Human Activity Recognition Using Two Millimeter-Wave Radars

ABDELRAHMAN ESHY, University of Twente, The Netherlands

Human activity recognition (HAR) aims to label, recognize, track human activities accurately, and it has been implemented through several approaches, such as ambient sensors, cameras, or wearable devices. However, in privacy-sensitive areas, a camera could collect extraneous ambient information that a user may not feel restful revealing. Therefore, millimeter wave (mmWave) radars have been proposed as an alternative for detecting and tracking human activity. The mmWave radars endure the unique advantage of being effective under non-line-of-sight scenarios, effectively capture a minimal subset of the ambient information using micro-Doppler spectrograms producing higher accuracy, and can track the user while preserving privacy. The article proposes an approach that can detect human activity recognition and track the human user accurately by using two-millimeter wave (mmWave) radars. The approach focuses on advanced machine learning algorithms, innovations in hardware architecture, and decreasing monitoring costs. This paper proposes RadHAR, a framework that performs accurate human activity detection using point clouds. The collected human activity data-set got evaluated, and a comparison of the accuracy of various classifiers on the data set found that the best-performing deep learning classifier achieves an accuracy of 97.71%. The evaluation shows the efficacy of using two mmWave radars for accurate HAR detection and reliable tracking.

Additional Key Words and Phrases: Activity Recognition, Classifiers, Millimeter Wave, Clusters, Tracking

ACM Reference Format:
Abdelrahman Eshy. 2022. Human Activity Recognition Using Two Millimeter-Wave Radars. In . ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

HAR aims to recognize the actions and movements of one or more users from a series of observations of the user's activities and label each action. Since human activity recognition is complex and highly diverse, many HAR systems have been intended and developed based on the recent development of sensor technologies and machine learning algorithms. For example, ambient sensors (e.g., cameras), wearable devices (e.g., smartwatches), and WiFi signals. These frameworks are usable and provide practical and accurate results. Despite that, radars have become a popular activity recognition process since they can operate in any lighting situation and various environmental situations, such as fog and rain. [2]. The mmWave radars enable a cost-effective sensing application, are compact, and have significant bandwidth and ideal range resolution; the new low-cost technology increased the popularity of mmWave-based solutions. There exist multiple frameworks regarding human activity recognition using one mmWave radar. The existing procedures and framework results using radars for HAR were points of interest for proposing different modifications and approaches using two radars to improve the detection results and provide accurate, reliable tracking. Recognizing human activity using two mmWave radars reveals more dense information than using one mmWave radar, as two mmWave radars provide higher resolution in the distance, velocity, and angle estimation of objects in the scene. Therefore, it effectively understands human status and motion and detects human activities more accurately. Hence, this paper will show the approach and framework of using two mmWave radars instead of the existing approaches for accurately recognizing human activities and achieving accurate tracking.

The Article Questions:

- What is the suitable approach for detecting human activities using two mmWave radars?
- What advancements in using two mmWave radars for HAR?
- What effect do two mmWave radars have on signal interference?
- What effect do mmWave radars have on tracking the user?
- What is the influence of noise on user tracking and detection?

The Article Objectives:

- Enhance the accuracy of HAR by utilizing two mmWave radars.
- Propose two mmWave radars framework and deep learning classifiers.
- Provide accurate, reliable tracking using two mmWave radars.

The Article Contributions:

- Proposes a suitable approach for HAR using two mmWave radars.
- Presents frameworks for using the two mmWave radars.
- Presents the advancements of two mmWave radars.
- Evaluates the performance and results of one mmWave radar and two mmWave radars.
- Presents the approach for tracking human activities using two mmWave radars.

2 RELATED WORK

Human activity recognition (HAR) gets widely researched through several sensing modalities, and numerous approaches have to get proposed in the literature. Researchers researched human activity recognition using sensors like cameras [22], WiFi [23], Wearable Devices [24], and millimeter-wave radar. First, Wearable devices have a practical approach to human activity recognition because they are attached to humans and provide incessant information regarding human activity. Wearable devices offer real-time feedback to the user based on the embedded processor for processing the data; thermal profligacy and power consumption are the main factors for powering the embedded platforms. It is unrealistic to presume that all people will use wearable devices compatible with the interface model [14]. Therefore, this drawback of wearable devices is the incentive to use the mmWave radar for recognizing human activities.

Second, a camera's recognition of human activity can provide dense information regarding human activities. Moreover, depth cameras exist, also known as three-dimensional (3D) cameras. These depth

TScIT 37, July 8, 2022, Enschede, The Netherlands

cameras provide distance information from the camera to the object besides the standard image and can capture three-dimensional information that leads to three-dimensional points cloud-based on the recognition of human activities[7]. Moreover, cameras can track moving objects and provide the most detailed dimensional information [11]. However, cameras are interfering, and it is a privacy issue, especially in privacy-sensitive areas, which would not let the user feel at ease with the possibility of capturing unnecessary ambient information. The mmWave tracking, the human activity approach, is based on the data points from the collected point cloud data set and avoids the privacy concern of tracking using the camera. Nonetheless, The accuracy of using a camera is equivalent to using millimeter wave radars [25]. Therefore, two mmWave radars can achieve higher accuracy than one mmWave radar and camera, besides avoiding camera concerns. Despite that, cameras can get used as ground truth for recognizing human activities using radars [17]. Hence, using two mmWave radars is adroiter and avoids all the camera's drawbacks.

Third, WiFi signals can recognize human motions and positions and detect if the human has fallen. The fundamental intuition is that different movements and positions introduce various multi-path distortions in WiFi signals and generate different patterns in the time series of channel state information (CSI) [1]. Nevertheless, WiFi needs more range resolution compared to mmWave radars regarding robust classification. Due to atmospheric attenuation, millimeter radars got limited to short-range applications of around 5 km for a 94 GHz transmission [10]. Since WiFi signals usually reach about 45 meters for a 2.4GHz frequency [9]. However, recognizing the human activities by two mmWave radars can provide more range resolution. Range resolution quantifies the capability to detect two objects separated in range along the same line of sight. Therefore, detecting human activities using two mmWave radars has more range resolution than WiFi and, based on higher resolution, indicates higher accuracy [26].

