Grouped People Counting Using mm-wave FMCW MIMO Radar

Ren, Liyuan; Yarovoy, Alexander; Fioranelli, Francesco

DOI
10.1109/JIOT.2023.3282797

Publication date
2023

Document Version
Final published version

Published in
IEEE Internet of Things Journal

Grouped People Counting Using mm-Wave FMCW MIMO Radar

Liyuan Ren, Alexander G. Yarovoy, *Fellow, IEEE*, and Francesco Fioranelli, *Senior Member, IEEE*

***Abstract*—The problem of radar-based counting of multiple individuals moving as a single group is addressed using an mm-wave multiple-input–multiple-output (MIMO) frequency-modulated continuous wave (FMCW) radar. This problem is challenging because the different individuals are closer to each other than the range/azimuth resolution, and their bulk Doppler signatures are difficult to distinguish, as they tend to move together. A processing pipeline is proposed, based on the combination of a multiple target tracking algorithm with a classifier to track each group and count the number of people within. Specific salient features are defined for the classifier and extracted from range–azimuth maps and cadence velocity diagrams (CVDs). The proposed pipeline has been experimentally validated in several outdoor scenarios with grouped people. The results show that the combination of tracking algorithm and classifier in the proposed pipeline outperforms alternative methods from the literature as well as a commercial toolbox for people counting.**

***Index Terms*—mm-wave radar, people counting, radar signal processing, tracking and classification.**

I. Introduction

PEOPLE Counting problem (also referred to as 'Regional People Counting' [1]) aims to sense the number or density of people in a Region of Interest (RoI). There are many application scenarios for this, such as epidemic prevention control, shopping malls, elevators, public transportation, security, and Internet of Things (IoT) functionalities in smart cities and urban environments [2].

Based on the existing literature, the problem of People Counting can be formulated with three different macro objectives relevant to different scenarios: 1) People Counting; 2) true People Counting; and 3) exact People Counting. Simple *People Counting* means estimating the density of people where/when knowing their exact number in the RoI is not necessary, e.g., at the airport entrance [3], or on a pedestrian road [4]. *True People Counting* means counting the specific number of people while the location (range, Angle of Arrival (AoA), and elevation information) of each person is not required, e.g., at the subway platform [5]. As the most advanced type of People Counting, the *exact People Counting* implies counting the accurate number of people and estimating the location of each person. This provides more information about people's behavior for business analysis, epidemiological investigation [6], and other IoT scenarios [7].

As manual counting is time consuming and prone to mistakes, many advanced technologies have been proposed to solve the People Counting problem, such as Wi-Fi/Bluetooth devices [8], thermal image sensors [9], video camera [10], radar [11], and LiDAR [12]. These technologies have different environmental requirements for proper working (e.g., the radiation of sunlight could influence the detection of the thermal image sensor [13]). LiDAR is expensive compared to radar, can be influenced by environmental light conditions, and some wavelengths may be harmful to human eyes, although it has the ability to obtain very fine and accurate detection of objects in space. In this area, radar is increasingly becoming a popular technology because of its robustness to light and weather conditions, contactless capabilities, and potential advantages in terms of privacy preservation.

There are two main categories of radar-based techniques for exact People Counting proposed in the recent literature. The first category includes the *feature-based counting* methods, where features extracted from the radar data are used to train a statistical or neural network (NN)-based classifier to estimate the number of people in the RoI [14], [15], [16]. In other words, this family of approaches defines the People Counting problem as a classification problem. These methods have generally low computational requirements and good performance for multiple stationary people, but they require large groups of diverse data for training the classifiers [17], [18]. The second category includes the *tracking for counting* methods. These are essentially multiple targets tracking (MTT) algorithms to estimate the motion of each person in the RoI, i.e., the number of people as well as their location. Depending on the wavelength and the distance from the radar, the target person can be considered as a point-like or extended target. The processing pipeline before the tracking algorithms may differ because of different radar types (i.e., frequency-modulated continuous wave (FMCW) radar or pulsed UWB radar) [19], [20]. Differently from feature-based counting methods, MTT-based algorithms do not rely on training data to address people counting and can be directly implemented. This offers more flexibility when counting multiple people in different environments compared with feature-based counting methods [20].

Radar-based people counting methods need to be robust against complex and diverse scenarios with multiple targets

Manuscript received 10 April 2023; accepted 1 June 2023. Date of publication 5 June 2023; date of current version 7 November 2023. This work was supported in part by NWO KLEIN *RAD-ART* Project. *(Corresponding author: Francesco Fioranelli.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by TU Delft HREC under Application No. 1387.

The authors are with the Microwave Sensing, Signals and Systems Group, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: f.fioranelli@tudelft.nl).

Digital Object Identifier 10.1109/JIOT.2023.3282797

moving. However, to the best of our knowledge, the grouping phenomenon is often not considered for radar-based people counting methods, i.e., the event when multiple individuals are moving together and close to each other as a single group. This is a challenging scenario for radar-based sensing and counting because these multiple grouped people are separated from each other by less than the range/azimuth resolution, but also their bulk Doppler signatures will be similar as they will tend to move together. This article proposes a dedicated pipeline to address specifically this problem, i.e., grouped people counting, by combining feature-based and tracking methods into a single framework. The main contributions of this work are as follows.

1) One of the most complex movements in the area of radar-based people counting, i.e., counting multiple individuals moving as single groups, is studied and addressed with a dedicated processing pipeline that uses multiple-input–multiple-output (MIMO) FMCW radar.
2) This pipeline was developed [21] by studying the specific characteristics of radar signatures of groups of people, leading to the intuition to combine Range–Azimuth maps and cadence velocity diagram (CVD) as salient features to count the number of people, together with an MTT algorithm.
3) The proposed pipeline has been experimentally validated in several outdoor scenarios with grouped people at the TU Delft campus. It is shown that the complementary strengths of feature-based and tracking-based methods in the proposed pipeline outperform alternative methods from the state-of-the-art (SOTA) literature as well as a commercial toolbox for people counting.

This article is organized as follows. The characteristics of radar signatures of grouped people are discussed in Section II to derive salient features from the radar data. Based on this study, the proposed pipeline combining tracking for counting and feature-based counting is introduced in Section III. The pipeline is experimentally validated with three data sets in the performance study presented in Section IV, with a comparison with alternative methods from the literature presented in Section V. Finally, conclusions are given in Section VI.

II. Radar-Based Characterization and Features for Grouped People Counting

In this section, the characteristics of radar signatures of grouped people are studied to derive features for the proposed pipeline later presented in Section III.

A. Definition of Grouped People and Its Modeling

Grouped People is defined as a cluster of people sharing neighboring locations and moving together, as shown in Fig. 1(a). This behavior is common in everyday life, such as friends walking in pairs or couples taking a walk together. From the perspective of analyzing radar data, separating grouped people is challenging, as the individuals are closer than the range/azimuth resolution, and their Doppler signatures are mixed together. In the existing literature on radar-based People Counting, the assumption on people's motion is often limited to walking without grouping, even when multiple people are present in the scene. Although some of the existing articles state that people move randomly, the grouping phenomenon seems to be not considered in their assumptions [18], or it is assumed that individuals do not always move as a group [15]. In other words, the people observed by the radar will always separate at a certain moment and move as separate individuals.

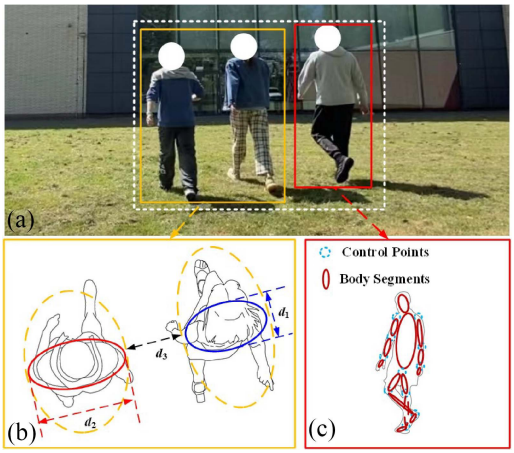


Fig. 1. (a) Ground-truth video for a group of three individuals, and schematic model of grouped people: (b) in the top view with the three parameters defining the size of the group, and (c) in the front view with control points and segments inspired by the Boulic model.

To help define the grouping scenario and study the grouped people's signatures in the radar's view, a model of Grouped People is developed [21]. First, a 3-D model of a walking individual is built inspired by the popular Boulic model [22] which considers 11 control points with ellipsoids to model the different body segments, as shown in Fig. 1(c). Then, the connection between different individuals closely located in a group is modeled with three parameters. These are shown in Fig. 1(b), namely, d_1 is the chest width of the individuals, d_2 is the shoulder width, and d_3 is the Euclidean distance between two neighboring individuals. The parameters d_1 and d_2 essentially define the dimensions of the ellipse with the area occupied by each individual in a group. These parameters could be derived from publicly available population data, or by enlisting volunteers to participate. d_3 is the distance between people in the group, which is related to the angular resolution cell of the radar. Simulated spectrograms for one individual and two grouped people derived from this model are shown in Fig. 2. It is visible how even with only two individuals it is difficult to isolate the envelope or individual body parts from the spectrogram. Therefore, in the next section, other radar data domains are analyzed for grouped people's signatures. It is important to note that this model was used to initially study the grouping scenario and possible features from relevant data domains, such as the examples of spectrograms in Fig. 2, but the data analysed for the subsequent sections with results are all experimentally collected.

