Classification of Human Activities with Distributed Radar Systems

Guendel, Ronny

DOI
10.4233/uuid:d2e1100d-af4e-4af7-8124-41ca5fa881c1

Publication date
2023

Document Version
Final published version

Citation (APA)
Guendel, R. (2023). *Classification of Human Activities with Distributed Radar Systems*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:d2e1100d-af4e-4af7-8124-41ca5fa881c1

Classification of Human Activities with Distributed Radar Systems

Classification of Human Activities with Distributed Radar Systems

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Thursday, 30 November 2023 at 15:00 o'clock

by

Ronny Gerhard GÜNDEL

Master of Science in Electrical Engineering,
Villanova University, Pennsylvania, United States
born in Greiz, Germany

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. DSc. A. Yarovoy	Delft University of Technology, promotor
Dr. F. Fioranelli	Delft University of Technology, promotor
Independent members:	
Prof. dr. ir. W. A. Serdijn	Delft University of Technology
Dr. ir. J. Dauwels	Delft University of Technology
Dr. ir. M. A. Zuñiga Zamalloa	Delft University of Technology
Dr. ir. J. Le Kernec	University of Glasgow, UK
Dr. ir. R. Roovers	NXP Semiconductors, NL
Reserve member:	
Prof. dr. M. Neerincx	Delft University of Technology / TNO, NL

TU Delft Delft University of Technology

This research has been carried out at the Delft University of Technology in the Microwave Sensing, Signals and Systems (MS3) group.

Keywords: radar signal processing, ultra wideband radar, radar sensor network, distributed radar, human activity recognition, micro-Doppler signatures, deep learning.

Printed by: Proefschriftspecialist, 1506RZ Zaandam, The Netherlands.

ISBN/EAN: 978-94-6366-769-2 (Paperback/softback)
ISBN/EAN: 978-94-6366-768-5 (E-book, PDF)

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

Author e-mail: r.guendel@icloud.com

To my family and friends, for their support and encouragement

Contents

List of Acronyms

(Bi-)GRU	(bidirectional) gated recurrent unit
(Bi-)LSTM	(bidirectional) long short-term memory
µD	micro-Doppler
ADAM	adaptive moment estimation
AoA	angle of arrival
CBD	consecutive block detection
CNN	convolutional neural network
CPI	coherent processing interval
DL	deep learning
DTL	derivative target line
EFS	exhaustive feature selection
FFT	fast Fourier transform
FMCW	frequency-modulated continuous-wave
FSST	Fourier synchrosqueezed transform
GAN	generative adversarial networks
GNN	graph neural network
HAR	human activity recognition
HOG	histogram of oriented gradients
HOTA	higher order metric for evaluating multi-object tracking
IoU	intersection over union
KNN/NN	(k-)nearest neighbors
LOGO	leave-one-group-out
LOPO	leave-one-person-out
LOS	line of sight
LR	linear regression
MAE	mean absolute error
MAL	multipath-assisted localization
MCC	Matthews correlation coefficient
MFCC	mel-frequency cepstral coefficients
MIMO	multiple-input multiple-output
ML	machine learning
MLP	multilayer perceptron
MOCAP	motion capture
MSE	mean squared error
MTI	moving target indication/indicator
OLS	ordinary least squares
OMP	orthogonal matching pursuit
PCA	principal component analysis

PRF	pulse repetition frequency
PRI	pulse repetition interval
PUT	phase unwrapping techniques
RBF	radial basis function
RCS	radar cross section
RD	range-Doppler
RDS	range-Doppler surface
RF	radio frequency
RMSE	root mean square error
RMSprop	root mean squared propagation
RNN	recurrent neural networks
RT/RM	range-time/range-time map
SDR	software-defined radio
SFS	sequential feature selection
SGDM	stochastic gradient descent with momentum
SISO	single-input single-output
SMOTE	synthetic minority over-sampling technique
SNR	signal to noise ratio
STAP	space-time adaptive processing
STFT	short time Fourier transform
ST-GCN	spatial-temporal graph convolutional network
SVD	singular value decomposition
SVM	support vector machine
t-SNE	t-distributed stochastic neighbor embedding
UWB	ultra-wideband
WVD	Wigner-Ville distribution

Summary

This thesis introduces the relevance of radar systems in the realm of human activity recognition (HAR) in Chapter 1. The study touches upon the complex understanding of continuous human activities and the existing challenges and gaps in current methodologies, hinting at the innovative technical approaches that are to be detailed in the following chapters.

The technical foundation of the research is given in Chapter 2 by introducing distributed ultra-wideband (UWB) radar systems. These systems, especially when spatially distributed, bring a depth of information by integrating data from multiple radar nodes and spatial perspectives. There is a significant emphasis on how different fusion techniques, both late and early, play a crucial role in harnessing data effectively, particularly in the context of HAR.

A critical contribution in the study is the potential to deviate from conventional radar data domains, such as micro-Doppler spectrograms for activity recognition. The research in Chapter 3 highlights an alternative approach, rooted in the radar phase information from a high-resolution range-time map, which bypasses the limitations of common FFT-based radar data domains. This methodology, paired with the histogram of oriented gradients (HOG) algorithm, showcases promising results that can be particularly interesting for real-time applications with computational constraints.

The research in Chapter 4 underlines the efficacy of employing a network of spatially distributed UWB radars for continuous HAR. These networks address the downsides of using a single sensor, like unfavorable aspect-angle observations. The study delves into fusion methodologies and their implementation in classifying activities, particularly using recurrent neural networks. To assess these continuous recognition systems, novel evaluation metrics are proposed, offering a deeper insight into the practicality and effectiveness of such systems with temporal classification capabilities.

Indoor radar networks often face multipath challenges. The study in Chapter 5 not only identifies this challenge, but also uses the multipath components by leveraging these typically unwanted phenomena to enhance classification capabilities. Through a pipeline that isolates, determines, and analyzes different propagation pathways, there is an evident boost in the network's perception. This novel approach showcases a significant performance upward trend, especially when employing convolutional neural networks.

Chapter 6 of the research focuses on the complexities of HAR in crowded environments. The study introduces the challenges of differentiating the activities of walking versus standing idle for multiple individuals simultaneously. The investigation shows initial promising results by using synthetic data generated from experimental recordings, by employing a regression-based approach and leveraging diverse techniques such as LSTM, CNN, SVM, and linear regression.

In conclusion, the research offers a reflective glance at the breakthroughs achieved in the domain of radar-based HAR in Chapter 7. The significant contributions and

advancements of the study are highlighted. Looking ahead, the chapter identifies research areas for exploration and further improvement.

Samenvatting

In dit proefschrift zal de relevantie van radarsystemen met betrekking tot human activity recognition (HAR) worden geïntroduceerd in Hoofdstuk 1. De studie raakt aan het complexe begrip van continue menselijke activiteiten en de bestaande uitdagingen en tekortkomingen van huidige methodes en verwijst naar de innovatieve aanpakken die in volgende hoofdstukken aan bod zullen komen.

De technische basis van het onderzoek wordt gegeven in Hoofdstuk 2 door de introductie van gedistribueerde ultrabreedband (UWB) radarsystemen. Deze systemen, met name wanneer zij ruimtelijk gedistribueerd zijn, bieden diepgaande informatie door middel van het integreren van data van verscheidene sensoren en perspectieven. Extra nadruk wordt gelegd op hoe verschillende fusietechnieken, zowel laat als vroeg, een cruciale rol spelen in het effectief benutten van beschikbare data, met name in de context van HAR.

Een cruciale bijdrage van de studie is het potentieel om af te wijken van conventionele radardatadomeinen zoals micro-Doppler spectrograms voor activiteitenclassificatie. Het onderzoek in Hoofdstuk 3 belicht een alternatieve aanpak, geworteld in de fase-informatie in een afstand-tijdweergave met hoge resolutie, welke de beperkingen van gebruikelijke FFT-gebaseerde radardatadomeinen omzeilt. Deze methodologie, in combinatie met het histogram of oriented gradients (HOG) algoritme, toont veelbelovende resultaten die met name interessant kunnen zijn voor realtime toepassingen met rekenkrachtbeperkingen.

Het onderzoek in Hoofdstuk 4 onderstreept de effectiviteit van de benutting van een netwerk van ruimtelijk gedistribueerde UWB radars voor continue activiteitenclassificatie. Zulke netwerken bieden uitkomst voor de nadelen van het benutten van individuele sensoren zoals observaties onder ongunstige aspecthoek. De studie verdiept zich in fusiemethodes en hun implementatie bij het classificeren van activiteiten, met name bij recurrente neurale netwerken. Ter beoordeling van deze continue classificatiesystemen worden nieuwe evaluatiemetrieken voorgesteld, welke een dieper inzicht bieden in de geschiktheid en effectiviteit van zulke systemen met tijdsafhankelijke classificatiemogelijkheden.

Overdekte radarnetwerken hebben vaak te kampen met uitdagingen op het gebied van multipath. De studie in Hoofdstuk 5 identificeert niet alleen deze uitdagingen, maar benut de doorgaans ongewenste componenten van het multipath-signaal ter verbetering van het classificatievermogen. Door een verwerkingsmethode die verschillende propagatieroutes isoleert, herkent en analyseert, is er een duidelijke vooruitgang in de perceptie van het netwerk te zien. Deze nieuwe aanpak toont een significante opwaardse trent in prestaties, met name bij het gebruik van convolutionele neurale netwerken.

Hoofdstuk 6 van het onderzoek legt de aandacht op de complexiteit van HAR in drukbezochte omgevingen. De studie introduceert de uitdagingen omtrent het onderscheiden tussen lopen en stilstaan voor meerdere personen tegelijkertijd. De veelbelovende eerste resultaten van de studie zijn behaald met een aanpak gebaseerd op

regressie, waarbij synthetische data, gegenereerd uit experimentele opnamen, is gebruikt en waar gebruik is gemaakt van diverse technieken zoals LSTM, CNN, SVM en lineaire regressie.

Concluderend biedt het onderzoek in Hoofdstuk 7 een terugkijkende blik op de doorbraken die behaald zijn op het gebied van radargebaseerde HAR. De belangrijkste bijdragen en vorderingen van het onderzoek worden benadrukt. Vooruitkijkend worden in dit hoofdstuk relevante onderzoeksthema's voor verkenning en verbetering geïdentificeerd.

Preface

As I embarked on my journey as a Ph.D. candidate at TU Delft, I entered with a solid foundational understanding, thanks to my Master's supervisor at Villanova University, Dr. Moeness G. Amin. Until this time, my academic focus had primarily revolved around fundamental radar theories, which left me with limited knowledge of the prevalent techniques and methods employed in radar machine learning, deep learning, and related fields. However, the first year proved to be a promising start, largely attributable to the exceptional support and guidance of both my promotors, Prof. DSc. Olexander Yarovyi and Dr. Francesco Fioranelli. As experts in fields spanning from machine learning to advanced radar signal processing, they provided extraordinary feedback. Their contributions went beyond technical expertise, as they also demonstrated commendable soft-skill capabilities.

Throughout this process, I have undergone significant personal growth and honed my skills as a discerning researcher. Despite facing numerous challenges and navigating the constraints imposed by the COVID-19 pandemic, my research steadily gained momentum over time. Gradually, I began to understand the intricate nuances of my research problem and gained a comprehensive understanding of its essential components.