Last is the low-cost single-chip mmWave radar for human activity recognition and tracking. The mmWave radar modality uses the minimum amount of ambient information using micro-Doppler spectrograms. The Doppler of a radar signal is a frequency shift due to the relative radial velocity between the radar and a target. Suppose the target consists of multiple scatterers, such as a person's limbs; their velocity differences lead to multiple Doppler frequencies, termed micro-Doppler[8]. First, the mmWave radar recognizes human activity. It produces a point cloud, then a conversion from the point cloud data to a micro-Doppler spectrogram before using a CNN to classify it[6]. Lastly, a voxelization representation of the point clouds for HAR using LSTM and CNN + LSTM classifier.

3 MILLIMETER-WAVE RADAR

This section investigates the potential of using mmWave radars for HAR and selects the FMCW (frequency modulated continuous wave) mmWave radars made by Texas Instruments(TI). The FMCW radars are a type of radar system where a known stable frequency continuous electromagnetic wave radio energy is transmitted and received from any reflecting objects and detect human activities in point clouds. By capturing the reflected signal, a radar system can determine the objects' range, velocity, and angle [5] in the scene, which effectively understands the human status, and motion and distinguishes the object of interest from the background cluster[4]. In this paper, the frequency modulated continuous wave(FMCW) mmWave radar is the Texas Instrument's IWR1443BOOST, with 76-81 GHz frequency. The mmWave radar contains a chip consisting of three transmitters and four receivers operating concurrently and equipped with multiple antennas besides integrated circuits and hardware accelerators for a complete on-chip data processing chain [4].

The mmWave radars have a superior range resolution based on the large bandwidth. The antenna size is inversely proportional to frequency. So, the higher the frequency spectrum, the smaller the antenna size. Accordingly, mmWave radars are compact, enclosing many antennas into minimal space, enabling highly directional beam-forming ($1 angular accuracy$). The narrow beam effectively allows the two mmWave radar to separate and recognize various environmental targets, such as intersections. The narrow beam in the angular direction and distance and velocity is necessary to detect target objects observed by radar separately [21]. The mmWave radars are resource-constrained; instead of providing raw data, their output is point clouds. The number of points in each frame captured by the two mmWave radars varies, increasing the complexity of constructing a neural network architecture that can process this data.

3.1 Intermediate Frequency

The mmWave radar transmitters send chirp signals Stx (a signal with the frequency increasing linearly with time) to detect objects in front of the radar. When the objects reflect Stx, the signal gets received as Srx. The radar combines the two signals Stx and Srx with a mixer and a low-pass filter to produce a mixed intermediate frequency (IF) signal [4]. The IF signal contains a frequency and phase equal to the difference between the transmitted signal Stx and the received signal Srx [5]. Then performing, a data processing chain over the IF signal to determine the presence of any objects.

3.2 Range Resolution

Micro-movements of minor extremities, such as the legs and arms, are challenging to recognize. However, configuring the radar resolution to make it higher improves the faithfulness of the measurements of such activities, besides improving the data quality of the classifiers for learning it. The range resolution (d_{res}) determines the minimum distance between two objects and distinguishes them. So, taking (c) the speed of light and (B) the bandwidth of the exhaustive chirp. So, the range resolution formula is ($d_{res} = \frac{c}{2B}$).

3.3 Velocity Resolution

The total bandwidth of a single chirp is the only factor for determining the final range resolution. The maximum bandwidth of the IWR 1443Boost radar is 4GHz, yielding a range resolution of 3.75 cm. Despite that, Velocity resolution(V_{res}) determines the minimum frequency difference between two discrete frequencies. However, the radar has limited control over the final wavelength, but its high frequency allows for a firm velocity resolution compared to lower-frequency radar. So, the radar frame time can get controlled.

4 SYSTEM DESIGN

The system's design approach starts by connecting the two mmWave radars via the serial ports to loads, sending the configuration files, receiving detection results, and filling them into data matrices. The mmWave radar transmits an RF signal and records its reflections off objects. Generating point clouds by computing sparse points and removing those corresponding to static objects, clustering the detected user by merging individual points into clusters, and removing noise. Afterward, voxelization is the representation of the point clouds. Lastly, recognize user activity from each user's sequential data, track the user in consecutive frames, and pass the collected data via the classifier to determine recognization accuracy. This section will endure the system design in further detail, following the approach to accomplishing the article's objectives.

4.1 Software Framework

This article presents an implemented software framework for managing the two mmWave radars and performing post-processing data. The system is written in python and includes the following modules.

- **Radar Handler:** Connects to the radar via the serial ports, loads, sends the configuration files, receives detection results, and fills them into data matrices.
- **Frame Processor:** Takes data matrices as input, performs customized data processing tasks, and outputs data matrices with the same format.
- **Visualizer:** Converting the continuous geometric information domain data structures into a rasterized image in 3D space. The visualizer manages several frame processors for a data processing chain and displays the final output in 2-D or 3-D formats.
- **Classifiers:** evaluating the mmWave radar data set through the machine learning implemented classifiers.

4.1.1 ***Radar Handler.*** The configuration file gives access to the user to specify the number of radars, the model of each radar, the antenna configuration, and the serial port number; the radar's two serial ports are accessible by the PC, one for configuring the radar and the other for transmitting the results. The threats get generated at startup, and each radar in use will have an independent thread generated. These threads handle the radar module by connecting to the serial ports to handle the communication between the host and the radar. The radar threads for reading the data are from the serial port constantly and parse them into a suitable structure, despite only pushing the result into the shared queue if the queue is empty. Thus, the framework supports multi-threaded environments.

