IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 19, 2021, accepted June 3, 2021, date of publication June 11, 2021, date of current version June 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3088452

On the Generalization and Reliability of Single Radar-Based Human Activity Recognition

ALI GORJI[1], HABIB-UR-REHMAN KHALID[1,2], ANDRÉ BOURDOUX[1], (Senior Member, IEEE), AND HICHEM SAHLI[1,2]

[1]Interuniversity Microelectronics Centre (IMEC), 3001 Leuven, Belgium
[2]Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), 1050 Brussels, Belgium

Corresponding author: Ali Gorji (gorjid81@imec.be)

This work was supported in part by Niko, and in part by the Flemish Agency for Innovation and Entrepreneurship.

ABSTRACT Identifying human activities using short-range and low-power radars has attracted much attention among the researchers and consumer electronics industry. This paper considers human activity recognition in the context of a single Frequency Modulated Continuous Wave (FMCW) radar as the measurement tool. A classification pipeline is proposed to handle the data pre-processing and feature extraction and a machine-learning based solution is devised to undertake the activity classification. The performance of the proposed architecture is evaluated under both unseen subjects and new room layouts. We show how the accuracy of the activity classification will be affected by situations such as poor aspect-angle and occlusions created by furniture that normally arise in realistic scenarios where an unseen layout is considered. A two-stage classifier will be then proposed to enhance the generalization of the model, especially, to unseen rooms. Besides, an extensive feature exploration will be conducted and the importance of features in the generalization will be studied. The results in this paper will conclude a machine learning pipeline that will generalize well to unseen subjects and new room layouts, which are two main difficulties that arise in most radar-based activity classification tasks.

INDEX TERMS Human activity recognition, multi-class classification, radar signal processing, robust activity recognition.

I. INTRODUCTION

Recognizing human activities in an indoor environment has recently received much attention among researchers [1]. Given the current revolution in the design and realization of smart-home solutions, the necessity for the indoor human activity recognition has raised, which brings with itself a need for studying the capability of the activity recognition under unseen human subjects and new room layouts. As a popular choice, radar sensing has been also considered as a major tool to tackle the indoor activity recognition [3], [7] due to its resilience to the harsh environmental conditions such as dim light and compatibility with the privacy concerns that may arise in many camera-based activity recognition systems.

Radar-based human activity recognition has been formulated as a multi-class classification problem and, therefore, classical machine learning techniques have been widely employed in the literature. The Support Vector Machine (SVM) has been used in [11] to classify and distinguish human activities from indoor targets using a through-the-wall radar system. SVM has been also applied in [22] to detect human activities by an Ultra Wide-Band (UWB) radar. Authors have extended their solution to a general activity classification by the UWB radar and have presented results in [23]. Authors in [20] have formulated the walking human body detection as a multi-class classification problem and have used a decision tree to undertake the classification task. A multi-radar solution has been also proposed in [21] to detect fine-grained human activities with a machine learning pipeline being formed that uses the random forest classifier as the backbone predictor. In [24], a novel class of Dynamic Range-Doppler Trajectory (DRDT) features has been devised for the purpose of continuous human motion recognition in living environments. Combined with a multi-stage machine learning model, authors in [24] have justified the superior performance of the new features in recognizing continuous

The associate editor coordinating the review of this manuscript and approving it for publication was Cheng Hu.

motions in various environments. Beside the supervised machine learning model, unsupervised feature-extraction by the Principle Component Analysis (PCA) has been also applied to the radar data in [13] for feature extraction and human activity classification. The feature-extraction module in [25] has been also employed to detect simple walking and standing activities of patients with a 4GHz radar and using unsupervised machine learning techniques.

Recently and with the emergence of deep Artificial Neural Networks (ANN), the application of deep learning in radar-based human activity recognition has attracted much attention given the capability of the ANN models in automatic feature extraction from the received radar echoes. A survey of recent ANN-based techniques for radar-based activity classification has been sketched in [7], [27]. It has been shown that the activity recognition as a multi-class classification problem can be well addressed by a variety of ANN models any of which can extract useful features from the radar raw echoes. For example, authors in [14] have used the radar point-cloud detection as the input to an ANN called point-net where the ANN model can tackle the multi-class classification task. A low-power solution is proposed in [6] to detect human activities in the kitchen using two mounted radars. There authors also show how the data from the radars are fused by a deep ANN model and the final performance is evaluated using the signatures obtained from the fusion layer. A Bidirectional Long Short-term Memory (Bi-LSTM) ANN model has been proposed in [38] for the continuous detection of human activities using an FMCW radar. Authors in [38] have used micro-Doppler images as sequence of inputs to a Bi-LSTM model that learns how to assign each radar frame to a pre-defined set of 6 activities. A joint segmentation and recognition model has been also devised in [39] for detecting both the transition between human actions and the type of executed activity. While statistical measures have been used in [39] to find the transition times, a Convolutional Neural Network (CNN) model has been proposed for detecting common human motions such as walking, sitting and running. In [9], activity classification has been addressed for indoor environments using both radar and camera as the sensing tools with a series of deep ANNs handling the fusion of data from different modalities. Real-time evaluation of HAR methods for indoor scenarios has been also performed in [10] where a Dynamical Neural Network (DNN) is proposed to combine the raw-features from both the videos and radar. In [4], activity classification has been tackled from a different perspective. Authors have used the video captions as teaching signals to train an ANN model that takes radar echoes as the input and produces captions that describe human interactions with the physical objects in a kitchen environment.

Although machine-learning based solutions have been widely developed for radar-based activity classification, the research has been recently steered towards studying the generalization of the classifiers to the environmental changes and unseen room layouts. In [28], authors have studied how walking at different aspect angles affects the performance of deep learning models in detecting human motions. A new CNN architecture has then been proposed in [28] for motion recognition and novel metrics have been devised to evaluate the generalization of the model to the angle variation. Joint motion recognition and person identification for multiple human subjects has been addressed in [29] using a single radar and in an indoor environment. The authors have also conducted a cross-room generalization study and have verified the performance of a trained Dynamical CNN (DCNN) in identifying walking patterns of multiple people when the trained model in an empty room is tested in a different room with furniture. The robustness of human activity recognition has been also studied in [30] where the authors have examined the efficiency of different classifiers in recognizing a set of six human activities. Data portability has been the topic of [31] where the classification models (both classical and deep learning models) have been tested across different rooms with different layouts. Authors in [31] show the consistent performance of activity classification when the tests running over four different rooms with Googlenet outperforming all other architectures.

While the previous articles (e.g. [30], [31]) have dedicated research to study the generalization of activity classification over unseen rooms, some important research questions are still unanswered. First, although a new recording scenario might mean different reflections or multi-path effects, activities such as sitting/standing are usually affected by the way they are performed by the human subject, as well as any furniture affecting the radar's line-of-sight. Note that both above cases are very common in any realistic activity classification problem where a trained model may be used to predict human activities in a new and unseen layout that could provide any variation of aspect-angles and occlusions. Therefore, apart from a general study on the generalization of activity classification over unseen environments, a deep investigation over the role of room layout on the accuracy of the classifier will help understanding the limitations of single-radar activity classification in real-world applications.

The following contributions significantly distinguish this work from the current state of the art:

- A pipeline for signal-processing and activity classification using FMCW radar has been built, including an algorithm for trajectory tracking of human subjects with a move-stop-move pattern. A set of novel tracking features has been created for the purpose of activity classification. It has been shown how a feature extraction pipeline can be designed using the results from the tracker and post signal-processing data-cubes such as micro-Doppler and point-clouds
- A two-stage classification algorithm has been developed to predict human activities in a generic room scenario, as the first stage, and in the presence of furniture, for the second stage of classification evaluation. It has been also shown that the new classifier shows much more superior generalization capability when compared to a single-stage activity classification.

- A deep analysis of hand-crafted features has been conducted by studying the importance of each feature category in the generalization of machine learning models. We have also shown how a good feature selection can improve the cross-room model generalization. One of the main contributions of this paper is then to study which combination of features provides the best model generalization capability, especially, when executed activities are tested in rooms with different layouts.

The rest of this paper is organized as follows. The human activity recognition problem will be discussed in Section II. Section III will present the measurement campaigns for the generalization assessment. Signal processing and machine-learning pipelines will be given in Sections IV and V, respectively. Experimental results including cross-room analyses will be provided in Section VI while deep feature exploration will be explained in Section VII. We conclude the paper in Section VIII.

II. PROBLEM STATEMENT

The scenario considered in this paper consists of a single human subject who enters an indoor environment and executes a set of pre-defined activities such as sitting and standing, etc.[1] A single radar is wall-mounted at a height of about 1.5m and is used as the sensing device to collect the reflections from the moving subject for the purpose of activity recognition. Let us define an **action primitive** as a certain human posture that could be detected within a certain amount of time and using received echoes from a sensing tool. To deal with the action classification problem, choosing a proper sensing tool, defining a well-formed set of action primitives, running well-defined measurement campaigns for data collection, and selecting a good model for activity classification are four essential components. Here, the measurement setup is first discussed and the set of action primitives is presented afterwards. Signal processing and the machine learning pipeline will be explained in Section IV.

A. SENSING TOOL

Measurements are collected using a single radar mounted in a wall-position (1.5(m) distance from the floor). We use the TI 60(GHz) FMCW mmwave radar with the device specifications being summarized in Table 1. The MIMO radar with 3 transmitters and 4 receivers provides 8 virtual azimuth bins and 2 elevation bins. While the azimuth angles could support the 2D tracking, the elevation data can be also leveraged to form some informative features for some of the activity classification tasks.

B. ACTION PRIMITIVES

Our scenario consists of a single human subject executing a sequence of action primitives in a living-room indoor environment with, potentially, different layouts. Table 2 lists the action primitives along with their descriptions and possible variations. Both SIT and STU primitives are executed in rooms with different layout designs to study the generalization of the proposed classification methodologies. A variable room layout may usually cause two difficulties for the classification task:

TABLE 1. System parameters for the 60(GHz) FMCW radar.