This has been a profound personal journey, replete with moments of triumph and tribulation. It is worth noting that the majority of this odyssey unfolded within the confines of the pandemic restrictions. However, I can now reflect on the invaluable lessons I have learned about myself, particularly in terms of tackling complex problems and effectively organizing my endeavors.

As I sit here writing this preface, I take solace in the knowledge that I have arrived at the culmination of my Ph.D. expedition. This dissertation serves as a comprehensive summary of my research findings. My fervent hope is that my contributions will inspire other researchers, spurring the development of novel ideas and approaches.

Ronny Gerhard Gündel
Delft, November 2023

1

Introduction to Distributed Radars for Activity Recognition

This chapter provides an overview of the context and motivation behind the utilization of radar systems for human activity recognition (HAR). It introduces the concept of a human ethogram which is related to the characterization of continuous sequences of human activities. Additionally, the chapter discusses various challenges, open problems, and the chosen research approach that will be addressed in subsequent chapters. Finally, the chapter concludes by presenting an outline of the upcoming content of the entire thesis.

Parts of this chapter have been published in:

R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Derivative Target Line (DTL) for Continuous Human Activity Detection and Recognition", in *2020 IEEE Radar Conference (RadarConf20)*, 2020.
I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli and A. Yarovoy, "A Survey on Radar-Based Continuous Human Activity Recognition", in *IEEE Journal of Microwaves*, 2023.

1

1.1. Radar Systems for Human Activity Recognition

Radar technology has emerged in recent years as a prominent and transformative force in diverse industries beyond the traditional domains of defense and security, such as automotive, human-machine interaction, and contactless healthcare provision. This significant progress can be attributed to the development of compact and cost-effective radar systems, which are more and more equipped with software-defined capabilities increasing their flexibility. As a result, radar applications have expanded beyond traditional domains, now encompassing crucial areas like medical radar, automotive radar, human-computer interaction, and smart environments. These noteworthy characteristics of modern radar devices, including their accessibility, reliability, portability, and affordability, make them highly suitable for seamless integration within living environments such as residences, educational institutions, and healthcare facilities, amongst others [1–4].

At its core, radar operates by emitting electromagnetic waves and analyzing the return signals to estimate the position of targets in the scene of interest. By measuring frequency shifts in consecutive radar waveforms caused by the Doppler effect, which occurs due to targets' motion, radar enables accurate estimation of targets' velocity. Direct Doppler & velocity estimation measurements play a crucial role in radar systems and have extensive applications in various domains [5]. Notably for this thesis, Doppler and micro-Doppler (μD) measurements with radar systems are essential in the context of human activity recognition (HAR), whereby μD defines the time-varying frequency shift over slow-time that can be extracted from the complex radar return signal by typically applying the short-time Fourier transform (STFT) [6–8].

The broad scientific literature has proposed a wide variety of approaches and sensing technologies for HAR. Nevertheless, based on the author's current knowledge, none of these approaches have succeeded in achieving optimal performance while simultaneously addressing crucial challenges such as standardization, universal adoption, and system robustness, resulting in the absence of a clear frontrunner. For all intents and purposes, the selection of a suitable sensing system for HAR depends on several factors, including budget constraints, user preferences, and the user's living situation. In essence, existing HAR approaches can be broadly categorized as wearable or contactless (non-wearable) devices, where the latter can be further categorized into ambient sensors and camera systems [9, 10].

Wearable solutions, such as those based on accelerometers, gyroscopic and magnetometer sensors, emergency buttons, or smart clothing, potentially involving smartphones, offer a certain ease in deployment. Positioned on anatomical sites like the chest, waist, armpit, or back, or even carried residing inconspicuously within pockets, their primary advantage lies in the minimization of interference from extraneous sources such as other individuals or pets. Yet, the limitations are apparent and in their intrusive nature, potential for physical damage, and the imperativeness of consistent usage or carriage render them less than ideal. Specifically, for populations grappling with cognitive impairments, the utility of emergency-button-based systems remains questionable given the potential oversight in their deployment during exigent situations [11–13].

Then, ambient sensors, such as pressure sensors and acoustic sensors, have established themselves as reasonable alternatives to the previously mentioned solutions. However,

pressure sensors cannot easily distinguish between pressure resulting from a person's weight and other factors, such as pressure caused by domestic animals. On the other hand, acoustic sensors are sensitive to loud or environmental noise. It is also important to note that the effectiveness of these sensors relies on the individual's consistent and proper use [14].

Remote/contactless sensors, given their non-intrusive nature, are often the choice for individuals who prefer an independent lifestyle even when aging. Their main advantage is that they do not depend on the user's direct interaction or compliance, thus facilitating uninterrupted daily routines. When faced with interference from multiple sources or subjects, strategically placing several sensors, each with its unique perspective, can effectively address this challenge. Among these, cameras stand out, offering high-resolution visuals, but not without their set of complications. They often confront problems like excessive details in the scene, occlusions, and variable light conditions. Not to mention, the inherent risk of breaching privacy is attached to camera-based solutions. Although some contemporary models use human body outlines for preserving privacy during motion detection, it is still a concern for many. Similarly, other sensor variants, whether they operate based on vibration detection or sound recognition, come with their specific sets of challenges. For instance, infrared sensors demand precise temperature configurations to deliver reliable results [15, 16].

In the sphere of contactless monitoring, radar systems, especially those that operate on microwave or millimeter-wave frequencies, present a distinctive proposition. Their defining feature revolves around their capability to detect subtle motions, respiratory patterns, and sometimes even heartbeats. What sets radar systems apart is their ability to detect through some obstructions, ensuring the detection of presence or motion even if hidden [17]. Unlike cameras, which may inadvertently infringe on privacy by capturing distinguishable images, radars sidestep this issue altogether. They are resilient to light variations, providing consistent monitoring whether in darkness or well-lit conditions. However, they are not devoid of challenges. Differentiating between closely placed objects or dealing with electronic disruptions, including radio interference by other radar or communication systems, often requires detailed calibration and adjustments. Yet, their diverse use-cases, from detecting falls to comprehensive health assessments, highlight their transformative potential in the realm of indoor monitoring [18–20].

Radar technology offers unique advantages that can complement other prior mentioned sensor modalities as a sensor network or operate on its own. For example, radar, being a non-contact device, remains unaffected by external lighting conditions, including darkness. Moreover, depending on the specific radar system, such as its operating frequency, resolution in ranges, and Doppler capabilities, it can even penetrate opaque objects like tables or walls. Furthermore, radar-based solutions ensure complete privacy protection for individuals under monitoring, as the backscattering signals of radar can detect human motion irrespective of clothing, making them particularly suitable for sensitive environments like hospitals, assisted living facilities, restrooms, and bedrooms, where the use of video cameras might cause discomfort [21]. Deploying radar-based in-home systems for HAR is an integral aspect of the ”aging-in-place” paradigm, offering enhanced security and safety for residents and their family members. Although wearable devices capable of identifying movements, especially falls, do exist, they are reliant on

battery operation and user compliance, whether they are push-button devices or based on accelerometers. In contrast, radar provides a non-obstructive motion sensor technology that can promptly alert caregivers and first responders to critical events related to the health and well-being of the individual under observation. Moreover, radar-based sensing for smart environments and gesture recognition for device control has the potential to revolutionize our way of life and improve global quality of life by transforming interactions with our surroundings. While radar's applications in civilian domains started approximately two decades ago with life-sign detection, ongoing research encompasses a wide range of applications, including monitoring of daily activities within smart environments, home security, assisted living, remote health monitoring, and human-computer interfaces. Notably, non-contact measurements of heart rate and respiration, including the detection of related conditions such as sleep apnea or sudden infant death syndrome, as well as fall detection, have been the primary focus in assisted living and remote health applications [22, 23].

The development of radar-based remote health monitoring technologies can have direct benefits for the aging population, which is projected to reach one billion worldwide by 2030. This can help mitigate the risks of critical events such as falls, but also provide a more general technology framework to monitor the well-being of vulnerable individuals. More in detail, according to the World Health Organization (WHO), falls are defined as "events that result in a person inadvertently coming to rest on the ground or at a lower level." Falls frequently lead to fatal injuries. In 2012, a report from the Centers for Disease Control and Prevention (CDC) revealed that 24,190 U.S. (United States) adults aged 65 and older died due to falls [24]. Similarly, concerning incidence rates are reported in Western Europe, with 13,840 falls per 100,000 people among those aged 70 years and older. The Netherlands has an average rate of 13,623 cases per 100,000 people and 145.5 deaths per 100,000 people. Notably, the Netherlands has a concerning case fatality rate $\left(\frac{\text{Incidence rate}}{\text{Death rate}}\right)$ of 0.011, ranking third after Switzerland and Norway in terms of higher case fatality rates [25]. The National Vital Statistics System also highlights the financial impact of falls. In 2015, the direct medical costs amounted to $637.5 million for fatal falls and $31.3 billion for non-fatal fall-related injuries among U.S. adults aged 65 or older [26]. Falls significantly contribute to morbidity and disability among the aging population. Another shocking statistic indicates that more than one-third of individuals aged 65 years or older experience falls annually, with half of these cases involving recurrent falls [27].

It is evident that prompt interventions following a fall can play a pivotal role in reducing subsequent complications and improving the autonomy and well-being of the elderly. This understanding has led to the ascent of fall detection as a crucial facet of health monitoring research. This rise is in tandem with advancements in sensing technologies tailored for telemedicine and the design of intelligent living spaces [11].

Recreating genuine fall scenarios for research and data acquisition presents its own set of challenges, as induced falls might not truly mirror real-life incidents, with fall events being rare and much less frequent than other activities. Recognizing this gap, a comprehensive study by Rantz et al. [28] was undertaken. The researchers incorporated stunt actors who were trained to mimic realistic fall patterns observed among seniors. They followed two meticulously crafted protocols: the first represented 21 diverse fall types, and the second portrayed 14 motions that could be potentially misinterpreted as

falls. Devised with clinical insights, these protocols aimed to encapsulate the myriad ways seniors might experience falls, from simple missteps to intricate scenarios such as unintentional descents from furniture. It is noteworthy to mention that these demonstrations spanned beyond the confines of clinical settings, extending to the natural living spaces of the elderly. Such an approach ensured the system's aptitude in accurately identifying authentic falls, fostering timely alerts for medical personnel.

The need to streamline healthcare expenses, coupled with aspirations to extend medical provisions to underserved rural territories, has kindled interest in the adoption of radar-driven gait analyzes. Such analyses find utility in diverse medical realms, from predicting fall susceptibilities, gauging recuperation post-injuries, tracking neuromuscular ailment trajectories, appraising therapeutic outcomes, conducting post-stroke evaluations, to deciphering anomalies in gait for therapeutic and rehabilitative intents [29].

Furthermore, broader applications of HAR are being explored in the context of security and energy-efficient smart home systems. For example, occupancy sensing, or alternatively, vacancy sensing, can be employed to intelligently control home systems, such as lighting or heating, ventilation, and air conditioning (HVAC) units [30]. In addition to conventional motion detectors that are triggered by fine-scale motion, such as turning the pages of a book, radar Doppler sensors can detect and classify even larger-scale activities and vital signs. The unique μD signature patterns exhibited by individuals can also be exploited to identify specific individuals [31].