4.1.2 ***Frame Processor.*** One visualization thread will produce a visualizer module and several frame processes to achieve customized post-processing on the received data. Several data queues will be created for each radar thread and shared with the visualization thread. The visualization thread fetched the data from individual queues. It fulfilled the user-defined post-processing tasks individually, depicting the combined results and fetching the next data cluster. The designed system operates so that the radar threads only push data once the visualization thread has finished the last frame. Hence, to avoid out-of-synchronization caused by different processing speeds of threads. Lastly, decoding the data from the data port starts when the radar searches for the data packet header, filters out unused packets, and extracts the detected object in the frame.

4.1.3 ***Visualizer.*** The visualizer's role is to load the data matrices from the entire radar threads, apply the user-defined frame processors, combine them into a single frame, and display the final output. The display can be configured as 2-D, 3-D, or both and supplies a suitable method to interpret the result. The frame processors define the operations to get performed on each radar frame. This article uses the clustering frame processors module. This module groups the data points in one frame into clusters according to their distance and filters out small clusters with a few points. The article uses the DBScan (density-based spatial clustering of applications with noise) algorithm [28]. This algorithm does not require prior knowledge of the scene and can extract all the qualified clusters. Lastly, this model significantly helps in reducing noise.

4.2 Voxelization

Voxelization converts data structures that store geometric information in a continuous domain into a rasterized image (discrete grid). Furthermore, a voxel represents a value on a three-dimensional grid. The points in the non-voxelized data in the three-dimensional grid consist of their x,y, and z coordinates. Although, the data can get represented as cubic elements with the ability to contain one or multiple points after voxelizing the data. Therefore, Voxelization converts the three-dimensional data into a voxelized grid for the processed data. Despite that, the Voxelization of the gathered data was according to the script of implemented repositories by Singh et al. [6]. The voxel size is a crucial aspect of Voxelization; the size of each voxel size inside the physical dimensions of what got registered inside a voxel is the size of each voxel.

4.2.1 ***Voxelization Approach.*** The Voxelization is responsible for visualizing the loaded data matrices from all the radar threads. It displays the final output by applying user-defined frame processors to combine them into a single frame. The implemented grid in this paper contains voxels, and each voxel size in the grid is (10 * 32 * 32). The data gets divided into 60 frames for creating workable samples, which implies a shape of (60 * 10 * 32 * 32) per instance among the voxelized representations leading to the classifiers. Moreover, the radar can have a consistent view of the scene from different location radars, and this is by combining the data and applying the appropriate rotation and translation to the coordinated from different radars. Hence, the display can be configured to 2D or 3D, providing a convenient way to interrupt the result. Figure 1 shows an example of the voxelized representation of the point cloud.

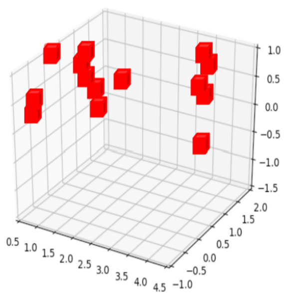

Fig. 1. Voxelized representation of point cloud

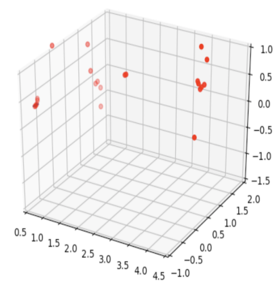

Fig. 2. Data point cloud

4.3 Merging Data

The radars detect human activities and store their point cloud in a data set. Despite that, using two radars for detecting human activities can function as two separate radars as every radar generates its point cloud data set. The aforementioned implies that these data sets can be previewed as single or combined two mmWave radars data sets into a combined point cloud data set since they detect the same scene. Therefore, this paper proposes an algorithm for merging any two-point cloud data sets generated from the MmWave radars for detecting the same scene and combined into one point cloud data set. Although, the algorithm requires the data sets to get generated from the same scene.

4.3.1 ***Preliminaries of merging data algorithm***. The point cloud data set contains two significant elements crucial for the program: frames and point IDs. So, the frame represents the time frame in the data set, the data set consists of different frames, and each frame represents the period of detecting human activity. For example, the radar starts detecting human activity, and the first 60 seconds get stored as frame one, and so on. However, each frame consists of coordinated points with their point ID. The program procedure starts by taking each frame from the two data sets and comparing the point IDs. Further, the program merges them depending on their IDs.

Some cases in their frame contain more Point IDs than the other frame. In this case, the extra point IDs get discarded to make it more reliable. Because if one radar frame contains more point IDs than the other frame IDs, it becomes difficult to detect the correct data set from which radar. Therefore, neglecting these IDs make it more reliable. Nevertheless, if a radar contains more frames than the other radar but in the correct sequence, the program adds this frame to the combined point cloud data set.

4.3.2 ***Program Approach***. The process starts by detecting the human activity by the two mmWave radars and outputs a point cloud. Nearby, containing two point clouds, need to merge to output a merged point cloud representing the detection of the two radars. The program procedure takes the frame of the two-point clouds, compares the two files by their point IDs, and merges them depending on their IDs. Afterward, the data file passes through the classifiers and Voxelization. Figure 2 shows the data point cloud.

5 EXPERIMENTAL SETUP

5.1 One MmWave Radar Experimental Setup

The experimental setup of one mmWave radar was configured to use all three transmitter antennas and all four receiver antennas to generate 3D point cloud data. The start frequency is 77GHz, and the bandwidth is 4GHz. The sensor was programmed to send 128 chirps every frame, and the number of frames per second was 33. The chirp cycle time is 162.14us, and the frequency slope is 70GHz/ms. So, these configurations provide a range resolution of 4.4cm and a maximum unambiguous range of 5m. In terms of velocity, it can measure a maximum radial velocity of 2m/s, with a resolution of 0.26m/s.