Symbol	Description	WM
(N_T, N_R)	Number of physical antennas	$(3, 4)$
f_0	Carrier frequency	60GHz
B_{eff}	Effective bandwidth	2.3GHz
v_{max}	Max. unambiguous velocity	3.87m/s
v_{res}	Velocity resolution	3cm/s)
r_{max}	Max. unambiguous range	15m
r_{res}	Range resolution	6.5cm
ΔT	Coherent processing interval	80ms
N_r	Number of range bins	230
N_d	Number of Doppler bins	256

TABLE 2. List of action primitives and their variations for indoor activity classification.

Primitive	Acronym	Variations
SitDown	SIT	Sitting down at different side angles
StandUp	STU	Standing up at different side angles
Walk_towards	WTW	Walking towards the radar
Walk_away	WAW	Waking away from the radar
Unknown	UNK	Any other primitive

- Furniture could be placed at different locations that will create different aspect angles with respect to the radar line-of-sight when the subjects intend to sit or stand. In this way, when a subject executes SIT or STU activity, the radar-based classification may fail to label the activity correctly given the poor aspect angle.
- Even in the case of a good aspect angle, subjects may be occluded by the furniture such as tables. In this case, some parts of the human body are usually masked by the medium and the radar does not receive the echoes from those parts of the human body that are involved in SIT/STU activities.

Note that both the poor aspect angles and occlusions are very common in a realistic scenario and a single-radar activity classification engine will suffer from the lack of data richness. One of the major contributions of this paper is to study the impact of the layout in the performance, especially, by analyzing the impact of individual features on the improvement of the generalization. In the following we refer to these primitives as activities.

III. GENERALIZATION ASSESSMENT

The main purpose of this paper is to study the generalization of a proposed ML-based pipeline for the activity classification in a general living-room indoor environment. For a general activity classification problem, we define **person-generalization** as the capability of the model in accurately

[1] This article deals with a single-target based activity recognition. Cases with multiple targets require a dedicated research that will tackle complex scenarios such as crossing or passing-by.

classifying the actions for an unseen human subject. The term **layout-generalization** describes the capability of the model in making accurate classifications for unseen room layouts. The generalization assessment is conducted by executing data collection campaigns with different participating human subjects and in three different layouts. The description for each campaign is given in the following sub-sections.

A. LAYOUT-FREE ROOM (DB1)

We first begin with an average-sized living-room (shown in Fig. 1) with a single chair being located in the middle of the room for the subjects to execute SIT and STU activities. The radar is located in the bottom-left corner of the room looking towards the top-right corner, which enables the sensing tool to capture the whole room given the 120^o azimuth beamwidth. Each subject is requested to start from any of the four corners of the room and walk to the chair to include walking aspect angles in the recorded measurements. To analyze the impact of walking patterns in the performance, each subject is asked to walk at different paces (slow/normal/fast), each case being repeated for four times. To study the person-generalization, we ask 15 subjects to execute the scenarios where people with different genders, heights and walking patterns are selected to enrich the dataset. As the room does not contain any furniture and subjects always keep a good distance from the wall, this campaign creates a layout-free recording dataset with partial aspect angles being added while no occlusion is expected.

B. ROOMS WITH THE FURNITURE (DB3 AND DB5)

To study the layout-generalization, we organize two more data campaigns in the presence of furniture. The first assessment is conducted in the same room of Fig. 1 but with added furniture (called DB3). As Fig. 2 shows, the radar is located at the same bottom-left corner while subjects can enter/leave the room from/to any two doors. The campaign is conducted by five new subjects any of whom executes SIT/STU activities at all the chairs and the sofa.

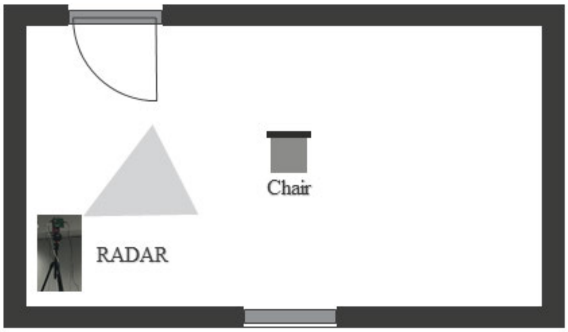


FIGURE 1. An empty room layout for running the first layout-free data-campaign.

The third set of measurements (called DB5) is recorded in a new room and in the presence of the furniture (shown by Fig. 3). It can be observed that chairs are located at different sides of the room to have the aspect angles for SIT and STU activities. In addition, the tables can model the occlusion in both SIT/STU and walking activities, especially, when a walking subject is masked by the table. This second campaign is also scheduled with 21 distinct subjects any of whom enters the room and executes SIT/STU at all the chairs, and the sofa.

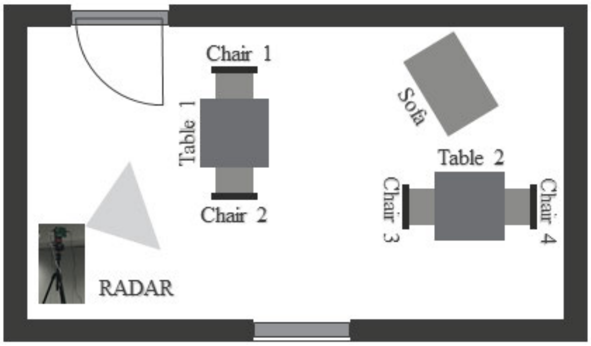


FIGURE 2. A room similar to the empty-room case but with added furniture for layout-generalization assessment.

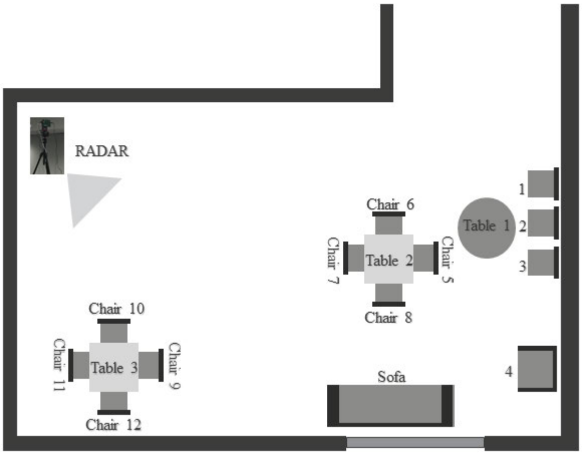


FIGURE 3. A new room with the furniture for the second layout-generalization assessment.

Remark 1: While all the experiments in this paper are done with a fixed radar location, a change in the location of the radar will affect the aspect angle of the subject to the radar's line-of-sight. In this case, by executing activities at chairs placed at different angles with respect to the radar and by studying the resulting performance, any possible impact imposed by changing the location of the radar will have been also analyzed.

IV. DATA PROCESSING AND ACTIVITY CLASSIFICATION

The activity classification in this work is modeled as a multi-class classification problem where a pipeline is designed to receive the raw radar measurements and conduct a number of pre-processing steps to produce a collection of features that will feed a machine-learning model aimed for

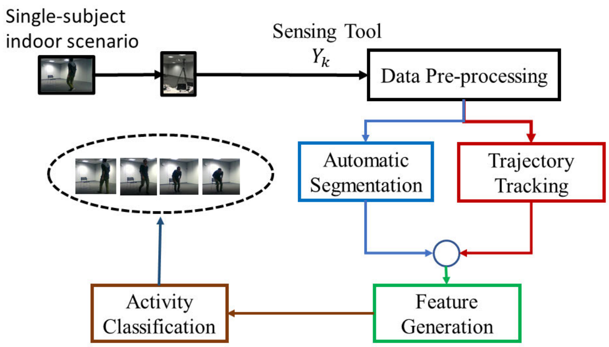


FIGURE 4. The HAR pipeline for radar-based activity recognition.

the activity classification. The activity classification pipeline is presented in Fig. 4 where the radar as the sensing tool outputs the raw-data Y_k at each kth radar-frame. After pre-processing, detection, and signal transformation, dynamic tracking is sought to predict the state of the target with an automatic segmentation being performed to extract those radar frames that correspond to the executed activities. The feature-generation pipeline then consumes the transformed signals and the tracker outputs to feed the machine-learning model that predicts the executed activity. The pipeline is now discussed in more details in the subsequent sub-sections.

A. DATA PRE-PROCESSING

Let us define the 4D $N_c \times N_T \times N_R \times N_f$ cube Y_k as the received complex echoes from the radar at the kth radar-frame where N_c, N_T, N_R and N_f are the numbers of chirps, fast-time samples, transmitters and receivers, respectively. Fig. 5 shows the data-processing pipeline that operates on the received Y_k data-cube and produces detection point-clouds and micro-Doppler (MD) cubes. The processing chain consists of the following blocks:

- Range-processing: the Fast Fourier Transform (FFT) is performed along with the fast-time samples of Y_k to form a new 3D cube RP_k of size $N_c \times N_v \times N_r$ where N_v denotes the number of virtual antennas ($N_v = N_R \times N_T$) and N_r is the number of range bins.
- Virtual-mapping: every pair of TX/RX antennas can be now projected into one of the elements in the 2D virtual antenna array whose shape is given in Fig. 6. The projected range-profile $\overleftrightarrow{RP}_k$ will be now a $N_c \times N_X \times N_Y \times N_r$ dimension with N_X and N_Y being the number of antennas in the 2D virtual array along the X and Y axes, respectively.
- Doppler-processing: the 4D cube $\overleftrightarrow{RP}_k$ is processed by another FFT along the slow-times (N_c) to give a 4D cube RD_k of size $N_d \times N_X \times N_Y \times N_r$ with N_d being the number of doppler resolution bins. Here, the doppler resolution can be also calculated by the following equation [32]:

