People Counting Using Low-cost FMCW MIMO Radar

Achieving tracking for counting and classification of groups of people using FMCW radar

PEOPLE COUNTING USING LOW-COST FMCW MIMO RADAR

ACHIEVING TRACKING FOR COUNTING AND CLASSIFICATION OF GROUPS OF PEOPLE USING FMCW RADAR

THESIS

to obtain the degree of Master of Science
in Electrical Engineering
at Delft University of Technology,
to be defended publicly on Monday, August 29, 2022 at 10:00 AM

by

Liyuan REN

born in Sichuan, China.

This thesis has been approved by the

Supervisor: Prof. Dr. Alexander Yarovoy
Daily supervisor: Dr. Francesco Fioranelli

Thesis Committee:

Prof. Dr. Alexander Yarovoy	MS3 TU Delft
Dr. Francesco Fioranelli	MS3 TU Delft
Dr. Marco Zuniga	ENS TU Delft

Keywords: People Counting, Automotive Radar, Grouping, Far Range

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

ABSTRACT

For the development of automatic People Counting systems, radar is increasingly becoming a popular technology because of the increasingly stringent privacy requirements for people demographic information and the requirement to operate in a challenging environment. Because of the complexity of multi-target movement and the diversity of application scenarios, Radar-based People Counting methods are required to have sufficient robustness. However, based on the review of the current literature, the grouping phenomenon (i.e., multiple individuals moving close together as a single group) was not often considered in the experimental scenarios.

This thesis aims to study one of the most complex motions of individuals, grouping, and address the People Counting problem more in general, including the cases of grouping of multiple individuals. After studying the characteristics of Group People (defined as a group of people sharing neighboring, adjacent locations and moving together), with the help of multiple-input-multiple-output (MIMO) frequency-modulated continuous wave (FMCW) Radar, the combination of the Range-Azimuth map and spectrogram/cadence velocity diagram (CVD) is proposed to solve Group People Counting.

Algorithm-wise, there are two categories of existing Radar-based People Counting methods, namely, tracking for counting methods and feature-based counting methods. It was found that these two categories of methods have complementary strengths. Thus, the proposed method combines the tracking for counting approach and feature-based counting approach into a new processing pipeline to estimate the number of people in each group in the scene while tracking each group. Based on it, the proposed method achieves "Beyond classification", which is the output the unlabeled classes not defined at the training stage. Moreover, compared with other state-of-the-art (SOTA) Radar-based People Counting methods, the proposed method outperforms them, and thus it is proved that the grouping problem can be solved in the Radar-based People Counting field.

ACKNOWLEDGEMENTS

At the time of writing this paragraph, it has been two full years that I have been living in Delft. Two years of graduate school have left me with a rusty bike chain. If I go back two years ago, to the plane ride from Chongqing to Amsterdam, I cannot believe that studying and living at Delft University of Technology was so happy and meaningful.

In my second year, I attended the graduation defense of many of my seniors and classmates. I still remember the scene at the airport, at the train station, when I saw them off. I cannot imagine that it is time for me to say goodbye to my graduate career and to Delft. There are so many people I would like to thank that I could write pages, and I hope you read until your name appears.

First, I would like to express my sincere gratitude to my advisor, Professor Alexander Yarovoy. I still remember when I was in the UWB course, Prof. Yarovoy was answering emails from students at 11:00 pm. In the weekly group meetings, Prof. Yarovoy was happy to share with us graduate students how to be a professional researcher and how to be a professional Radar researcher. He emphasized the importance of the presentation. I am grateful to Prof. Yarovoy for providing constructive comments on the general direction of my research. I also remember that during the midterm defense, the professor said that if the topic was very challenging, the whole group would give advice and share that risk. This reassured me to challenge this path with unpredictable results. I am also grateful to Prof. Yarovoy for telling me in general what places in the Netherlands are interesting and to enjoy April in the Netherlands. It showed me that in addition to studying, there is also life.

I sincerely thank my daily tutor Dr. Francesco Fioranelli for his guidance. From the eighth to the twentieth floor, the first place I spent the most time at EWI College was Room J. The second place must be the office where I had my meetings with Francesco. I thank him for supporting me in choosing this challenging topic. Every week of meeting, I was able to get professional guidance on the topic. Francesco was always very prompt in responding to my messages, and even when he was on vacation, he still checked his emails. I admire Francesco's ability to understand and summarize, always able to offer a lot of details and constructive comments in my academic writing. I thank him for his tireless revision of my drafts of manuscripts and for encouraging me. I remember we discussed the old Chinese saying that a strict teacher makes a good student. Thanks to Dr. Francesco's professionalism, which has benefited me greatly, I hope I can be a student that would make Dr. Francesco proud. I thank him and hope to have the opportunity to work together afterwards as well.

I would also like to thank all the professors, PhD, and master students in the Microwave Sensing, Signals and Systems (MS3) group, who together have created a harmonious and dynamic atmosphere for MS3. I hope the taste of the hotpot together in my house in 2021 did not disappoint :D. Thanks to Dr. Hans and Dr. Oleg, for their professional lectures and for giving me the information of FMCW Radar to read when using

the FMCW MIMO Radar. Thanks to Sen, Simin, Ignacio, Ronny, Yubin, for their tireless discussions with me and for always coming to our graduate office to ask if we had any problems with our research. I will also remember the happy memories of our outdoor recreation together. Thanks to all the members of "The Breakdown Club", I had a lot of fun on the 20th floor because of you. Thanks to Xingzhuo, Max, and Lucas who worked together in the office, without you, I would not be able to hear the bell when the EWI building is ready to close every day. Thanks to Zhongyuan, Chunchun, and Yongdian, the days of discussing academics and finding food in the Netherlands together are worth remembering.

Also, I sincerely thank the 15 volunteers who participated in my project, without them, I would not have been able to finish my project. Thanks to them for coming to EWI and participating in my experiment, rain or shine. I would like to thank professors from the Embedded and Networked System group for giving me a deep understanding of the value of the field of People Counting.

Also, I would like to thank the friends I met in the Netherlands. I would like to thank Xi for the care and attention that I received during my graduation. Thanks to my friends Chenyao, Chengming, Chenghao, Anxiang, Keyan, Haozheng, Beihong, Yixuan, Zhiheng, Xiao, Zhou, Willy, who I met after living together in Roland Holstlaan and became friends. Yaxuan, Qingru, Qie, Da, Ran, Fanhao, Xiaoyu, Xiong, Min, and Niels (names in no particular order). Thanks to them for their continuous encouragement and companionship during the Covid-19 period. I still remember the excitement of skydiving together in Interlaken, the excitement of seeing the aurora borealis together in Norways, taking a picture together on the Eiffel Tower in Paris, marveling at Gaudi's architecture together, and the happiness of celebrating each other's birthdays together. I wish you all the best in your future, cheers. Thanks to my friends Heqi, Junyan, Zhexin, Duiquan (in no particular order) who I met before coming to the Netherlands, and thanks to our first victory steak together in the Netherlands, I will remember the memories we made in Europe, in the Netherlands, and in Delft one after another. I would like to thank my friends, Jinglin, Renjie, Hanyun, Wenyu, Yidi, Songlei, Shangshang, Meiruo (in no particular order), whom I met at home before coming to the Netherlands, for the delicious Chinese food we made every time we had dinner together to warm my stomach. I express a special thank you to Prof. Zhiliang, Feiyang, Zuopu, Jianguo (in no particular order) in China for their inspiration and encouragement to continue my research life.

And a special thanks to my family, who always love and support me. Thank you to my parents, my grandparents, grandma and grandpa, for supporting me to study abroad. After living alone in the Netherlands for two years, I felt how hard it was to live on my own. I hope I can be the family that makes you all proud. Salute.

All the people mentioned above have brought lasting happiness to my life in the Netherlands. I wish them all the best and hope they have the best.

Liyuan Ren
Delft, August 2022

CONTENTS

List of Acronyms and Symbols

TI	Texas Instruments
Tx	Transmitter
Rx	Receiver
RoI	Range of Interest
FoV	Field of View
MIMO	Multiple-input-multiple-output
SISO	Single-input-single-output
HVAC	Heating, Ventilation, and Air Conditioning
RF	Radio Frequency
FMCW	Frequency-Modulated Continuous Wave
mmWave	Millimeter Wave
IR-UWB	Impulse Radio Ultra-wideband
SOTA	State-of-the-Art
CNN	Convolutional Neural Network
RCS	Radar Cross-Section
EMA	Exponential Moving Average
CFAR	Constant False Alarm Rate
CA-CFAR	Cell Averaging Constant False Alarm Rate
OS-CFAR	Ordered Statistics Constant False Alarm Rate
CHA-CFAR	Censored Harmonic Averaging Constant False Alarm Rate
DBSCAN	Density-based Spatial Clustering of Applications with Noise
OPTICS	Ordering Points to Identify the Clustering Structure
AoA	Angle of Arrival
KF	Kalman Filter
EKF	Extended Kalman Filter
ANN	Artificial Neural Network
LSTM	Long Short-term Memory

FC	Fully Connected Layer
KNN	*k*-nearest Neighbors Algorithm
SVM	Support Vector Machine
PSD	Power Spectral Density
CVD	Cadence Velocity Diagram
MSE	Mean Square Error
ATP	Average Probability of True Positives
MGTA	Multiple Groups Tracking Accuracy
PDF	Probability Density Function
PCA	Principle Component Analysis
MTT	Multiple Target Tracking

List of Figures

LIST OF TABLES

1

INTRODUCTION

1.1. BACKGROUND AND MOTIVATION

There are many applications for monitoring and counting the number of people in a given environment. First of all, the COVID-19 outbreak is a global emergency and its high spread rate and high mortality rate have caused worldwide disruptions, including more than 400 million confirmed cases and 6 million deaths [1]. To stop COVID-19 or other infectious diseases from spreading, it is necessary for all people to obey the rules of social distance and avoid large gatherings. With regard to sensing information on density / number of people, the study of the People Counting problem (also referred to as Regional People Counting [2]) has been reemphasized. In addition to the COVID-19 use case and related social distancing, application scenarios (in Fig. 1.1) such as shopping malls, elevator, public transportation, and security are all expected to benefit from automatic People Counting. Furthermore, the growth of the Internet of Things (IoT) and 5G technology allows for more application areas such as smart buildings and urban management (named City Brain [3]) where the study of People Counting can be relevant and provide useful information [4].

1

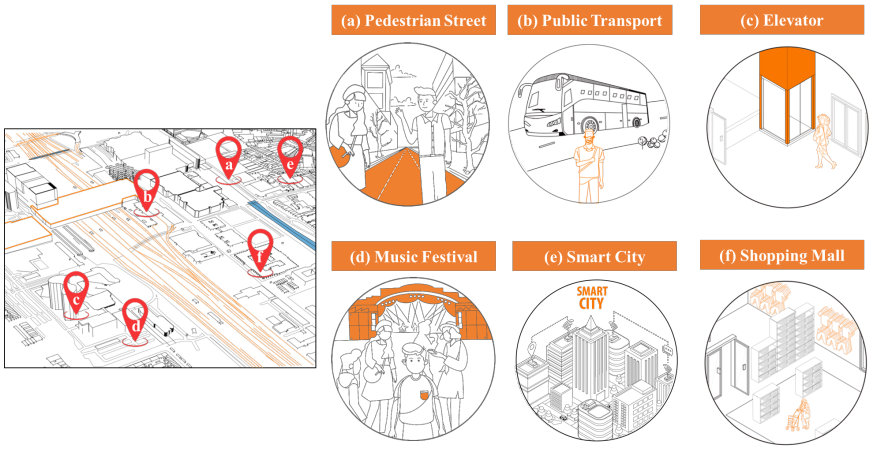


Figure 1.1: Example of People Counting scenarios in the axonometric city view.

Based on the existing literature, the problem of People Counting can be formulated with three different macro objectives in mind, which in turn influence the aim and the possible means to achieve it. They are *ranged* People Counting, *true* People Counting and *exact* People Counting. They are defined from various requirements of specific application scenarios.

- *Ranged* People Counting means counting the density of people where/when knowing the exact number of people in a space is not necessary, i.e. the airport entrance [5], the pedestrian street [6].

- *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, i.e., the subway platform [7].

- As the most advanced type of People Counting, the *exact* People Counting is counting the accurate number of people and knowing the location of each person. This type of People Counting enables more information about people for business analysis, epidemiological investigation [8], and other IoT scenarios [9].

The detailed introduction of these three types of People Counting problem and their application scenarios is provided in Section 2.1.

Manual Counting is commonly used in the traditional People Counting, but it ties up people's time and energy. Meanwhile, it is inevitable that people are prone to mistakes. Thus, many advanced technologies have been proposed to solve the People Counting problem. For example, WIFI/ Bluetooth device-based methods [10, 11], thermal image sensors [12, 13], video camera [14, 15], radar [16], and LiDAR [17]. These technologies have different environmental requirements for properly working (i.e. the radiation of sunlight could influence the detection of the thermal image sensor [18]), and both the advantages and disadvantages of all selected technologies are summarized in Table 1.1.

Due to the increasingly strict privacy requirements for statistical information of people and the requirement of operating in a challenging environment, Radar has some advantages compared with other technologies to solve the People Counting problem. There are mainly two categories of Radar-based People Counting methods proposed in the recent literature.

- The first category is the *feature-based counting* method. In this method, the characteristics of multiple people are collected and analyzed before outputting results (i.e. the histogram of amplitudes from received signals [19]). This category of methods generally has a low computing intensity requirement and good performance in multiple stationary people, but they require large groups of data for analysis.
- The second category is the *tracking for counting* method. This branch of methods can be regarded as Multiple Targets Tracking (MTT) methods, which apply tracking algorithms to estimate the motion of each person in the range-of-the-interest (RoI), so that the number of people and even their location can be known.

The pros and cons of both categories are summarized in Section 2.2.

After briefly summarizing the state-of-the-art (SOTA) in People Counting, there are still some research gaps, given as follows.

- Although existing literature mentioned that the motion of people is randomly walking, the grouping case (i.e. the case of multiple people walking or moving together close to each other) is not studied but does influence the robustness of Radar-based People Counting methods.
- *Feature-based counting* method: Although the commercial MIMO radar is used for People Counting, there is still a room to use the features (range, Doppler, azimuth angle, power etc.) provided by this Radar for better People Counting.
- *Tracking for counting* method: although current *tracking for counting* methods are able to track multiple individuals and also record their tracks, the limited motion of people for People Counting scenarios influences its scope of application.

1.2. CONTRIBUTION OF THE RESEARCH

The main contributions of this thesis research are summarized below:

- One of the most complex motions of individuals, grouping, is studied, and the Radar-based People Counting problem is addressed with a dedicated processing pipeline more in general, including the cases of grouping of multiple individuals.
- This pipeline was developed by studying the characteristics of Group People (defined as a group of people sharing neighboring, adjacent locations and moving together), with multiple-input-multiple-output (MIMO) frequency-modulated continuous wave (FMCW) Radar. Specifically, the combination of the Range-Azimuth map and spectrogram/cadence velocity diagram (CVD) is proposed to solve the People Counting problem including grouping case.

- Based on the fact that two categories of Radar-based People Counting methods, i.e., tracking for counting methods and feature-based counting methods, have complementary strengths, the proposed method combines these two approaches into a new processing pipeline to estimate the number of people in each group in the scene while tracking each group. Moreover, the proposed method can achieve "Beyond classification", which is the output the unlabeled classes not defined at the training stage. Additionally, compared with other state-of-the-art (SOTA) Radar-based People Counting methods, the proposed method outperforms them, and thus it is proved that the grouping problem can be well addressed in the Radar-based People Counting field.

Part of the results from this thesis work are being written up for publication in a high-quality IEEE journal such as IEEE Sensors, IEEE IoT Journal, and IEEE Transactions on Geoscience and Remote Sensing (TGRS).

1.3. STRUCTURE OF THE THESIS

The structure of the thesis is as follows:

First, the literature view on the Radar-based People Counting methods is introduced in 2 where the advantages and disadvantages of tracking for counting methods and feature-based counting methods are analyzed and summarized. Thus, the definition of grouping is introduced, and its characteristics are studied in Chapter 3. Based on the study, the proposed method that combines tracking for counting and feature-based counting is introduced in Chapter 4 to solve People Counting problem including grouping case. To test the performance, three Data Sets are used and the study of performance validation is provided in Chapter 5. Then, four SOTA Radar-based People Counting methods and a commercial product are used for comparison in Chapter 6. Finally, the conclusions and future work are given in Chapter 7.

Table 1.1: Selected technologies applied in People Counting and their brief analysis.

Technologies	How It Works	Environment Requirements	Comments
WIFI/Bluetooth device-based methods [10, 11]	It counts the number of connections	It depends on the device. Generally, the Environment requirements are less than those of device-less methods. [20]	Although it utilizes existing hardware on mobile devices, individuals are required to turn on the WIFI mode/ Bluetooth mode [21]. Meanwhile, people with multiple mobile devices can cause the error.
Thermal image sensor [12, 13]	It detects infrared radiation and can measure the surface temperatures of people.	The high environment temperature and the radiation of sunlight influence the detection.	Privacy and civil rights advocates have raised concerns about discrimination and loss of opportunity that can result from thermal imaging [22].
Video camera [14, 15]	It captures light from the visible spectrum as well others parts of the electromagnetic spectrum in the RoI.	Light and fog can influence the detection.	The good video resolution offers the ability to see what is taking place, and some video cameras even provide wide viewing angles to cover more detecting area. But it has privacy concerns for people.
Radar [16]	It radiates extremely short bursts of radio energy into space and monitors the reflected signal from people.	It can be operated in challenging environmental conditions like foggy or sunshine glare.	Radar used in People Counting is cheap compared to LiDAR, but it has poor resolution in angle compared with LiDAR.
LiDAR [17]	It uses the same principle as Radar except that it uses a laser instead of radio waves.	It does not have the ability to see through bad weather as well as the Radar does.	Although it has the ability to obtain very fine and accurate detection of objects in space, LiDAR is expensive compared to radar and some wavelengths may be harmful to human eyes (IEC 60825-1 standard) [23].

2

Literature Review on Radar-based People Counting Methods

People Counting is a complex and large field due to people's various motions and numerous application environments. When applying different technologies, challenges of People Counting and their solutions may have large differences. Thus, in this chapter, the scope of People Counting is narrowed and the literature review of the Radar-based People Counting is introduced. Section 2.1 is to define the types of People Counting. Section 2.2 defines the relevant scenarios for Radar-based People Counting and the types of radar systems used for this application, and summarizes the types of Radar used in the People Counting field. Then, in Section 2.3 two categories of Radar-based People Counting solutions and their pros and cons are analyzed. In the Section 2.4, the research gaps among these papers are summarized.

2.1. Types of People Counting and Their Requirements

From the existing literature, the problem of People Counting is a large and complex field. Inspired by Kouyoumdjieva et al. [24], People Counting can be divided into three types based on their application scenarios and requirements. They are *ranged* People Counting, *true* People Counting and *exact* People Counting.

- ***Ranged* People Counting** (also referred to density people counting [25]) is suggested as a basic type of People Counting when the exact number of people in a space is not necessary, e.g. the airport entrance [5], the pedestrian street [6]. In this case, the output is a measure of the density of people. For example, the density of people on the pedestrian street is 3 people per square meter. However, it does not mean that the *Ranged* People Counting is an easy task. Normally, phenomena such as multi-path, clutter, occlusion and mixing problems largely influence the accuracy of people detection [26].

- ***True* People Counting** is preferred for those scenarios where the accurate number of people in a given space is required. Meanwhile, each person's location is not required, e.g. the in-elevator counting. In the recent study, the SISO radar is popularly used in this type of People Counting to estimate the number of people [19, 27]. Thus, the potential limitations of this kind of radar-based solution are significant, and it is discussed in detail in Section 2.2.

- ***Exact* People Counting** is counting the number of people and also getting the location of each person, e.g. the in-vehicle passengers counting [28], smart office/house via Heating, Ventilation, and Air Conditioning (HVAC) control [29]. The term *location* means at least the azimuth and range of people are required to be known. This type of people counting is the highest level of People Counting, and the capability to know the location as well as the number of people is potentially appealing to a broad range of applications. To know the location of each person, the SISO radar can not be used, because it can not estimate the azimuth angle of targets.

2.2. SCENARIOS OF RADAR-BASED PEOPLE COUNTING

According to the relevant literature, People Counting has numerous application scenarios, and generally they are grouped into "Indoors" and "Outdoors". However, this division may not be suitable for Radar-based People Counting. Thus, Subsection 2.2.1 introduces the specific grouping of scenarios under Radar-based People Counting. Meanwhile, in Subsection 2.2.2 the types of Radar used in People Counting and their particular scopes and applications are discussed.

2.2.1. "OUTDOORS" OR "INDOORS"?

In several literature surveys on People Counting, the types of scenario are always divided into Indoors and Outdoors [30, 31]31], most likely related to the different lighting conditions that can affect camera-based systems. But in the radar. But in the radar case when dealing with People Counting, we are more interested in the static/dynamic people's behaviours, and related clutter and multipath effects. Therefore, it may be inappropriate to divide the scenarios into Indoors and Outdoors in the Radar-based People Counting problem. In this thesis, the application scenario are split into three types. They are *open* area, *narrow* area and *cluttered* area.

- ***Open* area** is defined as a place with little clutter and multi-path effect, i.e. outdoors lawn, indoors basketball playground. The common features of the *open* area are that there is no roof and wall, or the roof is high and the walls are far from the area under test.

- ***Narrow* area** is a challenging place when using Radar as there is an expected large multipath effect, that is, the elevator, the airport entrance. The *narrow* area generally has a low roof and is closely surrounded by walls. In this case, the Radio Frequency (RF) signal collected by a receiver via two or more propagation routes causes phase differences [32]. This could generate ghost targets and greatly increase the difficulty of People Counting.

- ***Cluttered* area** is also a challenging place where there is a large static/ dynamic clutter effect, i.e., the office with tables and chairs, the busy crosswalk with vehicles. In the People Counting, the clutter is defined as those unwanted detected objects that still appear as returns in the radar received signals. In general, the *cluttered* area has some static clutters (i.e., tables, trees) or dynamic clutters (i.e., cars, bicycles). How to effectively classify, separate, and suppress clutter is still a challenging research question [33].

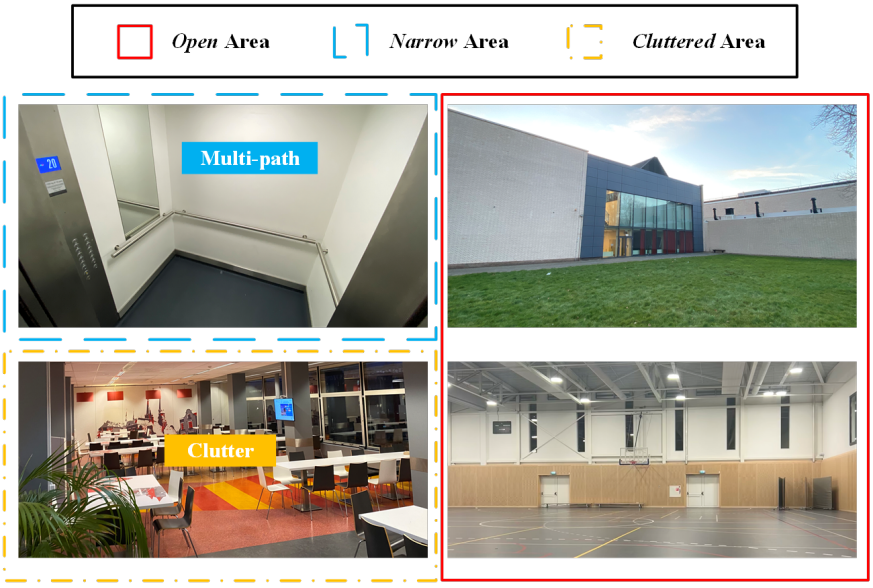


Figure 2.1: Examples of three types of scenarios for People Counting (*open* area, *narrow* area, *cluttered* area).

2.2.2. TYPES OF RADAR USED IN THE APPLICATION AREA

In the Radar-based People Counting field, the expression 'Radar' generally means a low-cost radar with limited number of channels. In recent studies, both the FMCW radar and the IR-UWB radar are applied in People Counting, while there is no systematic comparison between FMCW radar and IR-UWB radar [34, 35].

The major major performance-related difference among the different types of Radar is the number of transmitters and receivers. In other words, whether they provide azimuth/elevation information and the related resolution when counting people, influences which type of People Counting they are able to solve. As mentioned in 2.1, SISO radar is only applied for *Ranged* and *True* People Counting due to the lack of angular information, while SIMO and MIMO radar can be used for all cases. The literature review for this thesis shows that the SISO radar is still a popular choice for People Counting [35, 36, 37]. But the main weakness of the SISO radar is noteworthy and may cost avoidable expenses and consumption. The SISO radar can not provide the estimated azimuth (or even elevation) information of people in the Range of Interest (RoI), and it is questionable when dealing with multiple people standing/moving at the same range bin. Al-

though in some of those literature papers the accuracy of *Ranged*/ *True* People Counting is more than 80%, the limited information provided by the SISO Radar may increase the difficulty when analyzing and fixing the counting error.

2.3. TWO CATEGORIES OF "RADAR-BASED PEOPLE COUNTING" METHODS

In the existing literature in this field, solutions to "Radar-based People Counting" can mainly be divided into two categories: **tracking for counting** (introduced in Subsection 2.3.1) and **feature-based counting** (introduced in Subsection 2.3.2). Meanwhile, the pros and cons of both categories are summarized and explained in details in Subsections 2.3.3. Moreover, Some selected publications studying Radar-based People Counting using **tracking for counting** methods as well as **feature-based counting** methods are listed and described in Table 2.3 and Table 2.4.

2.3.1. THE STATE-OF-THE-ART SOLUTIONS OF TRACKING FOR COUNTING METHODS

The *tracking for counting* method is defined as using tracking algorithms to track each person's movement in the RoI for counting purpose. In particular, this category of method can be regarded as Multiple Targets Tracking (MTT) methods, while in this case "targets" means people. Depending on the wavelength, Radar cross-section (RCS) of people and the distance from the radar, the "target" can be point target or extended target. The pipeline before applying the tracking algorithm may differ because of different Radars (i.e. FMCW radar, IR-UWB radar) [38, 39]. In this subsection, the general pipeline of FMCW MIMO radar is discussed, and in this thesis people are extended targets because the wavelength is assumed to be in the mm-wave region, e.g. commercial 60 GHz radar. There are four steps for this general pipeline, which are *preprocessing* step, *detection* step, *clustering* step, and *tracking* step as shown in Fig. 2.2.

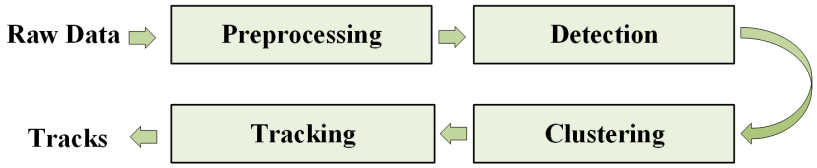


Figure 2.2: The general pipeline of Radar-based *tracking for counting*.

- ***Preprocessing* step**:

  The *preprocessing* step mainly includes clutter removal, angle, range, and Doppler estimation. The clutter removal includes static clutter removal and dynamic clutter removal.

  Hafner et al. [40] proposed the background subtraction method to suppress the static clutter. Will et al. [41] applied the exponential moving average (EMA) algorithm to remove stationary clutters. However, the moving clutter can also influence the association of tracking and even mask the presence of human beings.

Researchers work on the dynamic clutter removal method. Rizik et al. [42] performed the moving target and clutter segmentation based on the different ranges of Doppler and distinct micro-Doppler features. Nallabolu et al. [43] presented the improved intrinsic high-pass filter (HPF) with varying the weighting factor α in the EMA algorithm to differentiate moving targets and dynamic clutter.

The angle estimation is necessary since commercial radar has limited channels and thus the angle resolution is poor. Super-resolution methods are proposed to improve the angle estimation, i.e. MUSIC [44], MVDR [45]. Gurcan et al. [46] proposed the joint dimensional (range-azimuth-Doppler) estimation using MUSIC approach to improve the estimation accuracy. Meanwhile, the 2D-FFT is normally applied for the range and Doppler estimation.

- ***Detection* step**:

  From the related MTT paper, the *detection* step is of great importance since the clustering and tracking methods depend on the generated point cloud. Thresholding detection method is mainly applied and a group of constant false alarm rate (CFAR) processing (i.e. cell averaging CFAR (CA-CFAR) [47], ordered statistics CFAR (OS-CFAR) [48], censored harmonic averaging CFAR (CHA-CFAR) [49]) are applied to calculate a data adaptive threshold for better target detection. Since CA-CFAR has a requirement about the smallest distance between each target, OS-CFAR is popularly used in MTT scenario where targets are mixing [50]. Safa et al. [51] proposed an improved thresholding method featuring a higher probability of detection (P_D) for a certain probability of false-alarm (P_{FA}) which outperformed OS-CFAR in indoors scenario with a lot of connected clutters.

  However, the thresholding method is not robust in detecting weak targets [52]. Weak targets are common in the People Counting field, especially long-range detection (>10 meters). Thus, the track-before-detect (TBD) is proposed to deal with the weak detection issue. The key idea of TBD is that rather than declaring the presence of targets based on the measurements collected in a single frame, the measurements received in multiple frames are jointly processed, and the target is detected from the record of a number of candidate trajectories. Yan et al. [53] summarized some types of TBD method (i.e. *Particle filter* based TBD, *dynamic programming* based TBD, *Hough transform* based TBD, and *Optimization* based TBD) and proposed joint TBD to confirm more reliable candidate trajectories for target detection.

- ***Clustering* step**:

  When using the FMCW MIMO radar at mmWave, people are extended targets. In other words, more than one detection point are generated from one person. The *clustering* is to assign and combine all detections originated from one person at a given time step to one single input for the following *tracking* step. Density-based spatial clustering of applications with noise (DBSCAN) [54] and K-means [55] are commonly applied in the MTT *clustering* part, but this is not well suited for Radar because of the non-equidistant sampling density. In other words, the minimal azimuth distance between two detected points (originated from the same person)

increases with range. Thus, Kellner et al. [56] proposed the Grid-based DBSCAN to solve this. Meanwhile, Wagner et al. [57] proposed the improved ordering points to identify the clustering structure (OPTICS) to solve the case when two people comes too close to each other.

- ***Tracking* step**:

  After applying the clustering step, multiple people are represented as multiple point input (named 'plots') in each frame, and tracking is to decide reliable tracks among those frame-to-frame plots. Thus, the MTT has two challenging research questions. The first is how to decide the association of points, especially associating the existing track with plots at the next frame. The second research question is when to generate a new track, especially under the frequent missed detection and false detection situation.