5.2 Two MmWave Radars Experimental Setup

There are two experimental scenes for recognizing human activities using two mmWave radars. The first scene is setting up the radars for parallel beaming, and the effective horizontal detection angle of the devices is (60). Furthermore, configuring the two mmWave radars to different frequencies prevents signal interference; also, it is possible to set up the chirps, so that frequency bands are not shared [20]. The antenna size is inversely proportional to frequency. So, the higher the frequency spectrum, the smaller the antenna size. A higher frequency for a given antenna size allows the beam to be more closely focused [27]. The directional beams decline the signal impression of recognizing each human, allowing higher spatial reuse where multiple human activities can get detected in parallel within the scene[28].

The second scene is for the orthogonal beaming of the two mmWave radars, and the effective horizontal detection angle of the devices is (45). Moreover, the transmitted signals must be mutually orthogonal to avoid interfering with the other mmWave radar signal [28]. This dual orthogonal configuration secures more returns from static ground objects [18]. Therefore, beaming orthogonal improves the angular resolution of the radar system [27]. The wider the beam becomes, the more its power will be spread across the width.

The radar range is the distance determined along the sight line from the radar location to the target, and the configurations have a peak range of (8 m). The radar velocity is the component of the target's motion along the radar beam's direction, and the radar velocity peak is (1 m/s). The radar resolution is the ability of the radar system to distinguish between two or more targets on the same bearing but at different ranges, and the configured resolution range is (4 cm) [29]. The velocity resolution is the slightest velocity difference that can measure between two moving objects using a given spectrum, and it gets configured to be (0.1 m/s). The chirp per time is (125 us). The idle time for resetting the chirp is ten us and (115 us) chirp ramp time, (35 MHz/us) for the slope rate, and (4GHz) bandwidth utilization for the radar.

Table 1. Two mmWave Radars Configuration

Parameter	Value
Number of range samples	240
Number of chirps	16
Frequency	79.210 GHz
Bandwidth	2.55 GHz
PRI	64.2 us
Frame time	33.33 m/s
Max range	10 meter
Range resolution	0.045 meter
Max Doppler	+- 4.214 m/s
Doppler resolution	0.615 m/s

Table 2. One mmWave Radar Configuration

Parameter	Value
Number of range samples	120
Number of chirps	8
Frequency	77 GHz
Bandwidth	4 GHz
PRI	64 us
Frame time	0.162 m/s
Max range	5 meter
Range resolution	0.044 meter
Max Doppler	+- 4.214 m/s
Doppler resolution	0.26 m/s

5.3 Data Collection

Singh et al. materialized research indicating that to ensure human activities got performed at a constant tempo, each activity must perform for five trials in a row [6]. Three different activities got performed in front of the two mmWave radars in the two scenes. The activities were walking, jogging, and jumping. Moreover, the three different activities got performed in front of one mmWave radar with the same tempo and trails as for the two mmWave radar scenes. The table below shows the amount of data collected, and each activity trial's measured activities took 30 seconds to perform.

Data collection for scene 1 for the three performed activities showed that when the user is close to the mmWave radar, the radar cannot scan the entire body of the user because the sufficient vertical monitoring angle of the device is less than (20). Hence, the setup for scene 1 is that the two mmWave radars are 1 meter away from the user, the height of the mmWave radars is 1 meter, and the angle between the radars is (0). In scene 2, Two mmWave radars tripped at the height of 1 meter, with an angle between the radars of (45). For the scene of one mmWave radar, the height of the radar is 0.90 meters and 1 meter away from the user to illustrate a wide beaming angle for user activity recognition.

Table 3. Human Activity Dataset:

Activity	Records	Duration (seconds)
Walking	12	480
Jacking	12	480
Jumping	12	480

5.4 Data Training

The data training is the procedure after collecting the data from the two MmWave radars and getting performed on the mentioned machine learning classifiers. These classifiers contain different algorithms resulting in different accuracy results. Therefore, the accuracies of these classifiers will result in different previewing of the best classifier to achieve the research goal and objectives. The classifier parameters are a crucial aspect of achieving the best performance. So, for all the classifiers, each input data first frame flattened to a vector dimension of 16000. The classifiers used the Adam optimizer with a dropout ratio of 0.5. Despite that, the training-to-test sample ratio is 11:1, implying a balanced data set to decrease over-fitting. The model got trained for 30 epochs.

6 CLASSIFIERS

The different classifiers evaluated the mmWave radar data set. The authors of [4] presented and trained the following classifiers. First, there is the Support Vector Machine(SVM), Multi-layer perceptron (MLP), Long Short term Memory (LSTM), and Convolution neural network (CNN) combined with LSTM [4]. Therefore, this paper's accuracies compared to their accuracies gives fascinating insights.

6.0.1 ***SVM Classifier.*** First, the Support Vector Machine (SVM) classifier gets generated by flattening the time window voxelized representation by the provided frames (frames * 10 *32 *32)[4]. Second, the principal component analysis (PCA) reduces the training data's dimensions. The principal component analysis is a popular technique for analyzing large data sets with a significant number of dimensions/features per observation, enhancing data's interoperability while preserving the maximum amount of information and enabling the visualization of multidimensional data[12]. However, the number of components the PCA uses on the data set is 200 because PCA requires much computational performance from a computer. So, to train the SVM classifier, the PCA must be applied first.

6.0.2 ***MLP Classifier.*** First, the multilayer perceptron (MLP) comprises three fully connected layers and an output layer[4]. Second, MLP is a class of artificial neural networks (ANN). ANN model is an inspiration for the human brain. Last, the time window voxel representation can get sized by specifying the number of frames to create an input of the wanted dimensions.

6.0.3 ***Bi-directional LSTM Classifier.*** The bi-directional LSTM layer preserves information from the future and the past [4] and consists of two LSTM layers operating in parallel[4]. One layer runs from the past to the future, and the other layer from the future to the past. Hence, the two layers run as two ways input layers.