$$v_{res} = \frac{\lambda}{2N_c PRI}$$

where λ denotes the wavelength and PRI corresponds to the pulse-repetition-interval.
- Beamforming: we perform a 2D conventional beamforming [33] on the range-profile $\overleftrightarrow{RP}_k$ to form a 4D cube RA_k of size $N_c \times N_{elv} \times N_{azm} \times N_r$ where N_{elv} and N_{azm} denote the number of elevation and azimuth bins, respectively. Note that N_{elv} and N_{azm} are the number of points in the 2D FFT performed on $\overleftrightarrow{RP}_k$ along with any of X and Y directions.
- micro-Doppler processing: all the generated range-angle cubes RA_k are sent to a buffer that is consumed by a Short-Time Fourier-Transform (STFT) block [34]. The STFT performs a Fourier transform over consecutive time-windows (aka slow-time) to compute the range-doppler profiles in time. As shown by Fig. 5, the STFT block receives the previous cubes over a certain window of radar-frames (say Δk) to produce a 5D MD-cube of size $N_d \times N_{elv} \times N_{azm} \times N_r \times N_s$ where N_s denotes the number of slow-time samples (time-bins) in the MD signature. The MD image can be aggregated over the angles and range-bins to form a 2D MD cube MD_k of size $N_d \times N_s$ where the n_dth row of the matrix represents the amplitude of the n_dth doppler bin over time.
- Detection: we perform a 1D Constant False Alarm Rate (CFAR) [35] detection over the range on the range-doppler data-cube RD_k to generate a collection of $N_v \times 1$ detection vectors $\gamma_n = RD_k[n_d^*, :, n_r^*]$ with $(n_d^*, n_r^*)_n, n = 1, \ldots, N$ being the set of CFAR-detected range-doppler bins where N corresponds to the number of detections.
- Point-cloud generation: every $N_v \times 1$ complex vector γ_n is now consumed by Multiple Signal Classification (MUSIC) as a high-resolution angular estimation algorithm with both forward-backward and spatial smoothing [36] to estimate the azimuth angle ϕ_n and, possibly, the elevation θ_n (the elevation estimation is noisy given the 2 virtual antennas in the Y-axis). The point-cloud is then formed as a set of detected bins with the associated estimates of range, elevation, azimuth,

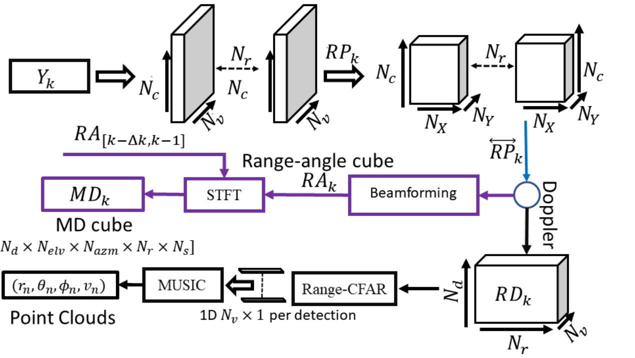
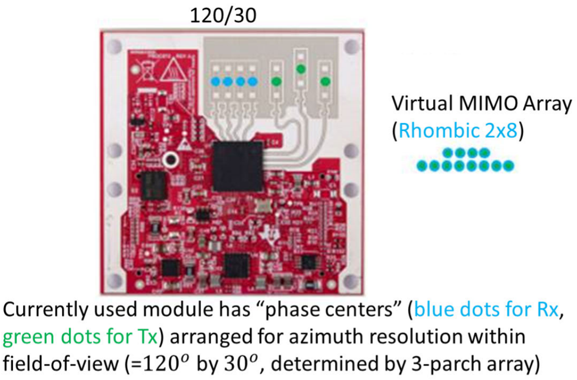


FIGURE 5. The data pre-processing and signal transformation pipeline for the wall-mounted radar.



Currently used module has "phase centers" (blue dots for Rx, green dots for Tx) arranged for azimuth resolution within field-of-view (=120^o by 30^o, determined by 3-parch array)

FIGURE 6. The wall-mounted radar chip and antenna on the PCB, with the resulting Rhombic virtual array.

and doppler-velocity, which is shown by the quadruple $(r_n, \theta_n, \phi_n, v_n)$.

B. TRAJECTORY TRACKING

One of the key elements of the activity classification pipeline is the dynamical tracking part whose architecture is shown in Fig. 7. Two main challenges in tracking a single subject in an indoor environment are as follows:

- Each subject usually executes a sequence of **move-stop-move** patterns when performing activities such as start walking, sitting, standing, stop walking, etc. We use an Interacting Multiple Model (IMM) filter designed for targets with move-stop-move patterns to mitigate the mentioned challenge [37].
- The moving subject generates multiple detection points per-frame, which makes the subject to be considered as an extended target in the radar measurements. This means that each estimated track can be associated with multiple received measurements that will complicate the data association problem. To design a data association solution for extended targets, the Generalized Nearest Neighbor (GNN) technique [16] is used to associate incoming measurements to the predicted state of the target.

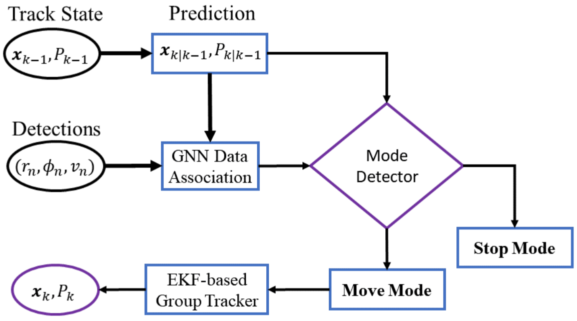


FIGURE 7. The dynamical tracking pipeline to perform single target tracking for subjects with move-stop-move patterns.

Let us now define $\{\boldsymbol{x}_{k-1|k-1}, P_{k-1|k-1}\}$ as the estimated state of the target at the $(k-1)$th radar-frame with $\boldsymbol{x}_{k-1|k-1}$ and $P_{k-1|k-1}$ being the estimated mean and covariance, respectively. Also, we assume $\{(r_{n_k}, \phi_{n_k}, v_{n_k})\}$, $n_k \in \{1, \ldots, N_k\}$ as the set of detected point clouds with N_k being the number of detected points at the kth scan. The following steps can be carried out to implement the tracking pipeline:

- **Track management:** to initiate a track or terminate an existing track, we use an m/n logic where a hypothesized track changes from a **tentative** state to **active** if the formed track is assigned to the received measurements in at least m out of n previous radar-frames. In the following, we assume that a track has been initiated by a proper m/n track management technique.
- **Dynamical target tracking:** the GNN-based group tracking is used to perform the data association and estimate the state of the target at each kth radar scan. The algorithm receives the raw data at each frame and detected measurements are associated with the initiated track using GNN. Note that, to discard false measurements, every measurement is assigned to a false track if its assignment cost falls below a threshold p_{fa}.
- **Mode update:** the mode of each target (MODE) is assumed to be chosen from the set {STOP, MOVE} with a binary indicator $\mu_k \in \{0, 1\}$ quantifying the state of the target at the kth radar-frame. The predicted radial velocity from each detection cluster is then used to update the target's state and make the mode detection. We also use the history of the binary mode indicator μ_k to determine the motion mode of the target. That is the mode of the target is STOP if at least δ scans of the last Δk radar frames attain $\mu_k = 0$. If the target is assigned to the stationary (STOP) mode, the mean and covariance are both set to $\boldsymbol{x}^s_{k|k}$ and $P^s_{k|k}$, respectively. If the target changes the mode from MOVE to STOP, the stationary mean is formed as

  $$\boldsymbol{x}^s_{k|k} = \left[x_{k|k-1} \; y_{k|k-1} \; 0 \; 0\right]^T$$

  with $\boldsymbol{x}_{k|k-1}$ being the predicted mean at the last MOVE step where $\{x, y\}$ correspond to the location estimates. The covariance for the stationary mode ($P^s_{k|k}$) is also set to a known P_0 matrix whose entries are chosen sufficiently large. This makes a larger gate for the predicted measurement at the stationary level and enables the tracking algorithm to capture the detections once the target makes a movement and, therefore, the mode is to be updated to MOVE.

Using the above-mentioned procedures, the dynamical trajectory tracking provides an estimate of the target state and associated covariance at each kth radar frame.

C. AUTOMATIC SEGMENTATION

The performance of the activity classification closely relies on the capability of the designed model to detect those radar frames that correspond to an executed activity (primitive).

Let us define N_k as the number of detections obtained at the kth radar frame. We also define $\hat{N}_k$ as the smoothed number of received detections that can be written as follows:

$$\hat{N}_k = (1-\alpha)\hat{N}_{k-1} + N_k$$

where $0 \leq \alpha \leq 1$. We also define Δk_l and Δk_f as two new parameters that determine the number of previous and future radar scans being reserved for start and end point detection, respectively. We, respectively, define the lagged and future number of buffered detection $\hat{L}_k$ and $\hat{F}_k$ as

$$\hat{L}_k = \hat{N}_{k-\Delta k_l+1:k}, \hat{F}_k = \hat{N}_{k:k+\Delta k_f-1}$$

The logic for the change-point detection requires to buffer the last Δk_l frames as well as the next Δk_f radar frames to determine the status of the current radar-frame. Each frame can be then assigned to one of **noAction**, **start**, **onGoing**, and **end** modes.

Given the data at the kth radar-frame, we specify the segment modes by analyzing the data at the $(k-\Delta k_f+1)$th frame as the radar-frames should be buffered within the next Δk_f frames to extract the required change-point statistics. The generated data for the frames with noAction status are used to estimate the background noise profile using the following equation:

$$\hat{\epsilon}_{k+1} = \frac{n_k\hat{\epsilon}_k + \hat{N}_k}{n_k + 1}$$

with $\hat{\epsilon}_k$ being the estimate of the noise at the kth frame and n_k as the effective number of noise samples up to the current frame.

The change-point detection for the automatic segmentation is now presented in Alg. 1. The function in Alg. 1 receives two buffers $(\hat{L}_k, \hat{F}_k)$, an estimate of the noise profile $(\hat{\epsilon}_k)$ and the latest state of the segments as the input and reports the current state of the frame. Upon the receipt of a start mode, a new segment is created and the segment may be terminated when an end signal is received by the algorithm.