  In the MTT modelling literature, there are various advanced filters used for tracking multiple extended targets (i.e. Multiple hypothesis tracking (MHT) [58], Gaussian-mixture probability hypothesis density (GM-PHD) [59], Poisson multi-Bernoulli mixture filter (PMBM) [60]). Those MTT filters have different definitions about the target birth model, extended target measurement model, and dynamic model. Based on them, the proposed filters are custom-designed and often tested with LIDAR in automotive related scenarios. However, due to the fact that the angular resolution of commercial radar is poorer than the LiDAR, the implementation of those filters in the radar case is complicated and requires a lot of modeling adjustments.

  In the Radar-based Tracking for counting, classic filtering methods (i.e. Kalman Filter (KF) [61], extended KF (EKF) [62], Bayesian filter [63]) are commonly been applied. Pegoraro et al. [64] proposed an extended object tracking Kalman filter and used a deep learning classifier to help estimate the position, shape and extension of people for better tracking. Nicolas et al. [65] proposed the improved Bayesian filter to track two people walking parallel to each other and mixing together. Ninos et al. [38] mentioned the order of clustering and tracking could influence the performance and the grouper tracker with clustering is proposed for mixing case at the close distance (< 3 meters). However, the scenarios of tracking for counting are limited, where a few people are testing in the RoI (less than 3) and people in the RoI keep moving for maintaining the existing trackers.

2.3.2. THE STATE-OF-THE-ART SOLUTIONS OF FEATURE-BASED COUNTING METHODS

The *feature-based counting* method is defined as using the prior knowledge of summarized features of received data with the help of the *artificial neural network* (*ANN*) or *non-ANN* methods to solve the People Counting. In other words, *feature-based counting* is defining People Counting problem as a classification problem. According to the existing literature, both ANN and non-ANN methods have a common pipeline including four blocks as shown in Fig. 2.3. The first block is to process the raw data received from Radar, and then pass them to the second block, which performs the feature extraction. In order

to jointly use the extracted features, the feature fusion block is proposed, which is represented as the third block. Finally, the associated classification is proposed to output the expected class (e.g. the number of people in the RoI).

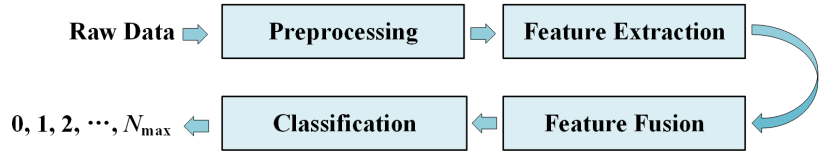


Figure 2.3: The general pipeline of *feature-based counting*.

- ***Preprocessing* step**:

  This step is similar to "Preprocessing" in the *tracking for counting* methods mentioned in Subsection 2.3.1. A series of algorithms are applied to obtain an "array" of inputs for following feature extraction. The "array" contain “vector” a one-dimensional array), “matrix” (a two-dimensional array), and "tensor" (three dimensions and above) for the remainder of this thesis.

  As Radar has evolved to provide multi-dimension information, the variety of input arrays available for People Counting has correspondingly increased. Based on the number of dimension, these inputs can be divided into three types: *signal*, *image*, and *point cloud*. *Signal* means a vector of inputs which is a one-dimensional array, such as the range profile [37], power spectral density (PSD) [27]. *Image* represents a two-dimensional array (also named as matrix). With the development of neural networks, *image* feature is popularly used for training [66, 67]. *Point cloud* means a high-dimensional array (more than three), and can be achieved when using a MIMO Radar [68, 69, 70]. According to the existing literature, there are few paper about directly process the point cloud for *feature-based counting*, while it is squeezed into lower dimension for the following classification process such as the range-azimuth map.

- ***Feature extraction* step**:

  This step is to find the "interesting" parts of input arrays, which are associated with the expected class, such as the number of people in the RoI. In this step, non-ANN and ANN methods have different directions to extract features from processed data. In non-ANN methods, features are devised manually from some sophisticated processes (such as Curvelet transform [26], CLEAN algorithm [71]) to form the feature vector, and they are named as handcrafted features [37]. In ANN methods, smarter tools are used to extract features than manual extraction, such as neural networks. But, as a disadvantage of these methods, such features are difficult to be interpreted.

- ***Feature fusion* step**:

  This step is applied when using multiple inputs for classification. Based on the fact that people move through scenarios that are very diverse and random, single type

of features cannot cover complex situations, such as people walking in the same range bin are less likely to be classified based only on the range-domain features [27]. In non-ANN methods, multiple feature vectors are first normalized separately to hold similar span level, and then concatenated to develop a single feature vector for following classification. In ANN methods, attention (which mimics cognitive attention) is used to do the feature fusion [72, 73]. As some features may contribute less for classification, attention is used to enhance some of these features while reducing the importance of others. In this case, the network is able to focus on a smaller but more important subset of features. Choi et al. [37] used long short-term memory (LSTM) to connect the information from multi-frame range profiles for People Counting purpose.

- ***Classification* step**:

  As summarized at the beginning of this subsection, People Counting is regarded as a classification problem in *feature-based counting* methods. The classification step is to assign a measurement to a given class based on trained data. In non-ANN methods, linear or nonlinear classifier (e.g., linear regression [74], k-nearest neighbors algorithm (k-NN) [75], naïve Bayes [76], random forest [77], support vector machine (SVM) [78]) is trained with feature vector and corresponding labels to optimize the decision boundaries, which enables the classifier to predict the number of people from newly measurements. In ANN methods, fully connected layer (FC) is commonly used to map the "distributed features" learned through the network to the expected classes and thus to achieve the role of a classifier [37, 79].

2.3.3. PROS AND CONS OF BOTH CATEGORIES OF METHODS

After summarizing the SOTA solutions of both categories, the benefits and drawbacks of both categories are summarized in Table 2.1 and explained in details.

Advantage 1 of *tracking for counting*: This category of methods does not rely on the pretrained data to solve People Counting.

Compared with the *feature-based counting* method, the *tracking for counting method* is able to be implemented directly without collecting and using the processed data for training in advance. This offers more flexibility when counting multiple people under various environments, in the comparison with *feature-based counting* methods [39].

Advantage 2 of *tracking for counting*: It is able to output and record the position of people in the RoI.

When the current radar evolves from SISO to MIMO, more information about the target is available, such as the azimuth, range, and even elevation information of people. This information of position is assigned to the status of target, and is used to predict and update the tracking algorithms [80]. In this case, we are able to be aware of "who is where" when applying *tracking for counting* methods.

Advantage 3 of *tracking for counting*: 3. It is able to use information from time series to predict and update the status of target, thus the accuracy of People Counting is improved.

In *tracking for counting* method, interframe information is used for predicting and

updating target status. Moreover, two major challenges when using the Radar to do the detection, which are multi-path effect and clutter effect, can be solved by tracking algorithms [81]. Thus, the accuracy of people counting is improved after connecting the relative information between each consecutive frame [82]. This can be used to deal with counting people with complex motion. For example, Person A is occluded by Person B in the i^{th} frame, but it shows up in the $(i+3)^{th}$ frame. In this case, the "tracking for counting" method could associate the past information and present information and update the number of people in the RoI [83].

Table 2.1: Summary of Pros and Cons of Both Categories of Radar-based People Counting Methods.

Methods	Advantages	Disadvantages
Tracking for Counting	1. This category of methods does not rely on the pretrained data to solve People Counting. 2. It is able to output and record the position of people in the RoI. 3. It is able to use information from time series to predict and update the status of target, thus the accuracy of People Counting is improved.	1. In the implementation stage, for newly received data, these methods require large computation intensity to output results. 2. It is not able to count the number of people when they are walking as a group (named as Group People) in the RoI.
Featured-based Counting	1. In the implementation stage, the pretrained model is able to output results with low computation intensity. 2. It is able to solve grouping problem, when treating this as a Classification problem for which an algorithm can be trained.	1. It requires large amount of data for training before outputting the expected labels. 2. From the existing method, outputs of pretrained model are limited to the number/density of people and can not know "who is where". 3. It ignores the close connection between each time step as people move in the RoI due to the fact that the outputs from the pretrained classification model are independent.

Disadvantage 1 of *tracking for counting*: In the implementation stage, for newly received data, these methods require large computation intensity to output results.

Compared with the pretrained model, *tracking for counting* method needs to predict and update target's status each time step for every new measurement. Meanwhile, interframe information is used to decide whether associating the existing track or creating a new track when more than 2 human targets move freely and randomly in a certain RoI [84]. As a result, large computation load is required to perform the processing and

make decisions for every newly received data.

Disadvantage 2 of *tracking for counting*: It is not able to count the number of people when they are walking as a group (named as Group People) in the RoI.

The assumption of applying *tracking for counting* methods is that the number of people equals to the number of tracks. However, it ignores that case when people are moving as a group. In this grouping case, the number of people are more than the number of tracks. This is the biggest disadvantage when applying *tracking for counting* methods in some certain scenarios, such as pedestrian street, music festival, where people are always moving in groups [85].

Advantage 1 of *feature-based counting*: In the implementation stage, the pretrained model is able to output results with low computation intensity.

In the feature-based counting method, classification models are trained before the implementation stage. In the implementation stage, the received data pass through this pretrained model, and the expected classes are output, such as the number/ density of people. Compared with *tracking for counting* methods, lower computation intensity is used [86, 87].

Advantage 2 of *feature-based counting*: It is able to solve grouping problem, when treating this as a Classification problem for which an algorithm can be trained.

As mentioned in the disadvantage 2 of *tracking for counting*, grouping is one of the most complex case of people's motion, and it does influence the accuracy of People Counting. *Featured-based counting* method is able to solve grouping problem when regarding it as classification problem. How to analyze and solve grouping problem is one of the contributions of this thesis.

Disadvantage 1 of *feature-based counting*: It requires large amount of data for training before outputting the expected labels.

Before applying the feature-based counting method to output the accurate result of counting, a large amount of data in many situations need to be collected. Multiple people with different motions in the RoI may have a large group of cases, e.g. three people walking separately, two people walking side by side. Some features may have bad performance of counting people in some specific cases, e.g. the accuracy of using micro-Doppler spectrograms for classifying two/ three people walking side by side is less than 80% [88]. It is a challenging research question of how to select a proper feature under those various cases. In a word, *feature-based counting* method can not work without collecting data in advance for a proper training of the classifier.

Disadvantage 2 of *feature-based counting*: From the existing method, outputs of pretrained model are limited to the number/density of people and cannot know "who is where".

According to existing studies, the labels of People Counting classifier are limited to the number/density of people in the RoI, and locations of people are not always available. There are two reasons of this fact. The first reason is that the use of MIMO Radar in People Counting is not yet not yet extensively studied, and thus the azimuth and elevation information of people is not available. The second reason is to be trained that if the classifier is trained for knowing "who is where", larger amount of data for training

process is needed. It is not a good choice for counting people in a wide range [27]. Therefore, currently *feature-based counting* methods focus on telling the number/density of people in the detected region instead of identifying "who is where".

Disadvantage 3 of *feature-based counting*: It ignores the close connection between each time step as people move in the RoI due to the fact that the outputs from the pretrained classification model are independent.

According to recent studies, the "feature-based" people counting methods ignore the information between each consecutive frame. In other words, each decision is made independently, while people's motion in the RoI is continuous. Ignoring temporal connection can output inaccurate results. For example, the estimated number of people in the elevator fluctuated at every time step, although people stayed static and nobody came in or went out of the elevator [19].

Table 2.2: Summary of Radar-based People Counting with related feature-based or tracking approaches.

Types of People Counting	Types of Radar	Categories of Radar-based People Counting Methods
Ranged People Counting	SISO, SIMO, and MIMO Radar	*Feature-based Counting*
True People Counting	SISO, SIMO, and MIMO Radar	*Feature-based Counting* and *Tracking for Counting*
Exact People Counting	SIMO and MIMO Radar	*Tracking for Counting*

2.4. SUMMARY OF RADAR-BASED PEOPLE COUNTING AND RESEARCH GAPS

After summarizing the Radar-based People Counting methods (shown in Table 2.2), there are still some research gaps, given as follows.

- Although existing literature mentioned that the motion of people is randomly walking, the grouping case (i.e. the case of multiple people walking or moving together close to each other) is not studied but does influence the robustness of Radar-based People Counting methods.
- *Feature-based counting* method: Although the commercial MIMO radar is used for People Counting, there is still a room to use the features (range, Doppler, azimuth angle, power etc.) provided by this Radar for better People Counting.
- *Tracking for counting* method: although current *tracking for counting* methods are able to track multiple individuals and also record their tracks, the limited motion of people for People Counting scenarios influences its scope of application.

In Table 2.4 and Table 2.3, some selected publications studying Radar-based People Counting using *tracking for counting* methods as well as *feature-based counting* methods are listed and described.

The contribution of this thesis is to fill the gaps identified above. The grouping problem in Radar-based People Counting is studied, and the proposed pipeline combines both advantages of tracking for counting and feature-based counting.

Table 2.3: Comparisons of the state-of-the-art radar-based people counting methods.

Authors	Radar Type	The Number or The Density	Experiment Set-up	Motion of People	Maximum Number of People	Method for Counting
Choi et al. [19]	IR-UWB radar	The Number	Elevator and indoors room with 5 m maximum range	Walking without mixing and staying static	10	Feature-based counting: generating a probability density function of the amplitudes from cluster features and deriving the maximum likelihood equation for people counting
Yang et al. [5]	TI mmWave FMCW radar	The Density	Narrow entrance	Walking and standing static	4	Feature-based counting: using the peak amplitude features from range-angle map, and applying a maximum likelihood classifier to identify the density
Abedi et al. [34]	TI mmWave FMCW radar	The number	In-vehicle with 3 m maximum range	Sitting still	5	Feature-based counting: using the range-azimuth heat map and applying machine learning method to count the number of people and indicate their location
Yang et al. [26]	TI mmWave FMCW radar	The density	Indoors with maximum 5 m range	Stand in a queue and multiple people gathering with different density	5	Feature-based counting: using curvelet features of multi-channel range-amplitude map and doing the feature fusion with amplitude features of range-amplitude signals
Choi et al. [27]	IR-UWB radar	The Number	Indoors with 10 m maximum range	Randomly walking	10	Feature-based counting: using the amplitude features from both range domain and frequency domain to efficiently address both dense and dispersed distributions of individuals

Table 2.4: Comparisons of the state-of-the-art radar-based people counting methods. (Continued)

Authors	Radar Type	The Number or The Density	Experiment Set-up	Motion of People	Maximum Number of People	Method for Counting
TI [61]	TI mmWave FMCW radar	The number	Open area and indoors conference table with 5 m maximum range	Moving without mixing and sitting still	5	Tracking for counting: clustering point clouds and applying extended kalman filter for tracking
Jia et al. [66]	SFCW through-the-wall radar	The number	Indoors with maximum 12 m range	Walking with mixing	6	Feature-based counting: using ResNet-based (a typical deep convolutional neural network) feature from range-time map
Yang et al. [89]	IR-UWB Radar	The Number	Narrow confined room and open lobby with 3 m maximum range	Randomly walking and standing in a queue	3	Feature-based counting: using CNN features from raw data. Input data with several signal processing methods (Clutter removing, DC removing etc.) were proved not necessary when applying CNN.
Bao et al. [90]	IR-UWB radar	The Number	Outdoors open area	Walking and standing static	10	Feature-based counting: using CNN features from multi-scale range-time maps
Huang et al. [91]	IR-UWB radar	The Number	Outdoors open area	Walking without mixing	5	Tracking for counting: using the density-based clustering method to get the point cloud of multiple people and applying the recursive Kalman filter tracking algorithm for counting purpose

3

RADAR-BASED CHARACTERIZATION OF GROUP PEOPLE COUNTING

In the existing literature in the Radar-based People Counting, the assumption of people's motion is limited to walking without grouping, even though they mentioned the motion is randomly walking. In this chapter, the characteristics of grouping are fully studied, which provide a valid idea for the design of the Group People classifier in the next chapter. In Section 3.1, the definition of Group People is introduced, and its model is proposed in Section 3.2. After that, in order to study the Grouping more systematically and to analyze whether the features currently used in the field have limitations, the study of Grouping with existing features is analyzed first in Section 3.3. Those features that are not significant for a different number of people grouping will be given the lowest priority in the Group Classifier design introduced in the next chapter. Then, the study of synchronization and cadence velocity diagram (CVD) is introduced in Section 3.4, which is a proposed manual approach to study the spectrogram. This study could support the Group classifier, and the conclusion is drawn in Section 3.5.

3.1. DEFINITION OF GROUP PEOPLE

Group People is defined as a cluster of people sharing neighboring locations and moving together, as shown in Fig. 3.1 with two people. This behavior is common in everyday life, such as friends travelling in pairs or couples taking a walk together.

Although the existing articles state that the target moves randomly, the motion of individuals does not include grouping in their assumptions [37]. Or they assume that individuals do not always move as a group [27]. In other words, the people observed by the Radar will always separate at a certain moment and move as separate individuals.

Grouping is one of the most complex motions of individuals. In contrast to using a camera for crowd segmentation, in the Radar's view, without prior knowledge, Group

People can not be separated individually when using collected information. Therefore, this is a challenge for one of the most popular Radar-based People Counting methods that uses the tracking algorithm to count (named as tracking for counting method, as discussed in Chapter 2). Group People can not be tracked individually, so they have to be treated as a single extended target. This fact breaks the implicit assumption of tracking for counting methods, which is that the number of tracks equal to the number of people.

From the study of performance comparison, grouping can not be solved properly when applying the current Radar-based People Counting methods. The results are shown in Chapter 6, and those results prove that grouping is an open problem in the Radar-based People Counting. It should be noted by the readers that in the rest of this thesis, the expression "the number of Group People" refers to the number of people in one group.

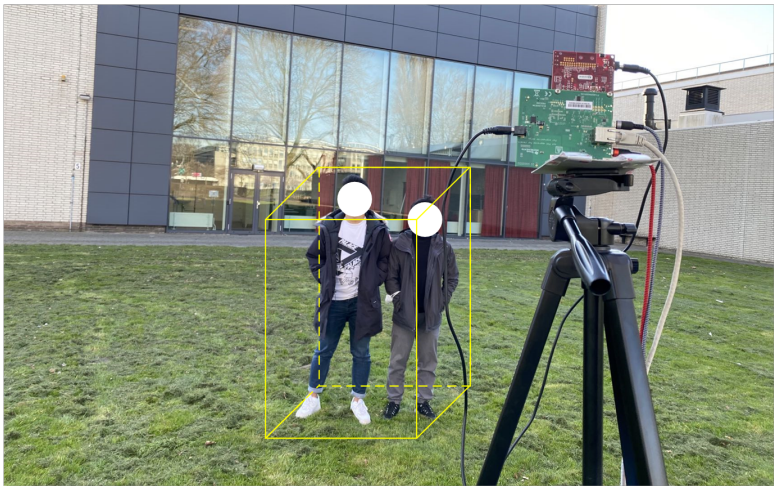


Figure 3.1: The ground truth of Group People with 2 individuals and the FMCW MIMO Radar is used.

3.2. GROUP PEOPLE MODELLING

After giving the definition of Group People, the section aims to build the model for Group People for future analysis.

The model of Group People is built from two parts. The first part is to build the 3D model of individual, then the second part is to link the relationship between each individual.

- **Part 1: model of individual**

  In the existing study, the Boulic model is popularly used for the motion model of human beings [92], which is similar to using 3D global marker in volunteers to make motion models of game characters. In this thesis, the Boulic model of individual is used with 11 control points and the ellipsoid is used as the body segments of group people, as shown in Fig. 3.2 (a).

Meanwhile, the locations of individual should also be taken into account. It is shown in the Radar equation Eq. 3.1:

$$R = \sqrt[4]{\frac{P_t G^2 \lambda^2 \sigma G_P}{(4\pi)^3 R^4 P_{Ni} L_S L_A L_{GP} SNR}} \tag{3.1}$$

where R is the range from transmitter antenna to the target and λ is the wavelength. P_t is the transmitted power, and P_{Ni} is the the noise power generated within the Radar. G represents the antenna gain, and G_P is the processing gain for noise interference. σ is the Radar cross-section (RCS), which indicates how detectable a target is by Radar. SNR is the signal-to-noise ratio. L_s is the system losses, L_A is the propagation path losses, and L_{GP} is the ground plane losses. Besides, the antenna gain is low when away from the boresight direction. Thus, the connection model must take angle and range attenuation into account.

- **Part 2: connection model of multiple people in the group**

  The connection model is required to consider the connection between each individual when they are in a group.

  As mentioned in Section 3.1, Group People are sharing neighboring locations. In this case, three parameters are used to represent this relation. In Fig. 3.2 (b), d_1 is the chest width of individuals. d_2 is the shoulder width of the individuals and d_3 is the Euclidean distance between two neighboring individuals. There are various ways of obtaining the values of d_1 and d_2, essentially the dimensions of the ellipse that define the area occupied by each individual in a group. For example, by looking up some publicly available population data, or by enlisting volunteers to participate in the data count. d_3 is the distance between people in the group. In this thesis, it is set at a maximum of 15 cm. The purpose of this is to prevent the distances between people from being so large that their boundaries can be distinguished by angle estimation (as shown in Eq. 3.2), which breaks the definition of group people.

$$\Delta\theta \approx \frac{0,886\lambda}{Nd\cos\theta}, \tag{3.2}$$

  where N is the number of virtual antenna elements, d is the distance between neighboring antenna elements, and θ is the angle between the target and the Line of Sight (LoS).

The result of the Group People model when two people walk as a group is provided in Section 3.3.

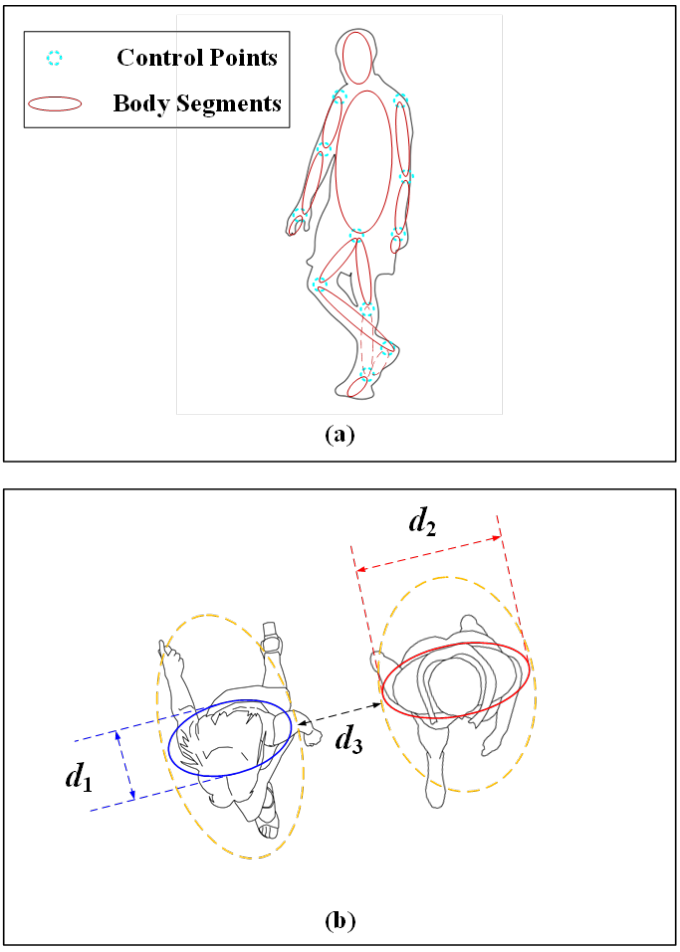


Figure 3.2: Schematic diagram of Group People model: (a) in the front view with the body points and segments related to the Boulic model, and (b) in the top view with the three important parameters defining the size of the group.

3.3. EXISTING FEATURES USED FOR PEOPLE COUNTING

Group People is first discussed in the Radar-based People Counting. As a starting point, several features and data formats which have already been used in this field are studied in this section. The aim is to find whether Group People also have distinct differences in these features. The readers should note that the data used here to generate example figures belong to the collected experimental Data Set I, which is discussed in more detail in Chapter 5.

- **Range-Time Map**

  Range-Time Map provides the range information of targets at a given time period. Jia et al. [66] found that the lines in the Range-Time map represent moving people

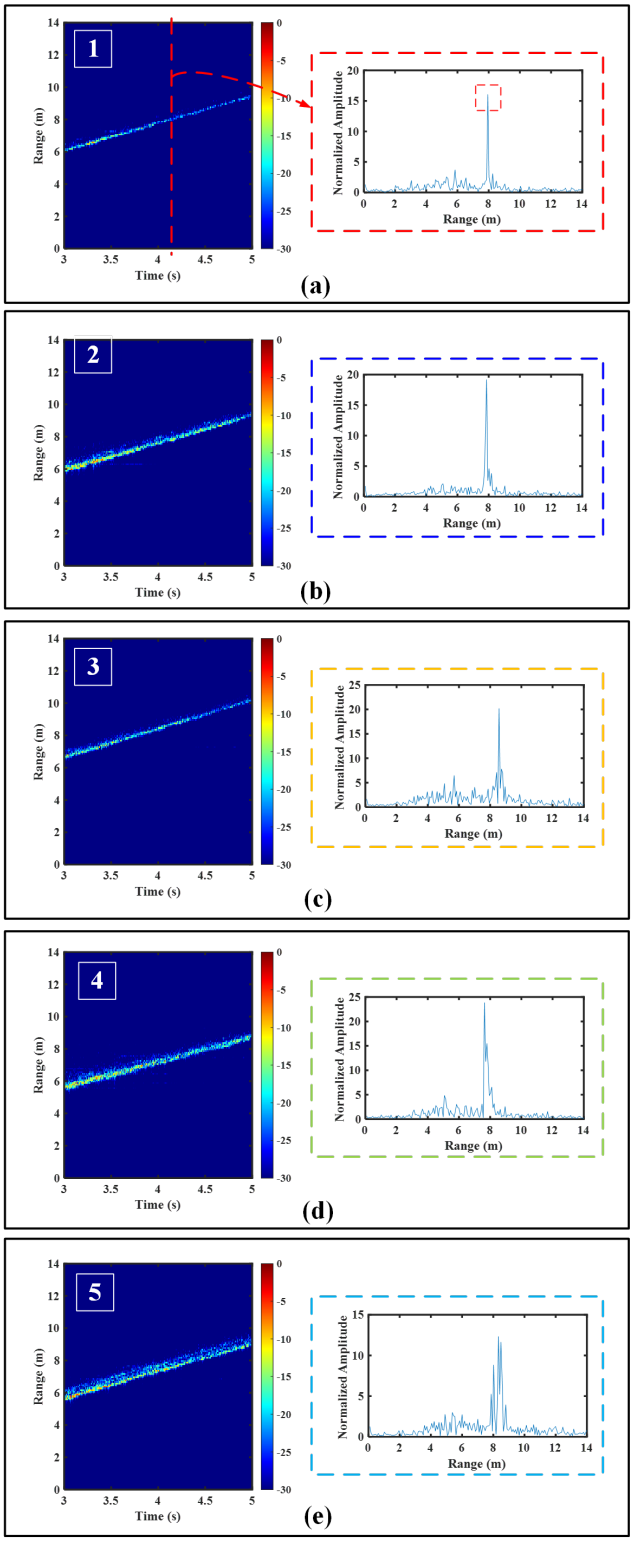


Figure 3.3: Range-Time maps and their range profiles of Group People for (a) one human, (b) two humans, (c) three humans, (d) four humans, and (e) five humans walking as a group.

and the number of lines is connected with the number of people. As shown in Fig. 3.4, in addition to the multi-path, the lines with high amplitude in Fig. 3.4 (a-f) relate to the number of people, and the number of continuous lines is roughly equal to the number of people.

However, this characteristic is not met in the Grouping case. In Fig. 3.3 (a-e), no matter how many people walking as a group, this characteristic is not obvious compared with previous studies. The number of continuous line does not associate with the number of Group People. Therefore, this morphological characteristics of Range-Time Map is not the best option for solving Group People Counting problem.

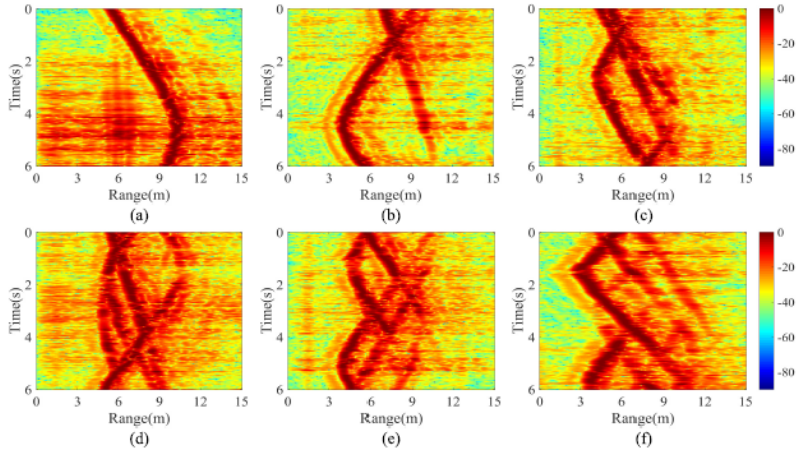


Figure 3.4: Range-time images used for Radar-based People Counting for (a) one human, (b) two humans, (c) three humans, (d) four humans, (e) five humans, and (f) six humans [66].

- **Range Profile**

  The Range Profile is the one-dimensional feature where targets are represented as the peaks, as shown in Fig. 3.5. Lee et al. [93] mentioned the probability density function of the range profile fits the Log-normal distribution. Based on it, Choi et al.[19] proved the number of people is associated with the number of peaks and the amplitudes, and thus applied the maximum likelihood ratio to do the People Counting. They mentioned the number of clusters which connected a group of peaks was necessary and important for People Counting.

  In the Grouping case, the characteristic in range profile does not match the conclusions drawn in the previous studies. In Fig. 3.3, a group of peaks appears in the Range profile. And as the number of people in the group rises, the peak amplitude of the signal does not follow a linear relationship. Besides, the number of clusters does not fit the number of Group People. Therefore, like the Range-Time map, the Range Profile is not the best choice when solving the Group People Counting.