6.0.4 ***Time-distributed CNN + Bi-directional LSTM Classifier***. The Time-distributed CNN + Bi-directional LSTM Classifier consists of five layers. The first three layers are the convolution layer, convolution layer, and max-pooling layer. Moreover, the fourth layer is the bi-directional LSTM layer, and the fifth layer is the output layer. Furthermore, every secular data part passes through the CNN layers [4].

7 TRACKING

The tracking system exploits the properties of the mmWave radar. The operation starts by transmitting an RF signal and recording its reflections off the objects. Then, analyzing the point cloud generated infers people's trajectory and identifies them from the database. The tracking system consists of three stages operating concurrently.

- **Point Cloud Generation:** The FMCW radar transmits millimeter waves and records the reflections from the scene. Then, it computes the sparse point cloud and removes those points corresponding to static objects. For example, the points that appeared in the previous frame.

- **Point Cloud Clustering:** Detecting potential human objects by merging individual points into clusters.

- **Tracking:** Tracking and identifying the exact human object in subsequent frames and using numerous object tracking algorithms.

7.1 Point Cloud Clustering

The mmWave-generated sparse point clouds need to be more informative for detecting distinct objects because the spare point clouds are scattered. Even though static objects get discarded via clutter removal, moving humans only inevitably all reflect the remaining points. However, the noise can be significant and leads to confusion with the points from a nearby human user.

Density-aware clustering method separates cloud points based on the Euclidean distance in the three-dimensional space. To determine which points in the scene are caused by human reflection, first merging points into clusters using DBScan algorithms. Moreover, the DBScan algorithm can automatically mark outliers to cope with noise without requiring the number of clusters to be specified.

The points clustering gets based on the Euclidean distance. In the study, the points of the human user are coherent in the horizontal (X-Y) plane [13]. The human user points are more scattered and difficult to merge along the (Z) axis. Therefore, modifying the Euclidean distance algorithm to place less weight on the contribution from the vertical z-axis in clustering [13]. According to Peijun Zhao, the modified Euclidean distance formula will be

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

[14]. Where p^i and p^j are two different points, the parameter α regulates the contribution of vertical distance.

Applying the Euclidean algorithm formula regarding this article experiment, the points in the range of 10 cm will get classified into one cluster. Despite that, The points with a distance of more than 10 cm will be treated as noise and discarded. The Kalman filter algorithm detects and removes the noise after applying the DBScan algorithm. Figure 3 shows how the DBScan algorithms get applied to the data.

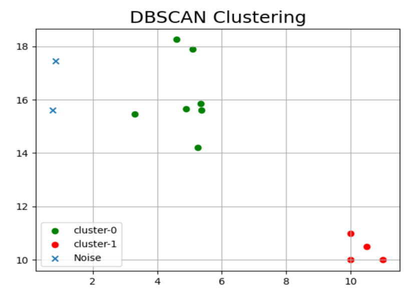

Fig. 3. Data clusters

7.2 DBScan Algorithm Parameters

DBScan algorithm has a parameter for indicating the maximum distance of two points in the same cluster. The parameter for this indication is (Eps). Moreover, the DBScan algorithm has a second parameter, which indicates the minimum point number in a cluster to cope with noise points. This parameter for this indication is (MinPts). This article experiment demonstrated the Eps to 0.05 and 20 as MinPts. Nevertheless, the α got set to 0.25 in the customized distance function.

7.3 Moving Object Tracking

Tracking a person requires capturing continuous individual point clouds, requiring an influential temporal association of detections, correction, and prediction of sensor noise. The procedure starts by creating ad maintaining tracks for object detection from each frame. Next, a new track gets created for each object detection from the first incoming frame or one that cannot be associated with an existing track. Hungarian algorithm gets dedicated to the Inter-frame object association. Nonetheless, if a track object is undetected for D continuous frames, the track receives marked inactive and is excluded from successive associations. Kalman filter algorithm gets applied for predicting and correcting tracks. Figure 4 displays the workflow of moving object tracking.

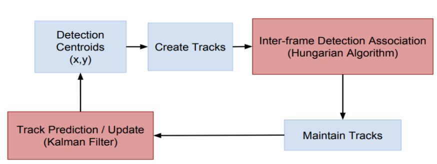

Fig. 4. Moving Object Tracking Workflow

7.3.1 ***Tracking Detection.*** The Hungarian algorithm is a practical combinatorial optimization algorithm. Minimizing the integrated distance loss is a purpose for constructing an association between each object detection and maintained track objects. However, the active tracks K_1 and the number of object detections at the current timestamp K_2 imply that the cost matrix can be non-square, directing considerable assignment problems [13]. The argument is that matrix M were (M_i, j) represents the distance of centers between track object i and object detection j in the current frame. So, giving that K as the greater of K_1 and K_2 implies the essential argument that the actual cost matrix with form entries constructs a KxK matrix M.

Additionally, for the case that (M_i, j) exceeded the step size threshold θ, then setting the cost to a large number L to avoid the association given the intuition that j should be joining person [13]. The case for ignoring such mappings and creating a new track for detection is in the case the detection is mapped to an augmented dimension or mapped to a correspondence with coast L [13]. Likewise, if a track object gets mapped to an increased dimension or a correspondence with cost L, then the tracked object is regaled as undetected to maintain tracks of detections successfully.

7.3.2 ***Tracking Correction.*** The Kalman filter intends to correct the sensor noise and offer predictive guidance in scenarios where tracked objects are undetected due to closure or temporary loss from the sensing region, starting by maintaining a state which consists of the location of the human user along the x and y axes. So, for each track, the initial state consists of the first detection location at each time frame. The Kalman filter updates the current state variables with a transition matrix by interconnected indecisiveness. The Kalman filter estimates the new position plus the covariance based on the current position. Thus, the Kalman filter algorithm calculates the cultivation more accurately than those established on a single sensor measurement.