Building upon these foundational applications, radar sensing in HAR extends beyond basic home automation, delving into advanced realms of data classification and monitoring models. These applications of radar sensing in the context of HAR require some form of classification algorithm to assign the measured/processed radar signature to a class of interest. Initially, such algorithms were typically chosen amongst the many options proposed within the supervised learning framework, such as nearest neighbor (NN), support vector machine (SVM), and classification trees, among others [32]. In recent years, there has been significant advancement in deep neural networks (DNNs), particularly in their applicability to state-of-the-art radar data domains. Specifically, these advancements have enabled the exploitation of the μD signature to enhance classification performance [15]. They allow precise tasks such as counting individuals in a room, recognizing individuals within groups, and even enabling vital sign monitoring at the same time [33], and identifying intruders based on known signature profiles [34]. Notably, the μD signature also captures distinctive characteristics when an individual carries a potential weapon, such as a rifle. This may allow for active shooter recognition and differentiation between armed and unarmed personnel, amplifying security capabilities [35, 36].

In summary, this initial section highlighted the importance of radar systems and their applications, particularly in the field of HAR. It emphasized the significant number of falling events and their dire consequences among individuals aged 65 years and above in the U.S., as well as among those aged 70 years and above in Europe. These statistics served as a compelling reminder of the potential impact that modern contactless sensing technology can have on enhancing the independent living lifestyle of the aging and most vulnerable segments of the population.

1

1.2. The Human Ethogram

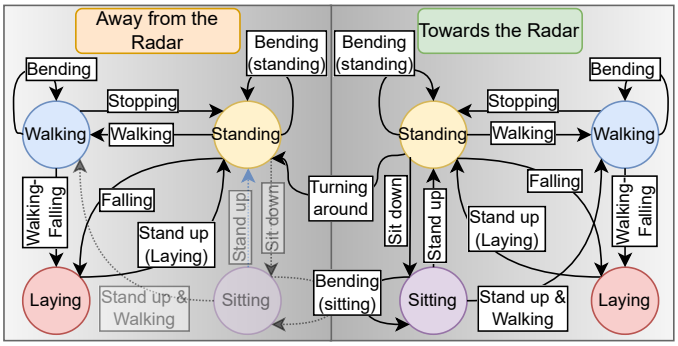


Figure 1.1: The *Human Ethogram* illustrates continuous two-way sequence recognition, detailing how sequential activities correspond to *states* like *Standing*, *Walking*, *Sitting*, and *Laying*. Certain pathways are highlighted as mutually exclusive. For instance, from a *Laying state*, one cannot transition into a fall, and consecutively sitting down does not establish a *Sitting state*. On the other hand, continuous bending leads to the *Standing state*. It is important to note that the *Sitting state* away from the radar is excluded in this study due to concerns about radar penetration through the chair's back [37].

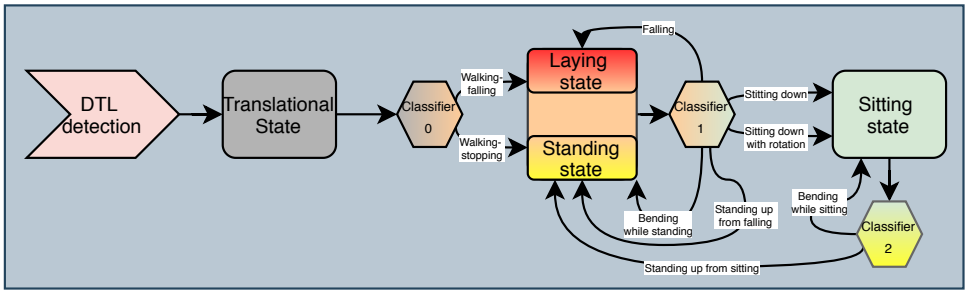
This section introduces the *Human Ethogram*, also known as the human motion state representation, which represents a Markov chain of possible movements based on activity sequences for human activity recognition (HAR). Human activities can be typically categorized into states, namely, *Walking*, *Standing*, *Sitting*, and *Laying*, amongst others, where the latter state also corresponds to the status after falling. An example of a state diagram illustrating these activities can be found in Fig. 1.1, which focuses on two-way gross-motor activities, where the individual is either facing the radar or turned away from it [38, 39].

The ethogram can be used to model transitions between different human activities in a sequence. For instance, a person can transition from walking to falling and subsequently stand up after a fall. However, it is not possible to stand up while already walking and being in a standing posture. Similar situations apply to in-place activities. For example, a person can transition from a sitting position to a standing position by performing a standing up action or vice versa from a standing position to a sitting position by a sitting down action. However in this regard, sitting down can only occur from a standing posture, while other postures are mutually exclusive.

These concepts found novel applications where possible activities are limited according to a-priori observed information, beneficial for classifying activity sequences. Furthermore, the aforementioned *Human Ethogram* enables the system to limit the options of possible classes for classification, which can boost its perception performance. An example can be found in Fig. 1.2 and more insights in [40].

The *Human Ethogram*, a Markov chain representation of a person's activity pattern, is deemed crucial for comprehensive monitoring. These patterns serve as measurements for logical sequences of activities outlined in the *Human Ethogram*, enabling the determination of (1) *States* and (2) *Actions*. *States* represent longer-lasting conditions, such as the *Standing state*, while *Actions* are transitions to another state or the same state, such as *standing up from*

(a) Example of a continuous sequence of human activities performed by a participant



(b) Flowgraph with restricted classifiers based on the Human Ethogram

Figure 1.2: (a) shows a continuous sequence of critical and non-critical activities: walking, falling, standing up from falling, sitting down, standing up from sitting, and walking again. (b) demonstrates a restricted classification process using multiple classifiers that can only be applied after specific states. The system includes a tracker to determine the person's *Translation state* (*Walking state*), with more insights provided in [40].

sitting or *bending from standing*, respectively. For example, the action of *standing up from sitting* entails a transition from a *Sitting state* to a *Standing state*. Conversely, the action of *bending from standing* involves starting and ending in a *Standing state*. Furthermore, statistical analysis of human behavior over a defined time span, including the duration of individual activities, centroid frequency (velocity), and tracked ranges, plays a significant role in evaluating and monitoring an individual's well-being. All these quantitative aspects can indeed contribute to forming a picture of the level of activity, fitness, and engagement that a person has with their own environment.

This novel technique has been utilized in [40] using a tracker named the *Derivative Target Line (DTL)*, triggering a state reset when detecting an acceleration of an individual. This reset leads to the determination of the person's *Translational state*. Subsequently, the following classifier comprises two classes: *walking stopping* and *walking falling*, as depicted in the flowchart in Fig. 1.2b, with a potential sequence example of a person *walking – falling – laying – standing up – sitting down – standing up – walking* shown in Fig. 1.2a.

Summarizing this section, the *Human Ethogram* introduces a valuable action-state representation for HAR inspired by the Markov chain. It categorizes activities into *states*, such as *Walking*, *Standing*, *Sitting*, and *Laying*. The action-state diagram illustrates possible transitions, highlighting mutually exclusive actions. The generic examples in Figs. 1.1 and 1.2 focus on two-way gross-motor activities with the individual facing or turned away from the radar. These concepts find applications in classifying activity sequences and improving perception performance. Monitoring a person's activity pattern is crucial, providing insights into states, actions, and well-being. Statistical analysis of behavior, including duration, velocity, and tracked ranges, plays a significant role. The Derivative Target Line (DTL) is an example of a tracker that triggers a state reset based on acceleration, determining the *Translational state*, and starts the classification procedure always using a binary classifier.

1.3. Challenges, Open Problems, and Approaches

This section introduces the challenges and open problems addressed in the conducted research, along with a brief overview of possible alternative methods and the chosen approaches discussed in subsequent chapters.

Advanced Radar Data Domains: The appropriate representation of radar data to effectively infer healthcare-related insights is paramount. In the state-of-the-art literature, studies predominantly rely on the use of micro-Doppler (μD) representations derived from the Short-Time Fourier Transform (STFT) [8, 41]. This approach enables the analysis of temporal Doppler patterns caused by different body parts in motion. By leveraging the direct relationship between Doppler modulations and body part velocities, algorithms operating on radar data can effectively characterize human movements, ranging from subtle vital sign-related motions to distinct gait patterns or complex activities. These μD patterns are indispensable inputs for classification pipelines, which can leverage machine learning techniques, including various architectures of neural networks, to extract and learn the relevant information. Alternatively, research conducted by different groups such as Li et al. [42], Liu et al. [43], and Zhang [44] have ventured into time-frequency methods apart from μD spectrograms generated by the STFT. These methods introduce the Gabor dictionary, before time-frequency features were extracted via the orthogonal matching pursuit (OMP) algorithm, suitable for sparse radar signal representations, then the mel-frequency cepstral coefficients (MFCC) for Doppler radar-based automatic fall detection was utilized, followed by the use of a pseudo-smooth Wigner-Ville distribution (WVD) based on the energy distribution for human gait analysis, respectively. Others avoid radar data domain investigations and rely on the capabilities of deep convolutional neural networks (DCNN) or other Deep Learning methods to extract information, requiring less emphasis on radar data domain exploitation [45]. However, alternative radar data domains beyond the μD signature are rarely explored in a systematic manner, so that non-time-frequency-based domains, and their distinctive characteristics remain unexplored [3, 9, 29]. Unlike the mentioned time-frequency domains, which necessitate frequency analyses, which often use a form of Fourier transform, this work presented in Chapter 3 deliberately avoids such time-frequency analysis altogether. Instead, it directly analyzes the complex phase of a range-time map, for which various challenges and opportunities were investigated.

Data Fusion for Distributed Radar Networks: The exploitation of radar networks in indoor scenarios remains a sparsely explored area in the current literature [46]. Challenges such as interference, synchronization issues, and cost considerations have limited their adoption [47]. Similarly to observations made by previous studies, employing a radar network introduces variations in the signal to noise ratio (SNR) conditions across the radar nodes due to factors such as target distance and radar cross section (RCS). However, deploying a distributed network of radar nodes offers unique advantages, including the ability to observe targets simultaneously from different aspect angles, and the capability to detect hidden targets or body parts obstructed by objects or multiple individuals within the room. By strategically designing the network layout, occlusions can be minimized or completely avoided, ensuring comprehensive coverage. Additionally, the inherent strength of radar lies in its ability to estimate the relative velocity from the Doppler shift, which can vary for different aspect/observation angles and, consequently, across radar nodes. This presents a specific challenge when dealing with extended targets, as observed in this work, where different body parts exhibit distinct scattering points with their own Doppler values. Addressing this challenge, the thesis focuses on developing advanced and robust fusion methods for distributed radar sensors that have not been adequately explored in the current literature on human activity recognition (HAR). Moreover, the fusion process extends beyond node-level fusion and delves into the investigation of fusing radar data across different domains, including signal level, feature level, and decision level. This research addresses this multi-level fusion problem throughout, with specific reference to Chapter 4.