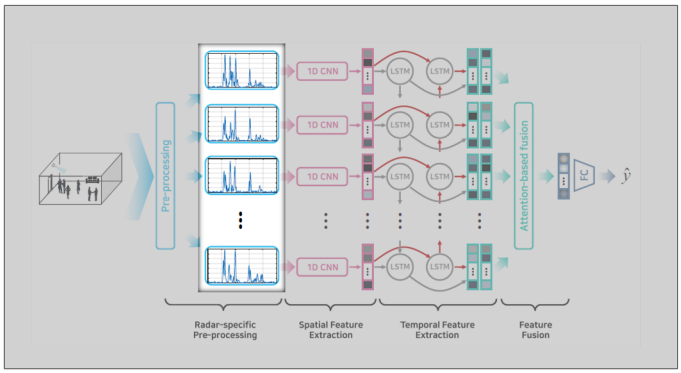


Figure 3.5: Range profile used for Radar-based People Counting [37]

- **Range-Doppler Map**

  Range-Doppler map has more information for People Counting than the Range Profile (in Fig. 3.6). As mentioned by Choi et al. [27], Range Profile performed good when individuals are sparse in the RoI, while it can not tell the difference when people are walking at the same range bin. Range-Doppler can be effective when individuals walking in different direction even if they are at the same range bin. For example, people walking towards the Radar has opposite sign of Doppler information compared with people walking back to the Radar.

  However, the Range-Doppler map becomes vulnerable in the Grouping case. In Fig. 3.7 (a-e), the width of Doppler information increases when the number of Group People increases, but this growth is not linear and robust. It can only be concluded that the larger number of Group People there are, the wider the width of Doppler is likely to be. The positive and negative signs of Doppler are only used for distinguishing different groups, which break the conclusion summarized in the previous study. Thus, the Range-Doppler map can be a feature to do the Group People Counting, but this feature is not robust enough.

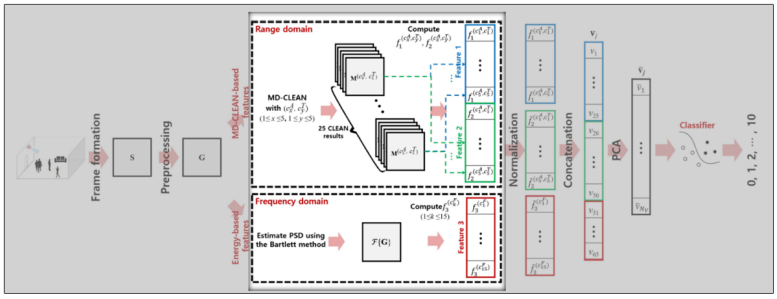


Figure 3.6: Range-Doppler features used for People Counting [27].

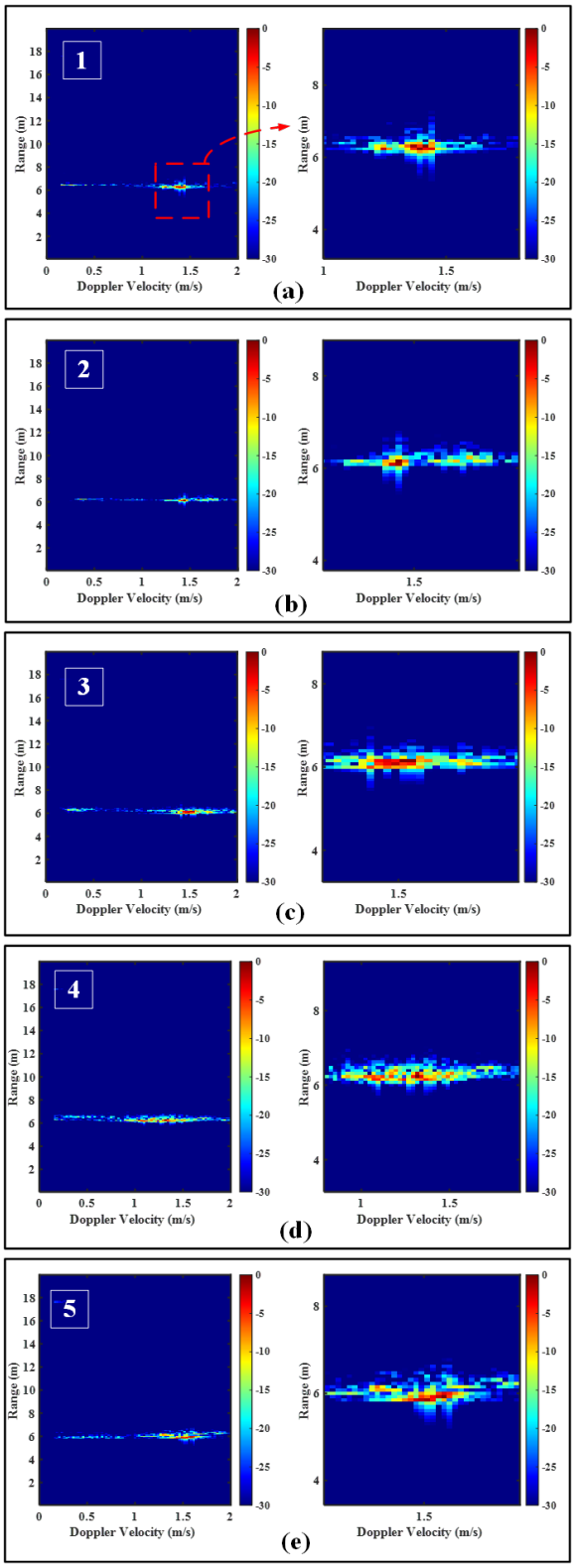


Figure 3.7: Range-Doppler maps and their zoomed versions of Group People for (a) one human, (b) two humans, (c) three humans, (d) four humans, and (e) five humans walking as a group.

- **Range-Azimuth Map**

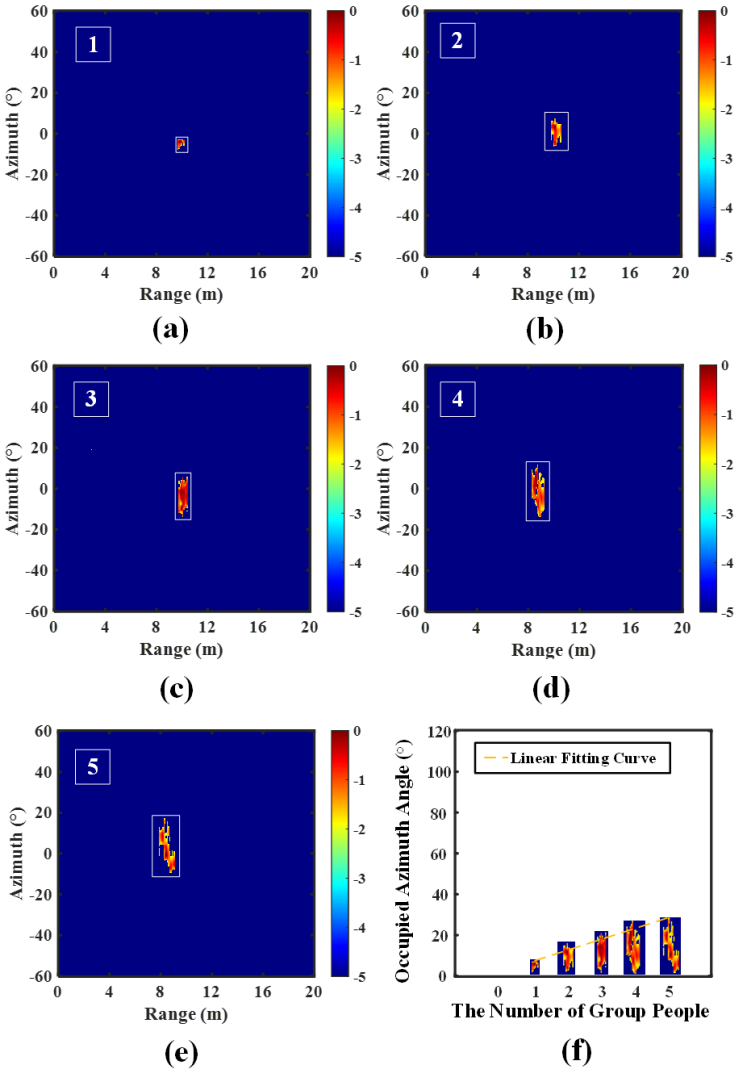


Figure 3.8: Range-Azimuth maps measured for (a) one human, (b) two humans, (c) three humans, (d) four humans, and (e) five humans walking as a group. (f) The summarized of occupied Range-Azimuth bins with Group People from (a-e).

Range-Azimuth map shows the spatial location of targets in the RoI. This feature is commonly used in narrow areas, such as vehicle, elevator, metal ramp [5]. Not all Radar systems are able to provide angle information, and thus the MIMO Radar is used to get azimuth or even elevation information of targets. But the weakness

of this feature is how to estimate the AoA due to the fact that the Radar has poor angle resolution compared to LiDAR, and the resolution degrades away from the boresight direction. Without the prior knowledge, extended targets can not be separated when the angle between them is less than the angle resolution.

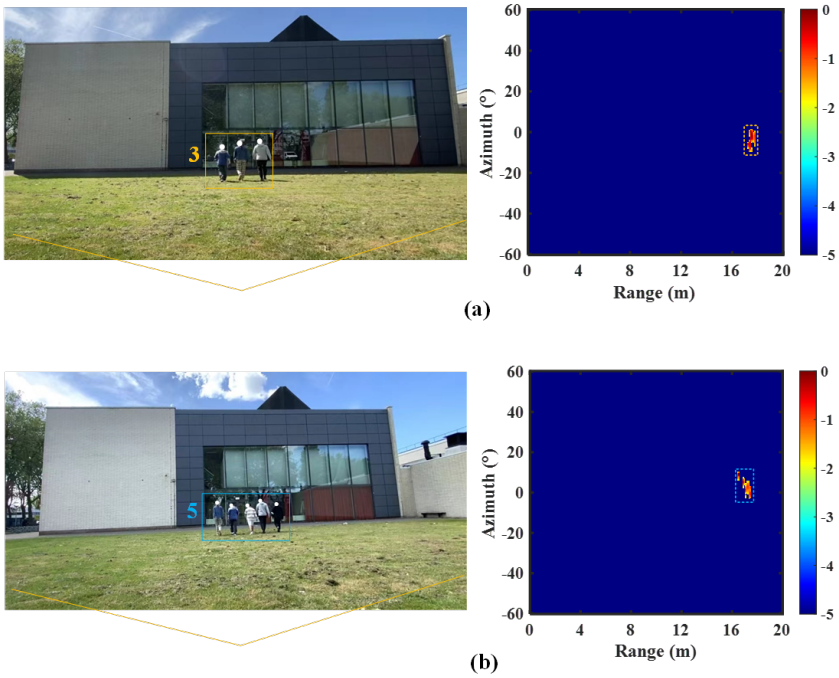


Figure 3.9: Range-Azimuth map in the far range for (a) three humans (b) five humans waling as a group.

Although it is challenging to estimate the accurate angle of targets, the Range-Azimuth map is promising for Group People Counting. Group People is a definition in spatial area, where Range-Azimuth exactly represents this characteristic. In Fig. 3.8 (a-e), when the number of Group People increases, the occupied width of the azimuth angle increases. This phenomenon is robust in the near range, while in the far range the Range-Azimuth map loses these advantages for Group People Counting. After extracting the Range-Azimuth bins from the figure and rank them by the number of occupied azimuth bins, it is found that when the number of Group people increases, more azimuth bins are occupied, while this relation is not linear according to the fitted curve in Fig. 3.8 (f). In Fig. 3.9 (a) and (b), where three and five people walk as a group in the 16 m range, the occupied bins seem the same. It can be explained by the Group People model. As people get further away from the Radar, their AoA gets smaller and smaller until it is less than the angular resolution. Near and far is a comparable term. In this thesis, based on the configuration of Radar and the Group People model, when the distance from the

Radar exceeds 12 m, this is called far range.

Therefore, Range-Azimuth map can be a good choice to solve the Group People Counting, but this feature is fragile when Group People exist in the far range. Other features which have not been used for Radar-based People Counting are required to be studied.

3.4. SYNCHRONIZATION AND CADENCE VELOCITY DIAGRAM

3.4.1. STUDY OF SPECTROGRAM AND CVD

After studying the existing features that have already been used in the literature on Radar-based People Counting, these features hardly reflect the periodic swaying of the limbs as the person moves. In the field of Radar, this periodic swinging feature is often represented in the micro-Doppler signature; hence, in this section the spectrogram, which represents the micro-Doppler information, is studied.

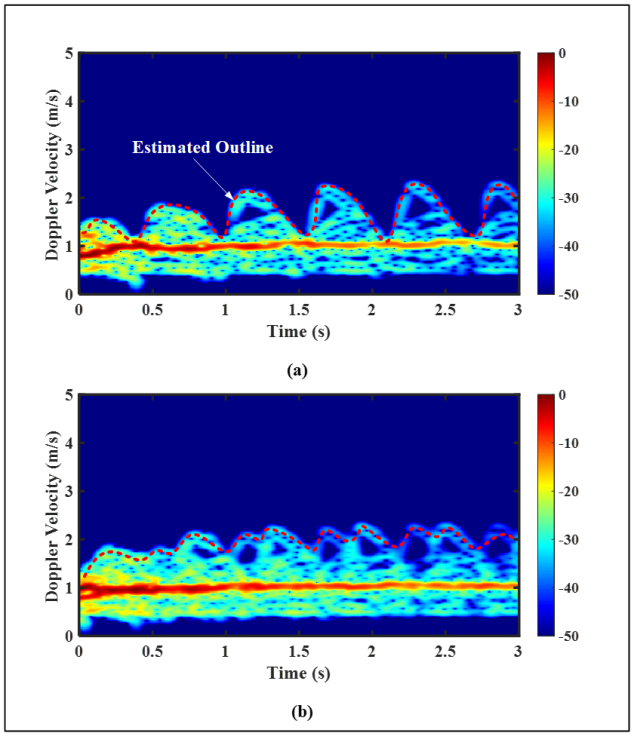


Figure 3.10: Simulated spectrogram based on the proposed Group People model (mentioned in Section 3.2)when (a) one person walks away from Radar, (b) two people walk away from Radar.

The spectrogram can represent well the characteristics of the gait of people. When a person moves away from the radar, we can clearly see the micro-Doppler generated by the movement of the limbs through the spectrogram (in Fig. 3.10 (a)). When analyzing more than one person moving together, we expect to be able to see the mixing of micro-

Doppler due to the inconsistent movements of multiple people, as shown in Fig. 3.10 (b). However, when experimenting in practice, it was found that not all spectrograms showed this "mixing" phenomenon. When comparing this with the video files recorded during the experiment, it was observed that even when volunteers were allowed to move in the way they were used to, there was some degree of consistency in their pace. Although this consistency is not perfectly uniform as for "soldiers walking in unison", the similarity in the frequency of their limb swings is what leads to the micro-Doppler being very similar. In other words, the image morphology of the spectrogram of their movement is similar to that of the spectrogram when only one person is moving.

Moreover, it was also found that the more people move together, the more likely we are to observe the phenomenon of mixing on the spectrogram. Therefore, this step consistency is related to the number of Group People and deserves to be studied in depth for the development of classifiers for Group People Counting. In this thesis, we refer to the "phenomenon that describes the consistency and inconsistency of pace when multiple people are moving" as the "*synchronization phenomenon*".

To study the synchronization phenomenon, the cadence velocity diagram (CVD) is introduced. The CVD $S_c(g_1, g_2)$ is calculated by taking the FFT of the spectrogram across the time axis:

$$S_c(g_1, g_2) = \sum_{l=0}^{N_w-1} |\hat{S}(l, g_2)| w_(l) \mathrm{e}^{-\mathrm{j}2\pi g_1 l/N_w}, \tag{3.3}$$

where $\hat{S}(l, g_2)$ is the spectrogram. 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 $g_1 = 0, 1, 2, ..., N_h - 1$ and $g_2 = 0, 1, 2, ..., N_w - 1$. By taking the FFT across the time axis, the CVD can represent the frequency with which the Doppler velocity of the target repeats. In other words, it exactly studies the periodic properties of people's motion. The ambiguity of the CVD frequency is $f_s/2$, where $f_s = 1/(N_w \sigma_l T_c)$. σ_l is the overlapping frequency and T_c is the duration of the chirp, which is defined by the Radar configuration. Due to the fact that the frequency of the limbs swing during normal human movement will not exceed 2.5 Hz and will not fall below 0.5 Hz, the CVD frequency is selected in this range for 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 5.1) were selected. Each set of data was collected by two different volunteers. Fig. 3.11 provides detailed information on the difference between the high synchronization case and the low synchronization case. They are analyzed in the following paragraphs. For the reminder of the reader, the Doppler velocity of spectrogram is displayed in the absolute value.

The synchrony of human movement was found to be related to the height of the person. 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. 3.11 (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. 3.11 (b)). In Fig. 3.11 (a-b), according to the control points and the estimated skeleton marked on the people, it is obvious that in the high synchronization case the two volunteers' feet are raised from the ground at all times at the same time and their arms are swung at 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.

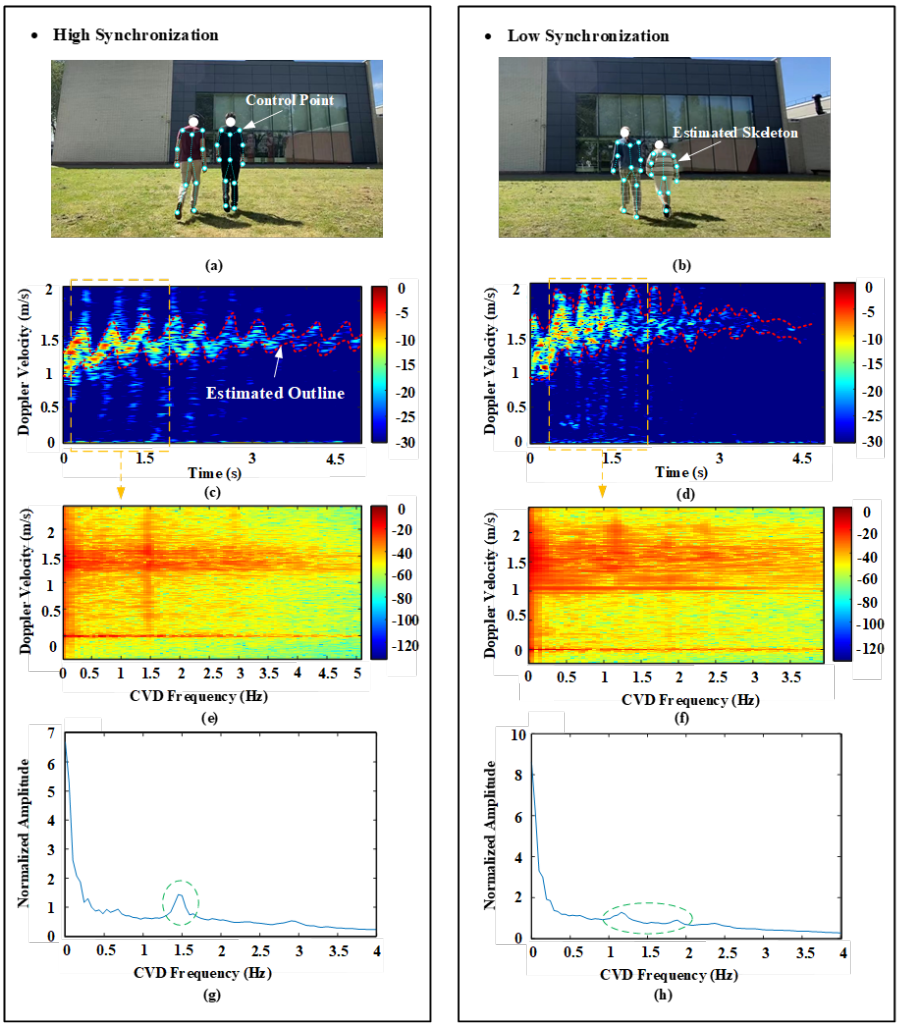


Figure 3.11: The analysis of synchronization. Ground truth with estimated skeleton in (a) high synchronization for people of similar height, and (b) low synchronization case for people of diverse height. Spectrogram in (c) high synchronization and (d) low synchronization case. The Doppler velocity is displayed in the absolute value. CVD map in (e) high synchronization and (f) low synchronization case. CVD profile in (e) high synchronization and (f) low synchronization case.

For the spectrogram, according to the estimated outline, the spectrogram is much less "clean" in the case of low synchronization (in Fig. 3.11 (c)) compared to that of high synchronization (in Fig. 3.11 (d)). Conclusions beyond this are difficult to obtain through empirical observation. Therefore, CVD is applied to extract the target's periodic movements. In Fig. 3.11 (e), it is seen that there is one vertical line with high ampli-

tude around 1.5 Hz along the CVD 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. 3.11 (g). This peak represents the periodic motion of people. In contrast, because in the case of low synchronization the inconsistency of the two gaits leads to "unclean" CVD, it is not possible to easily extract periodic information (in Fig. 3.11 (f)). After the same approach is applied to obtain the CVD profile for the low synchronization case, this peak is less pronounced in the 0.5 to 2.5 Hz interval than in the high synchronization case as shown in Fig. 3.11 (h). All in all, the CVD is good for studying the synchronization of people's motion, which is connected with the number of Group People.

3.4.2. SYNCHRONIZATION FACTOR δ

After concluding that the CVD profile is able to provide information on the level of synchronization, a parameter that indicates this is described in this subsection.

By comparing the difference between CVD profiles of high and low synchronization in Fig. 3.11 (g-h), we can see that the shape of the peaks is highly differentiated between 0.5 and 2.5 Hz. As Fig. 3.12 shows, the higher the synchronization, the larger the peak and the greater the width of the impulse will be. Conversely, the lower the synchronization rate, the smaller the peak and the wider the impulse.

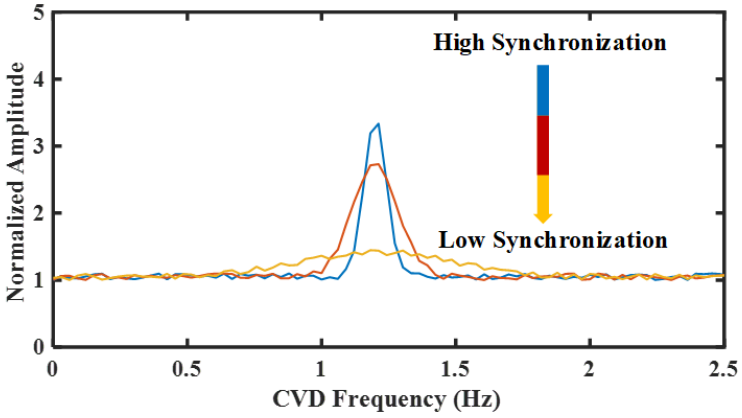


Figure 3.12: Example of CVD profile from the high synchronization case vs the low synchronization case.

To represent this difference, the synchronization factor δ is proposed in this thesis and its calculation is listed in *Algorithm 1*. This algorithm is similar to the calculation of the peak factor in [94]. To study the performance of the proposed synchronization factor, the training set of Data Set I (described in detail in Section 5.1) 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. 3.13, the fitted Gaussian distribution is proposed to study the synchronous phenomena in different numbers of Group People. It can be seen that as the number of people increases, the synchronization factor becomes smaller, while

for each multiple number of Group People, the reported higher value of this parameter is consistent and not an accidental phenomenon. This means that when designing a classifier, the difference between when the group gait is synchronised and when it is not needs to be taken into account by selecting suitable and sufficient features for this.

Algorithm 1 Propose synchronization factor based on CVD profile.

Require: CVD Profile $|Sc(g_1)|_k|$, in which $g_1 = 0, 1, ..., N_h - 1$, $k = 1, 2, ..., K$

Calculate ambiguity of CVD Frequency: $f_s = 1/(N_w \sigma_l T_c)$
Calculate CVD step: $x_c = \text{linspace}\,(-f_s/2, f_s/2, g_1)$
Select 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
$G_{max} = \max(|Sc_1(x_c)|_k|)$
$x_{c0} = \arg\max_{x_c}\,(|\mathrm{Sc}_1(x_c)|_k|)$
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

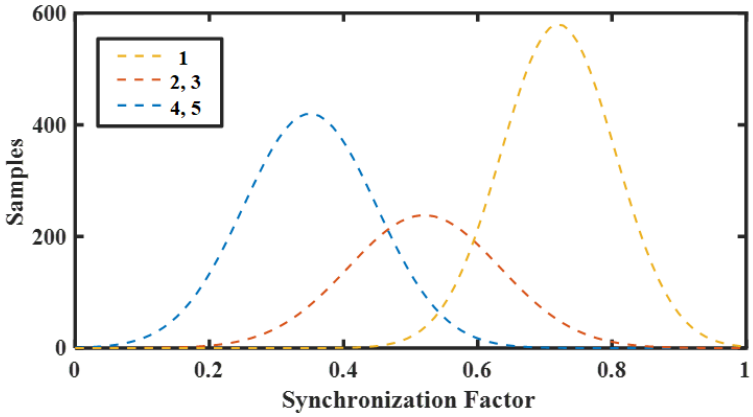


Figure 3.13: Fitted Gaussian distribution to the values of Synchronization factors when different number of Group People are moving in the RoI.

3.5. CONCLUSIONS

In this section, the characteristics of Group People have been studied.

Group people are defined as a group of people sharing neighboring, adjacent locations and moving together. As a starting point, existing data formats and features which have been already used in Radar-based People Counting in the literature, such as Range Profile, Range-Time map, and Range-Doppler map, all have similar signal or image mor-

phological characteristics regardless of how many people move as a group.

With the help of MIMO Radar, the angle information of the targets can be estimated. The Range-Azimuth map is more promising than the previous formats, as is shown by the fact that the more people move as a group, the more azimuth bins are occupied. But this feature on its own is vulnerable when Group People exist in the far range.

In addition to the Range-azimuth map, which can be used in the design of classifiers, spectrogram and CVD are found to be useful for Group People Counting. Existing features ignore that fact that people's movement is a periodic motion. The spectrogram of Group People that characterises this periodic motion is studied. Apart from the fact that the spectrograms are less clean when people walk at different paces than when one person walks, it is also possible for multiple people to walk at the same pace. With CVD, synchronization ratio is proposed to study whether people move in synchrony. It is concluded that as the number of people increases, the synchronization factor becomes smaller, while for each multiple number of Group People, people moving with relatively high synchronization is not an accidental phenomenon. This means that when designing the classifier, the difference between when the gait of groups of people is synchronized and when it is not, needs to be taken into account and suitable and sufficient features need to be selected for classification.

To conclude, features that are related to the Range-Azimuth map and spectrogram/CVD can be a good choice to solve Group People Counting.

4

Introduction of the Proposed Pipeline

After summarizing the advantages and disadvantages of the feature-based counting and tracking for counting methods, the proposed method combines them to deal with the grouping problem. This chapter is to describe in detail how those advantages are combined. In Section 4.1, the structure of the proposed pipeline is introduced. Then, how to generate and preprocess the Radar cube is given in Section 4.2. After that, the processed data is passed to the People Tracking and Counting block to estimate the number of people and their locations in the scene introduced in Section 4.3. In this section, for a complete exploration of the solution to Group People Counting, both non-ANN and ANN approaches are introduced to classify the grouping case. Finally, post-processing is provided in Section 4.4 to optimize the estimated results based on time information, and thus the exact people counting is achieved.

4.1. Overview of the Proposed Method

This section provides an overview of the proposed method, which combines the tracking for counting block and the feature-based counting block. In total, this method can be divided into four main parts, that is, Radar cube generation, pre-processing, people tracking and counting part and post-processing part shown in Fig. 4.1. The meaning of each section is briefly described in the following paragraphs, while the details are presented in the next sections of this chapter.

- The Radar cubes generation is the conversion of raw data from Radar data acquisition into a standard Radar cube (i.e., Channels × Fast Time × Slow Time × Frames) to facilitate the subsequent processing of the Radar cube.
- Pre-processing is to process the Radar cube and obtain the range, Doppler velocity, and angle information in the scene. Meanwhile, removing the static clutter is done to reduce the probability of miss detection and false detection.

- People tracking and counting part includes tracking for counting block and feature-based counting block. This part is designed to estimate the location and number of Group people. Thus, this is a coordinated application of tracking for counting and feature-based counting. It includes the counting unit which processes each decision window and estimates the number of Group People, and multiple target tracking (MTT) block to predict and update people's location in the scene.

- Post-processing is the process of updating the results for the current moment by combining the results before and after that point in time. In other words, the final estimated number of people is optimized before output based on the time information.

It is worth noting that there is a feedback loop in this method, where the counting unit and multiple target tracking interact and feed back into each other. That is, the decisions made by the counting unit or the multiple targets tracking algorithm are fed back to each other. For example, when one group stop moving in the scene, even though the counting unit thinks that there is no target in the scene because there is no Doppler information, the tracking algorithm will determine that the target has not left the scene based on its position and the predicted position information, and will continue to wait for the target to move, activating the counting unit to continue to update its estimate of the number of targets.

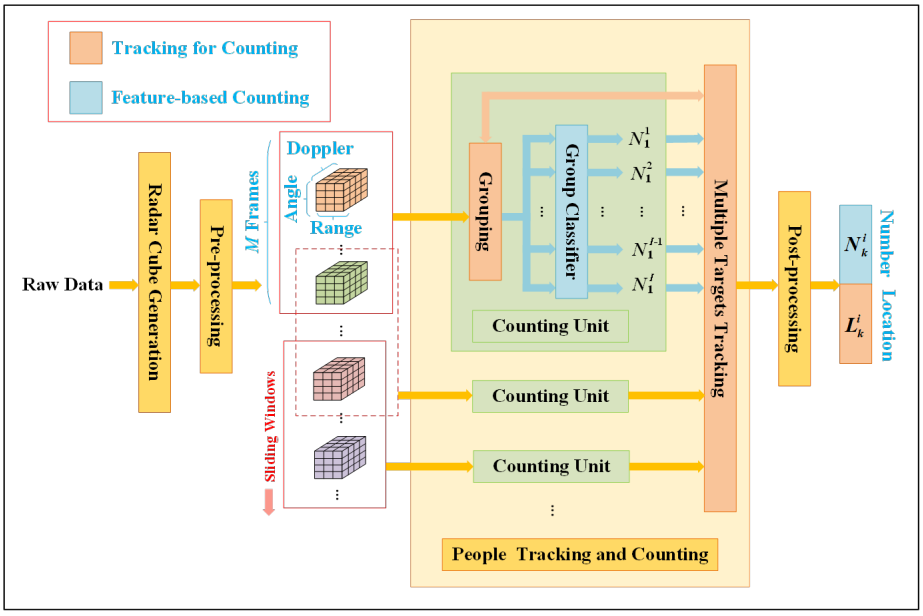


Figure 4.1: Overview of the proposed People Counting pipeline, which combines relevant element of the Tracking for Counting and Feature-based Counting approaches used in isolation in the literature.