B. Analysis of Radar Data Domains for Grouped People Counting

Noted from the previous section the challenge to use spectrograms for grouped people counting, different radar domains

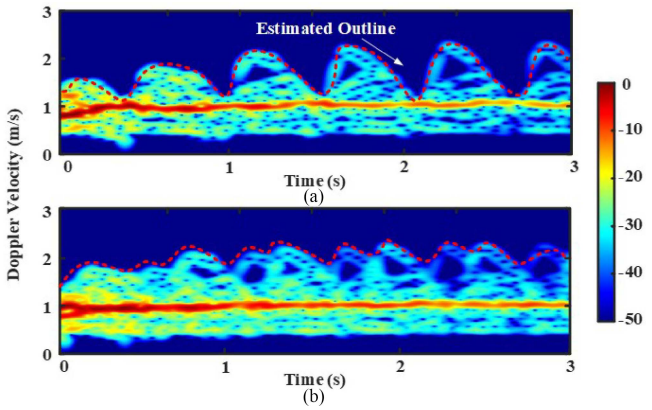


Fig. 2. Simulated spectrograms based on the proposed Group People model when (a) 1 person is walking and (b) 2 people are walking.

are considered. The aim is to find whether salient features can be extracted from specific radar domains. The readers should note that the data used to generate these example figures belong to the collected experimental Data Set I, which is discussed in more detail in Section IV.

First, the *Range–Azimuth* domain is considered. Range–Azimuth map shows the spatial location of targets in the RoI. This domain is commonly used in narrow areas, such as vehicle, elevator, and metal ramp [3]. Not all radar systems are able to provide angular information, as MIMO radar is needed to estimate azimuth and sometimes even elevation information of targets. However, a potential weakness of this representation is that radar has poorer angular resolution for extended targets, such as people compared to LiDAR, and the resolution degrades away from the boresight direction. Nevertheless, the Range–Azimuth map can provide some information for Grouped People Counting as it shows directly relevant spatial information for the area occupied by a group of people.

At farther distances however, the link between the number of occupied cells in the range–azimuth domain and the number of grouped people is found to be less pronounced on its own to estimate the number of people in a group [21]. For example, in Fig. 3, where three and five people walk together as a group at approximately 16 m of distance, the number of occupied cells seems to be the same. This can be explained by the Grouped People model. As people get further away from the radar, their angular footprint gets smaller and smaller until it is less than the angular resolution. The results when only using the features in range–azimuth domain for Grouped People Counting are provided later in Section IV-C.

C. Study of Synchronization Between People in Groups

Additional consideration is then given to the *Doppler domain*. Micro-Doppler spectrograms are typically used in the literature on radar-based human monitoring for gait and activity recognition. From the simulated spectrograms in Fig. 2, it was noted that it is difficult to separate micro-Doppler contributions from multiple people walking together as a group, even when they are just two individuals, because their signatures are mixed together. However, the simultaneous analysis of experimental recordings for grouped people [21] showed

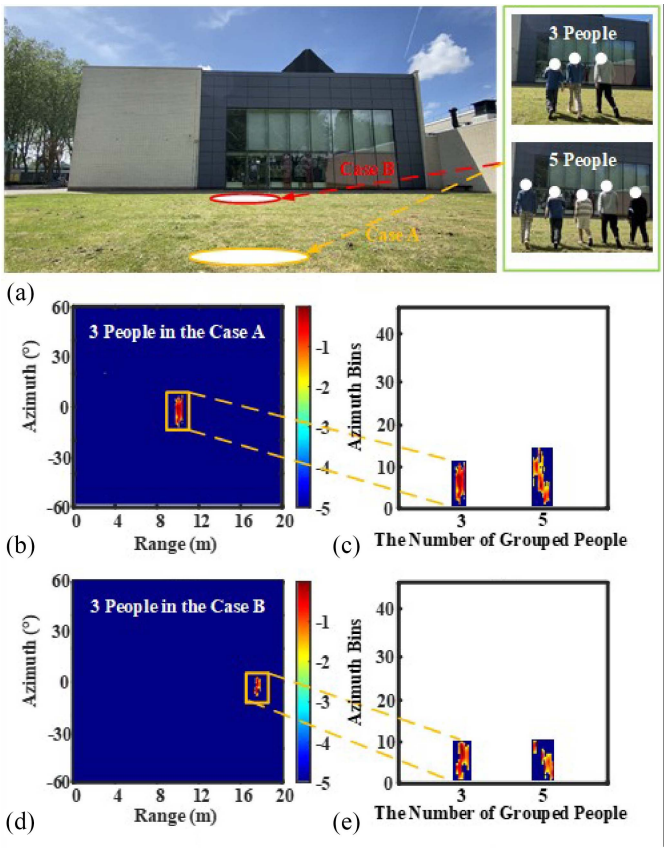


Fig. 3. (a) Video ground truth for three people and five people walking as a group in two cases denoted by A and B. (b) Range–Azimuth map in the case A at about 9 m, with comparison of the detected cells for groups of 3 versus 5 people (c). (d) Range–Azimuth map in the case B at about 16 m, with comparison of the detected cells for groups of 3 versus 5 people (e).

that this mixing can be less pronounced than in the simulated signatures. Essentially, what appeared to happen when the participants could walk in a natural way was that they would walk together with a relatively consistent, synchronized pace. While this "synchronization phenomenon" was not perfect (as for example the case of soldiers marching together), it was still noticeable in the spectrograms for a relatively small number of grouped people and tended to disappear for larger groups.

To quantify the synchronization phenomenon, the CVD is introduced starting from the spectrogram. The spectrogram is defined as

$$\hat{S}(l, g_2) = \sum_{n=0}^{N_b-1} s(n + l\triangle l)w(n)\mathrm{e}^{-\mathrm{j}2\pi g_2 n/N_b} \tag{1}$$

where $s(n)$ is the complex signal in the range–time domain, and $\hat{S}(l, g_2)$ is the spectrogram calculated from this STFT process. $w(n)$ is the Hanning window of the length N_b. $\triangle l$ is the number of overlapping samples between each two window.

The CVD $S_c\,(g_1, g_2)$ is then calculated by taking the FFT of the spectrogram across the time axis

$$S_c(g_1, g_2) = \sum_{l=0}^{N_w-1} |\hat{S}(l, g_2)|w(l)\mathrm{e}^{-\mathrm{j}2\pi g_1 l/N_w} \tag{2}$$

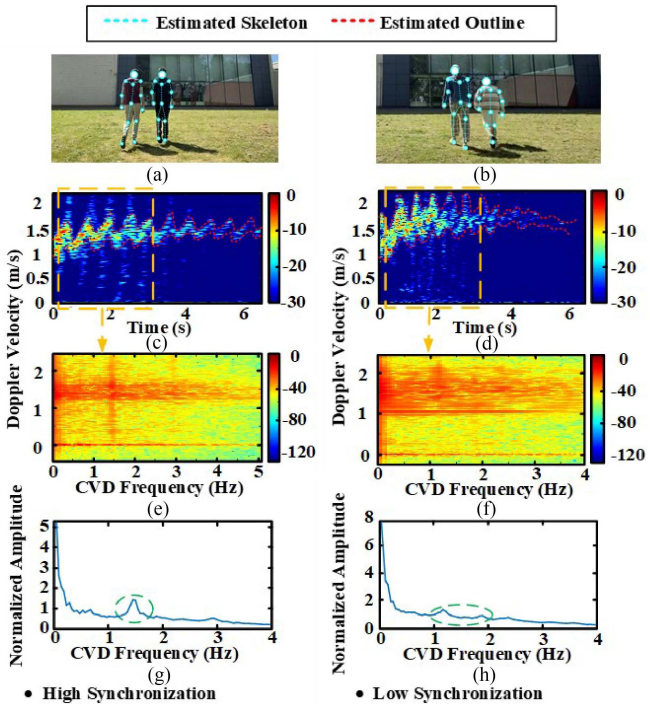


Fig. 4. Analysis of the synchronization for two people walking as a group. Ground truth with estimated skeleton in (a) high synchronization case for people of similar height, and (b) low synchronization case for people of diverse height. Spectrograms for (c) high synchronization and (d) low synchronization case, with Doppler velocity displayed in absolute value. CVD maps in (e) high synchronization and (f) low synchronization case, with CVD profiles in (g) high synchronization and (h) low synchronization case.

where $w(l)$ is the Hanning window with length N_h. N_w is the total number of windows when applying FFT. The indices are defined as $g_1 = 0, 1, 2, \ldots, N_h - 1$ for cadence frequency, and $g_2 = 0, 1, 2, \ldots, N_w - 1$ for Doppler.