Advanced Evaluation Metrics for Continuous HAR: The utilization of distributed radar networks in conjunction with continuously recorded data poses significant challenges. In particular, the analysis of continuous radar data for HAR that encompasses sequences of mixed activities remains an open problem in current research. Existing radar-based HAR approaches often focus on analyzing isolated motions and actions, treating them as windowed "snapshots", disregarding the continuous nature of human activities [3, 48, 49]. In reality, human activities unfold as seamless sequences with undefined transitions and duration between different motions. These sequences exhibit a diverse range of activities, including full-body actions such as walking, stationary movements involving limb motions (e.g., sitting, bending to pick an object from the floor), and even life-threatening scenarios such as falling. Also, in an evaluation process of such continuous sequences, conventional metrics, such as simple accuracy, account very little for such crucial interruptions, when, i.e., a sequence of walking is interrupted by any random prediction outliers. Again, the characteristics of such sequences vary significantly depending on factors such as the gender, age, and physical condition of the subject, and environmental constraints imposed by surrounding furniture and objects. Such complexity introduces additional challenges in capturing the salient information relevant to the performed activities. This was investigated briefly by Chen et al. [50], for recorded radar sequences, whereas with a focus on ensemble learning algorithms, a different aim than presented in this work. Consequently, treating radar data for HAR as isolated snapshots fails to reflect real-world scenarios and hinders progress in areas such as utilizing distributed radar networks for continuous data stream monitoring. Furthermore, the analysis of such data streams is well suited for classifiers with temporal capabilities, commonly referred to as recurrent neural networks (RNNs), with architectures like long

short-term memory (LSTM) being widely adopted [51, 52]. However, evaluating the performance of these classification algorithms with temporal capabilities necessitates the use of novel, proposed evaluation metrics that effectively capture temporal relations. Therefore, the evaluation problem of continuous radar sequences is addressed in the subsequent Chapter 4.

Exploiting Multipath for HAR: Multipath and clutter are well-established phenomena in radar systems. Multipath occurs when radar signals encounter reflections from multiple pathways beyond the radar-target line of sight, resulting in potential signal distortion and ambiguity. Conversely, clutter refers to undesirable radar echoes caused by stationary objects in indoor scenarios or environmental factors, significantly interfering with target detection and tracking processes. To mitigate the effects of multipath in radar systems, sophisticated signal processing techniques, such as adaptive beamforming, are commonly employed to suppress unwanted reflections and enhance the desired target signal [53, 54]. Additionally, the use of multiple antennas or antenna arrays with appropriate spacing and diversity techniques can aid in mitigating the adverse impacts of multipath propagation. However, the exploration of the potential benefits of multipath in radar remains limited, with most research focusing primarily on point-targets [55]. As such, the aim is to exploit multipath effects for extended targets. By harnessing the coherent signals reflected from various aspect angles thanks to multipath, a novel approach was formulated to generate artificial radar nodes that facilitate the observation of targets from diverse perspectives [56]. Furthermore, compelling evidence has been presented, demonstrating that even a single radar node, using only a single-input single-output (SISO) system when coupled with strategically positioned reflectors, can effectively capture an extended target from multiple aspect angles, resulting in significantly improved perception and classification performances, as presented in Chapter 5.

Multi-People Monitoring: Multi-people perception tasks using radar entail the challenging objective of accurately identifying and differentiating multiple individuals within the radar sensing field, relying on their distinctive radar signatures. State-of-the-art research has successfully shown vitals in combination with spatial tracking of people groups [33], and others have applied group classification to distinguish between walking and running groups [57]. However, none of the abovementioned work used a comprehensive sensor network with up to five radar nodes. In fact, this complex task necessitates the utilization of robust signal processing techniques, advanced machine learning algorithms, and effective feature extraction methods to overcome the inherent intricacies arising from overlapping radar echoes and the diverse nature of human movements and orientations. However, to the best of the author's knowledge, the literature has yet to explore the concept of treating this problem as a regression task, which offers the advantage of estimating approximate occurrences of human activities. Consequently, the formulated regression problem presents the potential benefits of providing continuous estimates, as opposed to discrete predictions offered by classifiers. Such an approach proves particularly valuable in scenarios where precise and discrete predictions are unnecessary, such as obtaining a rough estimation of individuals walking within a crowded environment [58]. The problem has been addressed by employing both conventional and temporal regression approaches, with scenes involving up to five individuals walking or standing stationary in an indoor environment, as addressed in Chapter 6.

1.4. Outline of the Thesis and Chapter Abstract

The research was organized into distinct work packages, each addressing specific challenges. Each of these work packages corresponds to a dedicated chapter in this thesis.

Chapter 2: Fundamentals of Distributed Radar Networks

This chapter serves as an introduction to the utilized pulsed ultra-wideband (UWB) radar system, which forms a spatially distributed system of monostatic operating radar nodes. A radar signal model is presented, followed by the description of the human locomotion model. Additionally, it introduces common state-of-the-art data fusion concepts, which play a pivotal role in integrating data from multiple radar nodes and diverse radar data domains. These fusion concepts find applications in the thesis from Chapter 3 onward.

Chapter 3: Radar Data Domains for Indoor Monitoring and Classification

The chapter describes research focused on human movement recognition, specifically targeting human activity recognition (HAR) and gesture detection. A novel method, which utilizes radar phase information directly extracted from a high-resolution range map (RM) is proposed, eliminating the need for traditional μD spectrograms. This alternative approach overcomes the limitations of time-frequency trade-offs and computationally intensive processing, offering comparable or superior classification results. The histogram of oriented gradients (HOG) algorithm is employed to extract distinctive shape and pattern features from the wrapped phase domains. The effectiveness of the method is validated on independent datasets, achieving a consistent classification accuracy exceeding 92 %. Additionally, feature fusion techniques that combine different data domains, such as the modulus of RM with the RM phase information, enhance robustness and accuracy in activity recognition tasks. These findings highlight the potential advantages of the proposed approach for low-latency, real-time applications, and computationally constrained scenarios.

Chapter 4: Data Fusion in Distributed Radar Systems for HAR

This chapter describes the application of a network of five spatially distributed pulsed UWB radars for continuous HAR. By employing multiple sensors, the network addresses challenges such as occluded perspectives and unfavorable aspect angles. Two fusion techniques are investigated: signal level fusion on range-time maps (RT) and radar node selection based on target location and velocity derived from multilateration processing and tracking. Recurrent neural networks (RNN) with and without bidirectionality are employed for activity classification using μD spectrograms obtained through sensor fusion. Novel evaluation metrics are proposed to account for the continuous nature of activity sequences and dataset imbalances, providing a comprehensive assessment beyond conventional accuracy metrics. The findings highlight the importance of considering these metrics for a more thorough evaluation of continuous HAR.

Chapter 5: Multipath Effect Exploited for Human Activity Recognition

This chapter addresses radar sensor multipath problems in indoor radar networks, which are often seen as unwanted phenomena to mitigate. Existing research has mainly focused on single radars and non-extended targets, neglecting the potential benefits of utilizing multipath components to enhance classification capabilities. In this chapter, a pipeline is proposed that isolates a target's line of sight (LOS), determines its position, and analyzes higher-order multipath components. The pipeline also accounts for edge cases, such as extended targets observed across multiple range bins and various aspect angles, even when using only one radar. By exploiting the advantages of higher-order multipath components along with LOS observations, the network's perception is improved. The pipeline is validated by comparing classification results to those obtained with a single radar and the radar network's LOS only. For perception tasks, the multipath components and LOS components are fused and fed into a 12-layer Convolutional Neural Network (CNN). The results are compared to a simple multi-layer perceptron (MLP) classifier that uses features extracted using Principal Component Analysis (PCA). The use of a multi-radar network with LOS & multipath demonstrates a significant improvement in test performance of approximately +11 % for continuous and consecutive recorded human activities.

Chapter 6: Multi People Monitoring using Regression

This chapter focuses on radar-based HAR in crowded environments, specifically addressing the challenge of continuous activity recognition involving multiple individuals moving in arbitrary directions within an indoor area. To tackle this problem, a regression-based approach is proposed, aiming to create robust and accurate systems for monitoring human activities. Novel techniques using LSTM or CNN regression models, along with linear regression (LR) and SVM regressors, are compared. These approaches leverage extracted features from radar data using methods such as HOG and PCA. The evaluation employs a rigorous Leave-One-Group-Out method, with performance assessed using common regression metrics such as root mean square error (RMSE). Promising results are observed for crowds of three and five individuals, with RMSE values around 0.4 and 0.6, respectively. These outcomes are achieved primarily by using the μD spectrogram or range-Doppler data domain.

Chapter 7: Conclusions and Further Work

In the concluding chapter, the key results of the research presented in this thesis are summarized, offering a comprehensive reflection on the achieved outcomes. The achieved results and significant contributions made throughout the thesis are highlighted, showcasing the advancements and added value brought to the field of study. Additionally, the author identifies untapped potential and areas for improvement, presenting intriguing opportunities for future research and the prospect of novel contributions.

2

Fundamentals of Distributed Radar Networks

This chapter provides a discussion of the technical intricacies surrounding distributed ultra-wideband (UWB) radar systems employed in this thesis, thereby setting the stage for the subsequent chapters. Additionally, it introduces the human locomotion model before it offers a detailed overview of the data fusion principles in the context of distributed radar networks. These fusion techniques, applied in a human activity recognition (HAR) context, span from late fusion techniques (decision-fusion) to the early fusion concept known as signal-level fusion, which involves the collaborative utilization of data from all radar nodes.

Parts of this chapter have been published in:

R. G. Guendel, M. Unterhorst, E. Gambi, F. Fioranelli, and A. Yarovoy, "Continuous human activity recognition for arbitrary directions with distributed radars", in *IEEE Radar Conference (RadarConf21)*, 2021.
I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli and A. Yarovoy, "A Survey on Radar-Based Continuous Human Activity Recognition", in *IEEE Journal of Microwaves*, 2023.
R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Exploiting Radar Data Domains for Classification with Spatially Distributed Nodes", in *SET-312 Research Specialist Meeting Distributed Multi-Spectral/Statics Sensing, Bled, Slovenia.*, 2022.

2.1. Applications: Pulsed Ultra-Wideband (UWB) vs. mm-Wave Radars in Human Activity Recognition

2

Radar systems utilized for human activity recognition (HAR) have undergone marked advancements due to the incorporation of new technologies. Pulsed ultra-wideband (UWB) radars and mm-wave radars stand out as primary candidates in this arena. Notably, mm-wave radars operating in the V-band (60 GHz) and W-band (77 GHz), associated with indoor monitoring and the growing automotive sector, respectively, have seen an enhancement in market value, a decrease in hardware costs, and consequently, rising interest from researchers. This section presents a comparative analysis of these radar modalities and applications, highlighting their significance for HAR and related sectors.

Pulsed Ultra-Wideband (UWB) Radars: Pulsed UWB radars are mainly distinguished by their extensive bandwidth, which offers high-resolution capabilities. This feature enables the precise detection of human movements, making these radars indispensable in HAR scenarios where subtle motion differentiation is vital [59]. The adaptability of UWB radars, especially when combined with software-defined radios (SDRs), makes them a top choice for various HAR tasks and vital sign monitoring applications. In [60], Lazaro et al.'s research group employs Acconeer's pulse coherent radar [61] for seat occupancy detection, leveraging the 60 GHz frequency band. This radar can detect submillimeter movements, allowing for the recognition of human presence by slight body movements. The system differentiates between a human and other objects on a seat by evaluating the standard deviation of peak amplitude variations. Moreover, it can measure breathing rates by examining changes in peak amplitude, using digital filters to eliminate noise. The system's ability to measure breathing rates is not compromised by car vibrations. However, it can be affected by pronounced body movements. Practical uses for this system encompass monitoring driver fatigue, identifying child seat occupancy, and potentially evaluating sleep quality and potential sleep apnea.