4.2. RADAR CUBE GENERATION & PRE-PROCESSING

For conciseness, this section presents the first two parts of the proposed methodology, that is Radar cube generation, and pre-procesing.

The generation of Radar cubes is the conversion of raw data from Radar data acquisition into a standard Radar cube to facilitate subsequent processing of the Radar cube. Based on the use of MIMO radar, angle estimation of the target is possible, and the virtual antenna is generated (mentioned in Appendix A).

When using the FMCW MIMO Radar, the received raw data is the complex signal of each receiver, for example if the radar used has three Txs and four Rxs, the length of the received raw data is $(4 \times n_s)$, where n_s is the total number of samples of each Rx. But by analyzing the raw data at this point, we cannot know when the data received by the receiver was transmitted by which transmitter. Therefore, it is important and necessary to generate the radar matrix, and how to calculate it depends on the parameters of the radar and the parameters of the chirp. The *Algorithm* 2 is applied to reshape the cube into (Channels × Fast Time × Slow Time × Frames), when using the Time Division Multiplexing (TDM) mode of the MIMO Radar.

Algorithm 2 4D Radar Cube Generation.

Require: $X_r(n_{Rx}, n_s)$,chirps per frame n_c, samples per chirp n_{adc}, total number of samples of each receiver n_s, the number of transmitters n_{Tx}, the number of receivers n_{Rx}

1: Calculate the number of frames $n_f = n_s/(n_{Tx} \times n_c \times n_{adc})$
2: Calculate the number of virtual arrays $n_v = n_{Tx} \times n_{Rx}$
3: Calculate the Read Cube R_e = linspace $(1, n_s, n_s)$
4: Reshape the Read Cube to capture the samples generate by chirps of each Tx $R_e(n_s/(n_{adc}, n_{adc}))$
5: Use FOR loop to extract the samples transmitted by each Tx, and build the virtual array
6: **for** each $i \in [1, n_t, (n_s/n_{adc})]$ **do**
7: $X_R(n_a, n_{adc}, n_c, n_f) = [X_r(1, n_{Rx}, Re(i), n_f), ..., X_r(\max(n_a), n_{Rx}, Re(i), n_f)]$
8: **end for**

Ensure: $X_R(n_a, n_{adc}, n_c, n_f)$

Subsequent pre-processing aims to obtain the Doppler velocity, Range and Angle information from the Radar cube. Meanwhile, the static clutter is removed for more accurate and effective feature extraction, which is applied in later steps. The algorithm that implements this pre-processing part is placed in *Algorithm* 3.

It is worth mentioning that the estimation method of angular information in this paper is different from simply performing an FFT across the channel dimension of the cube for all range and Doppler bins. This commonly used FFT beamforming is to do the Fourier transform along the channel dimension of the cube, but in fact the angle information obtained in this case has a high sidelobe, which is not very good for determining where the target is. Especially for the extended target in the case of Grouping, it is not very easy to confirm the boundary of the target just by the amplitude of the resulting

angular profile.

In this thesis, the Doppler-based angle estimation method is applied and is included in *Algorithm* 3. The base idea is to generate the Range-Doppler map first. Then the detection method (i.e., 2D CA-CFAR) is applied to find the range-Doppler bins. 2D CA-CFAR is an adaptive algorithm, and is widely used in the Radar systems. The detection points are collected when the cell under test (CUT) is higher than the adaptive threshold, as shown in Fig. 4.2. The threshold is called adaptive, because the threshold is continuously updated by the training cells, which are around the CUT. Meanwhile, the guarding cells are used to avoid energy leakage from the CUT to destroy the estimate of threshold [95]. In this case, the static clutter is removed, and targets are located at certain range-Doppler bins. Finally, the FFT beamforming is performed at those range and Doppler bins along the Channel axis.

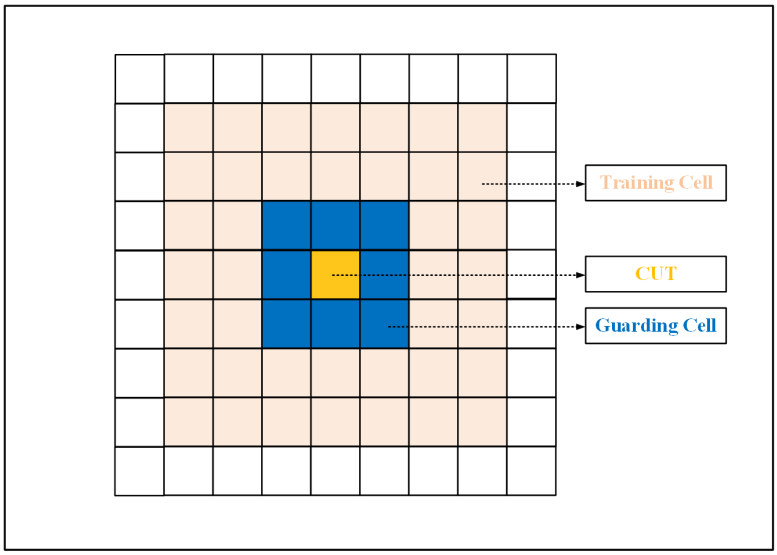


Figure 4.2: Example of the 2D CA-CFAR.

In Fig. 4.3, an example from Data Set III (later described in detail in Section 5.1) is used to show the results of the FFT beamforming and the Doppler-based angle estimation method. It can be seen that when the classic FFT beamforming is applied directly in all range bins, there is a lot of clutter, as shown in Fig. 4.3 (a). Thus, at low SNR conditions as in this case, it is difficult to determine the location and edge of targets when the detection method such as CA-CFAR is applied directly. However, according to 4.3 (b), only moving targets in the scene are focused and studied when the Doppler-based angle estimation method is applied.

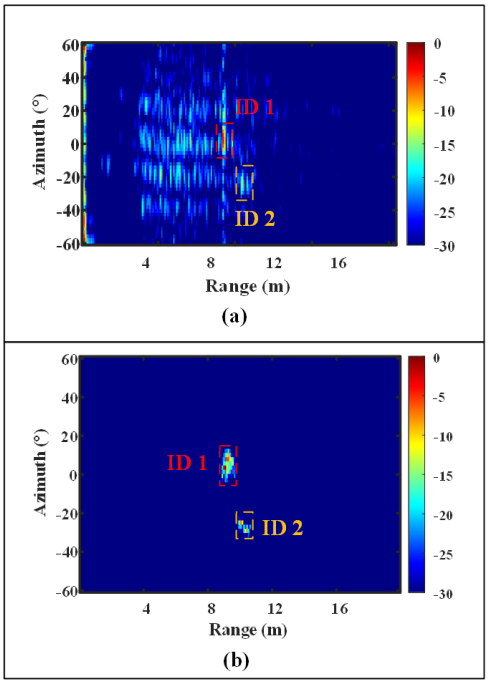


Figure 4.3: Results of angle estimation when applying (a) FFT beamforming, (b) Doppler-based angle estimation.

After the pre-processing part, moving targets are defined by values in the processed Radar cube (Angle × Range × Doppler × Frame). The M-frame sliding windows are applied to automatically transfer multiple frames to the people tracking and counting part, because the spectrogram required multiple frames to be generated, and the CVD map needs enough time window to extract the periodic motion. In this thesis, the length of the sliding windows M is set at 10. Meanwhile, the overlap ratio has to be determined. This represents the proportion of the mth window that overlaps with the $m-1$th window. In other words, this indicates that the number of frames is kept in the next decision window. In this thesis, the overlapping ratio is set at 90%. The decision window, as the name suggests, means that each time a window is entered, the number of people in the scene and their location are obtained by the people tracking and counting part (introduced in the next section).

4.3. PEOPLE TRACKING AND COUNTING

4.3.1. GROUPING

Grouping is the first block of the people tracking and counting part, and its mission 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 spectrogram/CVD map) to the Group classifier. In other words, clustering and assignment methods are used in this part, and its algorithm is provided in *Algorithm* 4.

Algorithm 3 Pre-processing steps of detection and estimation in range, Doppler, angle.

Require: $X_R(n_a, n_{adc}, n_c, n_f)$

1: **for** $n = 1, 2, 3, ..., n_f$ **do**

2: Range and Doppler estimation based on 2D-FFT $X_{\hat{R}}(n_a, n_R, n_D, i)$ =FFT2($X_R(n_a, n_{adc}, n_c, i)$)

3: Apply detection method CA-CFAR to determine the Range and Doppler bin for angle estimation $(n_{\hat{R}}, n_{\hat{D}}, i)$ =CA-CFAR($X_{\hat{R}}(n_a, n_R, n_D, i)$)

4: Apply angle estimation FFT beamforming at the selected range-Doppler bins along Channel axis $n_A = \arg\max_A(\text{FFT}(X_{\hat{R}}(n_a, n_{\hat{R}}, n_{\hat{D}}, i)))$

5: **end for**

Ensure: $X_{\hat{R}}(n_A, n_R, n_D, n_f)$

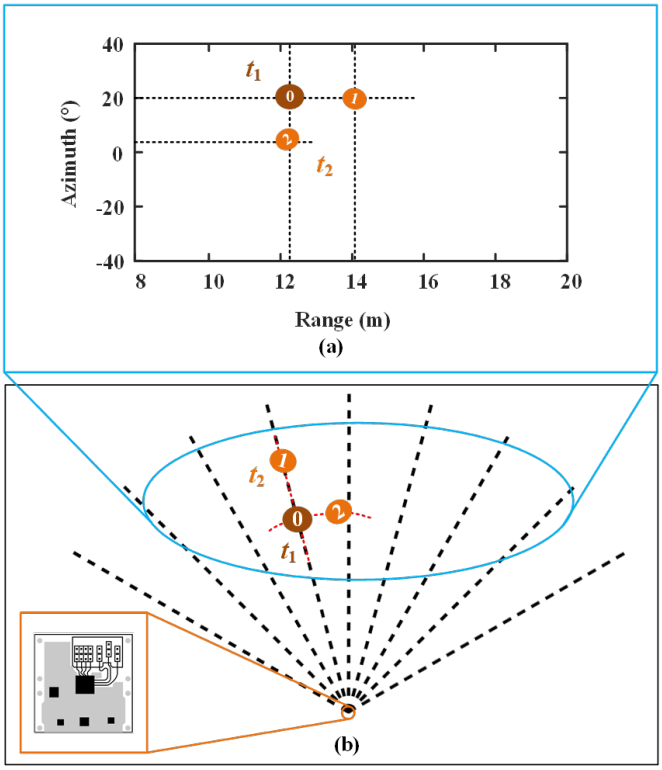


Figure 4.4: Difference between calculating the Euclidean distance (a) based on the Cartesian coordinate, (b) based on the principle of the Radar, where Target 0 at t_1 and Target 2 at t_2 locates at the same range; Target 0 at t_1 and Target 1 at t_2 locates at the same azimuth.

- **Clustering:**

  The use of clustering serves two purposes. Firstly, to calculate distances on the

Range-Azimuth map based on the density of measurements and thus to separate the different clusters within the scene, i.e., Group People. Secondly, to calculate the central position (cluster centroid) based on the measurements of each group, which is used in the tracking algorithm to reduce the complexity of the calculation.

In this thesis, the grid-based Density-based Spatial Clustering of Applications (DBSCAN) is applied [96]. Compared to standard DBSCAM, grid-based DBSCAN takes the relation between angle bin and range bin into account to calculate the distance that is suitable for the Radar principle.

As a result, different groups with their measurements are divided in the scene, and the central position of each group is calculated and passed to the MTT block.

- **Assignment:**

  The assignment algorithm has two purposes. The first is to assign the processed feature to the correct group. Let us discuss this with an example. Group 1 is detected walking from range bin 130 to range bin 137 in one window time, 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 (which means the hypothesis in the i th frame and the k th 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 thesis, the best association hypothesis is determined by the distance between each estimated position of the groups with the Gaussian model **??**.

  Compared to calculate the Euclidean distance based on the Cartesian coordinates, the principle of the radar, i.e. its data in a range-azimuth map rather than a Cartesian map, is taken into account. As shown in Fig. 4.4, there is a target (Target 0) in the area of 12 meters from the radar at an angle of 20 degrees to the Line of Sight (LOS) at t_1. At t_2, two groups appear at a distance of 14 meters from the radar at an angle of nearly 20 degrees to the LOS marked as Target 1 and at a distance of 12 meters from the radar at an angle of nearly 2 degrees to the LOS marked as Target 2. According to the Cartesian coordinate, Target 2 at t_2 is the selected point with the highest association hypothesis, as shown in Fig. 4.3 (a). However, in the actual case where Radar is used, sampling rates between different angle bins are different. In other words, even if the angle between the LOS of two targets and the radar differs by 15 degrees, when they are away from the Radar, the distance between the two is much larger in the actual range-azimuth map compared to the Cartesian coordinate. As shown in Fig. 4.4 (b), Target 1 at t_2 is the correct association target for Target 0 at t_1 with the highest association hypothesis.

4.3.2. NON-ANN GROUP CLASSIFIER

This subsection is an introduction to the proposed non-ANN classifier.

Through the analysis in Chapter 3, Range-Azimuth map and spectrogram are recommended as input to the classifier to deal with Group People Counting. However, when designing the non-ANN classifier, the challenge is that it is required that the features composing the feature vector be extracted artificially.

Algorithm 4 Grouping algorithm including a clustering and an assignment/association part.

Require: $X_{\hat{R}}(n_A, n_R, n_D, n_f)$,

1: Based on the sliding window, multiple frames are formatted into k windows, and each window includes ith frame.
2: **for** k=1, 2, ..., K **do**
3: **for** i=1, 2, ..., I **do**
4: Based on grid-based DBSCAN, determine the number of groups in the scene $E_(i, j)$, where j is the number of groups in the scene, and $j = 1, 2, 3, ..., J$.
5: Based on the Radar principle, select the nearest positions at $i+1$ with the best association hypothesis θ_k^i.
6: Assign the selected features to each group $f(E_{(i,j)}, k)$.
7: **end for**
8: **end for**

Ensure: $x_k^i(n_A, n_R, n_D), f(E_{(i,j)}, k), \theta_k^i$

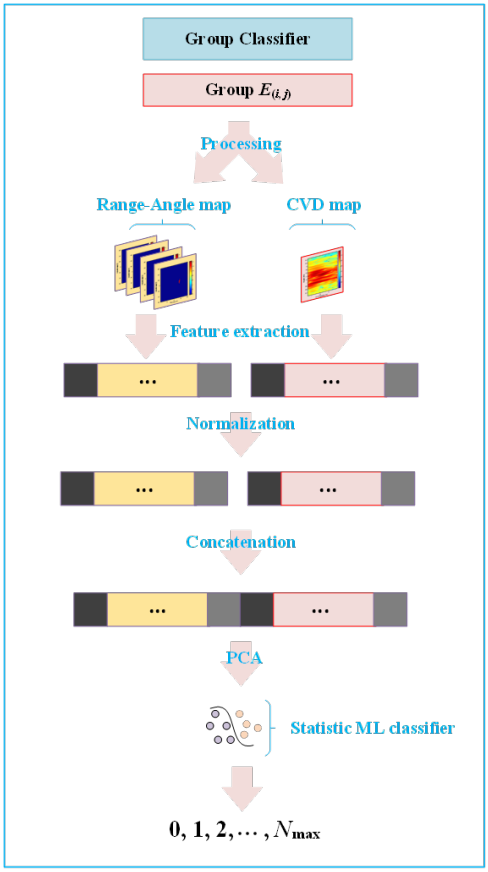


Figure 4.5: Overview of the proposed non-ANN group classifier.

The features of Range-Azimuth are very intuitive. based on the occupied azimuth bin, the difference in the number of Group People can be discerned intuitively in the near range. However, unlike the range-azimuth map, the features of group people are difficult to extract from the spectrogram. As concluded in Section 3.4, CVD is calculated from spectrogram to emphasize better the feature of periodic motion of people, because it is not easy to manually get features from spectrogram. So CVD map is used for the design of non-ANN classifiers where manually extracting features is a must.

As shown in Fig 4.5, Range-Azimuth map and CVD are used in the design of the non-ANN classifier. Range-Azimuth map and CVD features are described in detail in *Algorithm* 5. The features extracted from the Range-Azimuth map and the spectorgram are summarized below:

- **Range-Azimuth Map**

  When manually extracting features from the Range-Azimuth map for non-ANN classifiers, three kind of features are extracted, i.e., (1) the length of the occupied azimuth bin, (2) the peak amplitude, (3)the mean amplitudes are selected from each two frames. Deciding to use features each two frames is to reduce the overfitting, where the total number of features of Range-Azimuth map for classification can reach 30. So in this thesis, the length of the feature vectors extracted from the Range-Azimuth map is 15.

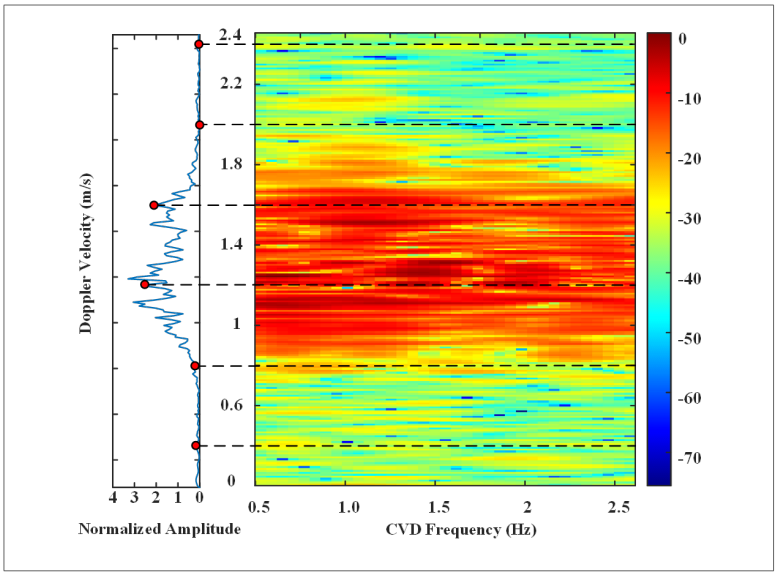


Figure 4.6: Example of the 4th types of the CVD map, where the length of selected mean amplitude is 6. The left figure is the mean amplitude from the CVD map along the Doppler velocity axis.

- **CVD Map**

  For the CVD map, four kinds of features are extracted for the classification, i.e., (1) the synchronization factor δ (studied in Section 3.4), (2) the maximum CVD

frequency of the CVD profile, (3) two half-height CVD frequencies from the CVD profile, and (4) the selected mean amplitudes of the CVD map along the Doppler velocity axis are used, inspired by Ricci et al. [97]. In this thesis, the length of the selected mean amplitude is 6, as shown in Fig. 4.6. So in this thesis, the length of the feature vectors extracted from the spectrogram is 10.

In total, 25 features are used for the proposed non-ANN classifiers. Then, the normalization is applied to each selected features from Range-Azimuth map and CVD map. It enables the eigenvalues of the samples are converted to the same scale, which in turn improves the speed of the gradient descent method to solve the optimal solution for ANN classifiers. Some non-ANN classifiers also require feature normalization, e.g., Support Vector Machine (SVM), k-nearest Neighbors Algorithm (KNN) [98].

Then, two sets of feature vectors are concatenated and passed to principal component analysis (PCA) to reduce the dimensionality of the predictor space. Thus, it is to make sure all components used for the classification are useful for classification. Reducing the dimensionality can help classification models prevent overfitting [27]. In this thesis, the PCA is required to keep only the principal components that explain 80% of the variance.

After the PCA, the feature vectors are passed to the statistical machine learning classifier. The selection of the classifiers and the study of the proposed non-ANN classifier performance are provided in Section 5.2.

Algorithm 5 Feature extraction of Range-Azimuth map and CVD map.

Require: $f(E_{(i,j)})$
for $j = 1, 2, ..., J$ **do**
  if Determining odd and even numbers of I, $\text{mod}(I, 2) == 1$ **then**
    $I = I - 1$
  end if
  for $i = 1, 2, ..., I/2$ **do**
    Select the maximum occupied azimuth bin of each two frame $f_1(E_{(i*2,j)})$ and record its range bin $f_2(E_{(i*2,j)})$
    Select the maximum amplitude of each two frame $f_3(E_{(i*2,j)})$
  end for
  M frames are used to generate CVD map, $f(E_{(:,j)})$ is used.
  Select the synchronization factor δ_j
  Select the maximum CVD frequency of the CVD profile $f_4(E_{(:,j)})$
  Select the half-height CVD frequency $f_5(E_{(:,j)})$
  Select mean amplitude of the CVD map along the Doppler velocity axis $f_6(E_{(:,j)})$
end for
Integrate the selected features into two sets of feature vectors $\hat{f}_r(E_{(i,j)}) = [f_1(E_{(i*2,j)}), f_2(E_{(i*2,j)}), f_3(E_{(i*2,j)})]$, and $\hat{f}_s(E_{(i,j)}) = [f_4(E_{(:,j)}), f_5(E_{(:,j)}), f_6(E_{(:,j)})]$
Ensure: $\hat{f}_r(E_{(i,j)}), \hat{f}_s(E_{(i,j)})$

4.3.3. ANN GROUP CLASSIFIER

This subsection introduces the proposed ANN Group classifier.

The previously described Non-ANN Group classifier uses Range-Azimuth map and CVD to solve the Group People Counting problem, while ANN Group classifier is required to be more "artificial". "Artificial" means that the ANN classifier is required to automatically extract the salient features, so the two matrices of the Range-Azimuth map and spectrogram are directly input to the proposed ANN group classifier. So in this case, compared to extracting the feature from the Range-Azimuth map in several frames separately when designing the non-ANN classifier, Range-Azimuth maps with M frames are averaged and then inputted to the classifier.

The pipeline of the proposed ANN group classifier is shown in Fig. 4.7. Inspired by Choi et al. [37], there are mainly three blocks of the pipeline.

The first is the feature extraction block. In contrast to manual feature extraction, CNN neural networks are applied to learn and extract features on such two-dimensional matrices. Rather than building CNN layers at random, this thesis chose to use some layers of ResNet18 [99] as a backbone. Then, after concatenating the two sets of outputs obtained through the CNN layers, the transformer learns feature, which are extracted from the two images, in parallel using the Multi-Head Attention mechanism [73]. Finally, the fully connected (FC) layer is used to output the likelihood of each class, which is the number of Group People. The detailed structure of the proposed ANN Group classifier is listed in Table 4.1.

Comments on each block of the proposed ANN classifier are provided.

- Feature extraction

  For Radar data, the number of Radar data is limited, which requires the network structure not to be too deep. Therefore, ResNet-18 is used for feature extraction of the 2D matrix. Compared with adding CNN layers arbitrarily, the degradation problem with increasing number of layers can be solved effectively and the fitting ability can be improved with the help of residual blocks [99].

  In this thesis, the first three blocks of ResNet-18 are used for feature extraction, because when using all the blocks of ResNet-18, overfitting occurred. The input channel of the first block of the ResNet-18 backbone is changed to 1, because the input is not a picture, there is no information of the three channels of RGB.

- Attention

  One of the advantages of Transformer is the ability to achieve parallel computation when solving sequences. In classical RNN structures (LSTM [72], GRU [100]), they process them one by one according to time. The hidden state h_t at time t is determined by the previous hidden state ht-1 and the current input. This results in the current moment having to wait for the historical results to be updated, and very early historical information is discarded because of the performance impact. But transformer is able to read the whole matrix at the same time, for example, the full motion of people can be processed in parallel rather than step by step. This broader inductive bias enables it to process generalized information [101].

Because of the input format requirements of the transformer, the spectrogram and Range-Azimuth map are extracted by the CNNs separately and integrated into a sequence, which is then input to the transformer.

- FC

  In the ANN classifier, the last layer of this classifier is the FC, which outputs the probability of each class. Thus, the estimated number of Group People is outputted by selecting the label with the largest probability.

  Due to the multi-head attention architecture in the transformer model, the output sequence length of a transformer is the same as the input length. Thus, the number pf the output channel of the FC is the maximum number of Group People $N_{max}+1$.

The performance study of the proposed ANN classifier are provided in Section 5.2.

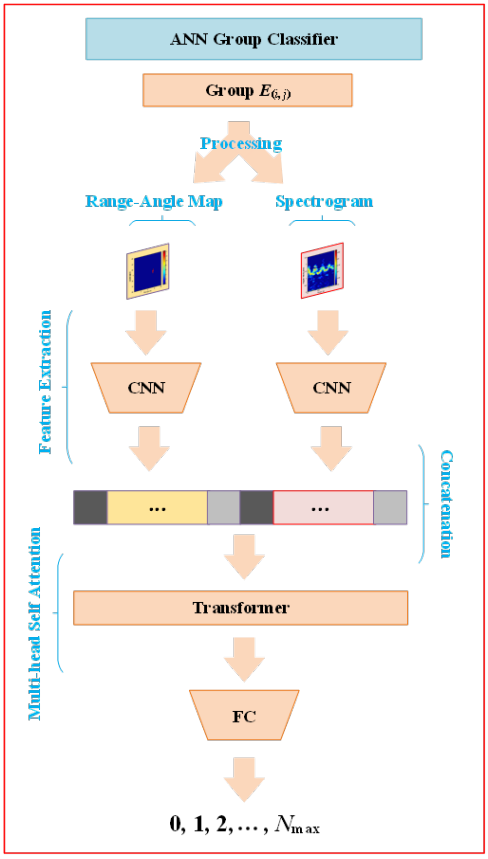


Figure 4.7: Overview of the proposed ANN group classifier.

Table 4.1: Detailed architecture of the proposed ANN Group classifier.

Part	Layer	Output Size	Kernel
Feature Extraction	Conv1	$64 \times 112 \times 112$	7×7, 64, stride 2
	Conv2_1	$64 \times 56 \times 56$	3×3 max pool, strid 2
	Conv2_2	$64 \times 56 \times 56$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
	Conv3	$128 \times 28 \times 28$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
	Ave1	$128 \times 1 \times 1$	average pool
Concatenation		$256 \times 1 \times 1$	
Attention	Transformer	$256 \times 1 \times 1$	$d_{model} = 256$, $n_{head} = 4$
Classification	FC	$(N_{max} + 1) \times 1$	$I_f = 256$, $O_f = N_{max} + 1$

4.3.4. MULTIPLE TARGETS TRACKING

After the best association hypothesis is selected, this part is to update all tracks from the Range-Azimuth map, and initiate new tracks if needed. This processing is part of the Global Nearest Neighbor (GNN) tracking, which is used for tracking N-object [102], and the motion model of target is the constant velocity model. The algorithm is listed in *Algorithm* 6. Compared to other MTT algorithms such as Gaussian-mixture probability hypothesis density (GM-PHD), Poisson multi-Bernoulli mixture filter (PMBM) [58, 59, 60], GNN tracking algorithms require fewer models of targets and motions to be built, and thus save the computation power, due to the fact that the complex motion in this thesis is only the grouping. The core of the GNN tracking is actually the extended Kalman filter which includes the prediction, and update step.

There are two important cases for the feedback loop of the proposed method.

- **False detection and miss detection**

  For possible false detections and missed detections, the thesis divides the trajectories into two categories, one is the confirmed trajectory and the other is the tentative trajectory.

  A confirmed trajectory means that the best association hypothesis θ_t for the target can always be found at t. The tentative trajectory means that the target with the maximum association hypothesis is not found at the next moment. For tentative trajectories, one of two reasons will make the algorithm decide to delete the trajectory. The first is that if the number of invisible trajectories exceeds a certain value, then the trajectory is deleted. This eliminates false detections due to environmental noise and ghost targets generated by the multi-path effect. The second is that the group classifier continuously determines that there is no person on these tracks. It can eliminate dynamic clutters in the scene.

- **Static group in the scene**

  For the case where the target moves for a period of time after entering the scene and then remains static in the scene, two reasons will allow the tracking algorithm

to maintain the tracks. One is that the Group classifier estimates that there is at least one person in this group at the previous time. The second is that the target has not left the scene by the prediction of the tracking algorithm. Later, if the target starts to move again, the grouping and MTT algorithm will continue to implement People Tracking and Counting using the position where it stayed as the target birth place.

From this we can also see that the feature-based counting block and the tracking for the counting block can help each other to make better decisions, for both position estimation and People Counting.

Algorithm 6 Tracking via GNN approach.

Require: $x_k^i(n_A, n_R, n_D), f(E_{(i,j)}, k), \hat{N}_k^i, \theta_k^i$
for $k = 2, 3, ..., K$ **do**
  for $i = 2, 3, ..., I$ **do**
    Initializing Tracks $L_k^i(x_k^{i|i})$
    Given θ_k^i
    Prediction: Chapman-Kolmogorov prediction
    $p_k^{i|i-1}\left(x_k^i\right) = \int \pi\left(x_k^i \mid x_k^{i-1}\right) p_k^{i-1|i-1}\left(x_k^{i-1}\right) \mathrm{d}x_k^{i-1}$
    Update:
    The exact posterior density is approvimated by GNN density parameterized by the object densities $p_{k|k}^{\theta_{1:k}^\star}(X_k)$
    $p_k^{i|i}\left(x_k^i\right) = P^{\mathrm{D}}\left(x_k^i\right) g_k\left(z_k^{\theta^\star,i} \mid x_k^i\right) p_k^{i|i-1}\left(x_k^i\right)$
    if $L_k^i(x_k^{i|i}) == 0$ **then**
      Assign the trajectory to tentaive trajectory $L_k^i.ID == 0$
      if the number of invisible trajectories exceed σ_{ID} & $\hat{N}_k^i == 1$ **then**
        Delete $L_k^i(x_k^{i|i})$
      end if
    end if
  end for
end for
Ensure: Locations and tracks of each group in the scene L_k^i

4.4. POST-PROCESSING

After the number of Group people and their location are estimated by the People Tracking and Counting block, with the help of tracking for counting block, the estimated number of people in each group are recorded in chronological order.

Then, the post-processing part is to combine results before and after this time to update the results for the current moment. In other words, the final estimated number of people is optimized based on time information. In the thesis, the five-point median filter is applied to average the estimation. The results of this part are shown in Section 6.3.

The algorithm of post-processing is provided as follows:

Algorithm 7 Post-processing to the estimated number of each group from the group classifier.