Algorithm 1 Real-Time Change-Point Detection for Identifying Segments in Radar-Based Human Activity Recognition.

```
Parameters: $\rho_{min}$, $\rho_{max}$ and $\tau_{min}$
Function: change-point-detector($\hat{L}_k$,$\hat{F}_k$,$\hat{\epsilon}_k$,$s$)
  1) Apply linear Least-Square (LS) fit to $\hat{L}_k$ and $\hat{F}_l$, derive the linear curve parameters and calculate the averages $\mu^L_k$ and $\mu^F_k$, respectively.
  2) Calculate the ratio $\rho_k = \frac{max(\tau_{min}, \mu^L_k - \hat{\epsilon}_k)}{max(\tau_{min}, \mu^F_k - \hat{\epsilon}_k)}$ with $\tau_{min}$ being a given minimum bound.
  3) If $s$ is *noAction*
      a) If $\rho_k \geq \rho_{min}$, return(*start*); Else, return($s$)
  4) Else If $s$ is *start or on-going*
      a) If $\rho_k \leq \rho^1_{max}$, return (*end*); Else, return (*on-going*)
  5) Else if $s$ is *end*
      a) return(*noAction*)
```

V. A MACHINE-LEARNING PIPELINE FOR ACTIVITY CLASSIFICATION

As shown in Fig. 4, the human activity recognition is formulated as a multi-class classification problem with a feature-generation block forming hand-crafted features for the machine learning model. Here we, first, present the list of features and, then, discuss the machine learning architecture for the activity classification in Section V-B.

A. FEATURE EXTRACTION

As the input to feature-extraction layer, we use both the transformed signals (list of range-doppler map, range and doppler profiles) and the tracking results. All features are generated over a window of W radar-frames where all the radar-frames within the interval $k_f \in [k-W+1, k]$ are used to form the features. Five main categories of features are defined in the feature-extraction pipeline. In the following, we detail the proposed features:

- **Tracking features (TRACK):**
  - Input: list of estimated states $\boldsymbol{x}_{k_f|k_f}$ and predicted measurements $(r_{k_f|k_f-1}, \phi_{k_f|k_f-1}, v_{k_f|k_f-1})$ for all $k_f \in [k-W+1, k]$.
  - Process: we first use 8 core tracking features by collecting estimates of, (1) locations (x, y), (2) velocities $(\dot{x}, \dot{y})$, (3) range, radial velocity and acceleration $(r, \dot{r}, \ddot{r})$, and (4) azimuth (ϕ). Let us define $\boldsymbol{T}^x_{k_f}$ as the extracted core feature per radar-frame with x being one of the 8 core features. The tracking features can be now defined by applying different operators over the core features. We define seven operations: **average** (avg), **median** (med), **maximum** (max), **minimum** (min), **standard deviation** (std), **linear least-square fit** (lsf), and **deviation from the linear fit** (dev). Tracking features can be now derived as follows:

    $$\mathcal{T}^x_k = g_{\mathcal{T}}(\boldsymbol{T}^x_{[k]}), \mathcal{T} \in \{\text{avg,med,max,min,std,lsf,dev}\}$$

    with $\boldsymbol{T}^x_{[k]} = \boldsymbol{T}^x_{k-W+1:k}$ and $g_{\mathcal{T}}(.)$ being the computing function for the operation.
- **micro-Doppler features (MD):**
  - Input: an $N_d \times N_s W$ MD cube MD_k over a window of W successive radar frames with N_d and N_s being the number of doppler and time bins, respectively.
  - Process: denoising is first performed over each received MD cube by subtracting MD_{k_f} from an estimated background noise over time. Two sub-categories of MD features are now computed from the raw MD data-cube as follows:
    - ∗ **MD_RAW:** A subsampled $N^r_d \times N^r_s$ matrix MD^r_k is first formed by performing an average pooling

over the MD cube as follows:

$$MD^r_k[n_d, n_s] = \frac{1}{w_d w_s} \times \sum_{i=n_d w_d}^{(n_d+1)w_d-1} \sum_{j=n_s w_s}^{(n_s+1)w_s-1} MD_k[i,j]$$

where $w_{\{d,s\}} = \left\lfloor \frac{N_{\{d,s\}}}{N^r_{\{d,s\}}} \right\rfloor$, and $n_d \in \{0, \ldots, N^r_d - 1\}$ and $n_s \in \{0, \ldots, N^r_s - 1\}$ correspond to the doppler and time bins in the subsampled image, respectively. The subsampled matrix is flattened and an $N^r_d N^r_s$ vector of MD raw features $\boldsymbol{M}^R_k$ is generated.

* **MD_ENV:** micro-Doppler envelopes are also extracted from the raw MD cubes by, first, calculating the average doppler-bin per time and, then, fitting the resulting time-series against different functions such as polynomial, exponential, linear and sinusoid. The coefficients of the curved graphs are used as the envelope features.

- **Range-Doppler region-of-interest features (RD_ROI):**
  - Input: after taking the absolute value of each RD_k entry and aggregation over the virtual antennas, a 3D matrix RD^*_k of size $N_d \times N_r \times N_f$ (with $N_f = W$) is formed by concatenating all RD_{k_f} terms and for $k_f \in \{k - W + 1, k\}$.
  - Process: denoising is first performed over each cube by subtracting RD_{k_f} from an estimated background noise that is calculated by thresholding the range-doppler imager at each radar-frame and obtaining those bins falling below the threshold. A subsampled $N^r_d \times N^r_r \times N^r_f$ matrix is first formed by performing an average pooling over the range-Doppler cube as follows [17]:

$$RD^r_k[n_d, n_r, n_f] = \frac{1}{w_d w_r w_f} \times \sum_{i=n_d w_d}^{(n_d+1)w_d-1} \sum_{j=n_r w_r}^{(n_r+1)w_r-1} \sum_{l=n_f w_f}^{(n_f+1)w_f-1} RD^*_k[i,j,l]$$

where $w_{\{d,s,f\}} = \left\lfloor \frac{N_{\{d,s,f\}}}{N^r_{\{d,s,f\}}} \right\rfloor$, and $n_d \in \{0, \ldots, N^r_d - 1\}$, $n_s \in \{0, \ldots, N^r_s - 1\}$ and $n_f \in \{0, \ldots, N^r_f - 1\}$ correspond to the doppler, range and time bins in the subsampled image, respectively. The subsampled matrix is flattened and an $N^r_d N^r_r N^r_f$ vector of range-doppler raw features $\boldsymbol{R}^R_k$ is generated.

- **PointCloud features (POINT):**
  - Input: set of detections for range $\boldsymbol{p}^r_{k_f}$, azimuth $\boldsymbol{p}^\phi_{k_f}$, elevation $\boldsymbol{p}^\theta_{k_f}$, and doppler $\boldsymbol{p}^v_{k_f}$ over W successive radar-frames with $k_f \in [k - W + 1, k]$ and $\boldsymbol{p}^x_{k_f}$ being the set of detections for $x \in \{r, \phi, \theta, v\}$ at a given k_f radar-frame.
  - Process: let us define $\boldsymbol{h}^x_{k_f}$ as the histogram of the detected points at each radar-frame with $\boldsymbol{h}^x_{k_f}$ being a $B \times 1$ vector of normalized weights where B denotes the number of bins and $\boldsymbol{h}^x_{k_f}[b_n]$ will be the number of detections falling in the b_nth bin. By concatenating all the hisograms over the W frames, we form a $W \times B$ matrix H^x_k. For any of four entities listed above, we compute the bin **average** (avg), **mode** (mod) and **standard deviation** (std) per radar-frame as follows:

$$\mathcal{S}^x_{k_f} = f_{\mathcal{S}}(\boldsymbol{h}^x_{k_f}), \mathcal{S} \in \{\text{avg,mod,std}\} \tag{1}$$

with $f_{\mathcal{S}}$ being a function that applies the operator $\mathcal{S}$ over the given input. Features are then derived by applying five operations on (1) as

$$\mathcal{F}^x_k = g_{\mathcal{F}}(\mathcal{S}^x_{[k]}), \mathcal{F} \in \{\text{avg,med,max,min,std}\}$$

where $g_{\mathcal{F}}$ is a function that applies the operator $\mathcal{F}$over the set of $\mathcal{S}^x_{[k]} = \mathcal{S}^x_{k-W+1:k}$. We then categorize features as **RNG**,**AZM**, **ELV** and **DOPP** with each term being defined as follows:

$$\{\textbf{RNG,AZM,ELV,DOPP}\} = \mathcal{F}^{\{r,\phi,\theta,v\}}_k(\mathcal{S}^x_{k-W+1:x}) \tag{2}$$

- **Meta features (META):**
  - Input: taking the absolute value of RD_k entries and aggregating over the virtual antennas, a 3D matrix RD^*_k of size $N_d \times N_r \times N_f$ (with $N_f = W$) is formed by concatenating all RD_{k_f} terms and for $k_f \in \{k - W + 1, k\}$.
  - Process: we extract three main core meta features as **centroid velocity** (VCENT), **dispersion in the velocity** (VDISS), and **range instantaneous energy** (RENG) from the range-doppler images (for more details see [17]). Let us define $\boldsymbol{m}^x_{k_f}$ as the derived core meta feature per radar-frame with x being one of the represented core features. The meta features are now computed by applying operations over the core features as follows:

$$\mathcal{M}^x_k = g_{\mathcal{M}}(\boldsymbol{m}^x_{[k]}), \quad \mathcal{M} \in \{\text{avg,med,max,min,std,lsf,dev}\}$$

with $\boldsymbol{m}_{[k]} = \boldsymbol{m}_{k-W+1:k}$.

B. ACTIVITY CLASSIFICATION

The human activity recognition problem in this paper is cast as a multi-class classification problem with the action primitives being the output classes. We use a Random Forest (RF) model [18] to undertake the activity classification. Two classification models are devised in this paper for the activity classification as follows:

- **Single-stage classification:** the feature-extraction is triggered once a new segment of W successive radar-frames is detected (see Section IV-C) and, then,

TABLE 3. Activity classification for the same-room performance evaluation with different types of classification results being reported in the table.