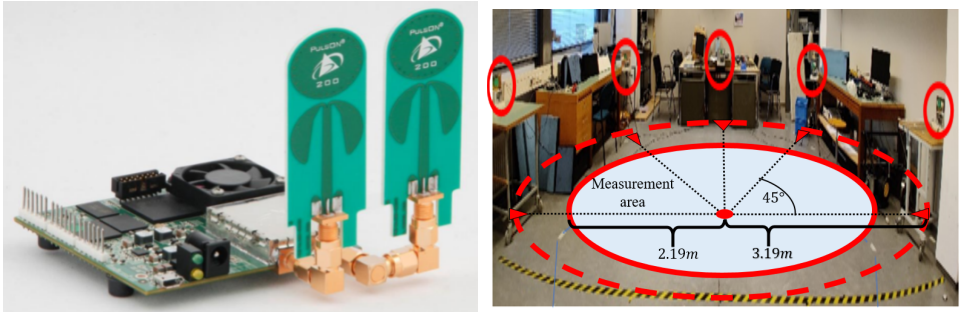
mm-Wave Radars: These radar systems, particularly those operating in the V-band and W-band, present a compelling case for a myriad of applications. These radars, provided by commercial players in the radar sector such as NXP [62], Infineon [63], and Texas Instruments [64], are renowned for their compactness, making them ideal for seamless integration into a diverse range of systems. This includes not only embedded monitoring systems, but also advanced driver assistance systems (ADAS) that require high-resolution, real-time data to function effectively [1, 65]. With the capability to achieve fine range and Doppler resolution, including fine angular resolution (depending on antennae pairings), they can detect subtle human activities, from small hand gestures to intricate body movements [66]. This precision is especially crucial in environments where the differentiation between activities depends on nuanced motion details. When considering the realm of HAR, the potential of these radars becomes even more evident. The 60 GHz radars, for instance, can be particularly effective in indoor monitoring scenarios, capturing detailed point cloud data and facilitating accurate activity classification. On the other hand, 77 GHz radars are often optimized for dedicated automotive use and can be more suited for outdoor or vehicular applications, providing a wider field of view and longer detection range including angular detection capabilities.

Moreover, the adaptability and versatility of these radars, combined with advancements in signal processing and machine learning algorithms, position them as front-runners in the evolution of HAR systems [67]. As research progresses, it is anticipated that the integration of these radars will pave the way for more sophisticated, reliable, and holistic HAR solutions, expanding their application beyond traditional domains.

Both pulsed UWB and mm-wave radars present distinctive advantages for HAR. While UWB radars are preferably selected for their very fine range resolution, adaptability, and comprehensive data representation, mm-wave radars are advantageous for their compactness, sensitivity, and potential for real-time monitoring. As the field of HAR continues to advance, the integration and synergy of these radar technologies are expected to lead to more sophisticated and effective HAR systems.

This section has briefly compared pulsed ultra-wideband (UWB) radars with frequency-modulated continuous-wave (FMCW) V/W-band radars and has shown some HAR and automotive applications. The Section 2.2 introduces the used radar from Humatics (former: Time Domain) of the type PulsON P410, together with the open dataset and the room layout where the experiments were conducted with multiple radar nodes employed. In the following, Section 2.3 shows a generic radar signal model assuming point-like scatterers. Then, in Section 2.4, human locomotion models are introduced with the extension of the generic signal model for point targets to the more comprehensive representations applicable for extended targets. Additionally, suitable radar data domains of the PulsON P410 radar are introduced in Section 2.5, and these domains are used throughout the subsequent chapters. The chapter ends with Section 2.6, a data fusion section introducing fusion modalities for combining data from different radar nodes, and finally, conclusions are provided in Section 2.7.

2.2. Introduction to Pulse Ultra-Wideband Radars



(a) Humatics PulsON P410 UWB radar node.

(b) Laboratory radar layout with 5 nodes.

Figure 2.1: (a) The Humatics (former: Time Domain) PulsON P410 ultra-wideband (UWB) radar with a center frequency of 4.3 GHz and 2.2 GHz bandwidth. (b) TU Delft laboratory of the Microwave Sensing, Signals & Systems Group (MS3) with a radar network setup consisting of five nodes marked by red circles. This setup was used for most of the research performed in this thesis.

This research focuses on continuous activity recognition beyond a single radar application case. To address this, a distributed radar network was introduced to capture

continuous activities in [68], and publicly available data was released [69]. The utilization of this distributed radar network confers various advantages, notably the ability to observe the person from multiple perspectives. Such multi-aspect observations are crucial for capturing a comprehensive representation of a person's kinematic, specifically encompassing the velocity components of different body parts in various spatial directions. Furthermore, the experiments are not restricted in terms of an individual's movement directions, aiming to represent real-world scenarios where people can move and perform activities in any direction and arbitrary trajectory. At the same time, this distributed layout also introduces the challenge of appropriately fusing the network's data. In this thesis, the distributed radar network consists of pulsed UWB radar nodes, as depicted in Fig. 2.1, with a single node presented in Fig. 2.1a and the typical room layout with 5 distributed radars placed in a semicircle shown in Fig. 2.1b. The PulsON P410 radar hardware from Humatics (previously Time Domain) was chosen due to its ability to synchronize the nodes within the network incoherently (or in a loosely coupled manner), according to *Class II* of the *Classification Scheme for Distributed Radar Systems* by Gottinger et al. [70].

2.3. Generic Radar Signal Model

Inspired by the mathematical derivation of Deudon et al. [71] and He [59], a simplified radar signal model for a single point-like scatterer will be derived in this section. Without any limitations on the radar type, it can be assumed that the transmitted signal can be represented as,

$$S_{tx}(t) = \sum_{m=0}^{M-1} p\left(t - mT_r\right) e^{j2\pi f_c t} \tag{2.1}$$

with t denoting the fast-time, $p(t)$ is the complex signal envelope, T_r is the pulse repetition interval (PRI), and f_c is the center frequency of the M pulses in the sequence corresponding to a coherent processing interval (CPI). When the transmitted signal is reflected by a moving point scatterer, a delayed, attenuated, and frequency-modulated signal follows as,

$$S_{rx}(t) = \alpha S_{tx}\left(t - \tau(t)\right) \tag{2.2}$$

where α is the complex amplitude response of the target, and $\tau(t)$ the round trip time delay, assuming a constant radial velocity v of a target during the CPI. The round trip time is computed as,

$$\tau(t) = \tau_0 - \frac{2v}{c}t = \frac{2R_0}{c} - \frac{2v}{c}t = \frac{1}{c}\left(2R_0 - 2vt\right) \tag{2.3}$$

with R_0 and τ_0 the target radial range and the constant round trip delay, respectively. Substituting Eqs. (2.1) and (2.3) into Eq. (2.2) leads to,

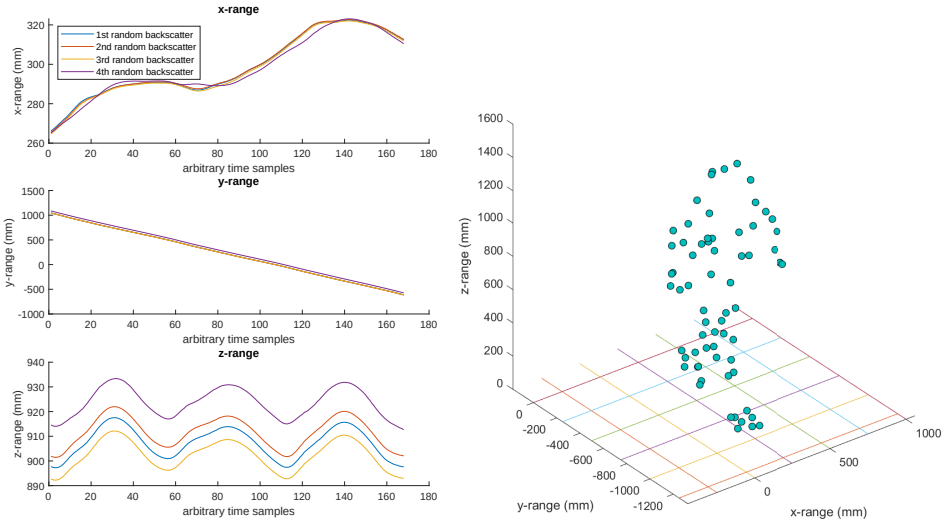
$$S_{rx}(t) = \alpha \sum_{m=0}^{M-1} p\left(t - \tau_0 + \frac{2v}{c}t - mT_r\right) e^{j2\pi f_c\left(t - \tau_0 + \frac{2v}{c}t\right)} \tag{2.4}$$

and with the constant phase terms τ_0 being absorbed into the complex amplitude α' as,

$$S_{rx}(t) = \alpha' \sum_{m=0}^{M-1} p\left(\left(1 + \frac{2v}{c}\right)t - \tau_0 - mT_r\right) e^{j2\pi f_c\left(1 + \frac{2v}{c}\right)t} \tag{2.5}$$

The derived equation shows the radar echo signal in fast-time domain, with more insights provided in the following references [59, 71].

2.4. Human Locomotion Model



(a) x-,y,-z-range change of the human locomotion model. (b) Human locomotion model, showing random radar scatterers.

Figure 2.2: (a) shows the x-,y,-z-range change of a human locomotion model which represents the action of walking continuously in the y-direction with some small variations in the x-direction and a sinusoidal change in the z-direction. (b) shows an example of the position of arbitrary scatter points related to a human body walking in a 3D Cartesian space [72].

This section yields an introduction of literature for further reading on human kinematic models, also known as human locomotion models, with models having their origin in the computer vision society. Furthermore, it is meant to illustrate the total radar scatter points that can be extended as a sum of multiple single-radar scatterers to serve for extended targets, such as a human with target sizes much larger than a range resolution cell, by using ultra-wideband (UWB) radar technology.

Numerous human locomotion models were presented in the literature with fundamental investigations provided by Chen et al. [6], Geisheimer et al. [73], Dorp et al. [74], or Vignaud et al. [75], where some authors underline the advantages of using the more realistic Thalmann model [76]. The latter model was modified by He et al. [77] to introduce a novel range-Doppler surface (RDS), a radar backscatter model designed for extended human targets. In addition to the aforementioned fundamental approaches, Erol et al. [78] proposed a kinect-based human micro-Doppler simulator, an advanced video-based motion capture (MOCAP) model. Both, He and Erol, have used the *Motion Capture Library* of Carnegie Mellon University (CMU) for their studies to reconstruct a human locomotion model [79], and from this the full target backscatter of coherent radar echoes is constructed.

Furthermore, a human motion analysis model from Wang et al. [72] has been modified to provide a visual example, as depicted in Fig. 2.2. In Fig. 2.2a, the range changes as a

function of time for the three spatial coordinate axes (x, y, z) are illustrated for a few randomly selected backscatter points. The entire human model, i.e., including all the assumed radar backscatter points of the human body, is presented in Fig. 2.2b. It is noticeable from these figures that in the y-direction there is a straight line over time, as the individual walks predominantly in that direction. Additionally, there is a minor shift in the x-direction orthogonal to the direction of the walking motion, and of particular note is the sinusoidal oscillation in the z-direction as the person moves forward.

