\pm 15^\circ$, and distances greater than 7 m. The mmWave data obtained after this post-processing step, without outlier removal, is shown in Figure 5e. Although this produced a reasonably accurate estimation of cluster positions, applying the DBSCAN algorithm for outlier removal further improved the results, as illustrated in Figure 5f.

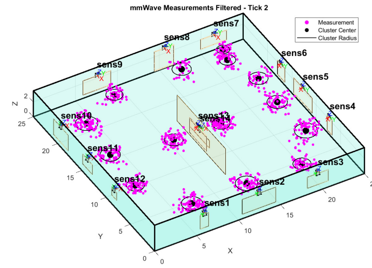
Beyond identifying clusters, the proposed solution can mon-



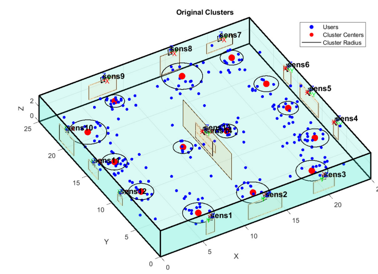
(a) Scenario 1: Visitors clustered in front of each painting



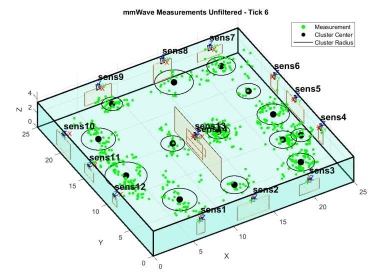
(b) Scenario 1: mmWave Measurements



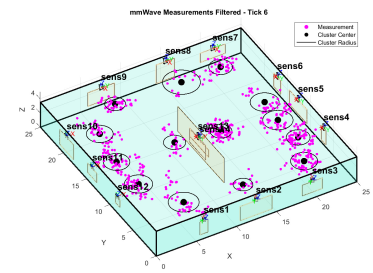
(c) Scenario 1: Filtered mmWave Measurements



(d) Scenario 2: Visitors clustered in front of each painting



(e) Scenario 2: mmWave Measurements



(f) Scenario 2: Filtered mmWave Measurements

Fig. 5: Comparison of visitor clustering and mmWave measurements.

itor user mobility over time and extract valuable insights into visitor distribution, artefact popularity, and areas prone to overcrowding. This information can help management make informed decisions about redistributing artefacts to enhance visitor satisfaction and overall experience quality. To demonstrate this concept, we developed a routine that divides the virtual space into cells of predefined size and accumulates the occurrences of mmWave measurements recorded within each cell over a specified period and plots a normalised. heatmap. An example of such a heatmap is presented in Figure 6, illustrating user mobility in Scenario 2 over a 60-second period.

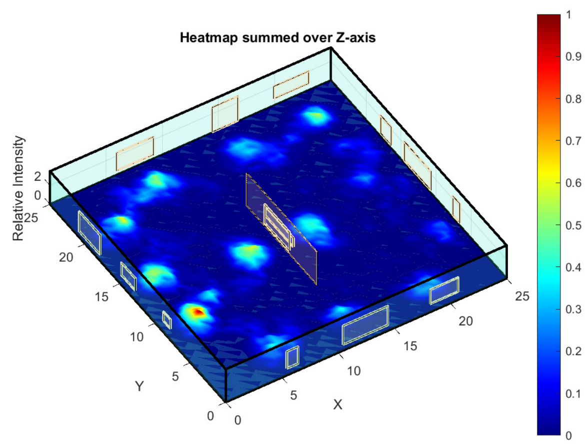


Fig. 6: Users Mobility Heatmap over 60 seconds

VI. Conclusion and Discussion

The use of mmWave technology in Metaverse systems presents significant potential for advancing how virtual and physical spaces are synchronized and experienced. The technology's' potential for high spatial resolution, low latency, accuracy across various lighting conditions, and the preservation of user privacy due to its non-reliance on visual imaging can enable real-time, non-intrusive crowd monitoring that can enpower a wide range of Metaverse applications ranging from digital twins of public spaces to interactive exhibitions, simulations, and smart city planning tools.

The deployment of mmWave-enabled crowd sensing addresses practical challenges in Metaverse development, such as maintaining synchronicity between physical and virtual environments with high scalability that can support the development of Cyber-Physical-Social Systems within the Metaverse. Our findings confirm that mmWave sensors could deliver reliable spatial data under dynamic, crowded conditions—an essential requirement for immersive and context-aware Metaverse applications. Combining DBSCAN and K-Means clustering further enables accurate detection and representation of crowd distributions, facilitating the generation of spatial heatmaps and mobility patterns. Beyond static crowd localization, the proposed system supports dynamic behavioural insights, such as identifying popular exhibits, crowd bottlenecks, and navigation trends. These insights are particularly valuable for use cases in cultural heritage, education, public safety, smart city planning, and XR-driven event management.

To translate these simulation insights into deployable systems, there is a need to quantify the practicalities of a museum-scale installation—namely the capital and operating cost of a dense mmWave grid, options for powering sensors over long horizons (PoE wiring vs. low-duty-cycle batteries), and the constraints imposed by local licence-free allocations of the 26-28GHz spectrum (mostly limited coverage due to high path loss and limited field of view). A preliminary techno-economic and energy-budget analysis is needed to identify the minimum viable sensor density that meets our requirements while keeping the total implementation and operational low for cultural-heritage operators.

Looking forward, several promising directions emerge. First, integrate a real-world mmWave hardware deployment in a realistic environment to validate our findings and refine our models and algorithms. Secondly, fusing mmWave data with complementary sensing modalities (e.g., audio, environmental, or inertial sensors) could enhance accuracy and enable richer behaviour classification.