Action	DB1(%)	DB5(%)		
Classifier	SS_noSITSTU	SS_noSITSTU	SS_withSITSTU	TwoStage
WAW	$96+/-5.2$	$88+/-4.4$	$87+/-2.4$	$82+/-4.2$
WTW	$94+/-9.3$	$89+/-2.8$	$89+/-2.4$	$90+/-3.5$
SIT	$96+/-5.3$	$75+/-5.1$	None	$86+/-3.6$
STU	$95+/-7.7$	$76+/-6.3$	None	$88+/-2.9$
SIT_STU	None	None	$91+/-2.4$	None
macro	$95+/-6.4$	$82+/-3.4$	$89+/-2.2$	$86+/-2.2$
micro	$95+/-7.1$	$82+/-3.5$	$90+/-2$	$86+/-2.4$
weighted	$94.5+/-7.2$	$82+/-3.4$	$90+/-2$	$87+/-2.2$

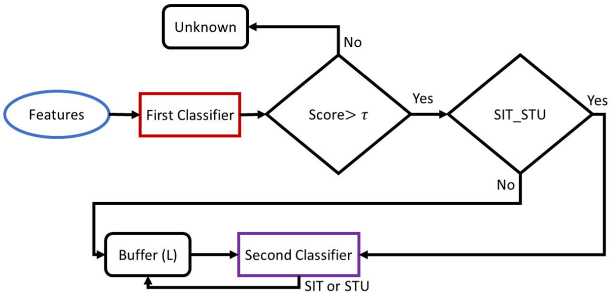


FIGURE 8. Two-stage classification for the activity classification problem.

the feature vector is sent to the RF model to classify the activity into one of {SIT,STU,WAW,WTW} classes.

- **Two-stage classification:** Fig. 8 shows the proposed architecture for the two-stage classification. The first stage takes as input the generated feature and uses RF to classify the activity into one the 3 {WAW,WTW,SIT_STU} classes where SIT_STU is a new composite label formed by merging SIT and STU activities. All the instances with a maximum score below a design parameter τ are projected to an Unknown label. If the predicted label is SIT_STU, a second RF model is used to classify the instance SIT vs. STU. The second classifier receives the assigned labels over the last L feature-frames and decodes SIT_STU based on the history of the executed activities. The two-stage classifier provides much more robust results in challenging environments and under the existence of occlusions and different aspect-angles where the performance of the single-stage classification for both SIT and STU labels is dramatically affected.

VI. EXPERIMENTAL RESULTS

We have created a data-base of features and labels by processing the recordings for all the campaigns explained in Section III. Each data sequence is created by processing W contiguous radar-frames with $W = 12$ is chosen for the living-room scenarios formed in Section III. The optimal value of W has been empirically chosen with details been omitted here for the sake of brevity and space limitation.

The activity classification is performed by the procedure given in Section V-B. The performance of the model is evaluated by the k-fold validation. As of the performance metric, the F1-score is used as the harmonic average of the precision (pr) and recall (rc) [18]. Given the existence of multiple labels, we adopt three scores, i.e. **weighted**, **macro** and **micro** averages of F1-scores to summarize the overall performance obtained by the machine learning model. As the measurements were collected in different rooms and layouts, we perform the evaluation per-room as well as cross-rooms to assess the generalization of the trained models.

A. SAME ROOM-LAYOUT EVALUATION

In this section, the performance of the model is evaluated for cases where the model is trained/tested on the same room and under the same layout. The single-stage RF classifier of Section V-B is trained using the records collected from the three data-bases (DB1,3,5). The performance of the model is evaluated by a 5-fold validation where participating subjects are divided into 5 equally-sized folds.

We perform the evaluation for each database separately. As DB3 contains only 5 subjects, the same-layout analysis is performed for DB1 and DB5 with a high number of participating subjects. The obtained F1-scores are presented in Table 3 with three types of classification strategies being evaluated, (1) single-stage classifier with no SIT_STU label (SS_noSITSTU), (2) single-stage classifier with SIT_STU composite label (SS_withSITSTU), and (3) two-stage classifier (TwoStage). When using the single-stage activity classification with no composite SIT_STU label, it is observed that the RF model has provided about 95% weighted F1-score for DB1 while the value has lowered by 20% in case of DB5 as the experimental data-base. The performance gap widens when SIT and STU labels are considered with DB1 proving 25% more accurate in recognizing SIT and STU activities in comparison to DB5. Note that the complex layout of DB5 as opposed to the simple empty layout designated to DB1 has led to some performance drop, especially, when subjects execute SIT and STU actions near tables and under different chair side-angles.

The impact of occlusion and side-angles is now analyzed by evaluating the performance of the model over any certain group of chairs in the layout of Fig. 3. We now

TABLE 4. The performance of single-stage activity classification for DB5 with the metrics being calculated per chair-type.

Action	SIT_F_C	STU_F_C	SIT_S_C	STU_S_C	SIT_FF_C	STU_FF_C	SIT_FB_C	STU_FB_C
Weighted F1-score (%)	89	93	67	67	60	68	74	74

define 4 classes of chairs as (1): **Free_Chairs (F_C)** being chairs 1-4 and the sofa in Fig. 3, (2) **Side_Chairs (S_C)** being chairs 5,7,9,11 in the layout of Fig. 3, (3) **Face_Backward_Chairs (FW_C)** being chairs 6 and 10 in Fig. 3, and (4) **Face_Forward_Chairs (FB_C)** being chairs 8 and 12 of Fig. 3.

The F1-scores are calculated for any of the above categories, and results are shown in Table 4. It is observed that the model has performed well in labeling SIT and STU activities for free chairs where nothing is occluding the subject against the radar line-of-sight. The performance degrades by around 15% when those chairs facing back to the radar are considered with the radar reporting the worst results when chairs are facing the radar. While the subjects are not occluded in case of side-chairs, the classifier suffers from the poor aspect-angle where the performance falls down to 67% when subjects sit or stand at side-chairs.

As a further test, we combine SIT and STU labels and create a new composite SIT_STU label that replaces all the previously created SIT and STU instances. The classifier is now trained with the new labels and the results are shown in Table 3 under SS_with SITSTU label. It can be seen that the classifier has been able to classify SIT_STU activity with more than 90% accuracy that also indicates the capability of the model to distinguish stationary activities (SIT_STU) from the walking primitives. Motivated by the reported results of the single-stage classifier with SIT_STU label, we now train a two-stage classifier whose architecture was given by Fig. 8. *To train the second classifier in* Fig. 8*, we also buffer* $L = 6$ *of previous detections reported by the first RF classifier.* The results for the two-stage classification can be now observed in Table 3 and under the name TwoStage. The obtained classification results outperform those reported by the single-stage activity classification where the F1-score for both SIT and STU activities show 10% improvement over those obtained by the single-stage classifier. Indeed, while the existence of the furniture has made the classification task much cumbersome for the single-stage RF model, the model still shows desirable generalization when detecting SIT_STU against walking actions. The two-stage classifier then exploits the aforementioned fact to build a classification strategy that shows a reasonable degree of robustness to the room layout and resulting challenges such as poor aspect-angles and occlusions.

B. UNSEEN ROOM-LAYOUT EVALUATION

In this section, we study how the trained model in a certain room-layout can generalize to a new environment with an unseen layout. The analysis is now performed in two cases with (1) the same room but with train//test being executed over different layouts and (2) two separate rooms with different layouts. Here the same-room part is first studied and, then, the metrics are extracted for a more challenging task of training on a certain room and testing on a new room with an unseen layout.

1) SAME-ROOM ANALYSIS

To study the generalization of the activity classification model to a the layout change within the same room, the next experiment studies how the trained model on DB1 performs on the unseen layout of DB3. For a more systematic evaluation, we define 5 categories for the furniture in Fig. 2 where new labels SIT_X and STU_X are created with X∈ {C1,C2,C3,C4,Sofa}. Here C{i} denotes Chair{i} in Fig. 2 and, therefore, label {SIT/STU}_C{i} refers to a SIT or STU activity executed at the corresponding chair.

In order to better understand the impact of layout on the model's performance, we now depict both micro-Doppler and range-Doppler images over the full course of both sitting and standing activities in Fig. 9. Graphs clearly show discrepancies in the micro-Doppler signatures normally used as the main set of features for the machine learning model. when a subject sits/stands, the radar echoes from the lower part of the subject's body show a possible range-deviation, which results in the micro-Doppler images with higher energies at positive/negative Doppler regions that helps the model label the action properly. While this signature is well captured in case of a Sofa, the occlusions caused by the tables and a change in the aspect-angle of each chair make the signatures less instructive for the machine learning model. In this case, the performance for both sit/stand activities will be dramatically affected.

As DB3 is created by 5 distinct subjects and to evaluate the generalization of the trained model, we incrementally add the subjects from DB3 to DB1 to form an extended training dataset. The single-stage classifier is then trained on the extended data and tested on the remaining subjects in DB3. The results are now shown in Table 5 for different portions of the data extension.

In case of no-extension **(nTrainInTest**= 0**),** the results in Table 5 show a dramatic performance drop of around 25% for both walking and stationary activities when compared toDB1 metrics listed in Table 3. Diving deeper into the metrics of Table 5, it can be observed that the trained model is still able to capture the activities executed at the Sofa with a reasonable accuracy while the results for the chairs significantly drop. It can be also observed that the model provides much better prediction results for Chair 2 as the chair is located almost in front of the radar and, therefore, is not occluded by the table. Both Chairs 1 and 4 that are masked