As in Section 2.3, the simple radar model of a point target was introduced. It can be used by considering the extended nature of a human body with multiple backscattered signals from each scatter point, as illustrated in Fig. 2.2b. Thus, the total echo of such human body surfaces can be computed as,

$$S_{total}(t) = \sum_{k=0}^{K-1} S_{rx}(t) \tag{2.6}$$

with $S_{rx}(t)$ from Eq. (2.5) and k, K indicating the backscatter points of a human body, with each providing a different radar cross section (RCS) [80].

Such point cloud concepts are the fundamental basis for applications of representing humans that can be further used for various usages, including perceptron-based recognition from gesture, posture, or finger motions. As an example of applications of this subject, the work of Lee et al. [81] is considered. They introduced a HAR model that leverages point cloud data, specifically addressing the challenge of radar data sparsity. The model utilizes 3D human joint coordinate estimates derived from a pre-trained model using radar data, with a Kinect serving as ground truth. It was determined that 25 joints offer optimal feature representations for human activities, and the extracted 3D human joint coordinates provide reliable features for sparse radar data. The proposed model using spatial-temporal graph convolutional networks (ST-GCN), a graph neural network (GNN), and the 3D joint coordinate estimates, achieves a classification accuracy exceeding 95 %.

2.5. Humatics PulsON P410 Radar and its Data Domains

Pulsed ultra-wideband (UWB) radar sensors from Humatics (former: Time Domain) of the type PulsON P410, as shown in Fig. 2.1a, were utilized in the thesis to form a distributed radar network. They were arranged in a semicircular baseline with a 45° separation, positioned 1 meter above the ground. The network consisted of five simultaneously operating monostatic nodes. The experiments were conducted in a circular space with a diameter of 4.38 m, as depicted in Fig. 2.1b, with the nodes highlighted in red [68]. The aforementioned open-source dataset [69] collected for this research comprises data from 15 subjects, with training and test data sequences split and activities performed in different orders compared to the training set.

The PulsON P410 radar can be used for monitoring human activities performed in realistic continuous streams. As a visual introduction, Fig. 2.3 showcases the common data domains utilized throughout the thesis, applied to a single-input single-output (SISO) radar. These domains include the (II,IV) range-time (RT), (III) range-Doppler (RD), and

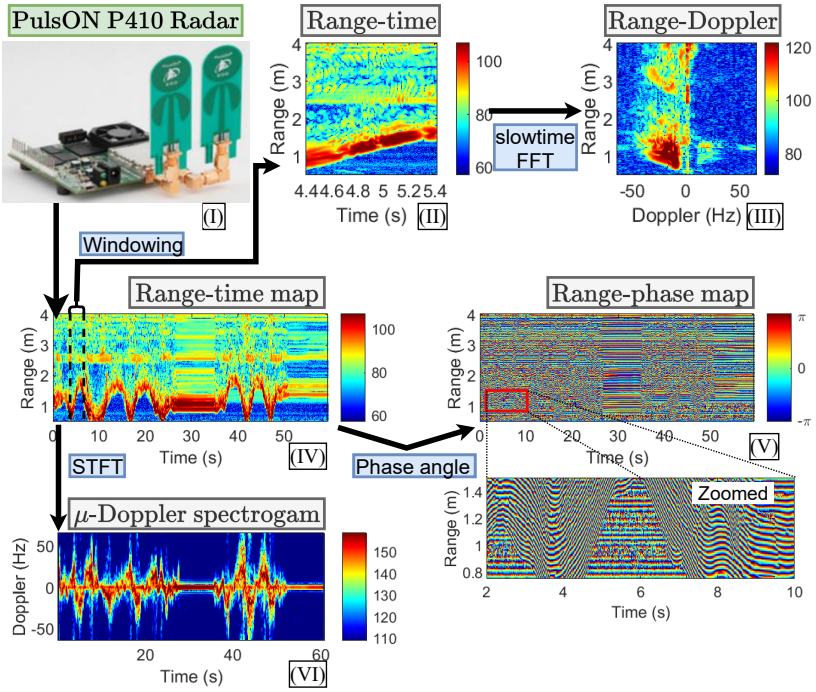
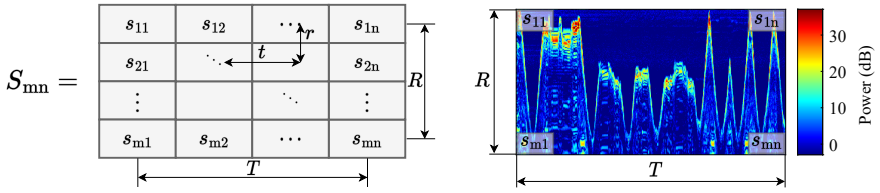


Figure 2.3: The flowgraph shows HAR radar data domains extracted from a (I) PulsON P410 radar node with, (II) the range-time obtained by windowing over (IV) the full range-time map, (III) the range-Doppler map by computing the FFT over slow-time of (II), (V) the observed range-phase map with a magnified view, and (VI) the micro-Doppler (μD) spectrogram obtained using the STFT.

the (VI) micro-Doppler (μD) spectrogram. Furthermore, Fig. 2.3.V illustrates the less common but potentially interesting phase information, a domain extracted directly from the RT map. This, in turn, necessitates feature extraction methods capable of capturing information from the fine line segments in such data, as depicted in the *zoomed* plot. Therefore, investigations were conducted using the histogram of oriented gradients (HOG) on phase data, which is a method capable of extracting features from the aforementioned contours in images. A comprehensive discussion of the data domain, specifically regarding the phase information, can be found in [82] and Chapter 3.

The following paragraphs discuss the radar's return signal of the Humatics (formerly Time Domain) P410, following up from the simple radar model, introduced in Section 2.3. This system is the primary radar used in this work and provides the range-time (RT) map immediately, denoted as S_{mn}, as depicted in Fig. 2.4a. A recorded RT map of a person performing a 2 min sequence of in-place and translation activities is shown in Fig. 2.4b with the same notation applied as in Fig. 2.4a. The matrix S_{mn} is the discretized and reshaped version from the radar model Eqs. (2.5) and (2.6) for single or extended targets, respectively, with the columns being the fast-time bins observed within the CPI. The matrix S_{mn} is provided by the radar system after the following internal processing steps of an optimal linear filter, coherent pulse integration for enhancing the SNR, time-interleaved



(a) Matrix representation.

(b) Range-time (RT) sequence.

Figure 2.4: The matrix representation for the received radar signal from the Humatics P410 radar is shown in (a), and (b) provides an example of one of these matrices of data depicted as an image. Specifically, this represents a 2min range-time (RT) sequence of a real recorded participant performing in-place and translation activities.

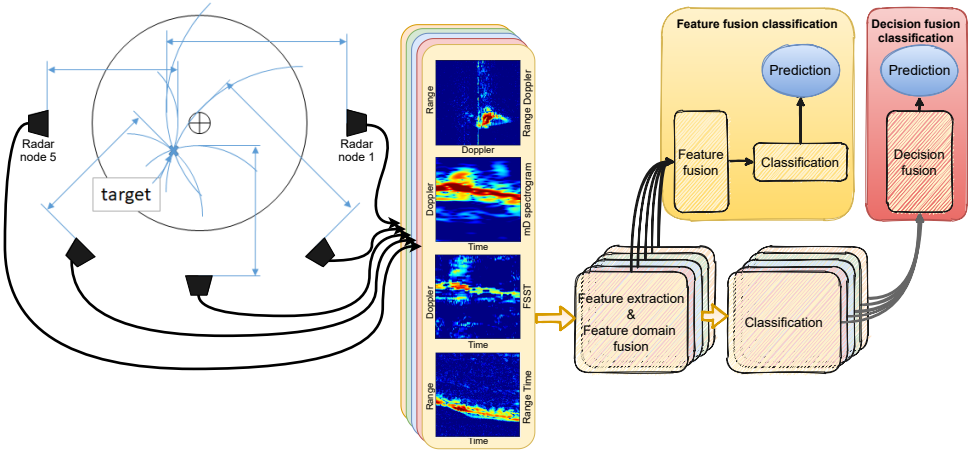
fast-time sampling, and RT map reshaping, whereas the user has limited freedom to engage in the preprocessing steps apart from modifying the desired number of coherent pulse integrations. Furthermore, s_{mn} represents the real samples (in-phase components) for each fast-time bin and slow-time bin with indices m and n, respectively. The backscattered signal is obtained through filter banks, and the quadrature component is generated using the Hilbert transform [82].

The time-of-flight separation, τ, between consecutive range bins in the m-direction is 61.024 ps [83]. The corresponding sampling resolution, r, is calculated as $r=\frac{\tau \cdot c}{2}$, resulting in a resolution of 9.153 mm. With this setting, the radar's coverage area, R, is 4.39 m with 480 received range bins, given by $m=1,\cdots,M$ for the open data provided [69]. The slow-time samples are denoted as n and spanning over the total time T [84]. To enhance the moving signature of human targets, cleaning steps such as MTI filtering, mean subtraction, and adaptive thresholding can be applied. The pulse repetition frequency (PRF), denoted as f_{PRF}, is typically set to 122 Hz, resulting in a pulse repetition interval (PRI) of 8.2 ms. This choice ensures unambiguous Doppler frequency results within ± 61Hz (equivalent to ± 2.2 m/s). The theoretical range resolution (r_{res}) achieved using a bandwidth of $B=2.2$GHz is 68 mm, as determined by $r_{res}=\frac{c}{2 \cdot B}$. For the chosen setting, the theoretical radar detection ranges up to 354 m [5, 84]. However, the data sheet [85] *Note 5 (Detection ranges)* mentioned that the theoretical detection range depends on the antennae used, the ambient environment, clutter, target size, and movement characteristics.

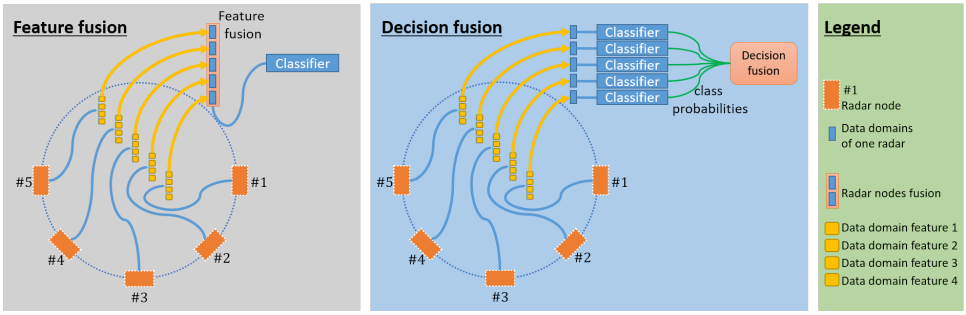
2.6. Introduction to Data Fusion and its Constraints

This section provides an overview of radar fusion concepts employed in this thesis, where different fusion techniques are investigated in conjunction with a distributed radar network. Radar data fusion can be categorized into three approaches: signal fusion (or pixel-level fusion), feature fusion, and decision fusion. Signal fusion occurs early in the process, before any feature extraction. Feature fusion involves concatenating feature samples from all nodes into a single feature vector, allowing fusion across data domains and sensor nodes. Decision fusion entails classifying for each node and combining the predictions, often through methods such as majority voting [86].