By taking the FFT across the time axis, the CVD represents the frequency with which the Doppler velocity in the target signature repeats, essentially the periodic synchronization in the Doppler signature. Due to the fact that the frequency of the limbs' swinging during normal human walking movement will not exceed 2.5 Hz and will not fall below 0.5 Hz [23], the CVD is selected in this range of cadence frequencies for subsequent study. To investigate the relationship between CVD and the synchrony of human movement, two sets of data from Data Set 1 (described later in Section IV) were selected. Each set of data was collected by two different volunteers. Fig. 4 shows detailed information on the difference between the high synchronization case and the low synchronization case.

The degree of synchrony of human movement was found to be related to the height of the people. When there was little difference in height between the two volunteers, they were more likely to move in high synchrony due to the similarity in their stride length [in Fig. 4(a)]. On the contrary, when there was a larger difference in height between the two volunteers, they were more likely to move asynchronously [in Fig. 4(b)]. In Fig. 4(a) and (b), according to the control points and the estimated skeleton marked on the people, it is shown that in the high synchronization case the two volunteers' feet are raised from the ground at the same time and their arms are swung at

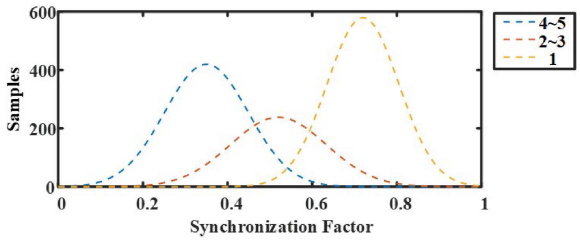


Fig. 5. Distributions of the synchronization factor parameter derived from the CVD for different numbers of grouped people.

a similar angle. However, in the low-synchrony situation, the two volunteers' feet landed on the ground at different times, and their arms swung at different frequencies.

In terms of spectrograms, this is much less "clean" in the case of low synchronization [in Fig. 4(d)] compared to that of high synchronization [in Fig. 4(c)]. As conclusions beyond this are difficult to obtain through empirical observation, CVDs are extracted. In Fig. 4(e), it is seen that there is one vertical line with high amplitude around 1.5 Hz along the cadence frequency. By calculating the mean value along the Doppler velocity of the CVD map, we can obtain the CVD profile and clearly find a peak at around 1.5 Hz as shown in Fig. 4(g). This peak represents the periodic motion of the two participants. In contrast, as in the case of low synchronization the inconsistency of the two gaits leads to "unclean" CVD, the resulting peak in the CVD map/profile is not very pronounced.

Based on this intuition, a quantitative parameter of *synchronization factor* is defined to be used as a possible feature for Grouped People counting. By comparing the difference between CVD profiles of high and low synchronization cases in Fig. 4(g) and (h), we can see that the shape of the peaks between 0.5 and 2.5 Hz is different. The higher the synchronization is, the higher the peak and the narrower its width will be. Conversely, the lower the synchronization is, the lower and wider the peak will be.

To represent this difference, the synchronization factor δ is proposed inspired by [24], and its calculation is listed in Algorithm 1. The synchronization factor is the highest when there is only one person walking and will decrease when more people will walk together as a group, indicating that it can be used as a suitable feature for Grouped People counting. To validate the contribution of CVD for Grouped People Counting, the training set of Data Set I (described in detail in Section IV-A) is used. Based on the fact that the calculated synchronization factor is the highest when there is only one person moving in the RoI, a normalization is proposed to constrain the synchronization factor from 0 to 1. In Fig. 5, the phenomenon of synchronization is studied by fitting Gaussian distributions to the values of the synchronization factor derived from the CVD for different number of grouped people. It can be seen that as the number of people increases, the synchronization factor becomes on average smaller. Moreover, the peak of the fitted distributions for different number of Grouped People is relatively well separated and consistent, suggesting that this can be a useful feature for grouped people counting.

Algorithm 1 Proposed Synchronization Factor From CVD Profile

Require: CVD Profile $|Sc(g_1)|_k|$ as defined in [28], with $g_1 = 0, 1, ..., N_h - 1$, $k = 1, 2, ..., K$, overlapping frequency factor σ_l, and sampling duration T_c

Calculate ambiguity of CVD Frequency: $f_s = 1/(N_w \sigma_l T_c) = 1/(N_w - \triangle l)T_c$

Calculate CVD step: $x_c = \text{linspace}\ (-f_s/2, f_s/2, g_1)$

Select the appropriate interval of CVD profile:

if $2.5 \geq x_c \geq 0.5$ **then**

$|Sc_1(x_c)|_k| = |Sc(x_c)|_k|$

end if

$G_{\max} = \max(|Sc_1(x_c)|_k|)$

$x_{c0} = \arg\max_{x_c}(|Sc_1(x_c)|_k|)$

Find the largest local minimum $G_1 = |Sc_1(x_c 1)|_k|$ around x_{c0}

Find the minimum $G_2 = |Sc_1(x_c 2)|_k|$

$\delta_k = ((G_{\max} - G_1)/G_1) + (G_{\max} - G_2)/G_2))/2$

Ensure: δ_k

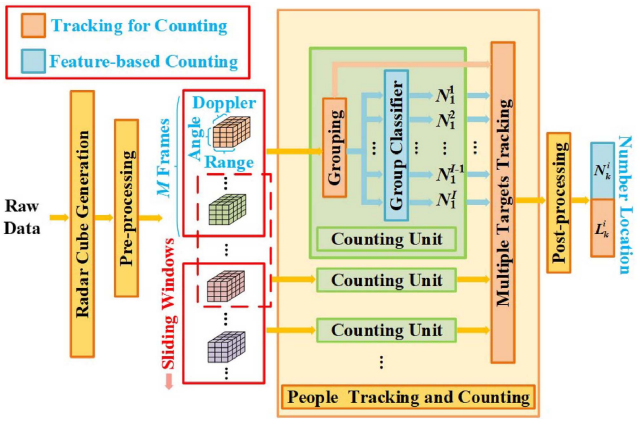


Fig. 6. Overview of the proposed people counting pipeline, which combines relevant elements of the tracking for counting and feature-based counting approaches typically used in isolation in the literature.

To conclude this section, it was shown that both features from the Range–Azimuth maps and the synchronization factor from CVD maps should be considered to address the problem of Grouped People counting.

III. Proposed Method

This section presents the proposed method which combines a "tracking for counting block" and a "feature-based counting block". The method can be divided into four main parts: 1) radar cube generation; 2) preprocessing; 3) people tracking and counting block; and 4) post-processing part, as shown in Fig. 6.

A. Radar Cube Generation and Preprocessing

The radar cube generation step converts the digitized raw data acquired from the radar into a standard cube format for subsequent processing. When using FMCW MIMO radars with multiple transmitter and receiver channels, the size of the resulting cube is Number of Channels × Number of Fast Time Bins × Number of Slow Time Bins × Number of Frames.

Subsequent preprocessing aims to obtain the Doppler velocity, Range and Angle information from the radar cube. Rather than applying an FFT across the three dimensions of the cube, in this article the Doppler-based angle estimation method is applied. The base idea is to generate the Range–Doppler map first with a 2-D FFT across fast and slow time samples. Then, a detection method (i.e., 2-D CA-CFAR [25]) is applied to find the range–Doppler bins containing targets and at the same time remove static clutter contributions near the zero Doppler. Finally, the FFT beamforming is performed along the Channel dimension only at those range and Doppler bins selected by the CFAR.

After preprocessing, moving targets are defined in the processed radar cube as sets of values of Angle × Range × Doppler × Frame. Sliding windows across M frames are then applied to select multiple frames and generate the complex signal in the range–time domain. In this case, the time of this complex signal in the range–time domain equals the number of frames (M frames) multiplied by the time per frame. Afterwards, after applying (1) and (2), spectrograms and then, CVD maps are generated in the subsequent people tracking and counting blocks of the pipeline. In this work, the length of the sliding window M is set at ten frames, with an overlap of 90% between consecutive frames. For each window, the number of people in the scene and their location will be estimated.

B. People Tracking and Counting Block

The people tracking and counting block is designed to estimate the location and number of Grouped People by a combination of tracking and feature-based counting. This includes a counting unit to process each decision window and estimate the number of Grouped People, and a multiple target tracking (MTT) algorithm to predict and update people's location in the scene.