8 RESULTS

This section presents the deep learning classifier accuracy results on the one and two mmWave radars for human activity recognition. Moreover, sensitivity and precision are necessary for evaluating the performance of the one and two-mmWave radars. To explore the best artificial neural network architecture for identification, a comparison of 4 different architectures got used. Each architecture model got trained under the same settings. The settings are 30 epochs and a dropout ratio. The table shows the accuracy of detecting human activities using mmWave radar based on these architectures and shows the sensitivity and precision of the mmWave radar.

Starting with the SVM classifier, as it performs poorly, achieving the lowest accuracy. There exist several reasons for this performance. The SVM classifier does not operate directly on the time window voxel data compared to the other classifiers that operate directly with the window voxel data. The SVM classifier needs to perform better with the data set that contains much noise (i.e., overlapping target classes). SVM is not suitable for classifying large data sets because the training complexity of SVM is highly dependent on the size of the data set. [15]

The MLP classifier is a multilayer perceptron and relies on an underlying neural network to perform the classification task. Its multiple layers and non-linear activation indicate MLP from a linear perceptron and can differentiate data that is not linearly separable. However, MLP includes too many parameters because it is fully connected. Each node gets connected to another in a very dense web. Hence, resulting in redundancy and inefficiency. Therefore, the MLP classifier achieves a reasonable accuracy but needs to be sufficiently high compared to the other classifiers. Regardless, MLP uses back-propagation for training the network.

The Bi-directional LSTM classifier uses LSTM variants as it showed to be performing for end-to-end learning of time sequence data. The used LSTM layer has the size of 256 and 128 hidden units. Bi-directional LSTM converges faster and significantly outperforms the other two architectures within fewer guesses [16]. Bi-directional LSTM can model the rich temporal correlations in a long sequence of frames from both ends. In contrast, a standard LSTM is essentially a feed-forward network that is difficult to encode the information from the beginning of a long sequence [16]. Such information loss degrades the identity inference performance.

The Time-distributed CNN + Bi-directional LSTM classifier performs better with higher accuracy for the following reasons. First, the used CNN classifier includes two convolution layers, with a max pooling layer after each convolution layer. The CNN is time distributed, implying the data of each frame is first sent into a two-layer 3D CNN for feature extraction, then the sequence data is sent into LSTM for classification [16]. Thus, this classifier's accuracy is the highest as it combines the functionalities of the LSTM classifier besides the CNN.

False detection is when the radar detects noise or other objects and is falsely detected as human. The radar's sensitivity is the ability to detect a human in the detection area reliably, and the radar precision is the ability to differentiate humans from false detection. Hence, an ideal system should contain high sensitivity and high precision. However, the experiment setup, human activities, and activities period were the same; for detection, using one mmWave radar and two mmWave radars. However, with one radar, the 46.9% precision indicates that more than half of the detections would be false detections. Two radars reduced the system sensitivity slightly, but the precision improved significantly to 98.6%. So this implies a greater than 50% probability that it is a false detection using one mmWave radar but highly precise detection with two radars.

Performance Evaluation of the Radars:

	Sensitivity	Precision
One mmWave radar	96.4%	46.9%
Two mmWave radars	90.4%	98.6%

Two mmWave radars accuracy test of different activity recognition classifiers trained on the human activity Dataset:

Classifier	Accuracy
SVM	65.32%
MLP	81.37 %
Bi-directional LSTM	91.17 %
Time-distributed CNN+ Bi-directional LSTM	97.71%

One mmWave radar accuracy test of different activity recognition classifiers trained on the human activity Dataset:

Classifier	Accuracy
SVM	63.74%
MLP	80.34 %
Bi-directional LSTM	88.42 %
Time-distributed CNN+ Bi-directional LSTM	90.47%

9 EVALUATION

9.1 Radars Evaluation

To evaluate the human activity recognition accuracy of using two mmWave radars. This article compared the usage of one mmWave radar to two mmWave radars. The collected data from the mmWave radar was for the same scene as the two mmWave radars and the same human activities. Nonetheless, the collected data went through the same framework and machine learning classifiers. Hence, comparing the results of using one mmWave radar and two mmWave radars, it evaluated that using two mmWave radars for detecting human activities achieves an improved accuracy than using one mmWave radar.

9.1.1 ***Performance Factors.*** False positive detection is a factor in the accuracy difference. The false positive detections can be noise or other objects in the detection area that got falsely detected as a human activity. However, the one mmWave radar indicates that more than the two mmWave radars detection would be false detections. Otherwise stated, the one mmWave radar detection is more likely to contain false detection than the two mmWave radars [3].

The one radar detection system reports many false alarms due to noise and flicking of the results [3]. The flicking got discovered due to the fast Fourier transform (FFT) process. The process begins when a slight shift in the signal can change the FFT bins once it arrives via the FFT. The angle FFT is when the FFT across the corresponding peaks in the series of antennas, and peaks in this angle FFT directly correspond to the angle of arrival of objects. So, a few meters of displacement on the object coordinates [3]. Further, this effect will expand, where a removal in the angle will result in a much larger displacement in the 3D space [3]. Conversely, utilizing two radars provides a system that permits two independent detections and can verify the results jointly.

9.2 Tracking Evaluation

A testing environment got created to observe the reliability of the tracking and to evaluate the tracking range and tracking error of mmWave radar. The testing environment got set to track the user from a marked place to a specific destination. Even though time gets synchronized through an NTP server, the system's coordinate transformation matrices got measured. The radars performance table showed that two mmWave radars construct a positive determination when both radars detect the person, reducing the impact of noise and increasing precision. So, evaluating the tracking error by comparing the trajectories between the two and one mmWave radars and measuring the trajectories by calculating the distance between all the clusters in the current frame and the previous frame using the Euclidean distance. The comparison of trajectories showed that the median tracking error of using one mmWave radar is 0.16 m and for two mmWave radars is 0.10 m. Besides, the tracking range of two mmWave radars is more significant as it can track at a distance of more than 5.5 m, while one mmWave radar can track at a distance of 4.5 m. The testing results show that the two mmWave radars can maintain reliable, accurate tracking on a large tracking area.