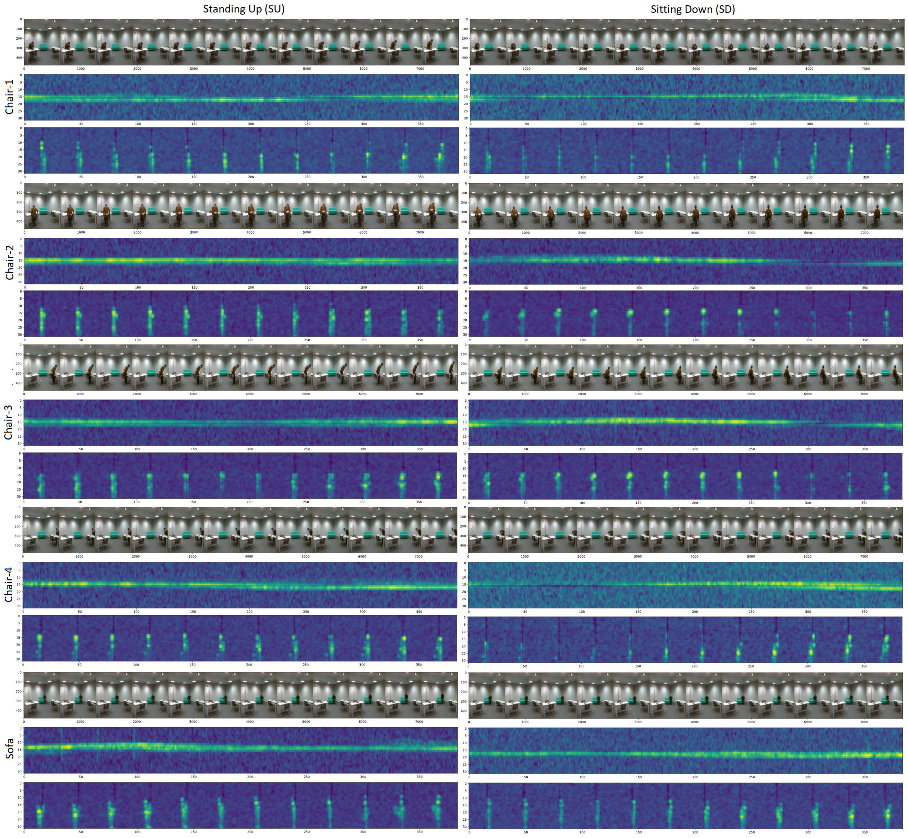


FIGURE 9. Variation of micro-Doppler and range-Doppler signatures for SIT and STU activities over the whole period of each action when each activity is executed at a sofa and chairs in different aspect angles. For each case, the middle graph denotes the micro-Doppler signature while the bottom one corresponds to the range-Doppler over radar-frames.

by the table show the lowest performance while Chair 3 is also hard for the model to capture given its poor aspect-angle, which makes the subject invisible to the radar. The model achieves the highest performance for Chair 2 with around 70% accuracy for both SIT and STU primitives. Note that as Chair 2 is fully visible by the radar and not masked by the table, the model is able to detect both SIT and STU activities with a reasonable accuracy although the metric is still lower than that of the Sofa since the subject's body is partially masked by the chair.

When data extension is taken (nTrainInTest> 0), it can be observed in Table 5 that extending DB1 data with a single object from DB3 can enhance the metrics by around 20% in average. While the metrics keep improving by adding more subjects from DB3, the gap tightens when the number of subjects from DB3 in the training set goes from 3 to 4. It can be also concluded that the model provides a satisfactory performance for the Sofa and Chair 2 by adding a single subject from DB3 while other cases still need to see more records from the new layout to adapt themselves to the changes in the environment. Note that this procedure may be used as an extra phase to calibrate the pre-trained model to the new layout in a room while adding some limited measurements to the existing model can improve the robustness and the generalization of the activity classification.

TABLE 5. Cross-layout generalization results for the model trained on the extended data from DB1 and DB3 and tested on the remaining DB3 subjects.

Activity	SS_noSITSTU (%)				
	nTrainInTest				
	0	1	2	3	4
WAW	75.4	88.4	90	96.7	96.3
WTW	66.4	86	87.5	91.7	92.8
SIT_C1	45.6	70.2	78.8	78.8	79.4
STU_C1	33.8	59.7	70.8	80.7	80.4
SIT_C2	65.3	72	87	90	91
STU_C2	76.8	82	91	92	93
SIT_C3	51.1	66	72	74	75
STU_C3	56.1	67	68	76	77
SIT_C4	38.1	62	76	80	79
STU_C4	45	62	74	78	78
SIT_Sofa	73.5	98	99	99	100
STU_Sofa	92	100	100	100	100
macro	66	79	83	83.6	86
micro	71.2	81.1	84.6	85	87
weighted	68.2	80.7	84.5	85	87

2) CROSS-ROOM ANALYSIS

In this section, we study how the trained model in a certain room can generalize itself to a new room with the furniture layout being different from the training campaign. For a rigorous generalization study, 6 different test cases are designed as follows:

- **Train_DB{X}_Test_DB{Y}:** the records collected by DB{X} are used as the training-set while DB{Y} will be held to evaluate the model's performance ({X,Y}∈ {1/3,5}).
  - **SS_noSITSTU:** the single-stage classifier with no composite-label is trained on DB{X} data and the performance is evaluated on DB{Y}.
  - **SS_withSITSTU:** the single-stage classifier is trained on DB{X} data using the newly defined label SIT_STU. DB{Y} data will be then used for performance evaluation.
  - **TwoStage:** we finally use the two-stage classifier and train the model on DB{X} data and, then, test on DB{Y}.

The results for all the test-cases can be found in Table 6. With DB1/3 being held as the training-set, the model shows a relatively poor generalization performance with the weighted metric being below 70% for all three test-cases. It is also observed that the two-stage classifier fails to boost the performance, especially, for SIT/STU activities as the single-stage classifier with the composite label shows inaccurate in distinguishing SIT_STU labels from walking activities. Note that as the two-stage classifier receives the history of assigned labels as the input, miss-classifications in SIT_STU activities are usually translated to a large miss-classification error in the two-stage classification results as many of SIT (STU) activities may be wrongly labeled as STU (SIT). Results in Table 6 also show poor F1-metrics for walking activities when compared to those of Table 3. Indeed, the existence of two tables and more variation in the aspect-angles of the walking subjects arising in the dataset of DB5 has made the poor generalization from a simpler DB1/3 scenario to the more complex DB5 layout.

Table 6 also verifies the enhanced generalization of the model when trained on DB5 and tested on DB1/3. In case of a single-stage classifier with no composite label, the trained model achieves a 15% improvement over the scenario with DB1/3 being used as the training-set. Results also show a significant improvement on the capability of the model in identifying both SIT and STU activities where the model tested on DB1/3 outperforms the one trained on DB1/3 and tested on DB5 by around 10% in both SIT and STU activities. When the single-stage model with the composite label is used, all the F1-scores are enhanced and the results show 86% accuracy for the new SIT_STU label. Such an improvement can be also observed in the metrics reported by the two-stage classifier where the model has classified SIT and STU activities by 78% and 86% F1-score, respectively, that outperforms the single-stage classification results by around 20%.

Reported metrics in Table 6 reveal a number of important observations. First, once the model is trained on a room that introduces a representative layout, the generalization is greatly enhanced when tested in a new room with an unseen layout. In other words, training the model on DB5 showed a desirable degree of generalization when the metrics are calculated using the data in DB3 that has a different layout than DB5. Also, we see the two-stage classification does not help in case of the model being trained on DB1/3 as the model does not show a high accuracy in classifying the new SIT_STU label.

As another experiment, we now study the impact of data extension in the classification performance and the generalization capability of the cross-room experiments. The regular single-stage model with no composite label is now used as the baseline and the macro-average is taken as the representative metric. The model is trained on the data from DB1/3 and tested by DB5 with the data being proportionally extended from DB5 to DB1/3. Fig. 10 shows the macro-average of F1-scores vs different extension ratios. With no-extension, all the cases show poor performance results as the trained model on DB1/3 does not generalize well to the unseen layout of DB5. The performance shows a significant improvement once a 10% data-extension is performed with the single-stage classifier trained on the sole DB1 proving a 40% enhancement with a small extension from DB5.

C. SUMMARY

Several experiments were conducted in this section to study the performance of different classification perspectives in identifying human activities in an indoor environment. The main observations can be summarized as follows:

- Compared to prior work in [30], [31], we have evaluated the generalization of the radar-based human activity recognition in a systematic way where the role of occlusion and aspect-angles has been deeply investigated in the capability of the model for detecting SIT and STU activities, something that has not been carefully discussed in the state-of-the-art.
- Our results indicate that the trained models show desirable classification accuracy once tested on the same training layout. Metrics in Table 3 verify the reasonable accuracy and generalization to the unseen subjects when the trained model is tested on the same layout.
- The accuracy of the activity classification is dramatically affected by the change in the room layout, especially, for SIT and STU labels. The existence of tables as the main source of occlusion and scenarios where subjects execute SIT/STU actions against chairs located at a poor aspect-angle to the radar's line-of-sight cause a significant drop in the classification accuracy.
- Our research offers two ways to enhance the generalization across different rooms and with unseen layouts. First, by training on a more representative room (DB5) and, then, using a two-stage classification, we showed in Table 6 that the metrics for the activity classification can be enhanced by 10% in average for all cases when trained on a certain layout and tested on an unseen layout.
- Beside the two-stage classification proposal, it was shown by both Table 5 and Fig. 10 that data-extension

TABLE 6. Cross-room generalization results for the different classification models when the model is trained on a room and tested on a different one.

	DB1/3_DB5 (%)			DB5_DB1/3(%)		
	SS_noSITSTU	SS_withSITSTU	TwoStage	SS_noSITSTU	SS_withSITSTU	TwoStage
WAW	67	64	63	88	84	78
WTW	70	67	66	92	88	87
SIT	49	None	52	58	None	78
STU	56	None	55	63	None	86
SIT_STU	None	75	None	None	86	None
macro	60	70	66	75	86	83
micro	60	70	67	82	86	83
weighted	61	70	66	81	86	84

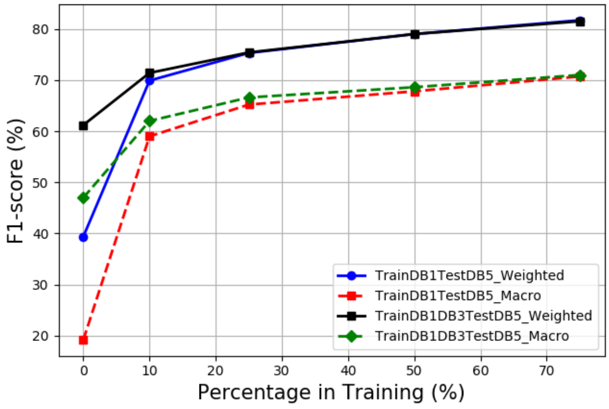


FIGURE 10. The macro-average results of the single-stage classification with different levels of the data-extension.

plays a vital role in boosting the model's performance where adding some limited data from an unseen layout to the existing model can significantly enhance the capability of the model in classifying the activities executed in an unseen layout.

The next section deals with a deep feature exploration where we will attain some improvement in the generalization of the model by choosing an optimal subset of features.