Drawing inspiration from the general concepts of information fusion in the context of a classification task, the author's work in [69, 87] proposes the following fusion approaches



(a) Scheme for multi-domain classification with four data domains extracted from each radar.



(b) Classifier block placement for *feature fusion* and *decision fusion*.

Figure 2.5: (a) displays four data domains extracted from each radar in the network, accompanied by a flowchart illustrating the prediction process for both *feature fusion* and *decision fusion*. (b) illustrates the classifier block placement in the processing pipeline for two fusion concepts.

for HAR, as shown in Fig. 2.5, within a distributed radar network:

Signal level fusion: This method involves summing the range-time (RT) maps of all radar nodes to form one signal fused RT map, then computing a micro-Doppler (μD) spectrogram from the fused RT map, and using the μD spectrogram for classification.

Feature level fusion: This approach involves utilizing the extracted features, such as the μD spectrogram features of each radar node, which are subsequently concatenated.

Weighted radar selection over time: Only one radar node's data is used for classification. At each time step, one radar node is selected out of the five based on the most suitable aspect angle and received power. To determine the most suitable node, multilateration processing followed by a tracking filter is implemented to determine the position, velocity, and acceleration of the target.

Orthogonal radar fusion: This method explores the theoretical capability of capturing arbitrary movement in space using two radars with orthogonal line of sight. Two different setups were investigated. The first setup involves combining two orthogonal radars through feature fusion, processing each one individually and concatenating the two spectrograms. The second setup first combines radar nodes 1 and 5 (which face each other as depicted in Fig. 2.1) through signal level fusion (i.e., summation), and then fuses the result with orthogonal node 3 (see Fig. 2.1). Fusing the orthogonal nodes is performed through feature fusion. Both approaches yield a two-dimensional spectrogram, displaying the x and y components of the two-dimensional velocity vector in space.

The classification performance of the fusion approaches was evaluated using several RNN architectures, with the simple signal fusion approach yielding the best results (with details presented in Chapter 4). Orthogonal radar fusion using radars 1, 3, and 5 achieved similar performance. All fusion approaches outperformed the use of a single, fixed radar. Given the success of signal-level fusion with the Bi-LSTM classifier, this fusion method was further employed to evaluate other types of classification networks. Gated recurrent units (GRU) were tested in both mono- and bi-directional modes, as well as mono- and bi-directional LSTM. All classifiers performed well with an F_1 score exceeding 84.4 %. Additionally, the author explored alternative evaluation metrics such as intersection over union and Jaccard index, which may be more suitable for assessing performances in scenarios with imbalanced classes, such as a higher frequency of walking compared to instances of in-place activities or falls. This aspect is also discussed with more details in Chapter 4.

2.7. Conclusion

This chapter introduced some common applications of pulsed ultra-wideband (UWB) radars versus frequency-modulated continuous-wave (FMCW) radars applied to HAR and related fields, such as the automotive sector. Following, the Humatics (former: Time Domain) of the type PulsON P410 is introduced. It outlines the gathered dataset and the experimental setup of this thesis, highlighting the deployment of multiple radar nodes in a semicircular arrangement. Afterward, a generic radar signal model is presented assuming point-like scatterers with the mathematical derivation inspired by Deudon et al. [71] and He [59]. Furthermore, an overview of human locomotion models is introduced with the mathematical extension to a radar signal model for extended targets when the radar's range resolution cells subceed the target size. Subsequently, an examination of radar data domains that can be extracted from the used pulsed SISO radar node is shown, with radar data domains ranging from the simple range time (RT) map to time-frequency representations extracted using the short-time Fourier transform (STFT), up to the phase information with its fine line segments. The following sections show data fusion concepts for combining different data domains extracted from various radar nodes as a counterpart to using only single radar sensors with only one domain. These concepts of the aforementioned radar domains and fusion models are used throughout the subsequent chapters.

3

Radar Data Domains for Indoor Monitoring and Classification

Micro-Doppler spectrograms are a conventional data representation domain for movement recognition such as Human Activity Recognition (HAR) or gesture detection. However, they present the problem of time-frequency resolution trade-offs of short-time Fourier transform (STFT), which may have limitations due to unambiguous Doppler frequency, and the STFT computation may be onerous in constrained embedded environments. We propose an alternative classification approach based on the radar phase information directly extracted from high-resolution range-time map (RM). This novel approach does not use the aforementioned micro-Doppler processing, and yet achieves equivalent or even superior classification results. This shows a potential advantage for low-latency, real-time applications or computationally constrained scenarios. The proposed method exploits the histogram of oriented gradients (HOG) algorithm as an effective feature extraction algorithm, specifically its ability to capture the unique shapes and patterns present in the wrapped phase domains, such as their contour intensity and distributions. Validation results consistently above 92 % demonstrate the effectiveness of this method on two independent datasets of arm gestures and gross-motor activities. These were classified with three algorithms, namely the Nearest Neighbor (NN), the linear Support Vector Machine (SVM), and the Gaussian SVM classifiers using the proposed phase information. Feature fusion of different data domains, for example, the modulus of the RM fused with the RM phase information, is also investigated and shows classification improvement specifically for the robustness of activity performances, such as the aspect angle and the speed of performance.

Parts of this chapter have been published in:

R. G. Guendel, F. Fioranelli, and A. Yarovoy, “Phase-based Classification for Arm Gesture and Gross-Motor Activities using Histogram of Oriented Gradients”, in *IEEE Sensors*, 2020.

3.1. Introduction to Radar Data Domains

The amount of research on human activity recognition (HAR) with radar sensors has increased tremendously over the past decades, with significant progress made in almost every area related to activities of daily living (ADL) [37, 88]. Closely associated are the areas of gesture and arm motion recognition, which attracted interest for their potential for remote control of smart devices [42, 89, 89, 90]. This field has seen the development of many different classification approaches, including those inspired by deep learning techniques, such as recurring neural networks (RNN) with their bidirectional implementations known as bidirectional long short-term memory (BI-LSTM) [51], frameworks for generating synthetic radar signatures via Generative Adversarial Networks (GANs) [4], and effective multi-frequency training for multiple radar sensors used for HAR [91]. These techniques for HAR and gesture classification also include multimodal frameworks where different sensing modalities can be combined with the radar. For example, in recent studies, magnetic induction systems and, more generally, wearables are also used for HAR in conjunction with radar applications [92, 93].

However, most of the research work on radar for HAR and gestures has focused on the modulus (magnitude) of the micro-Doppler (μD) spectrogram and, in part, on the range-time map (RM), the range-Doppler (RD), or the range-Doppler surface (RDS) as radar data domains to start the classification process [20, 38, 94–96].

We propose a different and innovative approach based on the usage of the phase information directly extracted from complex high-resolution RM matrices. To our knowledge, this data domain has been very marginally explored for radar-based HAR and gesture classification, while other researchers have, for example, applied phase unwrapping techniques (PUT) on the phase of the μD spectrograms [97]. Other research groups exploit both the wrapped and unwrapped phases of a radar's raw signal. These phases are often fused with other radar data domains, and novel masking techniques and adaptive thresholds are used before classification networks are utilized [98–102]. In this work, we compare conventional radar data domains, such as the μD spectrogram and the RM, with different formats of the proposed phase-based domain information, namely the phase of the μD spectrogram and the phase of the RM. In both cases, their original form (which includes all recorded range bins) and a "cropped" form (which only considers a spatial window, e.g., of approximately 1 m around the detected target for the RM) are considered.

Different features to be extracted from the aforementioned data domains are investigated. We tested that conventional, well-performing features extracted by the Two-Dimensional (2-D) Principal Component Analysis (PCA) were not able to capture the relevant information for HAR and gesture classification from the unique shapes in the phase of the RM matrices. On the contrary, the features derived from the histogram of oriented gradients (HOG) technique proved to be suitable when applied to phase matrices, as they are capable of capturing the salient patterns in terms of strength and orientation of the typical "line structures" in such plots, while still maintaining a relatively simple mathematical formulation compared to less easily explainable convolutional neural networks.

The remainder of the chapter is organized as follows. In Section 3.2, the radar signal model is presented along with the HOG feature extraction method and the data domains.

In Section 3.3, the experimental setup is outlined together with the detailed results on the gross-motor activities and the arm gesture datasets. Finally, the final remarks are provided in Section 3.4.

3.2. Data Domain Representation and Feature Selection

In this section, the different radar data domains will be introduced. In particular, the phase domain information is exploited, and its features are extracted using the histogram of oriented gradients (HOG) for the classification of human activities and arm gestures. Prior to this section, Chapter 2 has provided fundamental insights on the radar signal of the Humatics (former Time Domain) P410 system. Cleaning steps, such as moving target indication (MTI) filtering, mean subtraction, and adaptive thresholding, are applied to generate the RM, indicated as *RM-O* in Fig. 3.1.

3.2.1. Range-Time Map and Phase Angle Representation

The radar provides only the in-phase components, as presented in the matrix S_{mn} in Fig. 2.4. A common practice for reconstructing the complex signal along the range (R) is the Hilbert transform. The process is repeated across all column vector $[s_{1n}, s_{2n}, \cdots, s_{mn}]^T$ for n $= 1, \cdots,$ N [103]. The Hilbert transform creates a complex-valued causal function from the purely real-valued range profile with the property of a phase shift $\frac{\pi}{2}$ (90°), such that $\hat{S}_{\text{mn}} = Hil(S_{\text{mn}}) = Re(\hat{S}_{\text{mn}}) + i \cdot Im(\hat{S}_{\text{mn}})$ with $i = \sqrt{-1}$. Now, $\hat{S}_{mn}$ represents the complex signal matrix of the RM with in-phase samples, $Re(\hat{S}_{\text{mn}})$, and quadrature samples, $Im(\hat{S}_{\text{mn}})$ [104, 105].

From the Euler representation of complex numbers, the phase angle ϕ_{mn} of the signal $\hat{S}_{\text{mn}}$ can be computed as $\phi_{\text{mn}} = \angle\hat{S}_{\text{mn}}$, with ϕ_{mn} in the value domain of $\phi_{\text{mn}} \in \{\text{R}| -\pi < \phi_{\text{mn}} \leq \pi\}$. In fact, the matrix ϕ_{mn} has the same range and slow-time resolution as S_{mn} and $\hat{S}_{\text{mn}}$. The resulting RM phase plot can be seen in Fig. 3.1 indicated as, *RM-PO*, originating from the phase angle block in the flow chart.

We introduce the "cropped" RM, since studies have shown that the arm span to body height ratio is between 0.98 and 1.08, so that the range stretch of an arm toward the radar is approximately $^1/_2 \times$ bodyheight, which is roughly equal to $^1/_2 \times$ armspan. For this, the tallest test person in the dataset with 1.84 cm height can stretch their arms at a maximum of about 92 cm towards the radar [106]. To effectively capture the span of all motions, also including possible torso movements when performing fast gesture motions, we capture 20 % beyond the expected maximum range, resulting in 1.10 m. In other words, a "cropped" version of the RM matrices and their phases is considered using this spatial window centrally placed in the target range. The location of the subject to perform the cropping operation is provided by the derivative target line (DTL) [40] which can determine the person's distance from the radar (the DTL could also be replaced by other suitable target trackers [33]). This cropped phase plot is shown in Fig. 3.1 as *RM-PC*, originating from the cropping block of the flow chart. The cropped window of 1.10 m is also applied to the original RM, with an example shown in Fig. 3.1, denoted as *RM-C*.