9.2.1 ***Non-line-of-sight Conditions.*** The first experiment analyzes the robustness of mmWave radar under obstructed conditions because optical imaging-based tracking and identification methods, such as Depth cameras, cannot cope with obstructions [13]. Hence, an evaluation of the robustness of millimeter wave radar with some obstructions, such as aluminum foil sheet. The aluminum foil sheet thickness was approximately 4mm with a size of 10$\hat{5}$ mm$\hat{2}$, and placed this obstacle 1 cm away from the sensor so that the signals could not get transmitted in a line of sight condition. Then, let the user walk back and forth ahead of the mmWave radar while collecting sensor readings. Afterward, the generated 3D point cloud from the mmWave radar was used for tracking and identification to compare the percentage of change in point cloud density.

The impact of the aluminum foil sheet on the point cloud density of the mmWave radar was approximately 0.80%. Hence, the difference in point cloud density is below 1%. Therefore, that implies that the mmWave radar is robust against non-line-of-sight interference with less than a 1% change in point cloud density[13]. Thus, the robustness of the mmWave radar to thin obstructions can detect human activities and perform tracking and identification, for example, under furniture [13].

9.2.2 ***Impact of weighting the vertical axis in DBScan.*** An adequate clustering method produces a lower within-cluster variation, which indicates good clustering [18]. The evaluation of the compactness of clusters got based on distance measures. Such as the cluster-wise within average/median distance between observations [18]. Therefore, the cluster evaluation got based on similarity or dissimilarity, such as the distance between cluster points. So, the clustering algorithm performs well when separating dissimilar observations apart and similar observations [19].

DBScan algorithm requires defining the weighting parameter for improving the DBScan algorithm. In rehearsal, determining that (alpha = 0.25) results in a helpful clustering performance, as shown in figure 5. In contrast, points are effectively projected onto the x-y plane when (alpha =0), and the outliers merge into the cluster. Moreover, points corresponding to a person are split into two clusters when (alpha = 1) (standard Euclidean distance). Thus, finding a suitable value of (alpha) for obtaining a good clustering result [13]. Clustering results with a different (alpha), as shown in figure 5, indicates that a small (alpha) leads to loose clusters containing many noise points. A significant (alpha) splits a human object into two clusters. Therefore, setting (alpha =0.25) results in the best empirical performance.

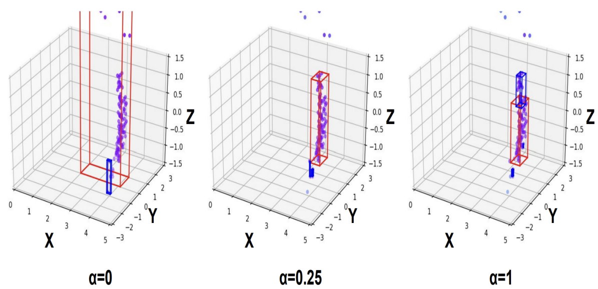

Fig. 5. Clustering results with different α

10 LIMITATION AND FUTURE SEARCH

This paper procedure performs pleasingly for tracking and identification. The paper elaborates on the approach limitations and enumerates future research directions.

10.1 Considerable Number Of Users

The paper exemplified the reliable performance of tracking and identification on one user. However, tracking and identifying over one user remains a genuine concern because the mmWave radar point cloud is sparse and periodically disturbs human detection and tracking significantly.

10.2 Monitoring Range

The range of the mmWave radar can be as extensive as 30 m, but the increased the range gets, the reduced spatial precision and worse the signal-to-noise ratio gets. If the user is far from the sensor, detecting and distinguishing the user from background noise is challenging.

10.3 Classifiers Dependencies

The classifiers' accuracy tables show that the MLP classifier performs poorly. Perhaps, the reason is that fully connected layers in the MLP classifier make no spatial and temporal assumption about the data. Conversely, the Time-distributed CNN + Bi-directional LSTM classifier assumes spatial and temporal dependency in the data and performs better.

11 CONCLUSION

The paper presented a two mmWave radar framework using a time window voxel representation of sparse mmWave radar point clouds. Furthermore, presented setups for avoiding signal interference and the influence of noise in tracking showed the advancements of using two radars for recognizing and tracking human activities to approach the research objective. The machine learning classifiers can learn the feature extraction transformation by directly training on the voxels and are designed to handle data's spatial and temporal dependencies. The evaluation of the deep learning classifier achieved an accuracy of 97.71%. Lastly, the article achieved the mentioned objectives and answered the research questions.

REFERENCES

(1) - Vrigkas, M., Nikou, C. and Kakadiaris, I. A. Vrigkas, M., Nikou, C., & Kakadiaris, I. (2015). A Review of Human Activity Recognition Methods. Frontiers In Robotics And AI, 2. doi: 0.3389/frobt.2015.00028

(2) - 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 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems (mmNets'19), October 25, 2019, Los Cabos, Mexico. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3349624.3356768

(3) - H. Cui and N. Dahnoun, "High Precision Human Detection and Tracking Using Millimeter-Wave Radars," in IEEE Aerospace and Electronic Systems Magazine, vol. 36, no. 1, pp. 22-32, 1 Jan. 2021, doi: 10.1109/MAES.2020.3021322.

(4) - Y. Qi, C. B. Soh, E. Gunawan, K.-S. Low, and R. Thomas, "Assessment of foot trajectory for human gait phase detection using wireless ultrasonic sensor network," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 24, no. 1, pp. 88–97, Jan. 2016.

(5) - Singh, A. D., Sandha, S. S., Garcia, L., Srivastava, M.(2019). RadHAR. Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems - MmNets'19, 51–56.

(6) - E. Kantoch, "Technical verification of applying wearable physiological sensors in ubiquitous health monitoring," in Proc. IEEE Comput. Cardiology, 2013, pp. 269–272.