VII. FEATURE EXPLORATION

We evaluate the importance of features in the model performance by formulating an optimization objective over a pre-defined taxonomy of features. Fig. 11 shows the defined taxonomy in this paper with all the leaves corresponding to each certain feature category. Let us define $\boldsymbol{x} = [x_i], i = 1, \ldots, I$ as a vector of binary variables $x_i \in 0, 1$ that quantifies the existence of the ith feature category in the pool of features. We also define$\boldsymbol{f_i}$ as the feature vector corresponding to the ith category. Assuming $\boldsymbol{x_n}$ as a certain vector created by a permutation of x_i values and $\mathcal{C}(\boldsymbol{x_n})$ as the associated performance metric generated by the RF model, the objective for feature optimization is defined as follows:

$$\boldsymbol{x}^* = argmax\ \mathcal{C}(\boldsymbol{x})$$

Solving the above problem requires to consider all possible subsets of feature categories (2^I different cases). Here we pursue a sub-optimal way to find the most contributing features with details being given by Alg. 2. The ranked list of features X^* and the corresponding metric $\boldsymbol{C^*}$ can be now used to plot the feature importance curve where for each ith entry of X^*, C_i^* denotes the metric obtained by training the RF model using all the features $X_{1:i}^*$. In the following, the importance curves are derived for each single scenario.

Algorithm 2 A Sub-Optimal Feature-Selection Procedure to Rank the Importance of Features Consumed by the RF Classifier

```
Init: Number of feature categories $I$, list of selected features $X^* = []$ and corresponding metrics $\boldsymbol{C^*}\ \boldsymbol{\mathcal{D}}$ [], and full set of features $F = \boldsymbol{f_i}, i = 1, \ldots, I$
While $F$ is non-empty
    • Set $\boldsymbol{C} = []$ as a temporary vector of metrics.
    For $\boldsymbol{f_i}$ in $F$
    1) Create a new set of features $X_i$ by appending $\boldsymbol{f_i}$ to $X^*$.
    2) Train the RF model using $X_i$, calculate the metric $\mathcal{C}_i$ and add the metric to $\boldsymbol{C}$.
    Choose the category $i^* = argmax_i \boldsymbol{C_i}$, add $\boldsymbol{f_{i^*}}$ to $X^*$, $\boldsymbol{C_i}$ to $\boldsymbol{C^*}$, and remove $\boldsymbol{f_{i^*}}$ from $F$.
Report $X^*$ and $\boldsymbol{C^*}$as the ranked set of features and corresponding cumulative RF metric, respectively.
```

A. SAME/cross-ROOM FEATURE ANALYSIS

We calculate the feature importance for both DB5 and DB1/3 data and rank the features according to the procedures given by Alg. 2. Results for every experiment are now shown in Fig. 12 where the y-axis in both graphs shows the macroAverage as the representative metric. It can be observed that tracking-related features (positional and velocity) are all ranked quite high in both cases. The experiment running on DB1/3 shows features from range-Doppler ranked on the top while these features have the model's performance degraded in case of DB5 data and ranked pretty low. MD_RAW features are also shown as the least important features for both DB1/3 and DB5 cases.

For the next experiment, we perform the feature selection on the same room but with different train/test layouts. To do this and for the case of DB5, we collect the captures in three distinct sets per subject where each set consists of action primitives executed at a certain set of chairs with the

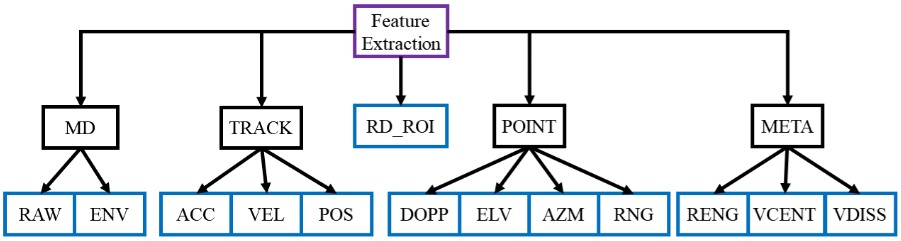
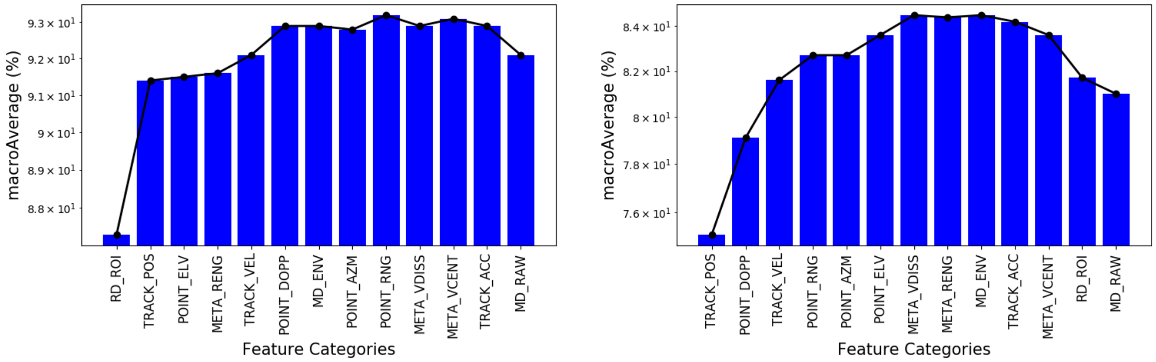


FIGURE 11. Taxonomy of features for exploration and performance optimization.



(a) (b)

FIGURE 12. Feature importance analysis for the experiment running on the same-room, (a) train/test on DB1/3 and (b) train/test on DB5. X-axis shows the categories of features as denoted by the taxonomy of Fig. 11 where all feature classes are ranked according to the procedure of Alg. 2.

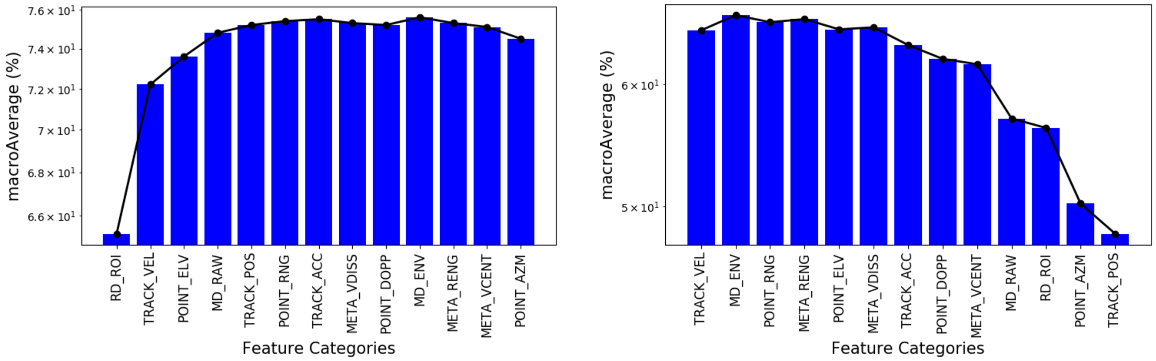


FIGURE 13. Feature importance analysis for the experiment running on the same room, but tested on different layouts, (a) train on DB5 with a different test layout, and (b) train on DB1 and test on DB3. X-axis shows the categories of features as denoted by the taxonomy of Fig. 11 where all feature classes are ranked according to the procedure of Alg. 2.

sets forming disjoint subset of chairs. The model is then trained on two sets and tested on the unseen set. In a separate experiment, we also train the model on DB1 and test on DB3 to evaluate the contribution of features when the model is exposed to a new layout. The results for feature selection are now depicted in Fig. 13. Unlike the graphs shown in Fig. 12, MD_RAW features are now ranked high by the model bringing around 2% improvement in macroAverage when MD_RAW features are appended to the list of existing features. The positional features still prove useful although their importance has downgraded when compared to Fig. 12.

As the last experiment, the features are examined in a more challenging task where the model is trained on DB5 and tested on a new room/layout (DB1/3). The results for feature-selection are now depicted in Fig. 14. It can be observed that a mixture of MD_RAW and positional features gives the best generalization capability with the best outcome being achieved when the MD_RAW features are combined with the elevation information and the position of the target obtained from the tracker.

B. SUMMARY

The feature exploration results presented in Section VII-A revealed important conclusions about the importance of the feature categories in the generalization capability of the trained activity classification models:

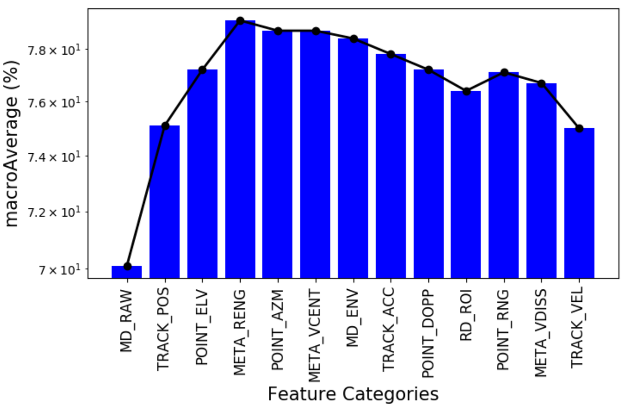


FIGURE 14. Feature important analysis for the cross-room experiment with the model being trained on DB5 and tested on DB1/3. X-axis shows the categories of features as denoted by the taxonomy of Fig. 11 where all feature classes are ranked according to the procedure of Alg. 2.

- First, positional features dominate other features for the case of same-room experiments. As Fig. 12 shows, the model gives a high weight to the positional features that decode the location of the furniture, which indicates an early sign of overtraining to the layout.
- When the model is trained on a certain room, but tested on a different layout, micro-Doppler features are elevated in the ranked list of features while velocity features that are more robust to the layout change go higher in the ranking. It can be also observed in Fig. 13 that the positional features are now ranked lower than the case of same-room analysis, especially, for DB5 experiment where adding the features dramatically degrade the metric.
- Finally, an analysis on the cross-room training/testing verifies the significant importance of micro-Doppler features in the generalization capability of the model. Results in Fig. 14 imply that a combination of micro-Doppler features along with those types of information that cannot be captured in MD such as the position and elevation has provided the best performance metric. Adding other features in Fig. 14 then degrades the metrics where the model trained with the full feature-set shows 5% performance drop compared to the best feature combination.