Require: Estimated results from the proposed group classifier $\hat{N}_k^i$

Given the length of time used for consideration m, m is odd number.
for k=1, 2, ..., K **do**
  for i=1, 2, ..., I **do**
    $m_h == \text{floor}(m/2)$
    $N_k^i = \text{mean}(\hat{N}_{k-m_h}^i, ..., \hat{N}_{k-1}^i, \hat{N}_k^i, \hat{N}_{k+1}^i, ..., \hat{N}_{k+m_h}^i)$
  end for
end for
$N_k^i = \text{Rounding}(N_k^i)$

Ensure: N_k^i

5

PERFORMANCE VALIDATION

In this chapter, the performance of the proposed Radar-based People Counting method is verified using experimental data sets collected from outdoors environment. In Section 5.1, the basic information about the equipment and its configuration are introduced. To evaluate the performance toward People Counting, three representative case studies are selected to show the performance of tracking part, Group People classification part, and the proposed method where the first two parts are integrated into a whole. The results are shown in Section 5.2. Then, the cause of incorrect results is studied in Section 5.3, which can be improved in future investigation. Finally, the conclusions of performance validation are drawn in Section 5.4.

5.1. EXPERIMENTAL SETUP AND DATA COLLECTION

This section aims to introduce the basic information about the experimental equipment and setup. In addition, some supporting explanation of TI Radar is provided in Appendix A.

In the thesis, the TI 60 GHz radar with associated hardware is chosen for People Counting purpose. The hardware in addition to IWR6843ISK Radar includes mmWave ICBOOST board, DCA1000EVM as shown in Figure 5.1. TI is a good tool among various commercial off-the-shelf radars. It makes the radar easy to control and record data. The TI Resource Explorer (`https://dev.ti.com/tirex/explore/node`) offers pre-defined functions to start with. Meanwhile, the TI radar has relatively cheap prices and offers off-the-shelf hardware.

IWR6843ISK is a 60 GHz mmWave sensor based on the IWR6843 device with long-range antenna. Three transmitters (Txs) and four receivers (Rxs) form a virtual 12-channel array. This board enables access to point-cloud data over USB interface. With the help of DCA1000-EVM, the raw data rather than point-cloud data is available. In this thesis, the raw data is directly processed instead of using point-cloud data processed by TI post-processing methods. Supported by the user manual of TI Radar [103], the custom configuration of TI Radar is designed for outdoor People Counting scenario, with parameters listed in Table 5.1. The maximum detection range is about 20 meters and a field of

view (FoV) of 120° region is chosen for experiments. To the best of my knowledge, this experimental region is wider than the existing Radar-based People Counting work [27, 37], setting a more interesting and challenging problem.

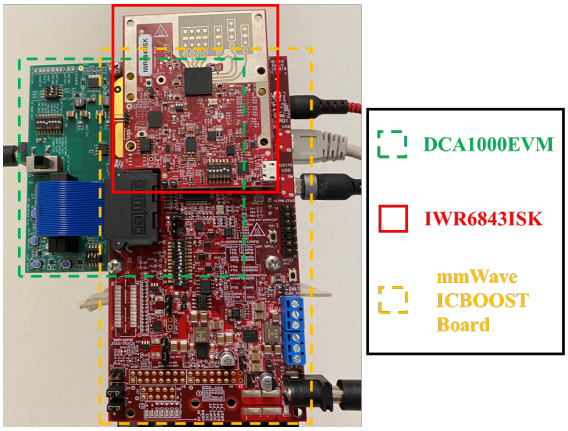


Figure 5.1: Setup of the People Counting equipment (1-mmWave ICBOOST board, 2-DCA1000EVM, 3-IWR6843ISK Radar).

The experiments were carried out in an outdoor *open* area where less multi-path and clutter effect occurred, mentioned in Subsection 2.2.1. The reason for choosing an outdoor scene is that the outdoor environment tends to have more rules and regulations on privacy protection than the indoor scene, such as the General Data Protection Regulation in the Netherlands [104]. Therefore, outdoor scenes are more common in Radar-based People Counting. The reason for choosing the open scene is that this paper focuses on solving the problem of Group People rather than removing the multi-path effect or clutter effect, so open scenes are used for experiments. A fan-shaped scene with a radius of 20 m and FoV of 120° was chosen for experiments as shown in in Fig. 5.2.

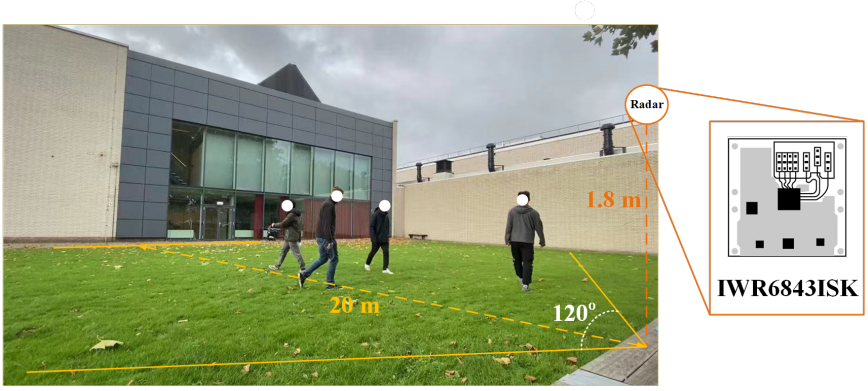


Figure 5.2: Overall measurement scenario of the proposed People Counting method in Case 2.

Three types of data sets are collected and constructed in the scene mentioned above (in Table 5.2). Due to the fact that the proposed method includes feature-based counting block and tracking for counting block, the three Data Sets have separate purposes. *Data set I* is used to testing the performance of the feature-based counting block and is also used for comparisons when applying SOTA methods. *Data set II* is used to evaluate the capability of tracking for counting block, and thus does not have training set as summarised in one of the advantages of *tracking for counting* methods in Table 2.1. *Data set III* is used to evaluate performance when combining feature-based counting block and tracking for counting block as a whole.

To ensure the reproducibility of the experiments, two approaches are followed. The first approach is to ensure the diversity of volunteers. Similarly to the experimental design of the authors Choi et al. [27], 15 volunteers were invited to participate in the experimental data collection shown in Table 5.3. Their heights varied from 1.65 m to 1.90 m, and their physiques basically covered the most common physiques in the current society. The second experimental approach was to ensure that the volunteers involved in the acquisition of the test set and the training set were not the same, while the acquisition of the test set and the training set was carried out at different times.

Table 5.1: Basic setting of IWR6843ISK and its configuration in this thesis.

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

To study Group People more systematically, three experimental assumptions were proposed. These assumptions help constrain the problem to a manageable complexity within the scope of the thesis.

- **Assumption 1: People in RoI will always enter and leave the scene under Radar surveillance.**

  In other words, people will not remain stationary in the scene. The effect of clutter is unavoidable when using Radar. The difference between a stationary clutter and

a person is that the person will always move in the scene. Based on this assumption, tracking algorithms were developed. A stationary clutter such as a tree in an outdoor scene will not be tracked and sent to classifier because the target has not moved in a limited time. Therefore, Assumption 1 is necessary.

- **Assumption 2: Group People in RoI will not be mixing, spawning, or occluded.**

  People's motion is complex and various. This thesis is focusing on study grouping. Thus, for Group People, other complex motions are not included. To better help distinguish different complex motions, mixing, spawning and occlusion events are explained. Mixing is an event when multiple targets share the same tracks at a certain time [105]. Spawning (also referred to as splitting) is an event where a single extended target separates into two or more extended targets, or vice versa [106]. Occlusion means that an existing target is obscured by a new target or obstacle, resulting in a miss detection [107]. This assumption of occlusion applies to the motions of people in the same group and the motions between different groups.

- **Assumption 3: When people move in RoI, they only walk instead of running.**

  Usually, the walking velocity is between a 1.3 meters per second and 1.6 meters per second [108]. Based on this fact, the configuration of Radar set 2.5 meters per second as the ambiguity of Doppler velocity.

Table 5.2: Description of Collected Data Sets (each group has maximum 5 people).

	Training Set	**Testing Set**
Data Set I	• Maximum one group in the RoI. • Approximately 6 minutes for each number of Group People	• Maximum one group in the RoI. • Approximately 1.5 minutes for each number of Group People
Data Set II	No training set for tracking for counting block	• Multiple people in the RoI without grouping. • Approximately 30 seconds for the testing data and maximum 4 people in the RoI
Data Set III	Using the Training Set of Data Set I	• Multiple groups (more than one and maximum three groups) in the RoI. • Approximately 1.5 minutes for each number of people in the RoI (ranged from 1 to 6).

Table 5.3: Physical characteristic of participants in the data set (in Gender, M represents male, and F represents female).

Individual	A	B	C	D
Gender	M	M	M	F
Height (m)	1.70	1.71	1.90	1.82
Weight (kg)	69	72	79	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	70	70	68
Individual	M	N	O	/
Gender	F	F	M	/
Height (m)	1.67	1.73	1.84	/
Weight (kg)	48	52	78	/

5.2. CASE STUDY

5.2.1. INTRODUCTION OF PERFORMANCE METRICS

The performance metrics are used to evaluate the capability of the proposed method. Due to the fact that the proposed method combines tracking for counting block, and feature-based counting block, it requires the performance metrics that can be used to jointly study performance of both blocks.

Three factors are included in the metrics, that is, Mean Square Error (MSE), Average Probability of True Positives (ATP), and Multiple Groups Tracking Accuracy (MGTA). MSE and ATP are used to evaluate the performance of the classifier.

Inspired by Choi et al. [37], MSE is able to jointly study how much a prediction differs from the true number of people in the RoI. Based on it, MSE reflects the accuracy and precision of the classification process, which is defined as

$$\mathscr{L}^{MSE} = \frac{\sum_k \sum_i (\hat{N}_k^i - N_k^i)^2}{\sum_k i_{max}}, \tag{5.1}$$

where i_{max} represents the number of Groups in a certain window. N_k^i represents the predicted number of people in ith Group at the kth window, and $\hat{N}_k^i$ is the true number of people at the same time. Thus, $\hat{N}_k^i \in \mathbb{N}$ and $N_k^i \in \mathbb{N}$ (set of non-negative integers). Unlike the general MSE used in Radar-based People Counting [37], the denominator is the number of decisions and is greater than the total number of people in the RoI (which is the denominator in Eq. 5.1). Therefore, the MSE used in the thesis is to focus on the hard case where numerous people are in the scene rather than reducing the weights of the case where there are a large number of people in the surveillance area.

ATP is used to study how well each label is trained. This factor is necessary for the Group People case, which is newly introduced in the Radar-based People Counting field. It is defined as

$$\mathscr{L}^{ATP} = \frac{\sum_{C_\gamma} TP}{|C_\gamma|}, \tag{5.2}$$

where $|C_\gamma|$ represents the number of trained labels. In this thesis, γ equals 6, and thus C_k is ranged from 0 to 6 and $C_k \in \mathbb{N}$. TP is the probability of true positives of the confusion matrix. Since each class of the training set is balanced, it is suitable to use ATP to study how well each class is classified.

Finally, MGTA is proposed to study the performance of tracking for counting block. In the general case of multiple targets, Multiple Objects Tracking Accuracy (MOTA) [109] is used to study the case of missed/false detection and mismatch of tracking methods. But in this thesis, grouping is introduced, and a single identity (ID) is used to represent one whole group instead of using different IDs to represent people in a group shown in Fig. 5.3. Therefore, general MOTA is not suitable in this thesis. Moreover, the value of MOTA can also result in negative values when the number of errors is larger than the number of targets in the RoI [110]. To overcome it, MGTA is proposed and defined as

$$\mathscr{L}^{MGTA} = \frac{\sum_k (MD + FD + IDS)}{\sum_k N_t^{total}}, \tag{5.3}$$

where N_t^{total} is the total number of people in the RoI during each windowing time. MD means the number of missed detections, and $MD \in \mathbb{N}$. FD is the false detection and $FD \in \mathbb{N}$. IDS is the identity switch and $IDS \in \mathbb{N}$. The identity switch (named miss match) influences the tracks of each group, and it could influence the accuracy when assigning different group's information to the Group classifier in the proposed method.

An example of MD, FD and IDS is shown in Fig. 5.3 where Fig. 5.3 (a) is the ground truth at t_1, and Fig. 5.3 (b) detection at a later time instance t_2. The identity of ID1 and ID2 switches at t_2, which is an example of the IDS event case. The MD occurs for ID2, where one person is missed from the cluster of the group. This kind of error can be automatically calculated by analyzing the points that are not clustered into the group with the help of the ground truth. ID3 in Fig. 5.3 (b) is an example of the FD, where the group truth does not have people at that predicted region.

(a) (b)

Figure 5.3: Example of calculating MGTA. (a) People Counting scenario in the real world (ground truth) at t_1. (b) Predicted identities of groups from tracking for counting methods at t_2.

5.2.2. CASE 1: GROUP PEOPLE COUNTING IN A SINGLE GROUP

In this case, the tracking for counting block in the proposed method is not available, and only the feature-based counting block is used (highlighted in Fig. 5.4). In this case, the proposed method can only output the estimated total number of people in RoI N_k^{total}. It is required that up to one group exist in the RoI. Based on it, Data Set I is used for study (mentioned in Section 5.1). The Data Set contains approximately six minutes of each number of Group People for training and 1.5 minutes of each number of Group People for testing, as discussed in Section 5.1.

This thesis provides two classifiers to solve the Group People Counting problem, i.e. the non-ANN group classifier and the ANN group classifier. In this subsection, the performance of both classifiers is tested as follows.

A. Non-ANN group classifier

There are two parts in which to study the performance of the proposed non-ANN classifier (introduced in Subsection 4.3.2). The first part is to study which non-ANN classifier can be used for the proposed non-ANN Group People Classifier. The second part is the "ablation study" of the proposed non-ANN classifier to examine how much each input contributes to the overall accuracy. For the remainder of the reader, the operation of the "ablation study" of the non-ANN classifier could also be called feature selection or input selection.

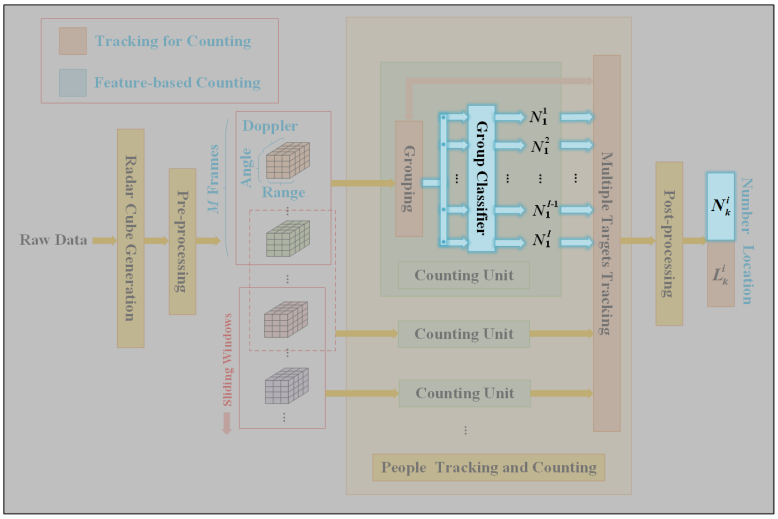


Figure 5.4: Overall scheme of the proposed People Counting method in Case 1 where the focus is on the Group People classifier block.

There are various types of classification algorithms, and the optimal classification algorithms are different depending on the feature vectors. For the best results, four common non-ANN classification algorithms are used and compared. They are: 1) Naïve

Bayes; 2) Random Forest; 3) Support Vector Machine (SVM) and 4)k-nearest Neighbors (KNN) classifiers. The parameters of each classifier are determined by the Hyperparameter optimization [111], and the design of classifiers are introduced as follows.

- **Naïve Bayes**

  The Naïve Bayes was created with Gaussian Kernel and calculated the probability that a measurement belongs to a label. Then the estimated number of people N is calculated by selecting the class which has the highest probability:

$$N = \underset{k\in\{1,\dots,K\}}{\operatorname{argmax}} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k), \tag{5.4}$$

  where N is the estimated number of Group People, and n features are applied. C_k is the classes with the number of k. In this thesis, k equals 6, and thus C_k is ranged from 0 to 6 and $C_k \in \mathbb{N}$. The probabilistic model is selected to fit the Gaussian distribution:

$$p(x \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}, \tag{5.5}$$

  where σ and μ represents the variance and mean of x.

- **Random Forest**

  The Random Forest was created with maximum 496 splits and 30 learners. The training algorithm for random forests applies the bagging to tree learners. Bagging repeatedly select a random sample with the replacement of the training set and fits tress to these samples. The predicted classes by averaging the predictions from all the individual regression trees:

$$N = \frac{1}{B} \sum_{b=1}^{B} f_b(x), \tag{5.6}$$

  where B is the number of individual regression trees and $B \in \mathbb{N}$.

- **SVM**

  The SVM classifier was created by using the Gaussian kernel and the scale of kernel is 0.75. The SVM is to find the best hyperplane that separates given data points of one labelled class from those of the other classes. The margin between two classes is used to study whether the hyperplane is promising or not, which is calculated the maximal width of the slab parallel to the hyperplane that has no interior points. Due to the fact that the data points are not linearly separated, the SVM is calculated by minimize :

$$\left[\frac{1}{n}\sum_{i=1}^{n} \max\left(0, 1 - N_i\left(\mathbf{w}^\top \mathbf{x}_i - b\right)\right)\right] + \lambda\|\mathbf{w}\|^2, \tag{5.7}$$

where $\max\left(0, 1 - N_i\left(\mathbf{w}^\top \mathbf{x}_i - b\right)\right)$ is the Hinge loss used for training the SVM classifiers. w and b are used to calculate the estimated labels, and N_i is the i th data points.$\lambda\|\mathbf{w}\|^2$ is to limit the margin size to let x_i assign on the correct side.

- **KNN**

  KNN is designed with a distance weight, where the weight is calculated by the squared inverse of the distance. The number of neighbors is set to 10. The data point is assigned to a class by a plurality vote of its k nearest neighbors. The distance is calculated by the Euclidean distance:

$$d_E\left(x^{(1)}, x^{(2)}\right) = \left(\sum_{i=1}^{n} \left|x_i^{(1)} - x_i^{(2)}\right|^2\right)^{\frac{1}{2}}, \tag{5.8}$$

  where $x^{(1)}$ and $x^{(2)}$ are two feature vectors from the input vectors. d_E is the Euclidean distance.

After introducing the design of four classifiers, Table 5.4 provides the results when using the Data Set I, and MSE and ATP are used for performance comparison, since this section does not include the tracking block. The confusion matrix of calculating MSE and ATP are listed in the Appendix B.

Table 5.4: Summary of the results of Data Set I when applying the feature-based counting block with different non-ANN classifiers. (RA represents the Range-Angle Map, CVD represents the CVD map)

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	<u>0.58</u>	<u>94.32</u>
RA + CVD	KNN	1.35	90.95

From the values of MSE and ATP, it can be analyzed to show that SVM has both the smallest MSE and the largest ATP, which proves that SVM is the best choice among four classifiers. This is attributed to two advantages of SVMs. The first is that the theoretical basis of the SVM method is a nonlinear mapping, and the SVM uses an inner product kernel function instead of a nonlinear mapping to a higher-dimensional space. The second point is that the final decision function of SVM is determined by only a few support vectors, and the complexity of the computation depends on the number of support vectors rather than the dimensionality of the sample space, which in a sense avoids the "curse of dimensionality".

Random Forest and KNN have similar performance, while the Naïve Bayes perform the worst. The main difficulty in estimating the posterior probability $p(C_k \mid x_i)$ based on the Bayesian formulation is that the conditional probability of the class $p(x_i \mid C_k)$ is a joint probability over all attributes, which is difficult to estimate directly from a limited number of training samples. To avoid this obstacle, the plain Bayesian classifier uses the assumption of conditional independence of features: for known classes, all features

are assumed to be independent of each other [112]. In other words, it is assumed that each feature independently affects the classification result. Therefore, the correlation between features are still strong, and thus Naïve Bayes is less effective in classification.

In the second part, an ablation study of the proposed non-ANN feature-based counting method is provided. This is to study whether both the Range-Azimuth Map and CVD are valuable in solving Group People Counting problem.

As summarized in Section 3.3, the Range-Azimuth Map is distinguishable for Group People in the near range where different number of Group People occupies distinct azimuth bins based on the angle estimation. However, Group People with similar number (i.e., three or five people walking as a group) occupies similar bins in the far range, empirically defined as approximately beyond 12m here. Because of it, the CVD is introduced and used for the proposed non-ANN Group classifier. To test the performance of each feature, Data Set I is selected in the far range and SVM is used for classification based on the study in the first part of this section. The border of the far and near range is defined based on the model of Group People and volunteers, and they are explained in Section 3.2.

From Table 5.5, when the CVD and Range-Azimuth maps are fused, MSE and ATP are improved. The fused results perform better than in the case where one of the features is used in isolation. Moreover, the MSE is really large when only CVD is used. This is based on the fact that CVD is processed from the spectrogram (in Fig. 5.5 (a)), which does not include the Range and Azimuth information of the targets. The energy-based features extracted from CVD are not robust enough due to the range and azimuth attenuation of Radar. In other words, CVD features are vulnerable to classify this case when Group People with a small number walk in the near range, while Group People with a large number walk in the comparatively far range or vice versa. This finding can be proved by Fig. 5.5 (b c), where the normalized energy could indicate this attenuation. Therefore, the Range-Azimuth information is necessary and import for Group People Counting.

Through the table, we can draw two conclusions. First, the features extracted from the Range-Azimuth map are promising to classify the number of Group people in the near range, while in the far range those features are not enough alone. Second, the features of the CVD map could improve the group counting accuracy in the far range, but only if they are linked to the location information (i.e., range, azimuth).

Table 5.5: Summary of the ablation study of the proposed non-ANN group classifier. (RA represents the Range-Angle Map, CVD represents the CVD map, and "Far" means features captured in the far range are used.)

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	<u>1.14</u>	<u>89.58</u>

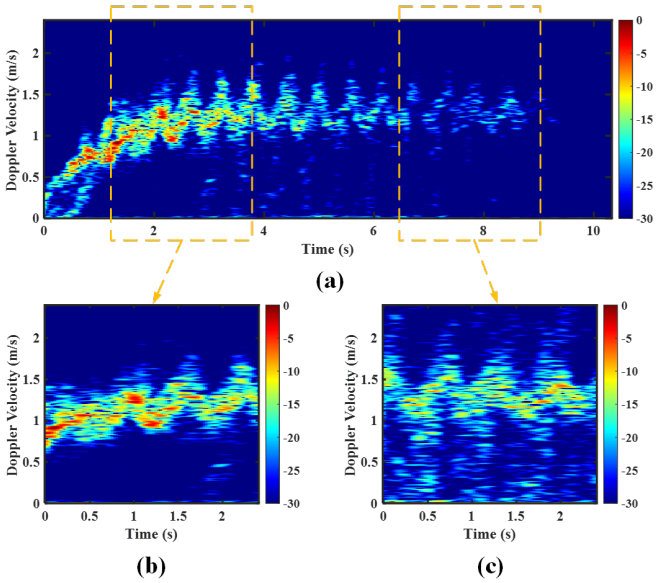


Figure 5.5: (a) Spectrogram of three people walking as a group. (b) Spectrogram at the near range. (C) Spectrogram at the far range.

B. ANN group classifier

As mentioned in Section 4.3, the proposed ANN group classifier is required to be more artificial, compared to the proposed non-ANN group classifier. Thus, compared to manually extracting features from the matrices, 2 matrices of the Range-Azimuth map and the spectrogram are directly inputted to the proposed ANN classifier.

The results of the ablation study of the proposed group classifier are provided in Table 5.6. Data Set I was implemented based on the Pytorch framework running with an GeForce GTX 1080 Ti (with 11 GB memory). This is to study whether Range-Azimuth map and spectrogram are capable in solving Group People Counting problem. The conclusion of this ablation study is similar to that when using the proposed non-ANN classifier. After using the Range-Azimuth map and the spectrogram jointly, ATP reaches the highest value among these three options, that is, 95.82%. It can also be found that the ATP exceeds 80% when either Range-Azimuth map or spectrogram is used alone. This also proves that the Range-Azimuth map or spectrogram is more discriminative than the other analyzed features under different number of people grouping, as summarized in Chapter 3. However, again, the MSE exceeds 12 when using the spectrogram, indicating that the classification is not precise enough in this case. Again, this is attributed to the fact that the number of people decaying and grouping at different locations can confuse the classifier using spectrogram. In other words, spectrogram needs to be paired with the location information of the target for more accurate population counts.

Compared to Table 5.5, it is found that using the network to extract features and classify them will have higher ATP than manually extracting them. In other words, the network will learn more hidden features and improve the accuracy of the population counts. Even though the estimated values are not that precise (MSE is not below 5), ANN shows the potential to be used to solve the People Counting problem. Based on this conclusion, it is worthwhile to investigate how to design a better network structure to solve Group People Counting problem in the future.

Table 5.6: Summary of the ablation study of Data Set I when applying the proposed ANN group classifier. (RA represents the Range-Angle Map; SPEC represents the spectrogram).

Input Features	Method	MSE ($*10^{-2}$)	ATP (%)
RA	Proposed ANN Classifier	5.45	88.12
SPEC	Proposed ANN Classifier	12.49	83.82
RA + SPEC	Proposed ANN Classifier	0.46	95.82

5.2.3. CASE 2: MULTIPLE INDIVIDUALS TRACKING

In this subsection, the performance of the tracking for counting block is tested. In this case, the feature-based tracking block in the proposed method is not available. Thus, the number of people in each person can only be 1 as shown in Fig. 5.6, and thus is named individual. The reason why the estimated number of people is not 0 is that the tracks of false detection are automatically deleted during the tracking process.

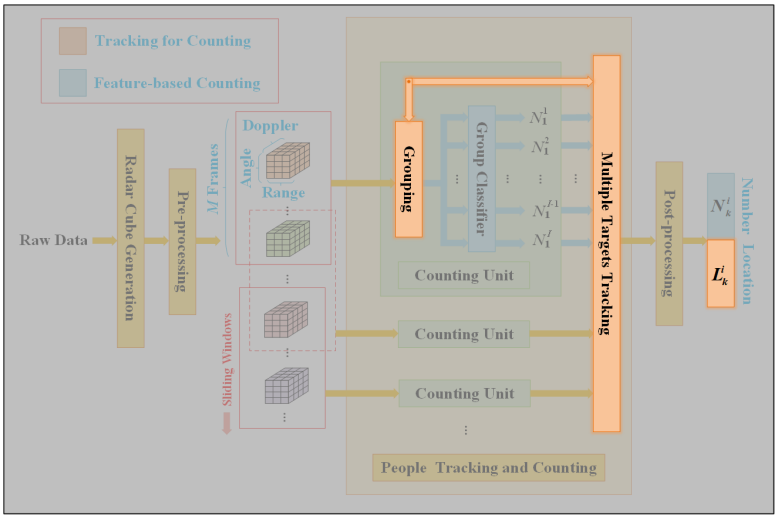


Figure 5.6: Overall scheme of the proposed tracking for counting block, where the focus is on the Multiple Target Tracking block.

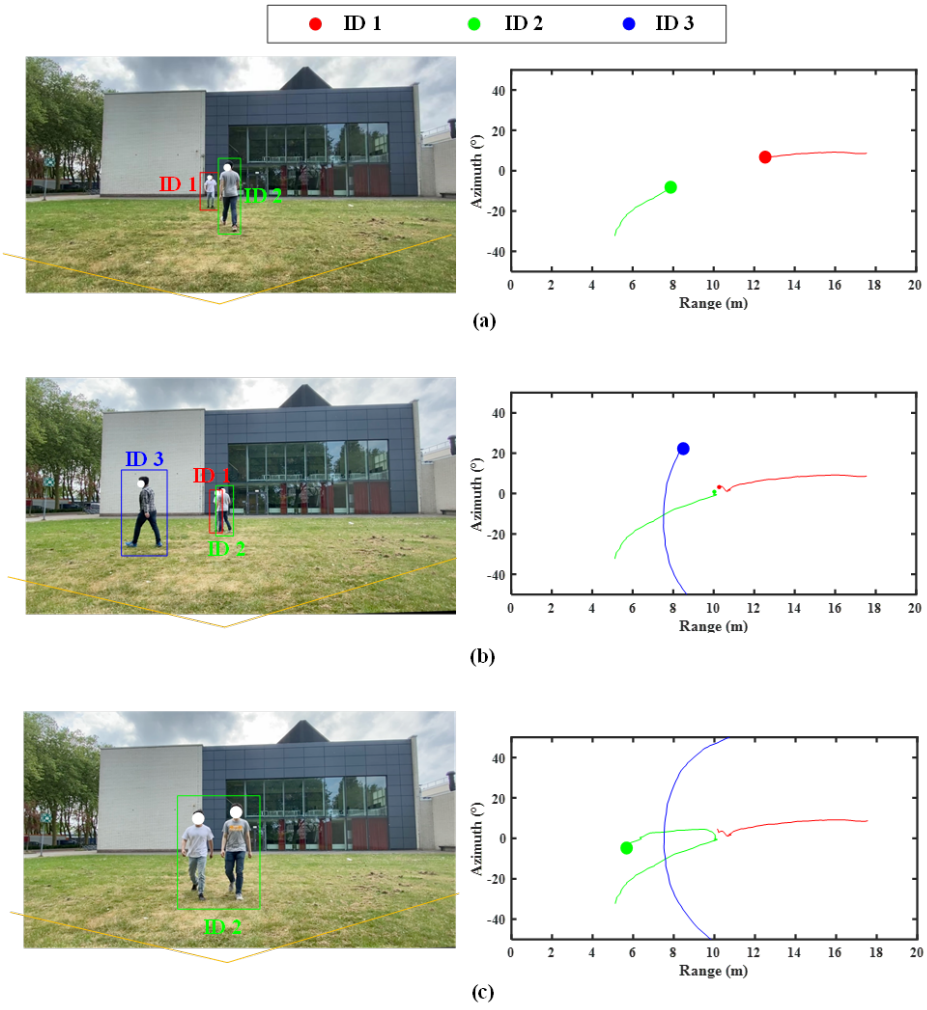


Figure 5.7: Results when applying the proposed tracking for counting block. The dot in the track is the instant corresponding to the picture. (a) the ground truth of two individuals in the RoI, and their estimated tracks. (b) the ground truth of three individuals in the RoI, and its estimated tracks. (c) the ground truth of two individuals in the RoI, and its estimated tracks. The orientation/geometry of picture and track plot are rotated

In this subsection, Data Set II is used where multiple individuals exist in the RoI, while there is no grouping case. Fig. 5.7 demonstrates the robustness of the performances when applying the proposed tracking for counting block. Fig. 5.7 (a) is the moment when two individuals were walking towards each other. From the estimated tracks, we could clearly see that there are two tracks, and they have a trend of moving towards each other. In Fig. 5.7 (b), two individuals finally met, and one individuals crossed in front of them. Although partial occlusion happened at ID 1 and ID 2, the tracking algorithms still hold the tracks. At this moment, three tracks are clearly shown in the Range-Azimuth map. In Fig. 5.7 (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 walking as a group. Therefore, the proposed tracking for counting block performs is well, even in the occlusion case.