1) Grouping: Grouping is the first operation of the people tracking and counting block, and its aim is to separate different groups in the scene from the Range–Azimuth map and pass each group's feature (i.e., the Range–Azimuth map, and the CVD map) to the subsequent Group classifier. This operation relies on clustering and assignment algorithms. The use of *clustering* serves two purposes. First, to calculate distances on the Range–Azimuth map based on the density of measurements, and thus to separate the different clusters within the scene corresponding to different groups of people. Second, to calculate the cluster centroid based on the measurements of each group, which is later used in the tracking algorithm. In this article, the grid-based density-based spatial clustering of applications (DBSCAN) is used [26]. Compared to standard DBSCAN, grid-based DBSCAN takes the relation between angle bin and range bin into account to calculate distances. As a result, different groups with their measurements are divided in the scene, and the centroid of each group is calculated and passed to the MTT block. The *assignment* algorithm has two purposes. The first is to assign the processed feature to the

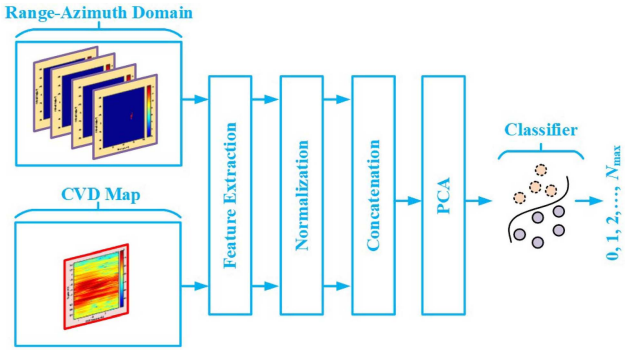


Fig. 7. Overview of the steps for the proposed group classifier.

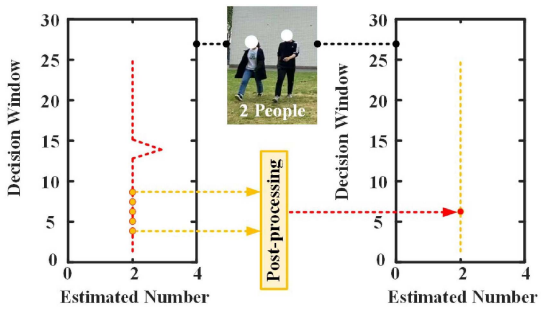


Fig. 8. Results when applying the post-processing block (ground truth with 1 group and 3 people present): estimated number of people per group before and after applying the post-processing block.

correct group. For example, Group 1 is detected walking from range bin 130 to range bin 137 in one window, and then the spectrogram calculated in that range is assigned to Group 1. The second purpose is to select the best association hypothesis θ_k^i (i.e., the hypothesis in the ith frame and the kth decision window) between each frame, and to prune all other hypotheses $\theta_k^i \in \Theta_k^i$. Thus, the single N-target hypothesis is applied. In this article, the best association hypothesis is determined by the distance between each estimated position of the groups with the Gaussian model [27].

2) Group Classifier: The group classifier estimates the number of people in each group. In this article, a statistical ML-based classifier is used as shown in Fig. 7. As previously discussed, features from the range–azimuth domain and CVD map are used. For features in the range–azimuth domain, three kind of features are extracted, i.e., 1) the length of the occupied azimuth bin; 2) the peak amplitude; and 3) the mean amplitudes calculated across two consecutive frames. As a consequence to calculate features every two frames, the length of the feature vectors extracted from the Range–Azimuth maps is 15 (three features for each pair of frames in a sliding window of ten frames). For the CVD map, four features are extracted inspired by [28], i.e., 1) the synchronization factor δ mentioned in Section II-C; 2) the maximum cadence frequency of the CVD profile; 3) two half-height cadence frequencies from the CVD profile; and 4) the selected mean amplitudes of the CVD map along the Doppler velocity axis. Normalization is applied to each selected feature extracted from Range–Azimuth map and CVD map. Then, the two sets of feature vectors are concatenated and principal component analysis (PCA) is used to reduce the dimensionality of the predictor space from the 25 features. Reducing the dimensionality can help classification models prevent overfitting [15]. In this work, only the principal components that retain 80% of the information are kept and passed to the classifier. The selection of the classifiers and the study of their performance are provided in Sections IV and V.

3) Multiple Targets Tracking: After the best association hypothesis is selected within the Grouping block, the MTT block updates all tracks from the Range–Azimuth map and initiates new tracks if needed. This is performed by the global nearest neighbor (GNN) tracking algorithm [29], with constant velocity target motion model. The core of the GNN tracking is the extended Kalman filter for the prediction and update step. Compared to other MTT algorithms, such as Gaussian-mixture probability hypothesis density (GM-PHD) or Poisson multi-Bernoulli mixture filter (PMBM) [30], [31], [32], GNN tracking algorithms require fewer target and motion models to be built and thus save computation power.

A "feedback loop" between feature-based counting and tracking for counting is also introduced for better estimation of the position and number of grouped people. The counting results from the proposed group classifier help the tracking part decide the confirmed trajectory and tentative trajectory to reduce the influence of false detection (FD) and missed detection. Meanwhile, the tracking for counting part also helps the feature-based counting block deal with the static targets in the scene.

The results when combining tracking for counting block and feature-based counting block are provided in Section IV.

C. Final Post-Processing Block

After the number and location of each group of people are estimated by the people tracking & counting block, they are recorded in chronological order. Then, the post-processing block combines and updates results over time to smooth any inconsistency. Specifically, in this article a five-point median filter is applied to average the estimation. Fig. 8 shows an example for this algorithm in the post-processing block.

IV. Experimental Results and Analysis

A. Experimental Setup and Data Collection

The TI IWR6843ISK radar is chosen to validate the proposed approach for grouped people counting. It is a 60-GHz mmWave radar with three transmitters (Txs) and four receivers (Rxs) forming a virtual 12-channel array. The radar is used with an mmWave ICBOOST board and a DCA1000EVM board enabling to record raw data rather than just point clouds. The radar is configured for outdoor People Counting scenarios, with parameters listed in Table I. The maximum detection range is about 20 m and a field of view (FoV) of 120° is chosen for experiments as shown in Fig. 9. To the best of our knowledge, this experimental region is wider than some existing Radar-based People Counting studies [15], [18]. The experiments were carried out in an outdoor open area of the TU Delft campus.

TABLE I
SETTINGS OF THE IWR6843ISK RADAR FOR THIS WORK

IWR6843ISK	
Carrier frequency	60 GHz
Number of Txs	3
Number of Rxs	4
Azimuth Field of View	+/- 60°
Azimuth Resolution at Boresight	15°
Elevation Angle Field of View	+/- 15°
Elevation Angle Resolution	58°
Configuration	
Maximum Detection Range	20m
Range Resolution	8cm
ADC Samples	256
Chirps per Frame	200
Frame Duration	0.12s
Maximum Doppler Velocity	2.5m/s

TABLE II
COLLECTED DATA SETS (EACH GROUP HAS MAXIMUM FIVE PEOPLE)

	Training Set	**Testing Set**
Dataset I	• Maximum 1 group in the RoI. • Approximately 6 minutes for each number of Grouped People.	• Maximum 1 group in the RoI. • Approximately 1.5 minutes for each number of Grouped People.
Dataset II	No training set needed for tracking for counting block.	• Multiple people in the RoI without grouping, i.e. individuals. • Approximately 30 seconds for the testing data and maximum 4 people in the RoI.
Dataset III	Using the Training Set of Data Set I to ensure training vs testing diversity.	• Multiple groups (more than 1 and max 3 groups) in the RoI. • Approximately 1.5 minutes for each number of people in the RoI (ranged from 1 to 6).

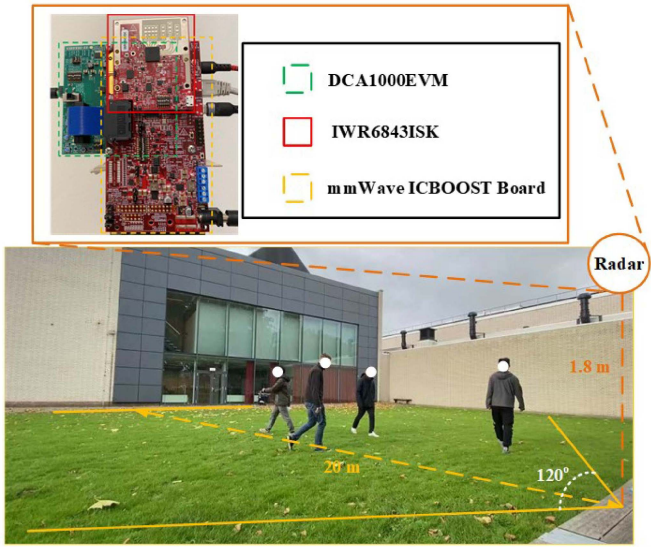


Fig. 9. Radar equipment and scene for the experimental setup to validate the proposed grouped people counting pipeline.