VIII. CONCLUSION

This paper proposed a systematic methodology to assess the generalization capability of a machine-learning based solution for activity classification in indoor environments. Using a well-defined tracking algorithm, a novel signal processing and feature extraction pipeline was devised to feed a dedicated two-stage classifier that detects well-defined human activities such as walking, sitting and standing. By running several measurement campaigns in rooms with different layouts, the performance of the machine learning model was evaluated under different layout variations such as aspect-angles and occlusions. After presenting an extensive feature exploration, the results in this paper concluded a mixture of micro-Doppler and tracking features provides the best generalization capability, especially when the trained model is employed to predict human activities in a new room with an unseen layout.

In a general setting, a single-radar module can still suffer from the insufficient generalization given poor aspect angles of subjects or variant room layouts. In this case and as a future work, a multi-radar solution is being investigated where the placement of multiple radars can potentially mitigate both occlusions and poor aspect angles. In addition, including more sophisticated human activities such as falling, cooking, opening or closing doors in the pipeline will be considered in the next extensions of this work.

REFERENCES

[1] S. R. Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition—A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, pp. 1–22, Jul. 2018.

[2] M. A. Al Hafiz Khan, R. Kukkapalli, P. Waradpande, S. Kulandaivel, N. Banerjee, N. Roy, and R. Robucci, "RAM: Radar-based activity monitor," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.

[3] M. S. Seyfioglu, A. M. Ozbayoglu, and S. Z. Gurbuz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1709–1723, Aug. 2018.

[4] L. Fan, T. Li, Y. Yuan, and D. Katabi, "In-home daily-life captioning using radio signals," 2020, *arXiv:2008.10966*. [Online]. Available: http://arxiv.org/abs/2008.10966

[5] G. Diraco, A. Leone, and P. Siciliano, "A radar-based smart sensor for unobtrusive elderly monitoring in ambient assisted living applications," *Biosensors*, vol. 7, no. 4, p. 55, Nov. 2017.

[6] F. Luo, S. Poslad, and E. Bodanese, "Kitchen activity detection for healthcare using a low-power radar-enabled sensor network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–7.

[7] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, May 2019.

[8] J. Monteiro, R. Granada, R. C. Barros, and F. Meneguzzi, "Deep neural networks for kitchen activity recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 2048–2055.

[9] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, T. Dhaene, and W. De Neve, "Indoor human activity recognition using high-dimensional sensors and deep neural networks," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12295–12309, Aug. 2019.

[10] H. Gu, "Real-time human activity recognition based on radar," M.S. thesis, Ball State Univ., Indiana, Muncie, IN, USA, May 2019.

[11] T. D. Bufler and R. M. Narayanan, "Radar classification of indoor targets using support vector machines," *IET Radar, Sonar Navigat.*, vol. 10, no. 8, pp. 1468–1476, Oct. 2016.

[12] B. Jokanovic, M. Amin, F. Ahmad, and B. Boashash, "Radar fall detection using principal component analysis," *Proc. SPIE*, vol. 9829, May 2016, Art. no. 982919.

[13] Y. Lang, Q. Wang, Y. Yang, C. Hou, D. Huang, and W. Xiang, "Unsupervised domain adaptation for micro-Doppler human motion classification via feature fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 392–396, Mar. 2019.

[14] M. Li, T. Chen, and H. Du, "Human behavior recognition using range-velocity-time points," *IEEE Access*, vol. 8, pp. 37914–37925, 2020.

[15] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 1, pp. 103–123, Jan. 1998.

[16] P. Konstantinova, A. Udvarev, and T. Semerdjiev, "A study of a target tracking algorithm using global nearest neighbor approach," in *Proc. 4th Int. Conf. Comput. Syst. Technol. e-Learn. (CompSysTech)*, New York, NY, USA, 2003, pp. 290–295.

[17] J. Lien, N. Gillian, P. Amihood, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.

[18] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2013.

[19] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation: Theory Algorithms and Software*. Hoboken, NJ, USA: Wiley, 2004.

[20] S. Abdulatif, F. Aziz, B. Kleiner, and U. Schneider, "Real-time capable micro-Doppler signature decomposition of walking human limbs," in *Proc. IEEE Radar Conf. (RadarConf)*, Seattle, WA, USA, May 2017, pp. 1093–1098.

[21] M. A. Al Hafiz Khan, R. Kukkapalli, P. Waradpande, S. Kulandaivel, N. Banerjee, N. Roy, and R. Robucci, "RAM: Radar-based activity monitor," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, San Francisco, CA, USA, Apr. 2016, pp. 1–9.

[22] J. Bryan and Y. Kim, "Classification of human activities on UWB radar using a support vector machine," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, Toronto, ON, Canada, Jul. 2010, pp. 1–4.

[23] J. D. Bryan, J. Kwon, N. Lee, and Y. Kim, "Application of ultra-wide band radar for classification of human activities," *IET Radar, Sonar Navigat.*, vol. 6, no. 3, p. 172, 2012.

[24] C. Ding, H. Hong, Y. Zou, H. Chu, X. Zhu, F. Fioranelli, J. Le Kernec, and C. Li, "Continuous human motion recognition with a dynamic range-Doppler trajectory method based on FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6821–6831, Sep. 2019.

[25] M. P. Ebrahim, M. Sarvi, and M. Yuce, "A Doppler radar system for sensing physiological parameters in walking and standing positions," *Sensors*, vol. 17, no. 3, p. 485, Mar. 2017.

[26] B. Erol, S. Z. Gurbuz, and M. G. Amin, "Motion classification using kinematically sifted ACGAN-synthesized radar micro-Doppler signatures," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 4, pp. 3197–3213, Aug. 2020.

[27] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, Jul. 2019.

[28] Y. Yang, C. Hou, Y. Lang, T. Sakamoto, Y. He, and W. Xiang, "Omnidirectional motion classification with monostatic radar system using micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3574–3587, May 2020.

[29] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 2994–3009, Apr. 2021.

[30] M. Jia, S. Li, J. L. Kernec, S. Yang, F. Fioranelli, and O. Romain, "Human activity classification with radar signal processing and machine learning," in *Proc. Int. Conf. UK-China Emerg. Technol. (UCET)*, Aug. 2020, pp. 1–5.

[31] S. A. Shah and F. Fioranelli, "Human activity recognition: Preliminary results for dataset portability using FMCW radar," in *Proc. Int. Radar Conf. (RADAR)*, Toulon, France, Sep. 2019, pp. 1–4.

[32] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," Texas Instrum., Dallas, TX, USA, Tech. Rep. SPYY005A, 2020.

[33] K. W. Masui, J. R. Shaw, C. Ng, K. M. Smith, K. Vanderlinde, and A. Paradise, "Algorithms for FFT beamforming radio interferometers," *Astrophys. J.*, vol. 879, no. 1, p. 16, Jun. 2019.

[34] E. Sejdić, I. Djurović, and J. Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digit. Signal Process.*, vol. 19, no. 1, pp. 153–183, Jan. 2009.

[35] H. Rohling, "Ordered statistic CFAR technique—An overview," in *Proc. Int. Radar Symp. (IRS)*, Leipzig, Germany, Oct. 2011, pp. 631–638.

[36] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 1, pp. 8–15, Jan. 1989.

[37] T. Kirubarajan and Y. Bar-Shalom, "Tracking evasive move-stop-move targets with a GMTI radar using a VS-IMM estimator," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 3, pp. 1098–1103, Jul. 2003.

[38] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from FMCW radar with bi-LSTM networks," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13607–13619, Nov. 2020.

[39] S.-W. Kang, M.-H. Jang, and S. Lee, "Identification of human motion using radar sensor in an indoor environment," *Sensors*, vol. 21, no. 7, p. 2305, Mar. 2021.

ALI GORJI received the B.Sc. and M.Sc. degrees from the Amirkabir University of Technology, Iran, in 2005 and 2008, respectively, and the Ph.D. degree from McMaster University, Canada, in 2012. Previously, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada. His research interests include statistical signal processing, detection and estimation theory, and target tracking and their applications in radar, sensor systems, and robotics.

HABIB-UR-REHMAN KHALID received the B.Sc. degree in electrical engineering from The University of Lahore (UoL), Pakistan, in 2013, and the master's degree in electrical and electronics engineering from the Katholieke Universiteit Leuven (KUL), Belgium, in 2018. He is currently pursuing the Ph.D. degree in engineering sciences with the Vrije Universiteit Brussel (VUB) in collaboration with the Interuniversitair Micro-Elektronica Centrum (IMEC), under the supervision of Prof. H. Sahli. His main research interests include digital signal processing, real-time embedded and control systems, multi-sensor data fusion, energy-efficient machine learning, and heterogeneous sensor-based human activity recognition.

ANDRÉ BOURDOUX (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from the Université Catholique de Louvain, Belgium, in 1982. In 1998, he joined IMEC. He is currently a Principal Member of Technical Staff with the Advanced RF Research Group of IMEC. He is also a System Level and Signal Processing Expert for both the mm-wave wireless communications and radar teams. He has more than 15 years of research experience in radar systems and 15 years of research experience in broadband wireless communications. He holds several patents in these fields. He is the author or coauthor of over 180 publications in books and peer-reviewed journals and conferences. His research interests include advanced architectures, signal processing and machine learning for wireless physical layer, and high-resolution 3D/4D radars.

HICHEM SAHLI received the degree in mathematics and computer science, the D.E.A. degree in computer vision, and the Ph.D. degree in computer sciences from the Ecole Nationale Superieure de Physique Strasbourg, France. Since 2000, he has been a Professor with the Department of Electronics and Informatics (ETRO) and a Scientist with the Interuniversitair Micro-Elektronica Centrum VZW (IMEC). He coordinates the Audio-Visual Signal Processing Laboratory (AVSP) within ETRO. AVSP conducts research on applied and theoretical problems related to machine learning, signal and image processing, and computer vision. The group explores and capitalizes on the correlation between speech and video data for computational intelligence where efficient numerical methods of computational engineering are combined with the problems of information processing.