References

[1] L. Nisiotis, C. Metra, J. Athavale, D. Milojcic, S. Dukes, J. Ranaweera, and S.-K. Hawkins, " The Emerging Metaverse: IEEE ISEMV 2024 Retrospective and Future Directions ," *Computer*, vol. 58, no. 04, pp. 58–64, Apr. 2025. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/MC.2025.3534445

[2] L. Nisiotis, "A system architecture for ai-driven interactive systems within a cpss framework for metaverse applications," in *2024 IEEE International Symposium on Emerging Metaverse (ISEMV)*, 2024, pp. 9–12.

[3] W. Wang, Y. Yang, Z. Xiong, and D. Niyato, "Footstone of metaverse: A timely and secure crowdsensing," *IEEE Network*, vol. 38, no. 2, pp. 171–178, 2024.

[4] C. Celes, A. Boukerche, and A. A. F. Loureiro, "Crowd management: A new challenge for urban big data analytics," *IEEE Communications Magazine*, vol. 57, no. 4, pp. 20–25, 2019.

[5] D. Koshnicharova, Nidhi, A. Mihovska, P. Koleva, and V. Poulkov, "Data-driven interactive crowd management systems for metaverse scenarios," in *2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2022, pp. 549–554.

[6] D. Darsena, G. Gelli, I. Iudice, and F. Verde, "Sensing technologies for crowd management, adaptation, and information dissemination in public transportation systems: A review," *IEEE Sensors Journal*, vol. 23, no. 1, pp. 68–87, 2023.

[7] D. Wang, M. Fattouche, and X. Zhan, "Pursuance of mm-level accuracy: Ranging and positioning in mmwave systems," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1169–1180, 2019.

[8] A. Sesyuk, S. Ioannou, and M. Raspopoulos, "Radar-based millimeter-wave sensing for accurate 3-d indoor positioning: Potentials and challenges," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, vol. 2, pp. 61–75, 2024.

[9] ——, "3d millimeter-wave indoor localization," in *2023 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2023, pp. 1–7.

[10] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 961–971.

[11] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, and P.-A. Muller, "Recent trends in crowd analysis: A review," *Machine Learning with Applications*, vol. 4, p. 100023, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666827021000049

[12] C.-H. Pu and X. L. Tan, "Lidar and image filtering and fusion techniques for real-time crowd monitoring system," in *2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA)*, 2024, pp. 1–6.

[13] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Laser-based detection and tracking of multiple people in crowds," *Computer Vision and Image Understanding*, vol. 106, no. 2, pp. 300–312, 2007, special issue on Advances in Vision Algorithms and Systems beyond the Visible Spectrum. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314206001937

[14] D. C. Duives, T. van Oijen, and S. P. Hoogendoorn, "Enhancing crowd monitoring system functionality through data fusion: Estimating flow rate from wi-fi traces and automated counting system data," *Sensors*, vol. 20, no. 21, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/21/6032

[15] A. Lesani and L. Miranda-Moreno, "Development and testing of a real-time wifi-bluetooth system for pedestrian network monitoring, classification, and data extrapolation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1484–1496, 2019.

[16] P.-F. Tsai, C.-H. Liao, and S.-M. Yuan, "Using deep learning with thermal imaging for human detection in heavy smoke scenarios," *Sensors*, vol. 22, no. 14, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/14/5351

[17] L. Ren, A. G. Yarovoy, and F. Fioranelli, "Grouped people counting using mm-wave fmcw mimo radar," *IEEE Internet of Things Journal*, vol. 10, no. 22, pp. 20 107–20 119, 2023.

[18] R. Zhang and S. Cao, "Real-time human motion behavior detection via cnn using mmwave radar," *IEEE Sensors Letters*, vol. 3, no. 2, pp. 1–4, 2019.

[19] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, ser. mmNets '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 51–56. [Online]. Available: https://doi.org/10.1145/3349624.3356768

[20] Y. Li, C. Zhang, J. Zhou, and S. Zhou, "Poi-gan: A pedestrian trajectory prediction method for service scenarios," *IEEE Access*, vol. 12, pp. 53 293–53 305, 2024.

[21] P. Miazek, A. Żmudzińska, P. karczmarek, and A. Kiersztyn, "Human behavior analysis using radar data: A survey," *IEEE Access*, vol. 12, pp. 153 188–153 202, 2024.

[22] Z. Huang, P. Yang, C. Zhou, W. Wu, and N. Zhang, "Joint sensing and communication for mmwave vr in metaverse: A meta-learning approach," *IEEE Internet of Things Journal*, vol. 11, no. 13, pp. 24 049–24 060, 2024.

[23] Y. Wang, M. Chen, Z. Yang, W. Saad, T. Luo, S. Cui, and H. V. Poor, "Meta-reinforcement learning for reliable communication in thz/vlc wireless vr networks," *Trans. Wireless. Comm.*, vol. 21, no. 9, p. 7778–7793, Sep. 2022. [Online]. Available: https://doi.org/10.1109/TWC.2022.3161970

[24] K. Zrar Ghafoor, L. Kong, S. Zeadally, A. S. Sadiq, G. Epiphaniou, M. Hammoudeh, A. K. Bashir, and S. Mumtaz, "Millimeter-wave communication for internet of vehicles: Status, challenges, and perspectives," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8525–8546, 2020.

[25] K. P. Sinaga and M.-S. Yang, "Unsupervised k-means clustering algorithm," *IEEE access*, vol. 8, pp. 80 716–80 727, 2020.

[26] K. Khan, S. U. Rehman, K. Aziz, S. Fong, and S. Sarasvady, "Dbscan: Past, present and future," in *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*. IEEE, 2014, pp. 232–238.