TABLE III
PHYSICAL CHARACTERISTICS OF PARTICIPANTS IN THE DATA SETS (M DENOTES MALE, AND F DENOTES FEMALE PARTICIPANTS)

Individual	A	B	C	D
Gender	M	M	M	F
Height (m)	1.70	1.71	1.90	1.82
Weight (kg)	69	72	85	62
Individual	E	F	G	H
Gender	M	M	M	F
Height (m)	1.69	1.72	1.75	1.66
Weight (kg)	70	68	72	50
Individual	I	J	K	L
Gender	F	M	M	M
Height (m)	1.65	1.77	1.80	1.78
Weight (kg)	45	75	75	68
Individual	M	N	O	/
Gender	F	F	M	/
Height (m)	1.67	1.73	1.84	/
Weight (kg)	48	52	83	/

Three different data sets are collected as listed in Table II, with the aim of validating different blocks of the proposed pipeline. *Data set I* is used to test the performance of the feature-based counting block and for comparisons with state-of-the-art methods. *Data set II* is used to evaluate the capability of the tracking for counting block, hence it does not need training data. *Data set III* is used to evaluate performances when combining both feature-based counting and tracking for counting blocks. As the focus of this study is the initial validation of the proposed pipeline to count grouped people, it is assumed that during the data collection the groups will not be mixing, spawning, or be occluded. Mixing happens when multiple target groups share the same tracks at a certain time [33]. Spawning (or splitting) occurs when a single extended target group separates into two or more extended targets [34]. Occlusion means that an existing target group is obscured by a new target or obstacle, resulting in missed detections [35]; this occlusion can happen within the same group or between different groups.

Two approaches are followed to increase the robustness of the validation. First, diversity in the body characteristics of the volunteers is ensured: similarly to the experimental design of [15], 15 volunteers were invited to participate in the data collection. As shown in Table III, their heights varied from 1.65 to 1.90 m, and weights from 45 to 85 kg. Second, it was ensured that the volunteers involved in the acquisition of the test set and the training set were not the same, while their acquisitions were carried out at different times.

B. Performance Metrics

Performance metrics that can assess the behavior of both the feature-based counting block and the tracking for counting block of the proposed pipeline are selected. These are the mean-square error (MSE), average probability of true positives (ATPs), and multiple groups tracking accuracy (MGTA). Detailed explanations for these metrics are introduced as follows.

Inspired by Choi et al. [18], MSE can characterize how much a prediction differs from the true number of people in the RoI, hence reflecting the accuracy and precision of the classification process. MSE is defined as

$$\mathcal{L}^{\text{MSE}} = \frac{\sum_k \sum_i \left(\hat{N}_k^i - N_k^i\right)^2}{\sum_k i_{\max}} \tag{3}$$

where $i_{\max}$ represents the maximum number of Groups in a certain window. N_k^i represents the predicted number of people in the ith Group at the kth window, and $\hat{N}_k^i$ is the true number of people, with $\hat{N}_k^i \in \mathbb{N}$ and $N_k^i \in \mathbb{N}$ (set of nonnegative integers).

ATP is used to characterize how well the group classifier in the proposed pipeline works for each trained class. It is defined as

$$\mathcal{L}^{\text{ATP}} = \frac{\sum_{C_\gamma} \text{TP}}{|C_\gamma|} \tag{4}$$

where $|C_\gamma|$ represents the number of trained classes. In this thesis, γ equals 6, and thus C_γ is ranged from 0 to 5 and $C_\gamma \in \mathbb{N}$. TP is the probability of true positives from the confusion matrix. Since each class of the training set used in this work is balanced, ATP is a suitable metric for classification performances.

Finally, MGTA is proposed to study the performance of the tracking for counting block. In the general case of multiple targets, multiple objects tracking accuracy (MOTA) [36] is used to study missed/FDs and mismatches of tracking methods. In this work, grouping is introduced, and a single identity (ID) is used to represent a whole group instead of using different identity switches (IDSs) to represent people in a group. It is defined as

$$\mathcal{L}^{\text{MGTA}} = \frac{\sum_k (\text{MD} + \text{FD} + \text{IDS})}{\sum_k N_t^{\text{total}}} \tag{5}$$

where N_t^{total} is the total number of people in the RoI during each window. $MD \in \mathbb{N}$ is the number of missed detections. $FD \in \mathbb{N}$ is the number of FDs. $IDS \in \mathbb{N}$ is the number of identity switch events. The identity switch (or mismatch) relates to the tracks' IDs of each group, and it could affect accuracy when information from different groups is passed to the Group classifier in the proposed pipeline. The ground truth is obtained from a simultaneous camera recording during the experiment. The values of FD and IDS are obtained by comparing the difference between the true location of people (i.e., the ground truth) and the estimated location of people at each decision window.

C. Case 1: Grouped People Counting in a Single Group

For this case, only the feature-based counting block of the proposed pipeline is used to estimate the total number of people in the RoI based on the data of *Data set I*. This contains approximately 6 min for each number of Grouped People for training the classifier, and 1.5 min for testing. Two aspects are explored: first the type of statistical classifier that provides the best results, and second the selection of the most suitable features to improve performances, out of those discussed in the previous section.

Four common statistical classifiers are used and compared. They are: 1) Naïve Bayes; 2) Random Forest; 3) support vector machine (SVM); and 4) k-nearest Neighbors (KNNs) classifiers. The parameters of each classifier are determined by a hyperparameter optimization [37]. The Naïve Bayes was created with Gaussian Kernel, and the probabilistic model is selected to fit the Gaussian distribution. The Random Forest was created with maximum 496 splits and 30 learners. The SVM classifier used a Gaussian kernel with scale equal to 0.75. The number of neighbors when applying KNN was set to 10. Table IV provides the results when using the Data set I, and the MSE and ATP metrics are used for performance comparison, as this case study does not include the tracking block, hence no MGTA.

TABLE IV
RESULTS OF DATA SET I WHEN APPLYING THE FEATURE-BASED COUNTING BLOCK WITH DIFFERENT CLASSIFIERS (RA DENOTES FEATURES FROM THE RANGE–AZIMUTH DOMAIN, CVD THOSE FROM THE CVD MAPS)

Input Features	Method	MSE ($*10^{-2}$)	ATP (%)
RA + CVD	Naïve Bayes	14.87	66.82
RA + CVD	Random Forest	1.53	89.10
RA + CVD	SVM	0.58	94.32
RA + CVD	KNN	1.35	90.95

From the values of MSE and ATP, it is shown that SVM provides both the smallest MSE and the largest ATP, hence it is the best of the four tested classifiers. Random Forest and KNN have similar performance, while the Naïve Bayes performs the worst. This may be due to its assumption of conditional independence of features [38], i.e., each feature is assumed to affect independently the classification results whereas there is still significant correlation between the features used here.

Furthermore, a study of the feature selection for the proposed feature-based counting method is provided. This is to characterize whether features from both the range–azimuth domain and CVD map are valuable in solving the Grouped People Counting problem. As summarized in Section II-B, Grouped People are distinguishable in the range–azimuth domain when they are located at range bins near the radar, where different numbers of grouped people occupy distinct azimuth bins. However, groups of people with similar number (e.g., three or five people walking as a group) occupy a similar number of azimuth bins in the far range, here empirically defined as approximately beyond 12 m, based on the model of grouped people [21]. To overcome this, features based on the CVD maps were introduced. Here, the classification is performed using an SVM classifier, based on the results in Table IV, for samples of Data set I where the groups of people were moving in the far range (above 12 m).

From Table V, when the CVD and Range–Azimuth maps are fused, MSE and ATP are improved and the fused results are better than in the case where one type of features is used in isolation. Moreover, the MSE is really large when only CVD is used to classify samples collected at far range. This is due to the fact that CVD maps (and spectrograms from which they are derived) do not include any information related to the distance and azimuth of the targets. This can generate confusions between smaller groups of people moving near the radar and larger groups moving in the comparatively far range. To summarize, two general conclusions can be drawn. First, the features extracted from the range–azimuth domain are promising to classify the number of Grouped People in the near range,

TABLE V
FEATURE SELECTION STUDY FOR THE PROPOSED SVM CLASSIFIER. (RA DENOTES FEATURES FROM THE RANGE–AZIMUTH DOMAIN AND CVD FEATURES FROM THE CVD MAP; "FAR" DENOTES SAMPLES OF DATA SET I CAPTURED IN THE FAR RANGE $> 12\,m$)

Input Features	Method	MSE ($*10^{-2}$)	ATP (%)
RA (Far)	SVM	1.29	85.95
CVD (Far)	SVM	15.78	61.17
RA + CVD (Far)	SVM	1.14	89.58