Fig. 5.7 (c) shows that although by applying the proper cluster method ID 1 and ID 2 is able to be separated with the prior knowledge of two tracks, the proposed tracking for counting is designed to better cluster Group People rather than separate them. It ensures that the features of Group People are fully passed to the feature-based counting block for classifying the number of people in each group.

The summary of results when processing the Data Set II are provided in Table 5.7, and the total number of three proposed errors are provided where there are 40 decision windows of each row in the table. It can be seen that performance is good, where the overall MGTA is 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. Second, there are few IDS errors, which means that features are well assigned to each group.

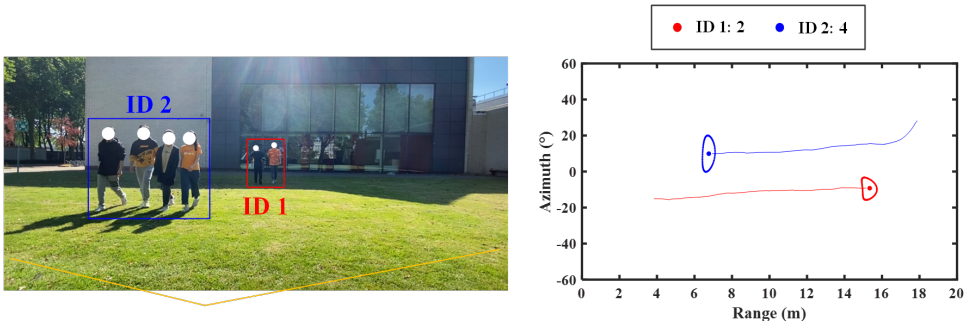
Table 5.7: Summary of the results of the proposed tracking for counting block. MD_{total} represents the sum of the MD at the decision windows. FD_{total}, and IDS_{total} also have the same meaning.

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

5.2.4. CASE 3: PEOPLE COUNTING IN MULTIPLE GROUPS ("BEYOND CLASSIFICATION")

After analyzing the results of the feature-based counting block and tracking for counting block separately, the result of the proposed method is provided, which combines the feature-based counting block and tracking for counting block.

As mentioned in Section 5.1, Data Set III is used in this subsection, where there are multiple groups in RoI, specifically training set of Data Set I and approximately 1.5 minutes for the testing data (introduced in Section 5.1). An example ground truth and result of Data Set III is shown in Fig. 5.8. There are two groups walking in the RoI and blue and red lines represent the tracks of each group. As mentioned in the assumption of experiments, there is no grouping, mixing, or spawning case of each group. Therefore, in the estimated tracks (in Fig. 5.8 (b)) there is no crossing of two tracks. Meanwhile, the predicted number of each group equals to the ground truth.



(a) (b)

Figure 5.8: Results when applying the proposed method. (a) the ground truth where two groups and six people in total in the RoI. (b) estimated tracks of two groups and predicted number of people in each group.

The overall performance in Data Set III when applying the proposed method is provided in Table 5.8.

1. The feature-based counting block performs well. MSE and ATP are even better than the performance when using Data Set I and only applying the feature-based counting method. Moreover, the scenario is much simpler than that of Data Set III. The reason is that, based on the tracking for counting block, the estimated number of each Group People can be arranged according to the decision window (time). Thus, the median filter can be applied to combine the estimated number from the feature-based counting block before and after this moment and output the median result for this moment. Fig. 5.7 (a) is the estimated number from the Group classifier where the ground truth is shown in Fig. 5.6 (a).A five-point median filter (post-processing block mentioned in 4.4 is applied and the time taken to obtain this result is only about 1.68 s, since each window is separated by only 0.12 s (mentioned in Section 4.2).

2. The MGTA is 1.67 %, which is acceptable. Among the three defined errors, almost no IDS occurs, and thus this identity switch would not influence the accuracy of the median filter mentioned above. Meanwhile, continuous MD did not occur on the entire track. This means that the features of Group People could be passed through the Group classifier successfully. In addition, although false detection may occur, the proposed group classifier had a high accuracy in classifying the zero targets in the group.

3. Moreover, it is worth highlighting that Data Set III includes the case where the total number of people is greater than the maximum number of expected classes. In the general classification problem, it is impossible to output unlabeled classes. For example, in Data Set III, the training set is using the training set of Data Set I, where only one group exists in the RoI and the maximum number is 5. Thus, for the group classifier, the expected outputs only range from 0 to 5. The tracking block links those classification results and makes it possible to count the case when the total

number of people in the RoI is more than 5. Therefore, it shows one of the potential advantages of the proposed method, and it is named "Beyond Classification".

Table 5.8: Overall results for Data Set III when using the proposed method (with proposed non-ANN group classifier).

	Data Set III
MSE ($*10^{-2}$)	0.13
ATP (%)	96.25
MGTA (%)	1.67

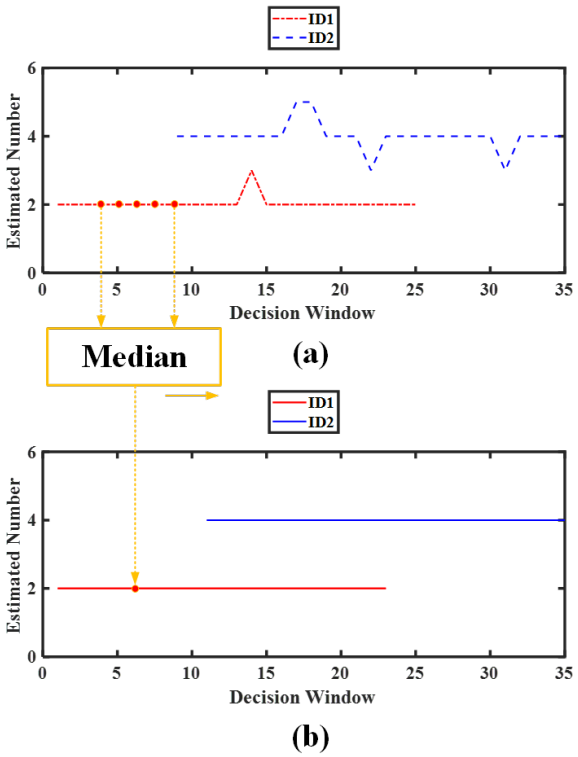


Figure 5.9: Results when applying the proposed method. (a) The ground truth where two groups and six people are present in total in the RoI. (b) Estimated tracks of two groups and predicted number of people in each group.

5.3. ERROR ANALYSIS

After summarizing the effectiveness of the proposed method for solving People Counting in the presence of grouping, this section focuses on the analysis of the causes that can lead the proposed method to estimate the wrong number of people.

- **Feature-based counting block:**

  Range-Azimuth maps and spectrogram-derived CVD maps are used for the design of the feature-based counting block.

  Due to the limitation of the detection method used in this thesis (i.e., CA-CFAR), not all features from the Range-Azimuth which are associated with people in the scene are used. Before generating the Range-Azimuth map, the CA-CFAR is applied to determine which Range-Doppler bins can be used for further angle estimation. In Fig. 5.10 (a b), the results before and after applying the CA-CFAR on the Range-Doppler map are provided. The red line in Fig. 5.10 (a) indicates the outline of the high-amplitude cluster on the Range-Doppler map that is likely to be connected with the moving people in the scene, while Fig. 5.10 (b) shows that after applying CA-CFAR a few bins have not been included. This can limit the performance of the group classifier to a higher level. Thus, an improved detection method can improve the performance of the feature-based counting block in future work.

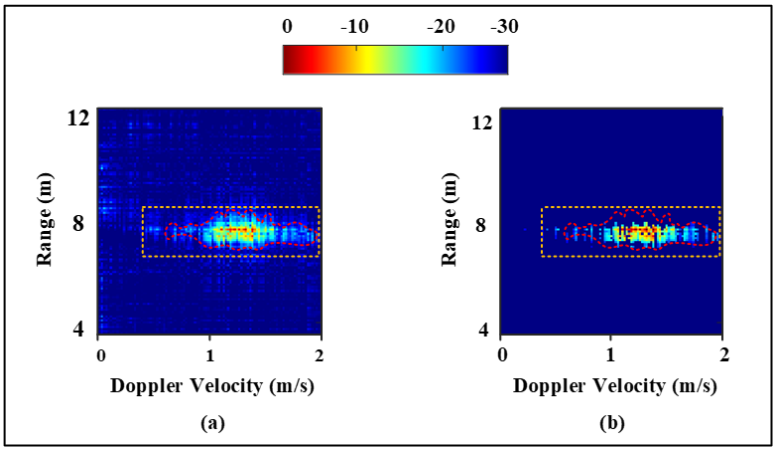


Figure 5.10: Example of Range-Doppler map (a) before and (b) after detection via CA-CFAR when three people are walking as a group.

- **Tracking for counting block:**

  The tracking for counting block tracks people in the scene and separates different groups in the scene. When there are multiple groups in the RoI, features from different groups are assigned separately to the feature-based counting block. Therefore, compared to when there is a single group in the scene, these blocks are of greater importance when there are multiple groups in the scene.

  Errors in the tracking for counting block are mainly caused by the case when features of groups are not fully sent to the feature-based counting block for classification. This is caused by the cluster method, i.e., grid-based DBSCAN. In Fig. 5.11, some detections of ID 1 are assigned to noise. This influences the completeness of Range-Azimuth map, and after passing this incomplete map to the classifier,

errors may occur. After carefully comparing this situation with the ground truth, some detections of the case when volunteers were not exactly walking side by side in the larger azimuth cluster are discrete, isolated, and thus not reachable for grid-based DBSCAN. An improved clustering method which uses the information from multiple frames to jointly decide if the point is unreachable may improve this specific case in future work.

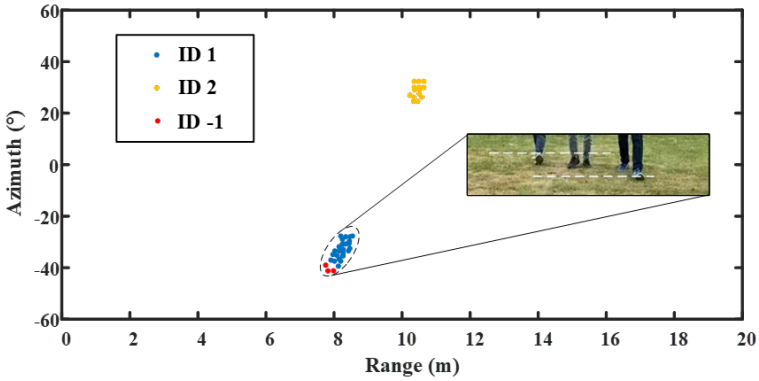


Figure 5.11: The example of the false clustering when applying grid-based DBSCAN, where ID 1 is three people not walking side by side in the scene, ID 2 is one person walking in the scene, and ID -1 is the label where DBSCAN falsely assign them to noise.

5.4. CONCLUSIONS

In this chapter, the performance of the proposed Radar-based People Counting method is verified with three different Data Sets. Since the proposed method combines the feature-based counting block and tracking for counting block, several performance metrics are proposed to jointly study the performance of both blocks, that is, Mean Square Error (MSE), Average Probability of True Positives (ATP), and Multiple Groups Tracking Accuracy (MGTA).

From the results, three conclusions are drawn.

- **The proposed feature-based counting block is able to solve the Group People Counting problem, where each input feature is important and necessary.**

  The features extracted from the Range-Azimuth map are relevant to classify the number of Group people in the near range, while in the far range those features are not enough. Features from the CVD map can improve the Group counting accuracy in the far range, but only if they are linked to the location information (i.e., range, azimuth).

- **The proposed tracking for counting block is robust enough to associate targets and estimate the position of targets in the RoI.**

  The aim of the tracking for counting block is to maintain the completeness of features applied to classify Group People is important for the Group classifier. According to the overall MGTA which is less than 5%, it can be concluded that the

proposed tracking for counting block meets the goal, i.e. it is robust enough to associate targets and estimate the position of targets in the RoI. Meanwhile, there are few IDS errors, which means that features are well assigned to the correct group.

- **The proposed method, which combines the advantages of tracking for counting and feature-based counting, can achieve "Beyond Classification", and enable more accurate classification than using individual blocks.**

  "Beyond Classification" is the case where the total number of people is more than the maximum number of expected classes from training. In the general classification problem, it is impossible to output unlabeled classes. But with the proposed method, counting the number of people that is larger than the trained labels can be achieved. Moreover, with the help of tracking for counting block, the estimated number of Group People among the time can be stored and pass through the median filter to output the overall estimation. This overall estimation combines the results before and after that moment, and according to the result, this estimation does improve the accuracy of Group People Counting.

The following chapter will compare the proposed method with the SOTA Radar-based People Counting methods and the commercial People Counting product.

6

PERFORMANCE COMPARISONS

After summarizing the benefits of the proposed method when dealing with grouping case, the comparisons study is introduced in this chapter. The existing People Counting product and the state-of-the-art (SOTA) Radar-based People Counting methods are reimplemented using Data Set I. As, to the best of my knowledge, grouping is not specifically studied in the field of radar-based people counting, I was unable to find a method suitable for a direct comparison on the ability to solve this problem. To demonstrate that grouping is indeed a problem in Radar-based People Counting, the most recent methods in the field of crowd counting were selected, including the tracking for counting method, non-ANN feature-based counting methods, and ANN feature-based counting methods. In Section 6.1, the TI "3D People Counting" software is reimplemented, which is developed based on tracking for counting method. After that, two non-ANN and two ANN feature-based counting methods are reproduced separately in Section 6.2 and Section 6.3. Finally, the summary of all SOTA methods and the proposed method are provided and discussed in Section 6.4.

6.1. COMPARISONS WITH THE EXISTING PEOPLE COUNTING PRODUCT

In this section, an existing People Counting commercial product, 3D People Counting, is chosen for performance comparison. 3D People Counting is one of the demonstration labs of the TI Industrial Toolbox [61]. When their counting tracking method is applied, moving targets are continuously tracked until they leave the scene. With 3D people counting software, 3D point cloud and estimated number of targets can be visualized using the visualizer module included in the toolbox.

This toolbox applies the "group tracker" to track individuals in the scene, where the "group tracker" means 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 investigate whether this existing product can solve the grouping problem and to study how robust the algorithm of tracking for counting is, the experimental design of collecting Data Set I with 1 to 3 people walking as a group and Data Set II was chosen. Because raw data are not available for this product, the data used are obtained from the same experimental field with reference to the ground truth of Data Set I and Data Set II. Fig. 6.1 is an example of the ground truth and visualizer results of Data Set I and Data Set II with multiple individuals without grouping in the scene. In Fig. 6.1 (a), there is one individual walking towards the Radar. The visualizer showed that the target was tracked, and the estimated target was 1 which is correct. However, when people walked as a group, the missed detection occurred when only one person was clustered, as shown in Fig. 6.1 (b). Thus, the number of estimated targets in the RoI was still one.

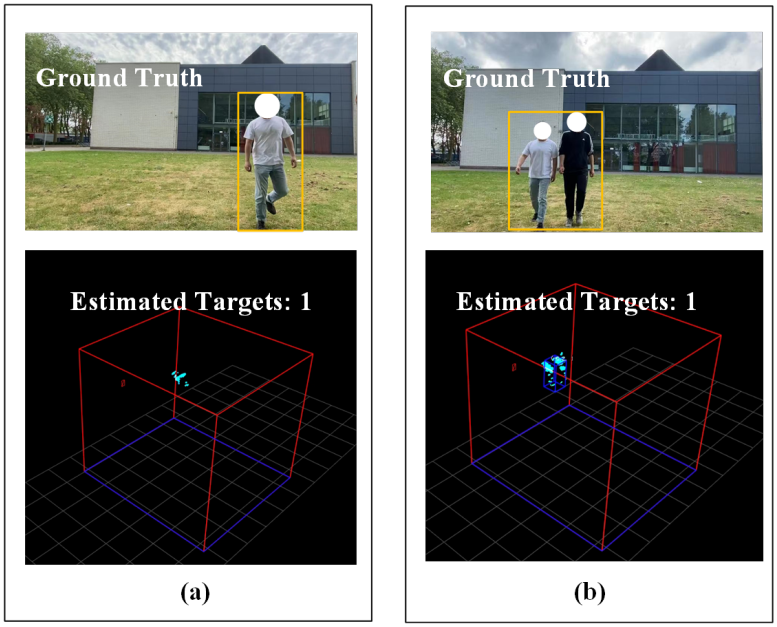


Figure 6.1: Ground truth and TI 3D People Counting visualizer results when (a) one person exists and (b) two people walking as a group.

The Data Set I used for testing included Group People ranged from 1 to 3 participants, and each class had 40 decision windows. The Data Set II used for tracking included maximum 3 people in the RoI, there were 40 decision windows among each number of people in the RoI. The results of Data Set I and Data Set II are provided in Table 6.1. The MGTA of Data Set I and Data Set II are higher than 40%, which means that at each decision window errors are most likely to occur. Meanwhile, the number of MD was really high when processing the Data Set I, which is the grouping case. This is because the tracking and localization algorithm within the TI toolbox did not include the block to deal with the grouping case. Although in the 3D People Counting implementation

guide, "Group tracker" is mentioned, this kind of tracker only means tracking the extended target, which is a group of detections, without attempting to estimate how many participants belong to each group. Meanwhile, the 3D People Counting is vulnerable to the case where multiple people enter the RoI at the same time. This situation requires tracking algorithms to predict and update multiple targets of birth at the same time, thus the number of FD was high.

Based on the results, two conclusions are drawn.

1. Currently, the TI 3D People Counting that used the tracking and localization algorithms cannot deal well with the Group problem.

2. the TI 3D People Counting performs well when people were entering the RoI one by one, whereas it is vulnerable to the situation when multiple people entered the scene at the same time, even though there was no mixing or grouping.

Table 6.1: Summary of the results for the TI 3D People Counting. MD_{total} represents the sum of MD for all the considered decision windows. FD_{total}, and IDS_{total} also have the same meaning.

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

6.2. COMPARISONS WITH NON-ANN METHODS

In this section, two SOTA non-ANN People Counting methods are chosen for comparisons. The advantage of non-ANN for People Counting is that its feature extraction method is sufficiently interpretable, and the method is simpler and more convenient for application compared to ANN methods. Due to the fact that both of the methods are feature-based counting methods, Data Set I is chosen for processing and comparisons where there is one group with different number of people existing in the scene.

The first SOTA non-ANN method for comparison is using the range profile for clustering and extract the main amplitude for the statistical study, as proposed by Choi et al. in [19]. In Fig. 6.2, an example of result when using their feature extraction method is shown, where the red line is the clustered amplitudes with distance information d_1. With the help of the probability density function (PDF) of the amplitudes of the major clusters that are associated with the number of people and their range, estimating the number of people in the scene is solved by using the maximum likelihood equation.

When reimplementing this non-ANN method, several shortcomings were found that may affect the effectiveness of Group People Counting.

1. The authors design the cluster detection algorithms based on the fact that the number of clusters was larger than the number of individuals because of the multipath effect, while in the grouping case it can be found that Group People only oc-

cupied several range bins, no matter how many people are in the group (as shown in Fig. 3.6). This can influence the performance of the cluster detection method.

2. The maximum amplitude of each cluster is chosen for analysis, while according to the study of the Range Profile in the grouping case in Section 3.3, the maximum amplitude is not enough to represent the number of Group People, because this amplitude shows a non-linear trend with the number of Group People.

According to Table 6.2, the ATP is 50.15 %, which means that using this method has a high probability of estimating the wrong number of people in the RoI. It proves that the maximum amplitude extracted from the Range Profile is not enough for reliable classification. The MSE is 0.2281, indicating that the estimated values are scattered and not very precise.This dispersed prediction indicates that the PDFs under different classes overlap a lot with each other. In other words, this classification method cannot distinguish different classes well based on the available features. It proves that using the log-Gaussian fitting only based on the relation between amplitude and the distance from the Radar is not robust to solve grouping problem..

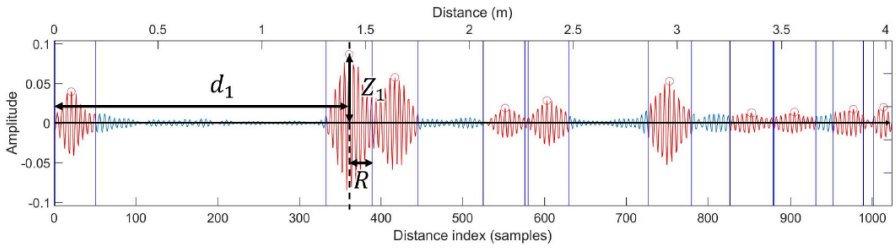


Figure 6.2: Example of the proposed non-ANN feature extraction by Choi et al. [19] where the red lines are the clusters. The data used in this example refer to the Range profile where individuals exist in different range bins.

The second SOTA non-ANN method for comparison was proposed by Choi at al. [27].This method fused the features of the Range Profile and the features of the Doppler domain to perform the People Counting. After concatenating those features, normalization and Principle Component Analysis (PCA) are performed to improve the speed of solving the optimal solution. In this approach, the feature extraction methods are all based on intensity or energy, as described in their paper. In the range domain, the MD-CLEAN method is proposed to eliminate multi-path effect and noise and extract amplitudes that reflect the number of individuals. In the Doppler domain, the mean value of the energy intensity is introduced into the feature vector, especially for solving the case when different numbers of people appear at the same range bin.

When reimplementing this SOTA method, several weaknesses were found that may affect the accuracy of Group People Counting.

1. This SOTA method used the mean of energy in the Doppler domain for feature extraction. But this ignores the case where people may not be moving in the scene, and thus the Doppler frequency is close to zero. The main energy obtained in this case differs significantly from the energy obtained when all people are moving.

2. The method used feature extraction on different domains separately and then applied the concatenation of the feature vectors, instead of fusing the information from both domains into 2D information and then applied the feature extraction method. This process loses more information compared to using 2D feature extraction for jointly processing. In other words, the feature extraction method proposed in this paper is different from using the Range Doppler map directly, which make it difficult to make a connection between the features of the target in the range domain and the features in the Doppler domain. Therefore, it may happen that although people are at different distances, they contribute the same Doppler information. In Table 6.2, the input of this method is written as "Range-Doppler Map" for naming convenience, but in fact the input is not a "map" for the remainder of the reader.

According to Table 6.2, the ATP is 66.88 %. Although the overall performance is still not good for solving Group People Counting, it is better than the non-ANN method that only uses range profile to do the People Counting. This proves the importance of using multiple-dimensional features for the Group People Counting problem. The MSE is 0.0687, which is much smaller than that of the first method used for comparison in this section. However, this does not mean that this method has higher precision. As mentioned for the drawbacks of the method, the number of stationary people in the scene can have a significant impact on the performance. In fact, the percentage of Group People that are not moving in Data Set I is not very high. Therefore, it is not possible to draw a solid conclusion about the advantage of this method in terms of precision.

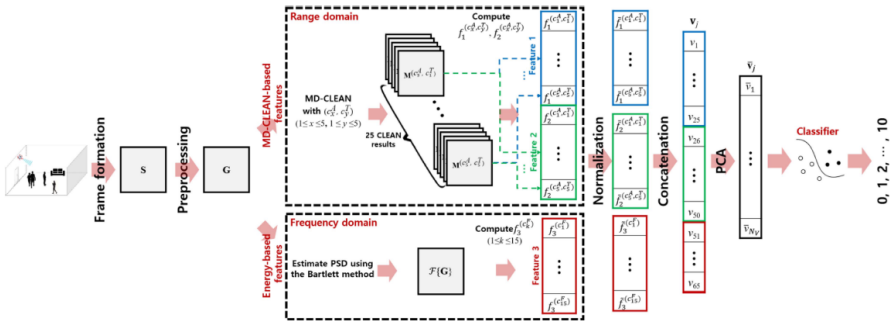


Figure 6.3: Overview of the non-ANN people counting method by Choi et al. [27] where features from range domain and Doppler domain are used.

6.3. COMPARISONS WITH ANN METHODS

Two SOTA ANN People Counting methods are reimplemented for comparison in this section. The advantage of applying ANN for People Counting is that, instead of manually extracting features, the network can directly determine which parts are best suited to this classification problem. In this case, the results have higher robustness when the training set is selected suitably and the network is properly designed. In this section, Data Set I is also used for comparisons where only one group exists in the scene.

The first ANN method for comparison is the method proposed by Jia et al. [66] that used the Range-Time map for classification with the help of ResNet-14 (in Fig. 6.4). The authors mentioned that, compared to AlexNet, the ResNet architecture enables deeper networks to be trained without overfitting. Residual units within ResNet play the key role in achieving this. This unit creates a direct connection between input and output, so that the new layer simply learns new features on top of the original input layer, that is, learning the residuals.

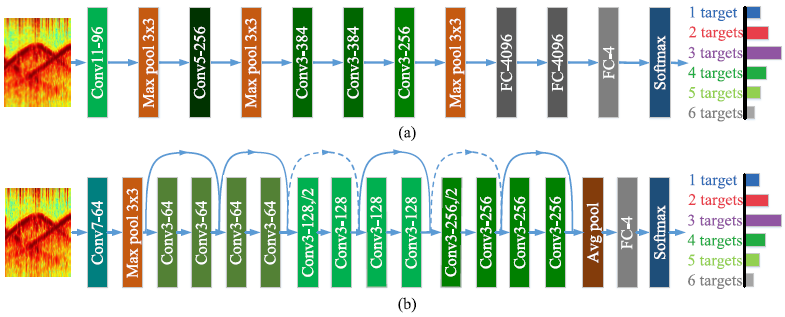


Figure 6.4: ANN People Counting pipeline of (a) AlexNet and (b) 14-layer ResNet [66].

When reimplementing this SOTA method, several drawbacks were found that can influence the accuracy of Group People Counting.

1. The Range-Time map is not sufficient to reflect the characteristics of Group People. Without the grouping case, the morphological features in the Range-Time map, that is, the continuous lines, are related to the number of people in the scene. However, as studied in Section 3.3, no matter how many people move as a group in the RoI, the number of continuous lines only is equal to the number of groups in the scene. Therefore, this morphological characteristic of the Range-Time Map is not the best option to solve Group People Counting problem.

2. The training set required in their article [66] is at least one order of magnitude larger than that of other ANN methods. Each class required 10,800 Range-Time images. In the reimplementation stage, in order to meet this condition, raw data from each of the 12 virtual channels is used to generate the Range-Time images. Even if data generated with this approach are correlated, by analyzing the losses and results, no overfitting or underfitting occurred. This method took longer to train the model than other ANN methods.

According to Table 6.2, the ATP is 60.42%, and the minimum probability of true positives is 40.2%. This proves that the Range-Time map is not enough to solve the People Counting problem. Although training the model took more time than other ANN methods, the MSE is only 0.0798, which means that the predicted result is not precise enough. In other words, there is a large error between the estimated value and the true value. In

summary, as the shortcomings of the previous analysis demonstrate, even using a training set containing tens of thousands of images and consuming much of the computer's performance, the method cannot solve the Group People Counting problem.

The second ANN People Counting method that is reimplemented for performance comparison is the method using CNN and LSTM for classification, as shown in Fig. 6.5. This paper used the same experimental scenarios and data as the one reproduced in Section 6.2. It improved the overall accuracy compared to that of the non-ANN method. It looks like this ANN method uses less dimensional features (Range Profile) compared to the non-ANN method, but in fact the LSTM applies the same information in the temporal dimension. Therefore, it is not surprising that this ANN People Counting method achieved better results.

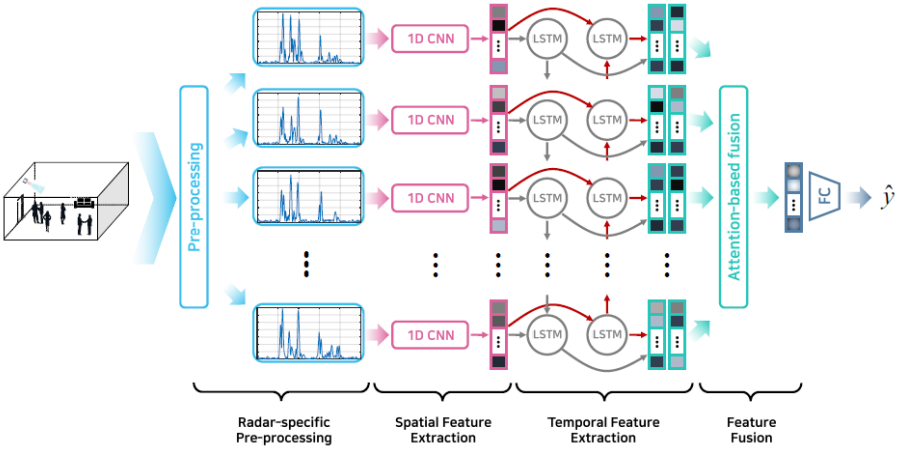


Figure 6.5: Overview of the ANN People Counting method with the CNN and LSTM by Choi et al. [37].

When reproducing this ANN method, for the remainder of the reader, in order to achieve the required output of the authors' proposed network, the stride size of all convolutional layers was modified to 1, instead of 2 as described in the original table.

One drawback was found that can influence the accuracy of Group People Counting. The method introduces an LSTM to relate the Range Profiles at different moments for classification. However, this module is only as useful as the variability of the range profiles at different time frames. Compared to the movement of multiple people at different distances, grouping provides less frame-by-frame information to reflect on the Range Profile. Therefore, this approach does not necessarily work better for the grouping problem.

According to the results in Table 6.2, as analyzed above, this ANN method also did not solve the Group People Counting problem, where the ATP is 63.48% and the MSE is 0.1240. By comparing the two ANN methods, it is found that both ANN methods have similar performance when using the range profile to do the Group People Counting. The second ANN method, which applied CNN and LSTM, performed slightly better. The input of the first ANN method is the Range-Time map. Although the Range-Time map

actually includes the frame-by-frame information, the self-attention Bi-LSTM which is used in the second ANN method is better to make the network focus on the frame-by-frame information. Therefore, the second ANN method outperformed the first ANN method.

6.4. CONCLUSIONS

After the reimplementation of four SOTA People Counting method, results are summarized in Table 6.2. Meanwhile, the results of only applying the proposed classifier, and the proposed method which combines the tracking for counting block and feature-based counting block are provided in this table.

Several conclusions are drawn on the basis of the results summarized in Table 6.2.