(7) - Ulrich, Michael & Yang, Bin. (2018). Short-Duration Doppler Spectrogram for Person Recognition with a Handheld Radar. 1227-1231. 10.23919EUSIPCO.2018.8553114.

(8) - R. Zhang and S. Cao, "Real-Time Human Motion Behavior Detection via CNN Using mmWave Radar," in IEEE Sensors Letters, vol. 3, no. 2, pp. 1-4, Feb. 2019, Art no. 3500104, doi: 10.1109/LSENS.2018.2889060.

(9) - How Far Will Your Wi-Fi Signal Reach? How Far Will Your Wi-Fi Signal Reach?. (2022). Retrieved 7 December 2022, from https://epb.com/get-connected/gig-internet/how-far-will-your-wi-fi-signal-reach/

(10) - Fukunaga, K. (2013). Terahertz applications in art conservation. Handbook Of Terahertz Technology For Imaging, Sensing And Communications, 615-623. doi: 10.1533/9780857096494.3.615

(11) - Poppe, Ronald. (2010). Poppe, R.: A Survey on Vision-based Human Action Recognition. Image and Vision Computing 28(6), 976-990. Image Vision Comput.. 28. 976-990. 10.1016/j.imavis.2009.11.014.

(12) - A Step-by-Step Explanation of Principal Component Analysis (PCA) A Step-by-Step Explanation of Principal Component Analysis (PCA). (2022). Retrieved 7 December 2022, from https://builtin.com/data-science/step-step-explanation-principal-component-analysis

(13) - P. Zhao et al., "mID: Tracking and Identifying People with Millimeter Wave Radar," 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2019, pp. 33-40, doi: 10.1109/DCOSS.2019.00028.

(14) - Zou, H., Zhou, Y., Yang, J., Gu, W., Xie, L., Spanos, C. J. (2018, April). Wifi-based human identification via convex tensor shapelet learning. In Thirty-Second AAAI Conference on Artificial Intelligence.

(15) - Cervantes, Jair & Li, Xiaoou & Yu, Wen & Li, Kang. (2008). Support vector machine classification for large data sets via minimum enclosing ball clustering. Neurocomputing. 71. 611-619. 10.1016/j.neucom.2007.07.028.

(16) - Chris Xiaoxuan Lu, Bowen Du, Hongkai Wen, Sen Wang, Andrew Markham, Ivan Martinovic, Yiran Shen, and Niki Trigoni.2017. Snoopy: Sniffing Your Smartwatch Passwords via Deep Sequence Learning. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1, 4, Article 152 (December 2017), 27 pages. https://doi.org/10.1145/3161196

(17) - Zhou, Yong & Dong, Yanyan & Hou, Fujin & Wu, Jianqing. (2022). Review on Millimeter-Wave Radar and Camera Fusion Technology. Sustainability. 14. 5114. 10.3390/su14095114.

(18) - 3D ego-Motion Estimation Using low-Cost mmWave Radars via Radar Velocity Factor for Pose-Graph SLAM 3D ego-Motion Estimation Using low-Cost mmWave Radars via Radar Velocity Factor for Pose-Graph SLAM. (2022). Retrieved 26 November 2022

(19) - Nemade, V., Pathak, S. Dubey, A.K. A Systematic Literature Review of Breast Cancer Diagnosis Using Machine Intelligence Techniques. Arch Computat Methods Eng 29, 4401–4430 (2022). https://doi.org/10.1007/s11831-022-09738-3

(20) - Clustering Validation Statistics: 4 Vital Things Everyone Should Know - Unsupervised Machine Learning - Easy Guides - Wiki - STHDA (2022) from http://www.sthda.com/english/wiki/wiki.php?id_contents=7952

(21) - OMRON Global. 3D Imaging of Outdoor Human with Millimeter Wave Radar using Extended Array Processing Retrieved 23 November 2022. from https://www.omron.com/global/en/technology/omrontechnics/vol54/005.html

(22) - C. Chen, R. Jafari and N. Kehtarnavaz, "UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 168-172, doi: 10.1109/ICIP.2015.7350781.

(23) - W. Wang, A. X. Liu, M. Shahzad, K. Ling and S. Lu, "Device-Free Human Activity Recognition Using Commercial WiFi Devices," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 5, pp. 1118-1131, May 2017, doi: 10.1109/JSAC.2017.2679658.

(24) - O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," in IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1192-1209, Third Quarter 2013, doi: 10.1109/SURV.2012.110112.00192.

(25) - Zhou, Yong & Dong, Yanyan & Hou, Fujin Wu, Jianqing. (2022). Review on Millimeter-Wave Radar and Camera Fusion Technology. Sustainability. 14. 5114. 10.3390/su14095114.

(26) - Golovachev, Y.; Etinger, A.; Pinhasi, G.A.; Pinhasi, Y. Millimeter Wave High Resolution Radar Accuracy in Fog Conditions—Theory and Experimental Verification. Sensors 2018, 18, 2148. https://doi.org/10.3390/s18072148

(27) - Researching Feasibility of mmWave solutions for Insect Detection Applications. - University of Twente Student Theses Researching Feasibility of mmWave solutions for Insect Detection Applications. - University of Twente Student Theses. (2022). Retrieved 7 December 2022, from http://essay.utwente.nl/89084/

(28) - W. -T. Wang, Y. -L. Wu, C. -Y. Tang and M. -K. Hor, "Adaptive density-based spatial clustering of applications with noise (DBSCAN) according to data," 2015 International Conference on Machine Learning and Cybernetics (ICMLC), 2015, pp. 445-451, doi: 10.1109/ICMLC.2015.7340962.

(29) - Range Resolution-Radartutorial Range Resolution-Radartutorial. (2022). Retrieved 7 December 2022, from https://www.radartutorial.eu/01.basics/Range%20Resolution.en.html