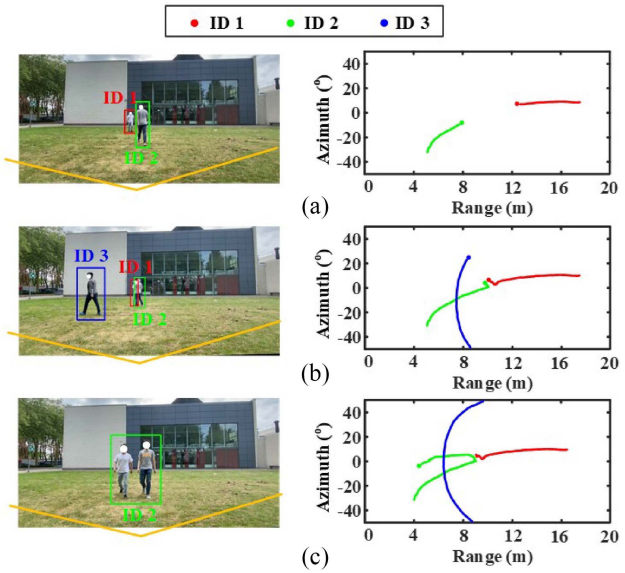


Fig. 10. Results when applying the proposed tracking for counting block. The dot in the track is the instant corresponding to the picture provided as a ground truth (note that the orientation/geometry of picture and track plot are rotated). (a) Tracks for two separate individuals in the RoI. (b) Tracks for two individuals coming closer and one individual walking across in the RoI. (c) Tracks for two individuals walking together as a group in the RoI.

while in the farther range those features can be complemented by CVD based features. Second, the features from the CVD map can improve the group counting accuracy in the farther range, but only if they are combined with location information (i.e., range and azimuth) and not on their own.

D. Case 2: Multiple Individuals Tracking

In this section, the performance of the tracking for counting block within the proposed pipeline is tested. Data set II is used where multiple individuals exist in the RoI, i.e., they are not grouped together while walking (hence, the feature-based counting block tested in Case 1 is not used here).

Fig. 10 demonstrates with an example the robustness of the performances when applying the proposed tracking for counting block. Fig. 10(a) is the moment when two individuals are walking toward each other. From the estimated tracks, it can be clearly seen that there are two tracks, and they have a trend of moving toward each other. In Fig. 10(b), the two individuals finally meet, and another individual crossed the scene in front of them. Although partial occlusion happens for the targets denoted by ID 1 and ID 2, the tracking algorithm still holds their tracks. At this moment, three tracks are clearly shown in the Range–Azimuth map. In Fig. 10(c), ID 3 walked out of the RoI, and ID 1 and ID 2 walk together. From the Range–Azimuth map, there is only one target existing in the RoI, and ID 1 is merged into ID 2 based on the fact that ID 1 and ID 2 are now walking as a group.

TABLE VI
SUMMARY OF THE RESULTS OF THE PROPOSED TRACKING FOR COUNTING BLOCK. MD_{TOTAL}, FD_{TOTAL}, AND $\text{IDS}_{\text{TOTAL}}$ REPRESENT THE SUM OF THE *MD*, FD, AND IDS EVENTS ACROSS ALL DECISION WINDOWS

The Number of People	MD_{total}	FD_{total}	IDS_{total}	MGTA (%)
1	1	0	0	2.5
2	0	2	0	2.5
3	2	0	0	1.7
4	1	2	1	2.5
Total	4	4	1	1.5

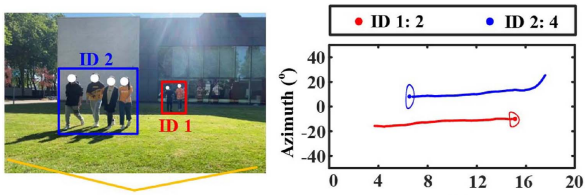
This example shows that the proposed tracking for counting block performs well even in case of events such as partial occlusion and merging of two individuals. Even if by using suitable clustering and prior knowledge of the two separate tracks the algorithm could distinguish target ID 1 and ID 2, they are instead clustered together when they move as a single group. This will ensure that this tracking block will pass the correct information to the feature-based counting block in the overall proposed pipeline in order to classify the number of people in each group.

Table VI provides a summary of the results for the tracking block when processing Data set II. The total number of the three possible errors and the MGTA metric are provided, counted over 40 decision windows for each number of people in the scene. The performance is good with overall MGTA less than 5%. Two conclusions are drawn from the results. First, as the number of people in the scene increases, the probability of MD and FD errors increases, as expected. Second, there are few IDS errors even with more people in the scene, which means that the assignment algorithm within the tracking block works well avoiding switching between groups.

E. Case 3: People Counting in Multiple Groups

After analyzing the results of the feature-based counting block and tracking for counting block separately, the result of the proposed pipeline combining them both is provided. Data set III is used in this section for the testing stage, where there are multiple groups of people in the RoI, whereas the training data for the feature-based classifier block comes from Data set I. An example of ground truth and results is shown in Fig. 11. There are two groups walking in the RoI, and blue and red lines represent the tracks of each group. As mentioned in the description of the experimental setup, there is no further grouping, mixing, or spawning events happening for a given group. Hence, the estimated tracks in Fig. 11(b) do not cross each other. Meanwhile, the predicted number of people in each group is equal to the ground truth.

The overall MSE when applying the proposed method is 0.0013, and the ATP is 96.25%, which are even better values than those obtained in Table IV for the simpler scenarios of Data set I (0.0058 MSE and 94.32% ATP, respectively). This good result is supported by the combined effect of the feature-based counting block and the tracking for counting block, with the additional effect of the median filter at the post-processing

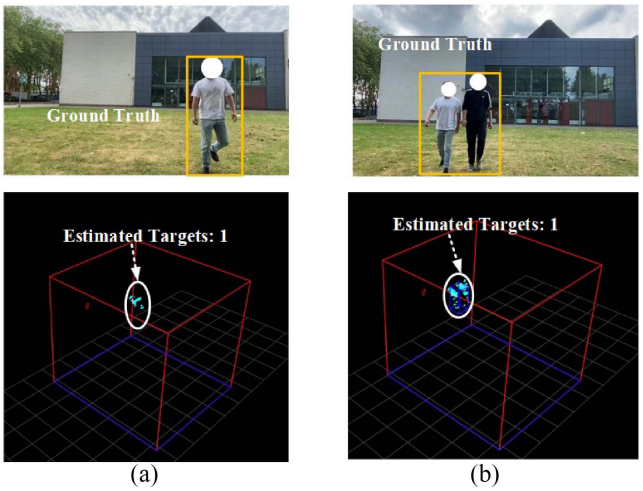


(a) (b)

Fig. 11. Example of results for the proposed pipeline. (a) Ground truth where two groups and six people in total are present in the RoI. (b) Estimated tracks of two groups and predicted number of people in each group.

TABLE VII
SUMMARY OF THE RESULTS FOR THE TI 3-D PEOPLE COUNTING. MD_{total}, FD_{total}, AND IDS_{total} REPRESENT THE SUM OF THE *MD*, FD, AND IDS EVENTS ACROSS ALL DECISION WINDOWS

	New Dataset I	New Dataset II
MD_{total}	65	29
FD_{total}	7	16
IDS_{total}	0	4
MGTA (%)	60.0	40.8



(a) (b)

Fig. 12. Visual ground truth and TI 3-D people counting visualizer results when (a) 1 person and (b) 2 people walking as a group are present.

stage (described in Fig. 8). The MGTA metric is 1.67%, which is an acceptable value. Among the three defined types of errors, almost no IDSs occur. Equally, MDs did not occur continuously on entire tracks. This means that the features of Grouped People could be passed from the tracking block to the Group classifier successfully. Moreover, although FD events may occur, the proposed group classifier retains a high accuracy in classifying zero targets in the group, i.e., rejecting FDs that the tracking part may have picked up.

Additionally, it is worth highlighting that Data set III includes cases where the total number of people in the RoI is greater than the maximum number of expected people per group, i.e., the number for which the feature-based counting classifier has been trained for. In a general classification problem, it is not possible for a classifier trained with conventional supervised learning approaches to output classes for which no training examples were provided. Specifically, in Data set III, the training set for the classifier within the proposed pipeline is Data set I, where only one group exists in the RoI with a maximum number of people equal to five. Thus, for the group classifier block, the expected outputs can only range from 0 to 5. However, thanks to the combination of the tracking block with this classifier bloc, it is possible to deal with situations when the total number of people in the RoI is more than five, i.e., higher than the number the pipeline has been originally trained for. This capability of "classification beyond training" is an advantage of the proposed pipeline compared to simply using a classifier in isolation.

V. PERFORMANCE COMPARISON

A. Comparisons With Existing People Counting Product

In this section, an existing People Counting commercial product is compared with our pipeline, namely, the 3-D People Counting demonstration lab from the TI Industrial Toolbox [39]. When their people counting method is applied, moving targets are continuously tracked until they leave the scene. This toolbox applies a "group tracker" approach to track individuals in the scene, essentially tracking a group of point clouds rather than a group of people. After applying the Range FFT, the minimum variance distortionless response (MVDR) beamforming is applied to estimate the angle. Finally, the detected points are determined by the Doppler FFT where the maximum amplitude in the power spectrum is selected.