Table 6.2: Summary of the Results of Data Set I in performance comparisons between the proposed approaches in this thesis (top four rows) and re-implementation of SOTA methods. (RA represents the Range-Angle Map. CVD represents the CVD map. SPEC represents the spectrogram.)

Input Features	Method	MSE ($*10^{-2}$)	ATP (%)
RA + CVD	Proposed Non-ANN Classifier (Only)	0.58	94.32
RA + SPEC	Proposed ANN Classifier (Only)	0.46	95.82
RA + SPEC	Proposed Method	0.42	96.05
RA + CVD	Proposed Method	0.26	98.42
Range Profile	Non-ANN Classifier [19]	22.81	50.15
Range Doppler Map	Non-ANN Classifier [27]	6.87	66.88
Range-Time Map	ANN Classifier [66]	7.98	60.42
Range Profile	ANN Classifier [37]	12.40	63.48

1. **Grouping is an open problem in Radar-based People Counting.**

   The SOTA Radar-based People Counting methods failed to solve effectively Group People Counting. Based on a comparison of the results in Chapter 6, it can be seen that the ATP of these reimplemented methods was not more than 70%. Meanwhile, the high MSE indicated that the precision of these estimations was low. Compared to the good results they got in their paper, most likely for experimental setups without the grouping phenomena happening, the problem of grouping affects the grouping problem affects the effectiveness of these methods.

2. **The proposed method has the best performance compared with these SOTA People Counting methods.**

   In terms of the grouping problem, the ATP using the proposed group classifier alone is more than 94%, and the MSE is two orders of magnitude better than other published methods. Furthermore, the ATP of the proposed method is even greater than 98%. This demonstrates the effectiveness of the proposed method in solving the grouping problem.

3. **When solving grouping problems, in addition to the need to consider the use of multidimensional information from collected data, the appropriate choice of features is even more important.**

   It was found that among these four SOTA methods, using only one dimension of information (Range Profile) has the worst results. All three other methods used two dimensions of information, but the non-ANN method which uses the Range Doppler map has the best result. This is because the Range Doppler map is more effective in distinguishing the number of Group People compared to the other features. Therefore, it is important and necessary to study the characteristics of the grouping and how these reflect in the different radar data formats and features before designing before designing the classification method.

7

CONCLUSIONS AND FUTURE WORK

7.1. SUMMARY AND CONCLUSIOS

This thesis focuses on studying and solving the Grouping problem in Radar-based People Counting. With the use of MIMO Radar, the proposed method takes advantage of *feature-based counting* methods and *tracking for counting* methods, to achieve the function of *exact* people counting.

The main content is summarized below as contributions to answer the formulated research questions, namely:

1. **Is Grouping a problem in Radar-based People Counting?**

   Grouping is a research problem in Radar-based People Counting. There are two reasons that can be used to justify this conclusion.

   First, according to my review of the literature on People Counting, even though different researchers mentioned that people's motion in their experiment setting is randomly walking, their experimental assumptions restrict the grouping case, e.g., they assumed that people did not always move at the same range bin, or that even if they were within the same range bin, they can always be distinguished by the direction of their movement (e.g., moving towards Radar, moving away from Radar) [19, 27, 37, 66].

   Second, the SOTA Radar-based People Counting method and commercial product such as TI 3D People Counting [61] did not solve the Group People Counting very well. Based on a comparison of the results in Chapter 6, it can be seen that the ATP of these reimplemented methods was not higher than 70%. Compared to the high performance they achieved in their paper, the grouping problem affects the effectiveness of these methods.

   Therefore, grouping is a research problem and deserves further study.

2. **What are the characteristics of Grouping in the Radar's view?**

As described in the definition of Group People in Chapter 3, a group of people is sharing neighboring locations and moving together. Existing features and data formats that have been already used in Radar-based People Counting, such as Range Profile, Range-Time map, and Range-Doppler map, all have similar signal or image morphological characteristics regardless of how many people move as a group.

With the help of MIMO Radar, the angle information of the targets can be estimated. The Range-Azimuth map is promising, as is shown by the fact that the more people move as a group, the more azimuth bins are occupied. But this feature is vulnerable when Group People exist in the far range.

The spectrogram of Group People that includes multiple frames of information is studied. Spectrograms have been found to be associated with the synchronization of crowd movements. The micro-Doppler becomes more irregular as the number of Group People increases as they do not walk or move in a synchronized manner, and the greater and the greater the number of people, the greater the degree of this irregularity.

Therefore, Range-Azimuth map and spectrogram/CVD in combination can be a good choice to solve Group People Counting.

3. **How to solve the counting problem when grouping?**

   As mentioned in Chapter 2, there are two categories of Radar-based People Counting methods, that is, tracking for counting methods and feature-based counting methods. It was found that these two categories of methods have complementary strengths.

   For tracking for counting methods, targets can be tracked continuously without the need for a training set to obtain information about the targets' position. They are not able to count the number of Group People based on the assumption that the number of tracks equals the number of people.

   For feature-based counting methods, they can solve the grouping problem when assigning this problem to the classification problem. But they ignore the close connection between each time step when people move in the RoI, due to the fact that the outputs from the pre-trained classification model are independent.

   Therefore, this thesis proposes a method that combines the tracking for counting block and the feature-based counting block to solve Group People Counting problem introduced in Chapter 4.

4. **What are the advantages of combining tracking for counting method and feature-based counting method?**

   There are two advantages of the proposed method that are explained in detail in Chapter 5.

   The first advantage is that the proposed method is able to achieve "Beyond Classification". "Beyond Classification" is the case where the total number of people is greater than the maximum number of expected classes. In the general classification problem, it is impossible to output unlabeled classes not defined at the

training stage. But with the proposed method, counting the number of people that is larger than the trained labels can be achieved.

The second advantage is that the proposed method has more accurate classification results than using one of the blocks individually. With the help of tracking for counting block, the estimated number of Group People over time can be stored and pass through the median filter to output the overall estimation. This overall estimation combines the results before and after that moment, and according to the result, this estimation does improve the accuracy of Group People Counting.

The main contributions of this thesis research are summarized below:

- One of the most complex motions of individuals, grouping, is studied, and the Radar-based People Counting problem is addressed with a dedicated processing pipeline more in general, including the cases of grouping of multiple individuals.
- This pipeline was developed by studying the characteristics of Group People (defined as a group of people sharing neighboring, adjacent locations and moving together), with multiple-input-multiple-output (MIMO) frequency-modulated continuous wave (FMCW) Radar. Specifically, the combination of the Range-Azimuth map and spectrogram/cadence velocity diagram (CVD) is proposed to solve the People Counting problem including grouping case.
- Based on the fact that two categories of Radar-based People Counting methods, i.e., tracking for counting methods and feature-based counting methods, have complementary strengths, the proposed method combines these two approaches into a new processing pipeline to estimate the number of people in each group in the scene while tracking each group. Moreover, the proposed method can achieve "Beyond classification", which is the output the unlabeled classes not defined at the training stage. Additionally, compared with other state-of-the-art (SOTA) Radar-based People Counting methods, the proposed method outperforms them, and thus it is proved that the grouping problem can be well addressed in the Radar-based People Counting field.

Part of the results from this thesis work are being written up for publication in a high-quality IEEE journal such as IEEE Sensors, IEEE IoT Journal, and IEEE Transactions on Geoscience and Remote Sensing (TGRS).

7.2. FUTURE WORK

With respect to the proposed method, in addition to the improvements in detection and clustering methods mentioned in Section 5.3 that can be used to increase the accuracy of People Counting, there are still some potential improvements that can be examined in future research. Specifically, attention can be paid to the interesting and challenging case of spawning and occlusion events.

- **Spawning case**

  Spawning (also known as splitting) is an event in which a single extended target separates into two or more extended targets, or vice versa [106]. It is one of the

most complex scenarios in the MTT field. In this field, multiple models are defined to solve this problem, such as the birth model of the spawning targets, the measurement model of the spawning target, and the main target measurement model [113].

The proposed method of this thesis, which counts the number of Group People and tracks each group separately, has the potential to deal with the spawning problem. As shown in Fig. 7.1, since the number of people in the group is known and there are three people walking as a group at t_1, the tracking algorithm could be set to know the possible birthplace around the large group, for example when one person is going to leave the large group, for example at t_4 in the figure. Thus, at t_5 the main group and the spawning target could be tracked separately and smoothly.

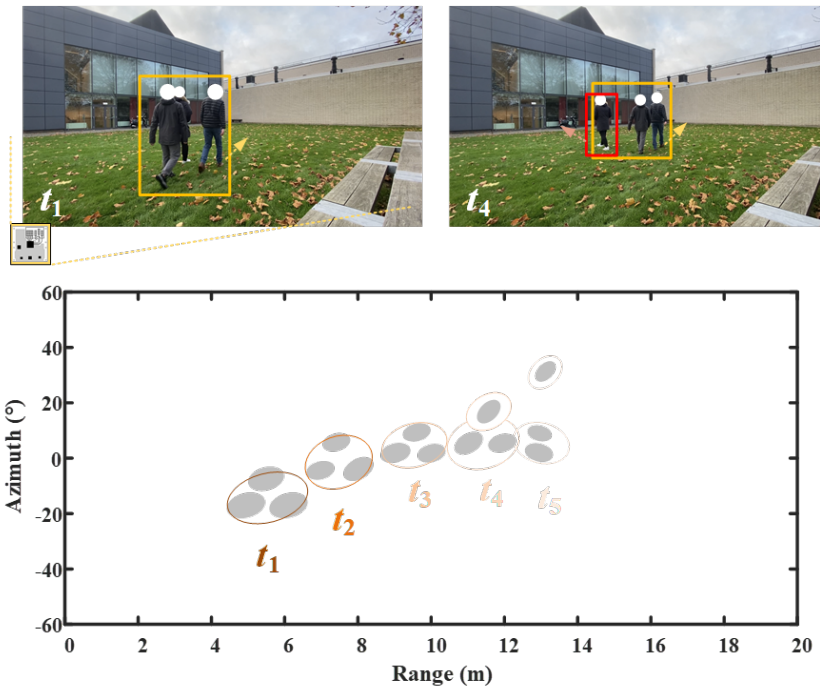


Figure 7.1: The example of spawning case. The gray spots represent the different people.

- **Occlusion case**

  In the occlusion event, people also gather and move together, but at this time there exists a situation where someone in the group is obscured by another individual and cannot be detected by the radar. An example of the occlusion case is shown in Fig. 7.2 where an individual is occluded away from the Radar.

  There are two solutions to this problem. First, multiple Radar sensors can be applied to build a Radar network that can look at the area of interest from multiple views. With this solution, because there are radar sensors in all directions, there

is no shadowing zone of detection. Second, MIMO Radar with the elevation information can be implemented to deal with this problem, when it is placed at a higher position with respect to the scene of interest. This is called 'overhead detection' and reduces the probability of occlusion by looking at the different people from above.

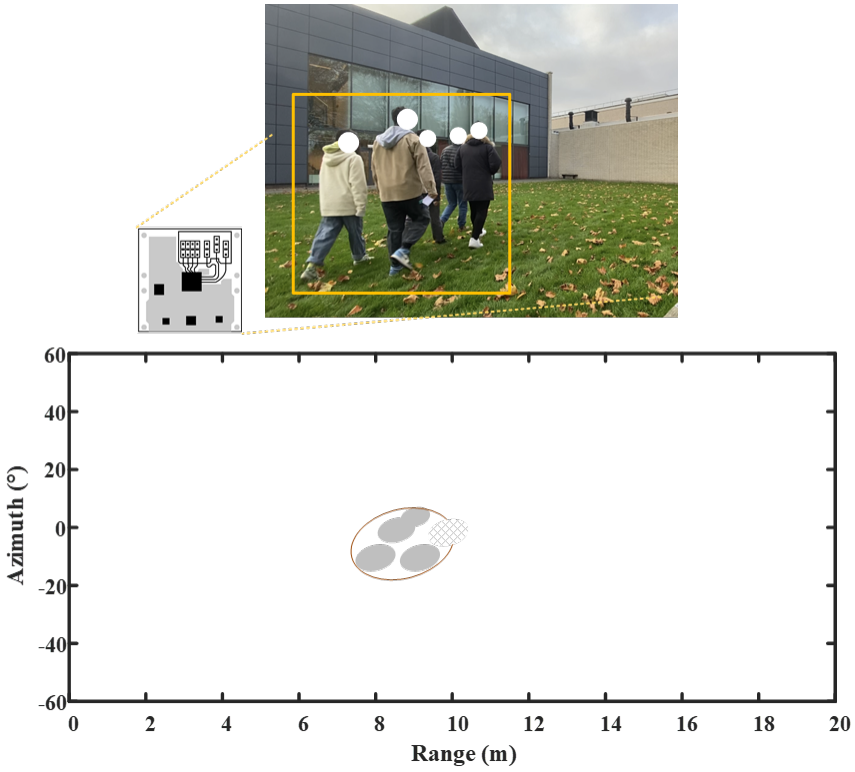


Figure 7.2: The example of occlusion case. The gray spots represent detected people, and the spot with the gray criss-crossed pattern is the missed detected person.

REFERENCES

[1] World Health Organization. *World Health Organization COVID-19 Dashboard*. 2022. URL: https://covid19.who.int/.

[2] X. Huang, Z. Yang, and C. Chen. "Regional People Counting Approach Based on Multi-Motion States Models Using MIMO Radar". In: *IEEE Geoscience and Remote Sensing Letters* (2022), pp. 1–1. DOI: 10.1109/LGRS.2022.3156793.

[3] F. Caprotti and D. Liu. "Platform urbanism and the Chinese smart city: the co-production and territorialisation of Hangzhou City Brain." In: *GeoJournal* (2020). DOI: 10.1007/s10708-020-10320-2.

[4] N. Cahyadi and B. Rahardjo. "Literature Review of People Counting". In: *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*. 2021, pp. 1–6. DOI: 10.1109/AIMS52415.2021.9466029.

[5] Z. Yang and X. Huang. "Cascaded Regional People Counting Approach Based on Two-Dimensional Spatial Attribute Features Using MIMO Radar". In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5. DOI: 10.1109/LGRS.2021.3103197.

[6] C. Tang et al. "Occupancy Detection and People Counting Using WiFi Passive Radar". In: *2020 IEEE Radar Conference (RadarConf20)*. 2020, pp. 1–6. DOI: 10.1109/RadarConf2043947.2020.9266493.

[7] J. W. Choi, X. Quan, and S. H. Cho. "Bi-Directional Passing People Counting System Based on IR-UWB Radar Sensors". In: *IEEE Internet of Things Journal* 5.2 (2018), pp. 512–522. DOI: 10.1109/JIOT.2017.2714181.

[8] Q. Guan et al. "[Epidemiological investigation of a family clustering of COVID-19]". In: *Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi* 41.5 (May 2020), pp. 629–633. ISSN: 0254-6450. DOI: 10.3760/cma.j.cn112338-20200223-00152.

[9] A. Zanella et al. "Internet of Things for Smart Cities". In: *IEEE Internet of Things Journal* 1.1 (2014), pp. 22–32. DOI: 10.1109/JIOT.2014.2306328.

[10] C. Wu, Z. Yang, and C. Xiao. "Automatic radio map adaptation for indoor localization using smartphones". In: *IEEE Transactions on Mobile Computing* 17.3 (2017), pp. 517–528.

[11] C. Martani et al. "ENERNET: Studying the dynamic relationship between building occupancy and energy consumption". In: *Energy and Buildings* 47 (2012), pp. 584–591. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2011.12.037.

[12] A. Taha et al. "Design of an Occupancy Monitoring Unit: A Thermal Imaging Based People Counting Solution for Socio-Technical Energy Saving Systems in Hospitals". In: *2019 11th Computer Science and Electronic Engineering (CEEC)*. 2019, pp. 1–6. DOI: 10.1109/CEEC47804.2019.8974311.

[13] InfraRed Integrated Systems Ltd. *Smart Building Technology*. 2022. URL: https://www.trueoccupancy.com/occupancy-sensing-technology.

[14] I. Ahmed et al. “A deep learning-based social distance monitoring framework for COVID-19”. In: *Sustainable Cities and Society* 65 (2021), p. 102571.

[15] J. García et al. “Directional People Counter Based on Head Tracking”. In: *IEEE Transactions on Industrial Electronics* 60.9 (2013), pp. 3991–4000. DOI: 10.1109/TIE.2012.2206330.

[16] Z. Peng and C. Li. “Portable microwave radar systems for short-range localization and life tracking: A review”. In: *Sensors* 19.5 (2019), p. 1136.

[17] X. Chen et al. “Range Image-based LiDAR Localization for Autonomous Vehicles”. In: *arXiv preprint arXiv:2105.12121* (2021).

[18] S. Ansari and S. Salankar. “An Overview on Thermal Image Processing.” In: *RICE*. 2017, pp. 117–120.

[19] J. W. Choi, D. H. Yim, and S. H. Cho. “People Counting Based on an IR-UWB Radar Sensor”. In: *IEEE Sensors Journal* 17.17 (2017), pp. 5717–5727. DOI: 10.1109/JSEN.2017.2723766.

[20] X. Lu et al. “Robust occupancy inference with commodity WiFi”. In: *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. 2016, pp. 1–8. DOI: 10.1109/WiMOB.2016.7763228.

[21] J. Weppner et al. “Participatory Bluetooth Scans Serving as Urban Crowd Probes”. In: *IEEE Sensors Journal* 14.12 (2014), pp. 4196–4206. DOI: 10.1109/JSEN.2014.2360123.

[22] S. Hannah, Z. Gabriela, and H. Rachele. *Thermal Imaging as Pandemic Exit Strategy: Liminations, Use Cases and Privacy Implications*. 2020. URL: https://fpf.org/blog/thermal-imaging-as-pandemic-exit-strategy-limitations-use-cases-and-privacy-implications/.

[23] International Electrotechnical Commission et al. “Safety of laser products-Part 1: Equipment classification and requirements”. In: *IEC 60825-1* (2007). URL: https://infostore.saiglobal.com/preview/98701189622.pdf?sku=861160_saig_nsai_nsai_2048777.

[24] S. T. Kouyoumdjieva, P. Danielis, and G. Karlsson. “Survey of Non-Image-Based Approaches for Counting People”. In: *IEEE Communications Surveys Tutorials* 22.2 (2020), pp. 1305–1336. DOI: 10.1109/COMST.2019.2902824.

[25] S. Di Domenico et al. “LTE-based passive device-free crowd density estimation”. In: *2017 IEEE International Conference on Communications (ICC)*. 2017, pp. 1–6. DOI: 10.1109/ICC.2017.7997194.

[26] X. Yang et al. “Dense People Counting Using IR-UWB Radar With a Hybrid Feature Extraction Method”. In: *IEEE Geoscience and Remote Sensing Letters* 16.1 (2019), pp. 30–34. DOI: 10.1109/LGRS.2018.2869287.

[27] J. H. Choi, J. E. Kim, and K. T. Kim. “People Counting Using IR-UWB Radar Sensor in a Wide Area”. In: *IEEE Internet of Things Journal* 8.7 (2021), pp. 5806–5821. DOI: 10.1109/JIOT.2020.3032710.

[28] L. Servadei et al. “Label-Aware Ranked Loss for robust People Counting using Automotive in-cabin Radar”. In: *arXiv preprint arXiv:2110.05876* (2021). DOI: 10.48550/arXiv.2110.05876.

[29] G. Violatto and A. Pandharipande. “Anomaly Classification in People Counting and Occupancy Sensor Systems”. In: *IEEE Sensors Journal* 20.12 (2020), pp. 6573–6581. DOI: 10.1109/JSEN.2020.2976547.

[30] I. Sobron et al. “Device-Free People Counting in IoT Environments: New Insights, Results, and Open Challenges”. In: *IEEE Internet of Things Journal* 5.6 (2018), pp. 4396–4408. DOI: 10.1109/JIOT.2018.2806990.

[31] G. Gao et al. “Cnn-based density estimation and crowd counting: A survey”. In: *arXiv preprint arXiv:2003.12783* (2020). DOI: 10.48550/arXiv.2003.12783.

[32] F. Kraus et al. “Using Machine Learning to Detect Ghost Images in Automotive Radar”. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. 2020, pp. 1–7. DOI: 10.1109/ITSC45102.2020.9294631.

[33] S. Yoo et al. “Adaptive Clutter Suppression Algorithm for Detection and Positioning using IR-UWB Radar”. In: *2018 9th International Conference on Ultrawideband and Ultrashort Impulse Signals (UWBUSIS)*. 2018, pp. 40–43. DOI: 10.1109/UWBUSIS.2018.8520164.

[34] H. Abedi et al. “AI-Powered In-Vehicle Passenger Monitoring Using Low-Cost mm-Wave Radar”. In: *IEEE Access* 10 (2022), pp. 18998–19012. DOI: 10.1109/ACCESS.2021.3138051.

[35] J. W. Choi, J. H. Kim, and S. H. Cho. “A counting algorithm for multiple objects using an IR-UWB radar system”. In: *2012 3rd IEEE International Conference on Network Infrastructure and Digital Content*. 2012, pp. 591–595. DOI: 10.1109/ICNIDC.2012.6418823.

[36] J. He and A. Arora. “A regression-based radar-mote system for people counting”. In: *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2014, pp. 95–102. DOI: 10.1109/PerCom.2014.6813949.

[37] J. H. Choi, J. E. Kim, and K. T. Kim. “Deep Learning Approach for Radar-based People Counting”. In: *IEEE Internet of Things Journal* (2021), pp. 1–1. DOI: 10.1109/JIOT.2021.3113671.

[38] A. Ninos et al. “Radar-Based Robust People Tracking and Consumer Applications”. In: *IEEE Sensors Journal* 22.4 (2022), pp. 3726–3735. DOI: 10.1109/JSEN.2022.3141202.

[39] V. H. Nguyen and J. Y. Pyun. “Location Detection and Tracking of Moving Targets by a 2D IR-UWB Radar System”. In: *Sensors* 15.3 (2015), pp. 6740–6762. ISSN: 1424-8220. DOI: 10.3390/s150306740. URL: https://www.mdpi.com/1424-8220/15/3/6740.

[40] S. Häfner et al. “Mitigation of Leakage in FMCW Radars by Background Subtraction and Whitening”. In: *IEEE Microwave and Wireless Components Letters* 30.11 (2020), pp. 1105–1107. DOI: 10.1109/LMWC.2020.3023850.

[41] C. Will et al. “Human Target Detection, Tracking, and Classification Using 24-GHz FMCW Radar”. In: *IEEE Sensors Journal* 19.17 (2019), pp. 7283–7299. DOI: 10.1109/JSEN.2019.2914365.

[42] A. Rizik et al. “Cost-Efficient FMCW Radar for Multi-Target Classification in Security Gate Monitoring”. In: *IEEE Sensors Journal* 21.18 (2021), pp. 20447–20461. DOI: 10.1109/JSEN.2021.3095674.

[43] P. Nallabolu et al. “Human Presence Sensing and Gesture Recognition for Smart Home Applications With Moving and Stationary Clutter Suppression Using a 60-GHz Digital Beamforming FMCW Radar”. In: *IEEE Access* 9 (2021), pp. 72857–72866. DOI: 10.1109/ACCESS.2021.3080655.

[44] M.A. Abou-Khousa et al. “High-resolution short-range wideband FMCW radar measurements based on MUSIC algorithm”. In: *2009 IEEE Instrumentation and Measurement Technology Conference*. 2009, pp. 498–501. DOI: 10.1109/IMTC.2009.5168500.

[45] R. Feger et al. “A four channel 24-GHz FMCW radar sensor with two-dimensional target localization capabilities”. In: *2008 IEEE MTT-S International Microwave Symposium Digest*. 2008, pp. 125–128. DOI: 10.1109/MWSYM.2008.4633119.

[46] Y. Gürcan and A. Yarovoy. “Super-resolution algorithm for joint range-azimuth-Doppler estimation in automotive radars”. In: *2017 European Radar Conference (EURAD)*. 2017, pp. 73–76. DOI: 10.23919/EURAD.2017.8249150.

[47] M. A. Richards. *Fundamentals of radar signal processing*. McGraw-Hill Education, 2014.

[48] H. Rohling and R. Mende. “OS CFAR performance in a 77 GHz radar sensor for car application”. In: *Proceedings of International Radar Conference*. 1996, pp. 109–114. DOI: 10.1109/ICR.1996.573784.

[49] R. G. Zefreh et al. “Robust CFAR Detector Based on Censored Harmonic Averaging in Heterogeneous Clutter”. In: *IEEE Transactions on Aerospace and Electronic Systems* 57.3 (2021), pp. 1956–1963. DOI: 10.1109/TAES.2020.3046050.

[50] A. Jalil, H. Yousaf, and M. I. Baig. “Analysis of CFAR techniques”. In: *2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*. 2016, pp. 654–659. DOI: 10.1109/IBCAST.2016.7429949.

[51] A. Safa et al. “A Low-Complexity Radar Detector Outperforming OS-CFAR for Indoor Drone Obstacle Avoidance”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), pp. 9162–9175. DOI: 10.1109/JSTARS.2021.3107686.

[52] E. Grossi, M. Lops, and L. Venturino. “A Novel Dynamic Programming Algorithm for Track-Before-Detect in Radar Systems”. In: *IEEE Transactions on Signal Processing* 61.10 (2013), pp. 2608–2619. DOI: 10.1109/TSP.2013.2251338.

[53] B. Yan, A. Giorgetti, and E. Paolini. “A track-before-detect algorithm for UWB radar sensor networks”. In: *Signal Processing* 189 (2021), p. 108257. ISSN: 0165-1684. DOI: https://doi.org/10.1016/j.sigpro.2021.108257.

[54] T. Wagner, R. Feger, and A. Stelzer. “Modification of DBSCAN and application to range/Doppler/DoA measurements for pedestrian recognition with an automotive radar system”. In: *2015 European Radar Conference (EuRAD)*. 2015, pp. 269–272. DOI: 10.1109/EuRAD.2015.7346289.

[55] H. Chen and D. Kasilingam. “k-means classification filter for speckle removal in radar images”. In: *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*. Vol. 2. 1999, 1244–1246 vol.2. DOI: 10.1109/IGARSS.1999.774592.

[56] D. Kellner, J. Klappstein, and K. Dietmayer. “Grid-based DBSCAN for clustering extended objects in radar data”. In: *2012 IEEE Intelligent Vehicles Symposium*. 2012, pp. 365–370. DOI: 10.1109/IVS.2012.6232167.

[57] T. Wagner, R. Feger, and A. Stelzer. “Modifications of the OPTICS Clustering Algorithm for Short-Range Radar Tracking Applications”. In: *2018 15th European Radar Conference (EuRAD)*. 2018, pp. 91–94. DOI: 10.23919/EuRAD.2018.8546579.

[58] S.S. Blackman. “Multiple hypothesis tracking for multiple target tracking”. In: *IEEE Aerospace and Electronic Systems Magazine* 19.1 (2004), pp. 5–18. DOI: 10.1109/MAES.2004.1263228.

[59] D. E. Clark, K. Panta, and B. Vo. “The GM-PHD Filter Multiple Target Tracker”. In: *2006 9th International Conference on Information Fusion*. 2006, pp. 1–8. DOI: 10.1109/ICIF.2006.301809.

[60] K. Granström, M. Fatemi, and L. Svensson. “Poisson Multi-Bernoulli Mixture Conjugate Prior for Multiple Extended Target Filtering”. In: *IEEE Transactions on Aerospace and Electronic Systems* 56.1 (2020), pp. 208–225. DOI: 10.1109/TAES.2019.2920220.

[61] TI document. *People Counting Demonstration Using TI mmWave Sensors*. 2021. URL: https://training.ti.com/people-counting-%20demonstration-using-ti-mmwave-sensors.

[62] J. Shen et al. “Evaluation of Unscented Kalman Filter and Extended Kalman Filter for Radar Tracking Data Filtering”. In: *2014 European Modelling Symposium*. 2014, pp. 190–194. DOI: 10.1109/EMS.2014.49.

[63] S. Oh, S. Russell, and S. Sastry. “Markov Chain Monte Carlo Data Association for Multi-Target Tracking”. In: *IEEE Transactions on Automatic Control* 54.3 (2009), pp. 481–497. DOI: 10.1109/TAC.2009.2012975.

[64] J. Pegoraro and M. Rossi. “Real-Time People Tracking and Identification From Sparse mm-Wave Radar Point-Clouds”. In: *IEEE Access* 9 (2021), pp. 78504–78520. DOI: 10.1109/ACCESS.2021.3083980.

[65] N. Knudde et al. “Indoor tracking of multiple persons with a 77 GHz MIMO FMCW radar”. In: *2017 European Radar Conference (EURAD)*. 2017, pp. 61–64. DOI: 10.23919/EURAD.2017.8249147.

[66] Y. Jia et al. “ResNet-Based Counting Algorithm for Moving Targets in Through-the-Wall Radar”. In: *IEEE Geoscience and Remote Sensing Letters* 18.6 (2021), pp. 1034–1038. DOI: 10.1109/LGRS.2020.2990742.

[67] H. Li et al. “Human Target Detection Based on FCN for Through-the-Wall Radar Imaging”. In: *IEEE Geoscience and Remote Sensing Letters* 18.9 (2021), pp. 1565–1569. DOI: 10.1109/LGRS.2020.3006077.

[68] I. Alujaim, I. Park, and Y. Kim. “Human Motion Detection Using Planar Array FMCW Radar Through 3D Point Clouds”. In: *2020 14th European Conference on Antennas and Propagation (EuCAP)*. 2020, pp. 1–3. DOI: 10.23919/EuCAP48036.2020.9135381.

[69] G. Lee and J. Kim. “Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-Trained 3D Human-Joint Coordinates”. In: *Applied Sciences* 12.4 (2022). ISSN: 2076-3417. DOI: 10.3390/app12042168. URL: https://www.mdpi.com/2076-3417/12/4/2168.

[70] A. D. Singh et al. “Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar”. In: *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*. 2019, pp. 51–56.

[71] J. Misiurewicz et al. “Radar detection of helicopters with application of CLEAN method”. In: *IEEE Transactions on Aerospace and Electronic Systems* 48.4 (2012), pp. 3525–3537.

[72] Y. Yu et al. “A review of recurrent neural networks: LSTM cells and network architectures”. In: *Neural computation* 31.7 (2019), pp. 1235–1270.

[73] A. Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[74] I. Jordanov, N. Petrov, and A. Petrozziello. “Supervised radar signal classification”. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2016, pp. 1464–1471.

[75] Z. Guo and P. Shui. “Anomaly based sea-surface small target detection using K-nearest neighbor classification”. In: *IEEE Transactions on Aerospace and Electronic Systems* 56.6 (2020), pp. 4947–4964.

[76] H. Li et al. “Sequential human gait classification with distributed radar sensor fusion”. In: *IEEE Sensors Journal* 21.6 (2020), pp. 7590–7603.

[77] Y. Lin et al. “Human activity classification with radar: Optimization and noise robustness with iterative convolutional neural networks followed with random forests”. In: *IEEE Sensors Journal* 18.23 (2018), pp. 9669–9681.

[78] T. D. Bufler and R. M. Narayanan. “Radar classification of indoor targets using support vector machines”. In: *IET Radar, Sonar & Navigation* 10.8 (2016), pp. 1468–1476.

[79] M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz. “Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities”. In: *IEEE Transactions on Aerospace and Electronic Systems* 54.4 (2018), pp. 1709–1723.

[80] A. Sinha et al. “Track quality based multitarget tracking approach for global nearest-neighbor association”. In: *IEEE Transactions on Aerospace and Electronic Systems* 48.2 (2012), pp. 1179–1191.

[81] M. Sebt, A. Sheikhi, and M. Nayebi. “Robust low-angle estimation by an array radar”. In: *IET radar, sonar & navigation* 4.6 (2010), pp. 780–790.

[82] J. Pegoraro, F. Meneghello, and M. Rossi. “Multiperson Continuous Tracking and Identification From mm-Wave Micro-Doppler Signatures”. In: *IEEE Transactions on Geoscience and Remote Sensing* 59.4 (2021), pp. 2994–3009. DOI: 10.1109/TGRS.2020.3019915.

[83] M. Heuer et al. “Pedestrian tracking with occlusion using a 24 GHz automotive radar”. In: *2014 15th International Radar Symposium (IRS)*. 2014, pp. 1–4. DOI: 10.1109/IRS.2014.6869181.

[84] P. Zhao et al. “mID: Tracking and Identifying People with Millimeter Wave Radar”. In: *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. 2019, pp. 33–40. DOI: 10.1109/DCOSS.2019.00028.

[85] G. Gennarelli et al. “Multiple Extended Target Tracking for Through-Wall Radars”. In: *IEEE Transactions on Geoscience and Remote Sensing* 53.12 (2015), pp. 6482–6494. DOI: 10.1109/TGRS.2015.2441957.

[86] C.T. Pham et al. “Convolutional neural network for people counting using UWB impulse radar”. In: *Journal of Instrumentation* 16.08 (2021), p. 8031. DOI: 10.1088/1748-0221/16/08/p08031.

[87] H. Li et al. “Robust Human Targets Tracking for MIMO Through-Wall Radar via Multi-Algorithm Fusion”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.4 (2019), pp. 1154–1164. DOI: 10.1109/JSTARS.2019.2901262.

[88] R.P. Trommel et al. “Multi-target human gait classification using deep convolutional neural networks on micro-doppler spectrograms”. In: *2016 European Radar Conference (EuRAD)*. 2016, pp. 81–84.

[89] X. Yang, W. Yin, and L. Zhang. “People Counting Based on CNN Using IR-UWB Radar”. In: *2017 IEEE/CIC International Conference on Communications in China (ICCC)*. 2017, pp. 1–5. DOI: 10.1109/ICCChina.2017.8330453.

[90] R. Bao and Z. Yang. “CNN-Based Regional People Counting Algorithm Exploiting Multi-Scale Range-Time Maps With an IR-UWB Radar”. In: *IEEE Sensors Journal* 21.12 (2021), pp. 13704–13713. DOI: 10.1109/JSEN.2021.3071941.

[91] X. Huang et al. “Indoor Detection and Tracking of People Using mmwave Sensor”. In: *Journal of Sensors* 2021 (2021). DOI: 10.1155/2021/6657709.

[92] P. Van Dorp and F. Groen. “Feature-based human motion parameter estimation with radar”. In: *IET Radar, Sonar & Navigation* 2.2 (2008), pp. 135–145.

[93] S. Lee et al. “Radar cross section measurement with 77 GHz automotive FMCW radar”. In: *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. 2016, pp. 1–6. DOI: 10.1109/PIMRC.2016.7794738.

[94] J. A. Balderrama, F. J. Masters, and K. R. Gurley. “Peak factor estimation in hurricane surface winds”. In: *Journal of wind engineering and industrial aerodynamics* 102 (2012), pp. 1–13.

[95] C. Kuang et al. “An improved CA-CFAR method for ship target detection in strong clutter using UHF radar”. In: *IEEE Signal Processing Letters* 27 (2020), pp. 1445–1449.

[96] D. Kellner, J. Klappstein, and K. Dietmayer. “Grid-based DBSCAN for clustering extended objects in radar data”. In: *2012 IEEE Intelligent Vehicles Symposium*. 2012, pp. 365–370. DOI: 10.1109/IVS.2012.6232167.

[97] R. Ricci and A. Balleri. “Recognition of humans based on radar micro-Doppler shape spectrum features”. In: *IET Radar, Sonar & Navigation* 9.9 (2015), pp. 1216–1223.

[98] Zhenyu Liu et al. “A method of SVM with normalization in intrusion detection”. In: *Procedia Environmental Sciences* 11 (2011), pp. 256–262.

[99] K. He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[100] K. Cho et al. “On the properties of neural machine translation: Encoder-decoder approaches”. In: *arXiv preprint arXiv:1409.1259* (2014).

[101] J. Bai et al. “Radar transformer: An object classification network based on 4d mmw imaging radar”. In: *Sensors* 21.11 (2021), p. 3854.

[102] P. D. Konstantinova, A. Udvarev, and T. Semerdjiev. “A study of a target tracking algorithm using global nearest neighbor approach.” In: *Compsystech*. Vol. 3. 2003, pp. 290–295.

[103] Texas Instruments Incorporated. *mmWave studio | TI.com*. 2022. URL: https://www.ti.com/tool/MMWAVE-STUDIO.

[104] Landbouw en Innovatie Ministerie van Economische Zaken. *Regels Cameratoezicht | Rijksoverheid.nl*. 2022. URL: https://www.rijksoverheid.nl/onderwerpen/overvallen-straatroof-en-woninginbraak/regels-cameratoezicht.

[105] K. Granstrom, C. Lundquist, and O. Orguner. “Extended Target Tracking using a Gaussian-Mixture PHD Filter”. In: *IEEE Transactions on Aerospace and Electronic Systems* 48.4 (2012), pp. 3268–3286. DOI: 10.1109/TAES.2012.6324703.

[106] K. Granstrom and U. Orguner. “On Spawning and Combination of Extended/Group Targets Modeled With Random Matrices”. In: *IEEE Transactions on Signal Processing* 61.3 (2013), pp. 678–692. DOI: 10.1109/TSP.2012.2230171.

[107] K. Dai et al. “Intertarget Occlusion Handling in Multiextended Target Tracking Based on Labeled Multi-Bernoulli Filter Using Laser Range Finder”. In: *IEEE/ASME Transactions on Mechatronics* 25.4 (2020), pp. 1719–1728. DOI: 10.1109/TMECH.2020.2994066.

[108] A. Seifert, M. Grimmer, and A. M. Zoubir. “Doppler Radar for the Extraction of Biomechanical Parameters in Gait Analysis”. In: *IEEE Journal of Biomedical and Health Informatics* 25.2 (2021), pp. 547–558. DOI: 10.1109/JBHI.2020.2994471.

[109] K. Bernardin, A. Elbs, and R. Stiefelhagen. “Multiple object tracking performance metrics and evaluation in a smart room environment”. In: *Sixth IEEE International Workshop on Visual Surveillance, in conjunction with ECCV*. Vol. 90. 91. Citeseer. 2006.

[110] Anton Milan, Konrad Schindler, and Stefan Roth. “Challenges of ground truth evaluation of multi-target tracking”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2013, pp. 735–742.

[111] C. Thornton et al. “Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms”. In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013, pp. 847–855.

[112] K. M. Ting and Z. Zheng. “A study of adaboost with naive bayesian classifiers: Weakness and improvement”. In: *Computational Intelligence* 19.2 (2003), pp. 186–200.

[113] D. S. Bryant et al. “The CPHD filter with target spawning”. In: *IEEE Transactions on Signal Processing* 65.5 (2016), pp. 13124–13138.

[114] Texas Instruments Incorporated. *IWR6843ISK Evaluation board | TI.com*. 2022. URL: https://www.ti.com/tool/IWR6843ISK.

[115] Texas Instruments Incorporated. *MMWAVEICBOOST mmWave sensors carrier card platform | TI.com*. 2022. URL: https://www.ti.com/tool/MMWAVEICBOOST.

[116] Texas Instruments Incorporated. *DCA1000EVM Real-time data-capture adapter for radar sensing evaluation module | TI.com*. 2022. URL: https://www.ti.com/tool/DCA1000EVM.

A

CONFIGURATION OF TI RADAR IN DETAILS

In Appendix A, the detailed information of Radar equipment and the basic principle of MIMO mode of TI Radar are given.

The TI hardware includes IWR6843ISK, mmWave ICBOOST board, DCA1000EVM. The supporting software is mmWave Studio (listed in Table A.1).

IWR6843ISK is a 60 GHz mmWave sensor based on IWR6843 device with long-range antenna. This board enables access to point-cloud data and power via the USB interface. Additional boards may be used to enable additional functionality. DCA1000EVM enables access to sensor raw data via LVDS interface. The mmWave ICBOOST board enables software development and trace capabilities through TI's packaged labs. This equipment is supported by mmWave Studio. mmWave Studio enables you to build manual frames and chirps in different user cases. Meanwhile, the MIMO mode and TX Beamforming can also be designed when using mmWave Studio. In this thesis, the MIMO mode is applied, since the beamforming mode can not provide enough Doppler information for classification, which is necessary for Group classification.

MIMO refers to a radar with multiple TX and multiple RX antennas. According to Fig. A.1, when using 3 TX and 4 RX antennas, users can generate a virtual array of 12 antennas. Thus, employing MIMO radar techniques results in an increase in the number of (virtual) antennas. In this case, the angle resolution is improved.

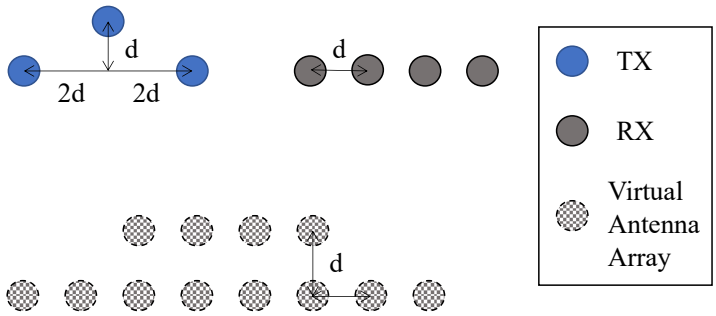


Figure A.1: Example of generating a MIMO virtual antenna array.

In the MIMO application, the transmitted electromagnetic waves generated by different TX antennas must be distinguishable. In other words, it should be achievable to separate different signals corresponding to different TX antennas. Two common techniques are time division multiplexing (TDM) and binary phase modulation (BPM). In this thesis, TDM-MIMO is used.

In TDM-MIMO, the orthogonality is in time. Each frame consists of several subframes, each subframe consisting of one TX. TDM-MIMO is the simplest way to separate signals from multiple TX antennas and is therefore widely used. However, the weakness of TDM-MIMO is that only one TX is active at any time. It does not use the full transmission capabilities of the device. Meanwhile, there is also an implication on the velocity of the object in the range of interest. It cannot go too fast since it needs to be observed with all TXs one after the other.

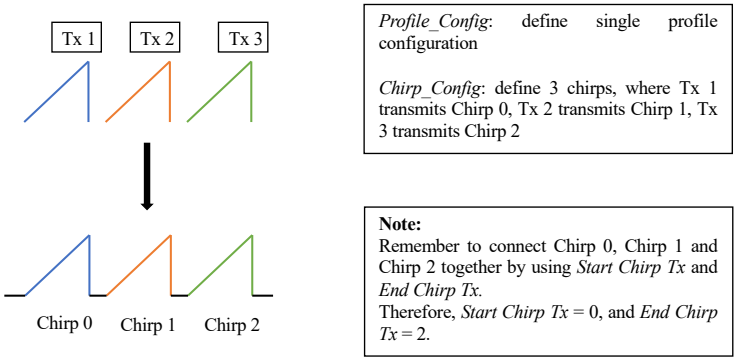


Figure A.2: Steps of setting TDM-MIMO mode with the TI 60 GHz radar used in this thesis.

In the mmWave Studio, the TDM-MIMO can be programmed as the flow chart shown in Figure A.2. There are three concepts when programming a TX signal in MIMO mode:

profile, chirp, and frame. Each of these concepts is briefly described as follows.

• Profile: A profile is a template for a chirp. This includes TX parameters such as the start frequency, slope, duration, and idle time, and RX parameters such as ADC sampling rate. Up to four different profiles can be defined and stored. In the TDM-MIMO mode, all chirps have the same parameters. Therefore, a single profile configuration is needed, which is named as Profile 0.

• Chirp: Each chirp type is associated with a profile and inherits all the properties of the profile. Additional properties that can be associated with each chirp include the TX antennas on which the chirp should be transmitted and any binary phase modulation that should be applied. In the TDM-MIMO mode, since 3 TX are used, 3 Chirp configurations are programmed. Chirp 0 enables TX 1, Chirp 1 enables TX 2 and Chirp 2 enables TX 3.

• Frame: Frame is constructed by defining a sequence of chirps using the previously defined chirp types. The number of frames transmitted in a single experiment can be defined. If the number of frames is 0, it means continuously transmitting signals until users press "Stop Frame" button.

Note: If you enable 3 TX one after another in different chirp ID (for example, chirp 0 enables TX0, chirp1 enables TX1 and chirp 2 enables TX2), then it is called TDM MIMO. And the total chirp number will be num_loop $*$ numChirp = 128$*$3. But if you just enable 3 TX simultaneously in one chirp ID (chirp 0 enables all three TX0, TX1 and TX2), then you only have one chirp ID (chirp 0), and the total chirp number should be num_loop $*$ 1 = 128.

Table A.1: Description of hardware and software used for the experimental work of this thesis.

Hardware/software	**Description**
IWR6843ISK [114]	mmWave sensor standard antenna plug-in module
mmWave ICBOOST [115]	mmWave sensors carrier card platform
DCA1000EVM [116]	Real-time data-capture adapter for radar sensing evaluation module
mmWave Studio 2.0.1.0 [103]	Configure and control mmWave sensor modules and collect analog-to-digital (ADC) data for offline analysis

B

Ablation Study

Appendix B is to provide the confusion matrix of the classifiers. Appendix B.1 provides the confusion matrix of the far range study of the proposed non-ANN classifier. Appendix B.3 provides the confusion matrix of different classifiers of the proposed non-ANN classifier.

B.1. Far Range Study of the Proposed Non-ANN Classifier

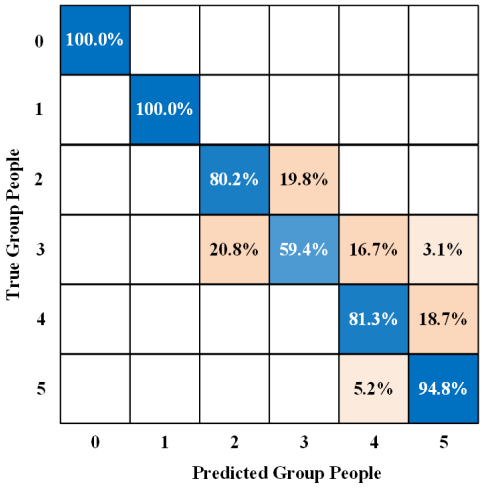


Figure B.1: Confusion matrix of non-ANN classifier when the input is the Range-Azimuth map.

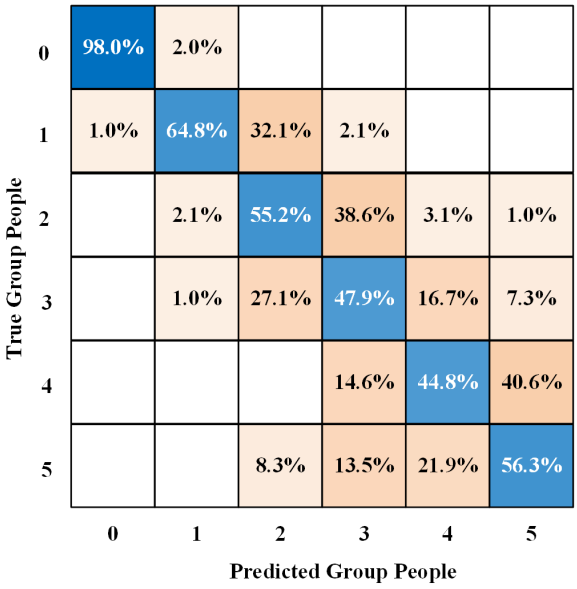


Figure B.2: Confusion matrix of non-ANN classifier when the input is the CVD map.

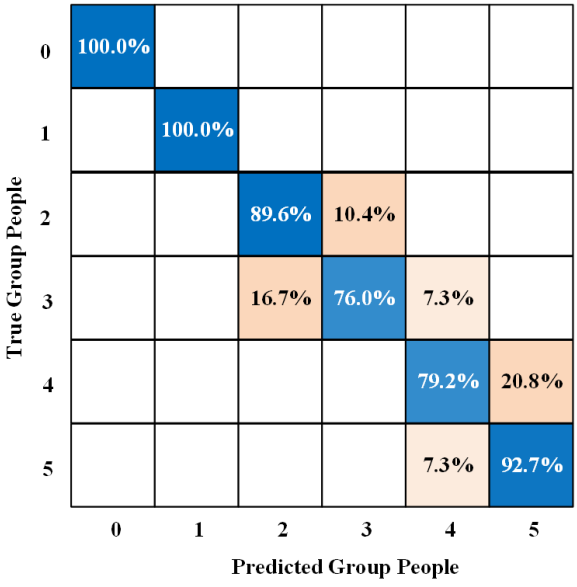


Figure B.3: Confusion matrix of non-ANN classifier when inputs are the Range-Azimuth map and the CVD map.

B.2. DIFFERENT CLASSIFIERS OF THE PROPOSED NON-ANN CLASSIFIER

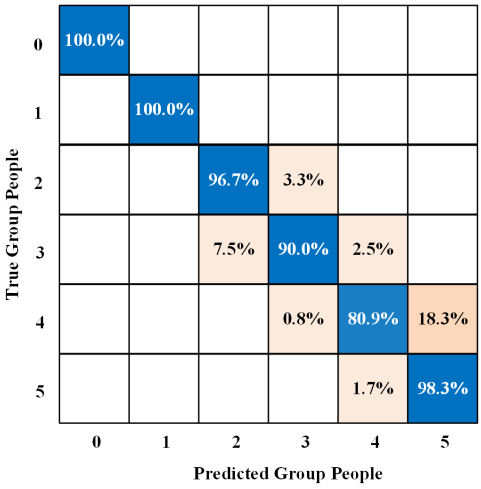


Figure B.4: Confusion matrix of non-ANN classifier when the classifier is SVM.

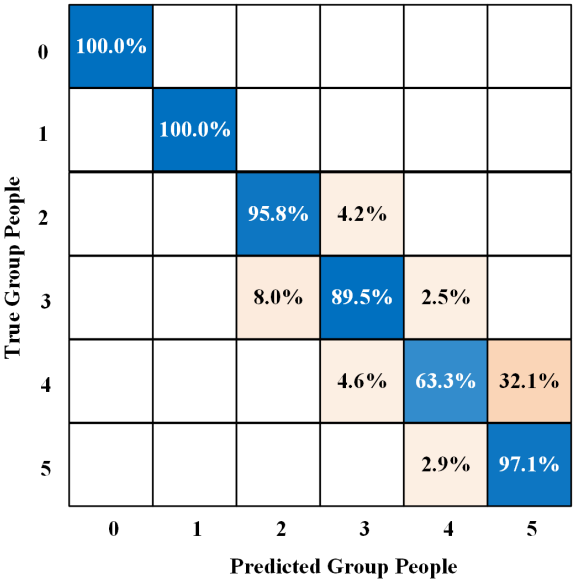


Figure B.5: Confusion matrix of non-ANN classifier when the classifier is KNN.

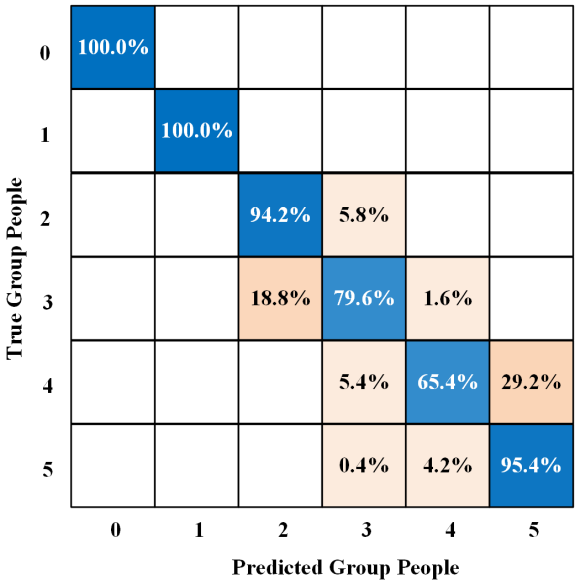


Figure B.6: Confusion matrix of non-ANN classifier when the classifier is the random forest.

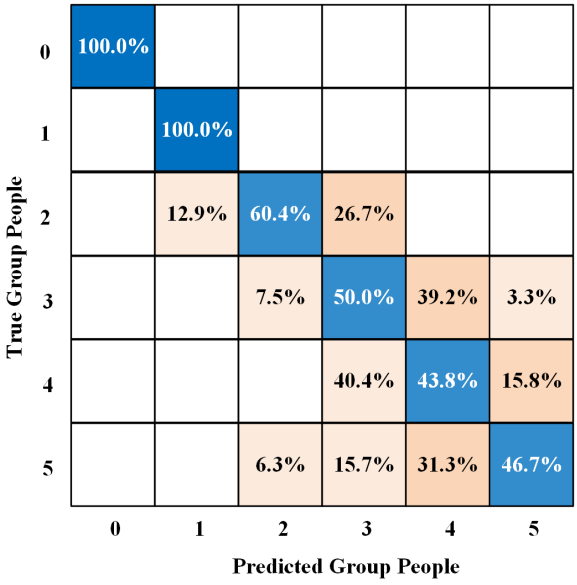


Figure B.7: Confusion matrix of non-ANN classifier when the classifier is naive Bayes.

B.3. ABLATION STUDY OF THE PROPOSED ANN CLASSIFIER

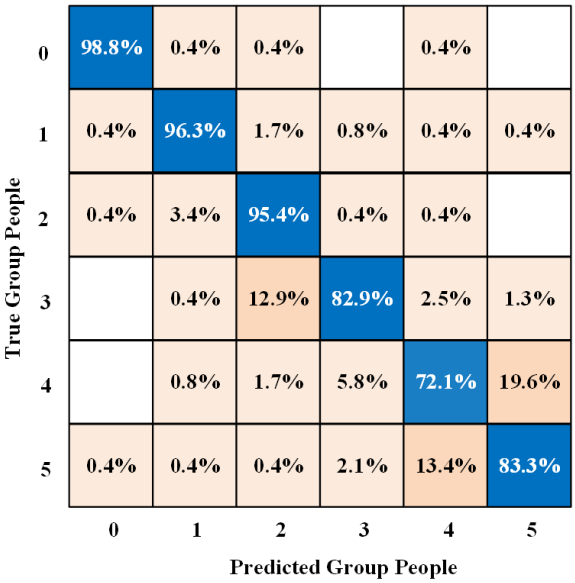


Figure B.8: Confusion matrix of proposed ANN classifier when the input is Range-Azimut map.

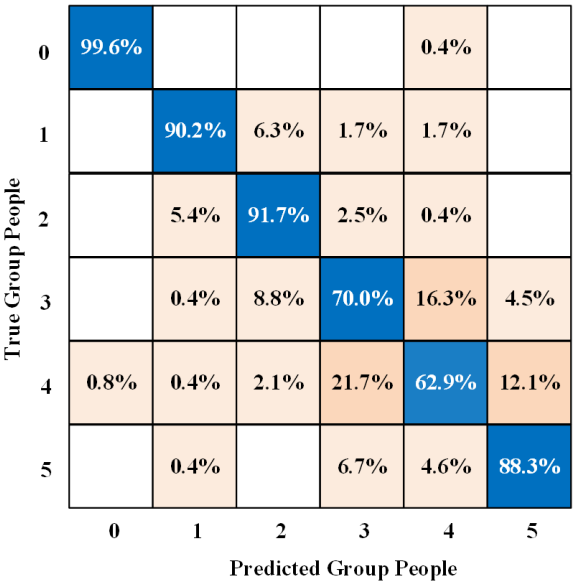


Figure B.9: Confusion matrix of proposed ANN classifier when the input is the spectrogram.

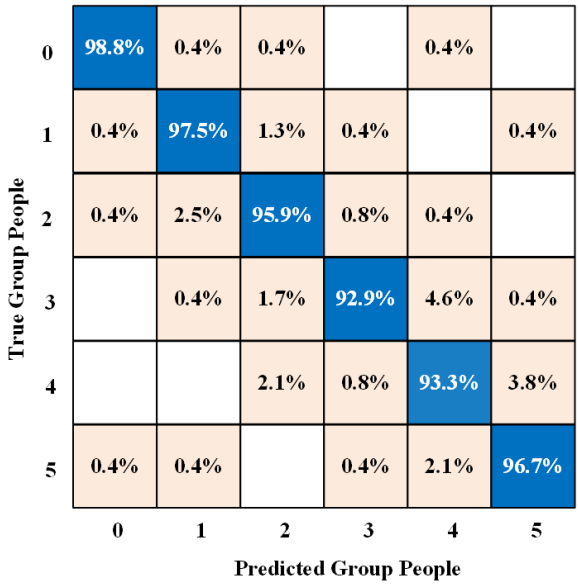


Figure B.10: Confusion matrix of proposed ANN classifier when the inputs are the spectrogram and Range-Azimuth map.

C

Performance Comparison

This chapter provides the results when applying different the state-of-the-art (SOTA) methods.

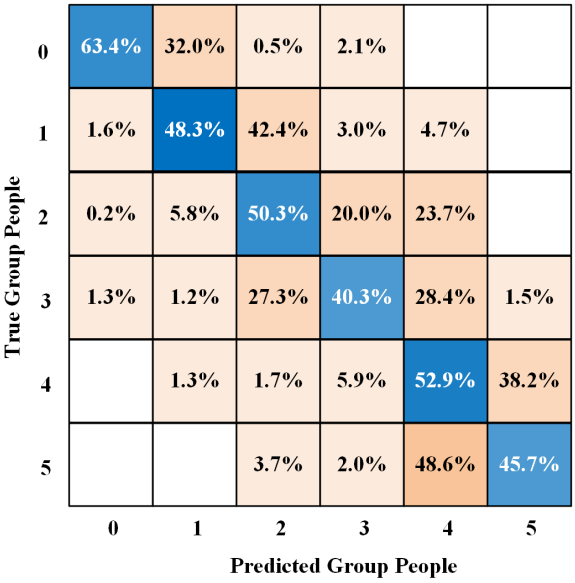


Figure C.1: Confusion matrix of the non-ANN method proposed by Choi et al. [19].

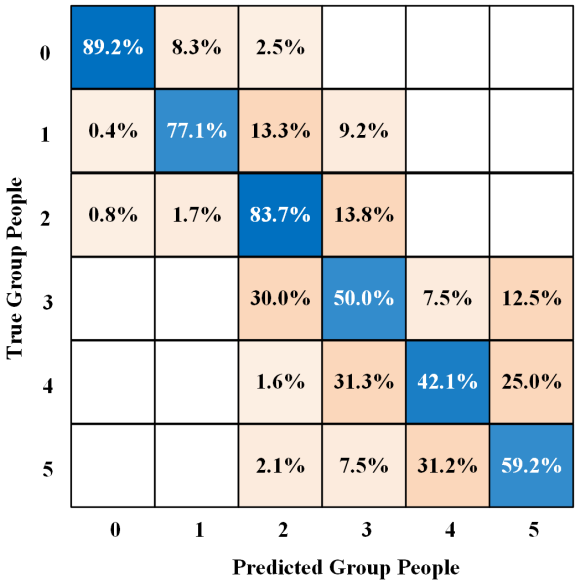


Figure C.2: Confusion matrix of the non-ANN method proposed by Choi et al. [27].

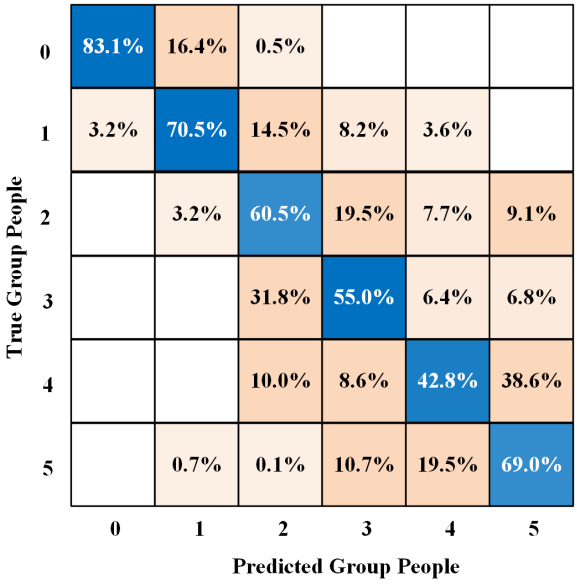


Figure C.3: Confusion matrix of the ANN method proposed by Choi et al. [37].

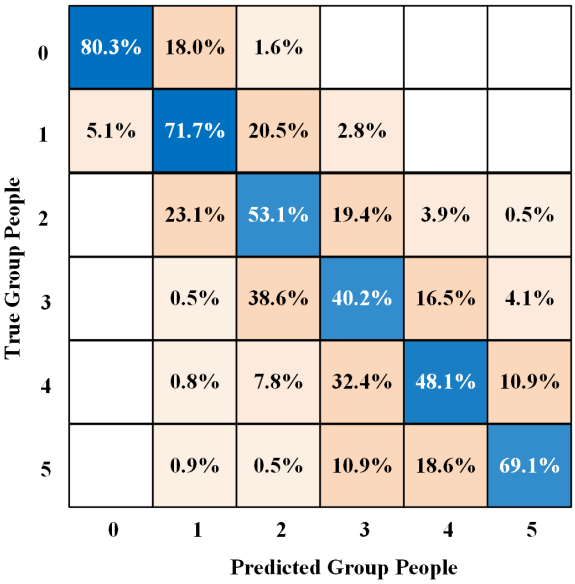


Figure C.4: Confusion matrix of the ANN method proposed by Jia et al. [66].