To establish a comparison, a repetition of the experiments performed for the collection of Data set I and II was carried out. It was not possible to use exactly the same data, as the radar does not record raw data but only point clouds when this 3-D People Counting demo lab is activated. Fig. 12 shows an example of the ground truth and visualizer output of the TI counting demo lab for two scenarios. In Fig. 12(a), there is one individual walking toward the radar. The visualizer showed that the target was tracked, and the estimated target number was 1, which is correct. However, when two people walked as a group, a missed detection occurred whereby only one person was detected in the cluster and the number of estimated people in the RoI was incorrectly set at 1, as shown in Fig. 12(b).

The performance metrics for the TI 3-D People Counting approach are reported in Table VII. Here, the expressions "New Data set I and II" are used to indicate that the data are similar to the Data set I and II described in Table II, but not exactly the same. The New Data set I focuses on groups of people and contains groups with numbers ranging from 1 to 3 participants; the New Data set II focuses on single individuals to be tracked and contains up to 3 individuals walking in the RoI. Both new data sets includes data for 40 decision windows for each number of people in the scene. The overall MGTA metric is higher than 40% for each data set, much worse than the MGTA obtained by our proposed pipeline in the case studies analyzed in the previous section. Particularly significant is the number of MD events for the New Data set I which is focused on the grouping case. This happens since the TI toolbox does not have in its localization and tracking algorithm a block to deal with the grouped people phenomenon. Even if in the TI 3-D People Counting implementation guide, a "Group tracker" is mentioned, this refers to only tracking an extended

target, i.e., a group of detections, without attempting to estimate how many distinct participants belong to each group. Additionally, the TI toolbox under-performs when multiple people enter the RoI at the same time, even without mixing or grouping behavior. Managing this situation would require the tracking algorithms to predict and update multiple targets' birth at the same time; the difficulty in this case is reflected by the high number of FD events in Table VII.

B. Comparisons With Alternative People Counting Methods

In this section, four radar-based people counting methods from the literature are compared to our proposed pipeline, namely, two methods based on a statistical classifier, and two methods based on NNs. As these methods do not use tracking, Data set I is used to compare their performances as its focus is on distinguishing the number of people in the only group present in the RoI. The results are reported in Table VIII.

The first alternative method [17] extracts features from the range profile for clustering, and then with the help of a probability density function (PDF), the clustered major amplitudes with their distance information are associated to the number of people in the scene. As in Table VIII, the resulting ATP is 50.15%, meaning that there is a high probability of estimating the wrong number of people in the RoI. This suggests that the maximum amplitude extracted from the range profile is not enough for reliable people counting. The MSE is 0.2281, suggesting that the estimated values are scattered and not very precise. This could be due to the fact that the PDFs for different classes overlap a lot with each other, leading to a not so robust classification performance in case of different people in a single group.

The second alternative method was proposed in [15]. This method fused features of the range profile and features of the Doppler domain to perform People Counting. As in Table VIII, the ATP is 66.88%. Although the overall performance is still not good for addressing Grouped People Counting, it is better than the previous method that only uses manually extracted features from the range profile. This proves the importance of using multidimensional features for the Grouped People Counting problem. The MSE is 0.0687, which is much smaller than that of the previous method [17], but still possible to improve. For example, it is hypothesized that a limitation of this method [15] is the extraction of features from separate range and Doppler domains, rather than doing this jointly.

The first NN-based method for comparison is proposed in [16]. It used the range–time matrix for classification with the help of ResNet-14. As in Table VIII the ATP is 60.42%, with a minimum probability of true positives equal to 40.2%. This indicates that using only features from the range–time domain is not enough to solve the People Counting problem. The MSE is 0.0798, which means that the predicted result is not precise enough. As an additional drawback, the proposed network was rather deep, requiring a significant computational training load to achieve the aforementioned results.

The second NN-based People Counting method that is reimplemented for performance comparison uses CNN and LSTM networks [18]. As shown in Table VIII, even this method could not address the Grouped People Counting problem, with the reported ATP equal to 63.48% and the MSE to 0.1240. By comparing the two NN-based methods, it is found that both methods have comparable ATP performances using features from the range profile or range time maps, but the second method [18], which combined CNN and LSTM, performed slightly better. This is attributed to the self-attention Bi-LSTM network which is used in such method to focus and learn frame-by-frame information, which is instead implicit if only range–time maps are used as in [16].

TABLE VIII
PERFORMANCE COMPARISON BETWEEN THE PROPOSED APPROACHES (TOP 2 ROWS) AND IMPLEMENTATION OF ALTERNATIVE METHODS FROM THE LITERATURE. (RA DENOTES FEATURES FROM THE RANGE–ANGLE DOMAIN, CVD FEATURES FROM THE CVD MAP)

Input Features	Method	MSE ($*10^{-2}$)	ATP (%)
RA + CVD	Proposed Group Classifier	0.58	94.32
RA + CVD	Proposed Pipeline	0.26	98.42
Range Profile	Statistical Classifier [17]	22.81	50.15
Range+Dop	Statistical Classifier [15]	6.87	66.88
Range Time	NN Classifier [16]	7.98	60.42
Range Profile	NN Classifier [18]	12.40	63.48

It should be noted that the performance results reported in the original papers for the alternative methods [15], [16], [17], [18] were higher than those achieved here with their reimplementation. This is caused by the fact that the methods were tested with grouped people data, hence a more challenging scenario that appears to be too complex for such methods, but that can be addressed by our proposed pipeline.

Comparing the four methods with each other, it is interesting to observe that the best results are achieved by the statistical classifier using combined range and Doppler features [15], which appears to also outperform the NN-based methods. This can be attributed to the usage of features from multiple, combined radar domains (i.e., range and Doppler) rather than features from a single domain, and reinforces the idea in our proposed pipeline to combine features from range–angle plots and CVD plots. Compared with the alternative methods reimplemented from the literature, our proposed pipeline achieves much better results to address the grouped people counting problem. The ATP using the proposed group classifier alone is higher than 94%, and the MSE is two orders of magnitude better than the alternative methods. Furthermore, the ATP of the full proposed pipeline (group classifier & tracking block combined) is even above 98%.

VI. CONCLUSION

This article proposes a radar-based people counting pipeline that can deal with the complex situation of multiple individuals moving together as a single group. The proposed pipeline is based on the complementary combination of a tracking algorithm and a feature-based classifier that estimates the number of people of each group tracked in the scene of interest. It is shown that a combination of features extracted from the range–azimuth domain and CVD maps of a 60 GHz MIMO FMCW radar provides robust results. The proposed pipeline has been experimentally validated with diverse data sets collected in an outdoor environment at the TU Delft

campus, showing good results and outperforming alternative methods from the literature as well as the 3-D People Counting toolbox provided by Texas Instrument.

ACKNOWLEDGMENT

The authors are grateful to the volunteers who participated to the data collection.

REFERENCES

[1] X. Huang, Z. Yang, and C. Chen, "Regional people counting approach based on multi-motion states models using MIMO radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[2] N. Cahyadi and B. Rahardjo, "Literature review of people counting," in *Proc. Int. Conf. Artif. Intell. Mechatronics Syst. (AIMS)*, 2021, pp. 1–6.
[3] Z. Yang and X. Huang, "Cascaded regional people counting approach based on two-dimensional spatial attribute features using MIMO radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[4] C. Tang, W. Li, S. Vishwakarma, K. Chetty, S. Julier, and K. Woodbridge, "Occupancy detection and people counting using WiFi passive radar," in *Proc. IEEE Radar Conf. (RadarConf)*, 2022, pp. 1–6.
[5] J. W. Choi, X. Quan, and S. H. Cho, "Bi-directional passing people counting system based on IR-UWB radar sensors," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 512–522, Apr. 2018.
[6] Q. Guan et al., "Epidemiological investigation of a family clustering of COVID-19," *Zhonghua Liu Xing Bing Xue Za Zhi*, vol. 41, no. 5, pp. 629–633, May 2020.
[7] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.
[8] C. Wu, Z. Yang, and C. Xiao, "Automatic radio map adaptation for indoor localization using smartphones," *IEEE Trans. Mobile Comput.*, vol. 17, no. 3, pp. 517–528, Mar. 2018.
[9] "Smart building technology." InfraRed Integrated Systems Ltd. 2022. [Online]. Available: https://www.trueoccupancy.com/occupancy-sensing-technology
[10] I. Ahmed, M. Ahmed, J. J. P. C. Rodrigues, G. Jeon, and S. Din, "A deep learning-based social distance monitoring framework for COVID-19," *Sustain. Cities Soc.*, vol. 65, Feb. 2021, Art. no. 102571.
[11] Z. Peng and C. Li, "Portable microwave radar systems for short-range localization and life tracking: A review," *Sensors*, vol. 19, no. 5, p. 1136, 2019.
[12] X. Chen, I. Vizzo, T. Läbe, J. Behley, and C. Stachniss, "Range image-based LiDAR localization for autonomous vehicles," 2021, *arXiv:2105.12121*.
[13] S. Ansari and S. Salankar, "An overview on thermal image processing," in *Proc. RICE*, 2017, pp. 117–120.
[14] R. Bao and Z. Yang, "CNN-based regional people counting algorithm exploiting multi-scale range-time maps with an IR-UWB radar," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13704–13713, Jun. 2021.
[15] J.-H. Choi, J.-E. Kim, and K.-T. Kim, "People counting using IR-UWB radar sensor in a wide area," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5806–5821, Apr. 2021.
[16] Y. Jia et al., "ResNet-based counting algorithm for moving targets in through-the-wall radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, pp. 1034–1038, 2021.
[17] J. W. Choi, D. H. Yim, and S. H. Cho, "People counting based on an IR-UWB radar sensor," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5717–5727, Sep. 2017.
[18] J.-H. Choi, J.-E. Kim, and K.-T. Kim, "Deep learning approach for radar-based people counting," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7715–7730, May 2022.
[19] A. Ninos, J. Hasch, M. Heizmann, and T. Zwick, "Radar-based robust people tracking and consumer applications," *IEEE Sensors J.*, vol. 22, no. 4, pp. 3726–3735, Feb. 2022.
[20] V.-H. Nguyen and J.-Y. Pyun, "Location detection and tracking of moving targets by a 2D IR-UWB radar system," *Sensors*, vol. 15, no. 3, pp. 6740–6762, 2015.
[21] L. Ren, "People counting using low-cost FMCW MIMO radar: Achieving tracking for counting and classification of groups of people using FMCW radar," Dept. Electr. Eng., M.S. thesis, Delft Univ. Technol., 2022. [Online]. Available: http://resolver.tudelft.nl/uuid:a7450fad-43ff-446e-ba8e-d7a10fc50029
[22] P. Van Dorp and F. C. A. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar Sonar Navig.*, vol. 2, no. 2, pp. 135–145, 2008.
[23] A. Ghaleb, L. Vignaud, and J. M. Nicolas, "Micro-Doppler analysis of wheels and pedestrians in ISAR imaging," *IET Signal Process.*, vol. 2, no. 3, pp. 301–311, 2008.
[24] J. A. Balderrama, F. J. Masters, and K. R. Gurley, "Peak factor estimation in hurricane surface winds," *J. Wind Eng. Ind. Aerodyn.*, vol. 102, pp. 1–13, Mar. 2012.
[25] C. Kuang, C. Wang, B. Wen, Y. Hou, and Y. Lai, "An improved CA-CFAR method for ship target detection in strong clutter using UHF radar," *IEEE Signal Process. Lett.*, vol. 27, pp. 1445–1449, Aug. 2020, doi: 10.1109/LSP.2020.3015682.
[26] D. Kellner, J. Klappstein, and K. Dietmayer, "Grid-based DBSCAN for clustering extended objects in radar data," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 365–370.
[27] A. Sinha, Z. Ding, T. Kirubarajan, and M. Farooq, "Track quality based multitarget tracking approach for global nearest-neighbor association," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1179–1191, Apr. 2012.
[28] R. Ricci and A. Balleri, "Recognition of humans based on radar micro-Doppler shape spectrum features," *IET Radar Sonar Navig.*, vol. 9, no. 9, pp. 1216–1223, 2015.
[29] P. D. Konstantinova, A. Udvarev, and T. Semerdjiev, "A study of a target tracking algorithm using global nearest neighbor approach," in *Proc. Compsystech*, vol. 3, 2003, pp. 290–295.
[30] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 5–18, Jan. 2004.
[31] D. E. Clark, K. Panta, and B.-N. Vo, "The GM-PHD filter multiple target tracker," in *Proc. 9th Int. Conf. Inf. Fusion*, 2006, pp. 1–8.
[32] K. Granström, M. Fatemi, and L. Svensson, "Poisson multi-bernoulli mixture conjugate prior for multiple extended target filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 1, pp. 208–225, Feb. 2020.
[33] K. Granstrom, C. Lundquist, and O. Orguner, "Extended target tracking using a Gaussian-mixture PHD filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3268–3286, Oct. 2012.
[34] K. Granstrom and U. Orguner, "On spawning and combination of extended/group targets modeled with random matrices," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 678–692, Feb. 2013.
[35] K. Dai, Y. Wang, J.-S. Hu, K. Nam, and C. Yin, "Intertarget occlusion handling in multiextended target tracking based on labeled multi-Bernoulli filter using laser range finder," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 4, pp. 1719–1728, Aug. 2020.
[36] K. Bernardin, A. Elbs, and R. Stiefelhagen, "Multiple object tracking performance metrics and evaluation in a smart room environment," in *Proc. 6th IEEE Int. Workshop Vis. Surveillance*, 2006, pp. 11–15.
[37] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2013, pp. 847–855.
[38] K. M. Ting and Z. Zheng, "A study of AdaBoost with naive Bayesian classifiers: Weakness and improvement," *Comput. Intell.*, vol. 19, no. 2, pp. 186–200, 2003.
[39] "People counting demonstration using TI mmWave sensors." TI Document. 2021. [Online]. Available: https://training.ti.com/people-countin%20demonstration-using-ti-mmwavesensors

Liyuan Ren was born in Xichang, Sichuan, China. He received the first B.Eng. degree in electronic information engineering from the University of Electronic Science and Technology of China, Chengdu, China, and the second B.Eng. degree (Hons.) in electrical and electronics engineering from the University of Glasgow, Glasgow, U.K., in 2020. He is currently pursuing the M.Sc. degree in signal and systems with Delft University of Technology, Delft, The Netherlands.

His research interests include high performance computing, signal and system, millimeter-wave radar, and multisensor navigation.

Alexander G. Yarovoy (Fellow, IEEE) received the Diploma degree (Hons.) in radiophysics and electronics, and the Candidate Phys. & Math. Sci. and Doctor Phys. & Math. Sci. degrees in radiophysics from Kharkov State University, Kharkiv, Ukraine, in 1984, 1987, and 1994, respectively.

In 1987, he joined the Department of Radiophysics, Kharkov State University as a Researcher and became a Full Professor in 1997. From September 1994 to 1996, he was with the Technical University of Ilmenau, Ilmenau, Germany, as a Visiting Researcher. Since 1999, he has been with Delft University of Technology, Delft, The Netherlands, where he leads there as a Chair of Microwave Sensing, Systems and Signals, since 2009. He has authored and coauthored more than 450 scientific or technical papers, six patents and 14 book chapters. His main research interests are in high-resolution radar, microwave imaging, and applied electromagnetics (in particular, UWB antennas).

Prof. Yarovoy is the recipient of the European Microwave Week Radar Award for the paper that best advances the state-of-the-art in radar technology in 2001 (together with L. P. Ligthart and P. van Genderen) and in 2012 (together with T. Savelyev). In 2010 together with D. Caratelli, he got the Best Paper Award of the Applied Computational Electromagnetic Society. He served as the General TPC chair of the 2020 European Microwave Week (EuMW'20), as the Chair and TPC Chair of the 5th European Radar Conference (EuRAD'08), as well as the Secretary of the 1st European Radar Conference (EuRAD'04). He served also as the Co-Chair and TPC Chair of the tenth International Conference on GPR (GPR2004). He served as an Associated Editor for the *International Journal of Microwave and Wireless Technologies* from 2011 to 2018 and as a Guest Editor of five special issues of the IEEE TRANSACTIONS and other journals. From 2008 to 2017, he served as the Director of the European Microwave Association.

Francesco Fioranelli (Senior Member, IEEE) received the Laurea (B.Eng. cum laude) and Laurea Specialistica (M.Eng. cum laude) degrees in telecommunication engineering from the Università Politecnica delle Marche, Ancona, Italy, in 2007 and 2010, respectively, and the Ph.D. degree from Durham University, Durham, U.K., in 2014.

He is currently an Associate Professor with Delft University of Technology, Delft, The Netherlands, and was an Assistant Professor with the University of Glasgow, Glasgow, U.K., from 2016 to 2019, and a Research Associate with University College London, London, U.K., from 2014 to 2016. He has authored over 140 publications between book chapters, journal, and conference papers, edited the books on *Micro-Doppler Radar and Its Applications* and *Radar Countermeasures for Unmanned Aerial Vehicles* (IET-Scitech), in 2020. His research interests include the development of radar systems and automatic classification for human signatures analysis in healthcare and security, drones and UAVs detection and classification, automotive radar, wind farm, and sea clutter.

Dr. Fioranelli received three best paper awards.