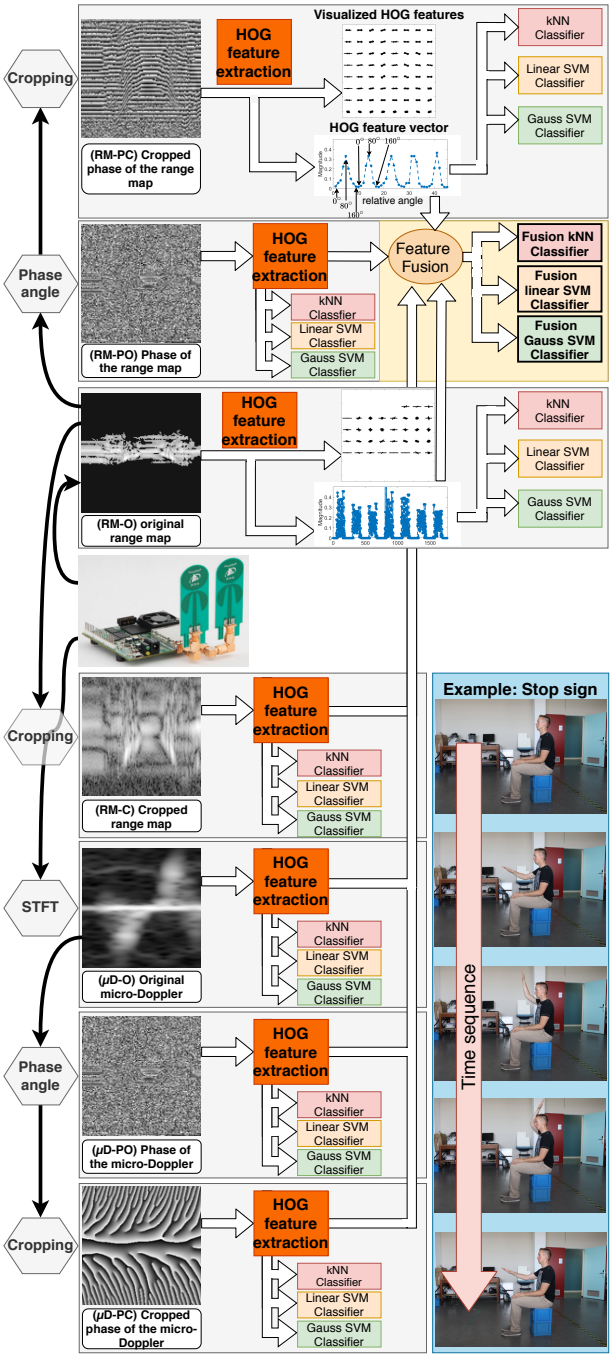


Figure 3.1: Schematic representation of the feature extraction by the histogram of oriented gradients (HOG) algorithm for the individual data domains, with feature fusion shown in the yellow box.

3.2.2. Spectrogram Representation

From the Hilbert transformed signal, $\hat{S}_{\text{mn}}$, the fast Fourier transform (FFT) across each scan is computed. Then, the short-time Fourier transform (STFT) is applied to the vector of the 4 GHz frequency for the computation of the µD spectrogram [107, 108]. Using a f_{PRF} of 121.95 Hz yields to an unambiguous Doppler frequency of ±60.97 Hz. Thus, the unambiguous velocity is ±2.17 m/s and is calculated as $\pm v_{un} = c_0 \cdot f_{\text{PRF}}/(4 \cdot f_o)$, with, c_0, the speed of light and, f_0, the center frequency of 4.2 GHz (operational frequency band: 3.1–5.3 GHz) [94]. An example of the µD is shown in Fig. 3.1 as *µD-O*. The phase angle and the cropped phase angle of the µD are computed from the RM explained in Section 3.2.1. The phase information of the micro-Doppler (µD) spectrograms is also computed. The original phase angle and the cropped phase angle of the µD can be seen in Fig. 3.1 as *µD-PO* and *µD-PC*, respectively. The selected phase map, denoted as *µD-PC*, captures only 25 % of the original unambiguous Doppler frequency, resulting in a Doppler frequency extent of ±15.24 Hz. Using this method, the fine Doppler shifts near the Doppler center are more discernible. This increased level of details in the region of interest leads to improved classifiers' performance due to enhanced feature extraction capability.

In summary, the data representation domains in Fig. 3.1 are:

RM-PC	Cropped phase of the range-time map
RM-PO	Phase of the range-time map
RM-O	Original range-time map
RM-C	Cropped range-time map
µD-O	Original micro-Doppler
µD-PO	Phase of the micro-Doppler
µD-PC	Cropped phase of the micro-Doppler

3.2.3. Histogram of Oriented Gradients (HOG)

The histogram of oriented gradients (HOG) is a powerful tool for edge and contour detection and has been widely used in the computer vision and optical character recognition fields due to the ability to characterize strength and regularities of line patterns and contours in images [109]. This method first determines the gradients, g_x, and g_y, by the partial derivative as, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, so that the gradient vector is defined as,

$$\nabla f(x,y) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} f(x+1,y) - f(x-1,y) \\ f(x,y+1) - f(x,y-1) \end{bmatrix} \tag{3.1}$$

for a matrix $f(x,y)$, where x and y represent the individual samples or pixels. From the $\nabla f(x,y)$ two important attributes are extracted:

- The magnitude of the vector by computing the L_2-norm as,
  $g = \|\nabla f(x,y)\|_2 = \sqrt{{g_x}^2 + {g_y}^2}$
- The directional orientation as, $\theta = \arctan\left(\frac{g_y}{g_x}\right)$

It is noted that $f(x, y)$ defines only an area of the whole image matrix, where in our experiments the HOG sizes are $[16, 16]$ or $[32, 32]$ pixels. Examples of the visualized HOG features are shown in Fig. 3.2 at the bottom, where the top shows the related *RM-PC* matrices. Specifically, the contours in the *RM-PC* appear to be mainly horizontally orientated with an approximate variance of up to $\pm 45°$. Therefore, vectors that are vertically oriented have almost zero length, which relates to the magnitude of g. The extracted features are represented by histograms, examples are shown in the flow chart of Fig. 3.1 for *RM-PC* and *RM-O* and denoted as "HOG feature vector". Then, the "HOG feature vectors" are the input of the classifiers [109–112].

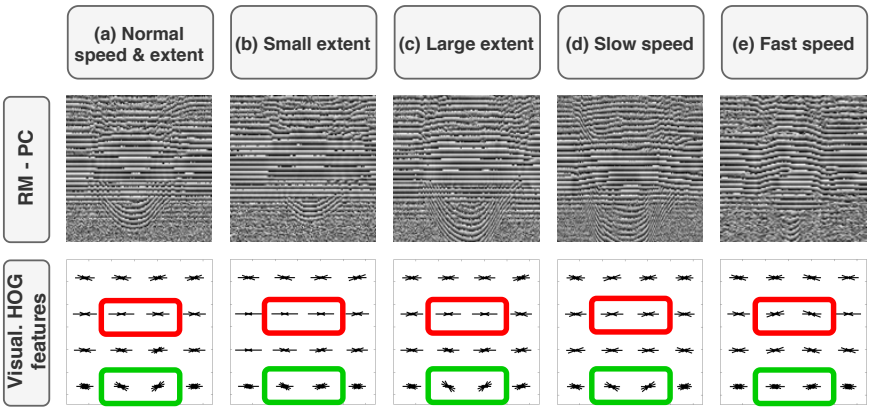


Figure 3.2: The impact on the HOG features illustrated for the extent and speed variation for the *push and pull arm* activity.

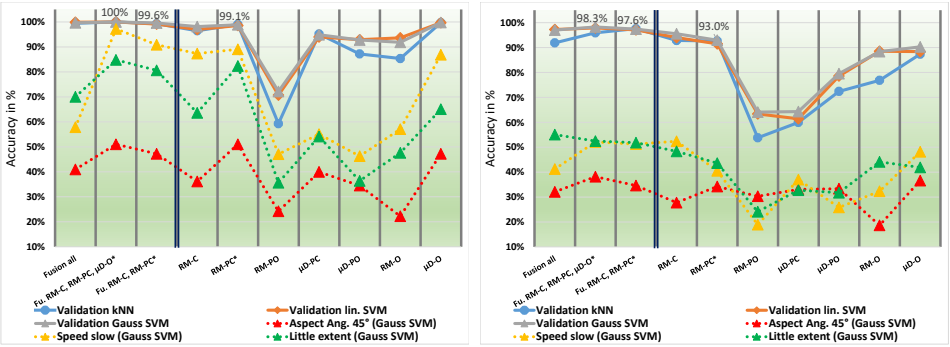
In Fig. 3.2, we show the HOG feature for the activity of *push and pull arms*. Here, we illustrate the differences in performing the motion with (b) small arm extent, (c) large arm extent, (d) slow speed, (e) fast speed, and compare those with the (a) normal performed motion. Specifically, two regions were deliberately selected, the red and green circled areas to characterize the features. Body movements occur typically in the middle regions, as those marked with the red rectangles. It can be seen that the small extent leads to little body movements, since the torso is mainly static on a fixed position, thus resulting in mostly horizontal HOG vectors. Furthermore, the green-framed areas contain HOG vectors mostly originating from the arm movement towards the radar. As a result by performing the activities with large extents, the HOG feature vectors become more steeply orientated and diagonal in contrast to the small extent case. Another important difference can be seen by performing the motion with fast speed, where the inertia of the torso and the arms inevitably lead to a larger backward movement of the torso than for slower speed. As a result, the HOG features emphasize the backward movement of the torso in a very distinctive way.

3.2.4. Feature Selection and Classification

In the previous section, the HOG feature extraction was explicitly described. Now, the orientated gradients are extracted for the 2D detection windows ($[16, 16]$ or $[32, 32]$ pixels)

Table 3.1: Average performance of gesture experiments for two different f_{PRF} for the feature fusion case (3 left side columns) and individual domains (7 right side columns). Average test performance presented for AA of 45°, slower speed and smaller extent.

Features / Classifier	Fusion all	Fu. RM-C, RM-PC, µD-O	Fu. RM-C, RM-PC	RM-C	RM-PC	RM-PO	µD-PC	µD-PO	RM-O	µD-O
Validation performance for f_{PRF}=122 Hz, HOG=[16,16]										
kNN=5 (euclidean)	99.46	100.00	99.29	96.49	99.05	59.29	95.18	87.20	85.36	99.70
SVM linear	99.88	100.00	99.17	96.96	98.57	70.71	93.99	92.92	93.63	99.70
SVM Gauss	99.64	100.00	99.58	98.16	98.87	72.14	94.88	92.80	91.85	99.76
Test performance for SVM Gaussian classifier, f_{PRF}=122 Hz, HOG=[16,16]										
(1) Aspect angle: 45°	40.96	51.04	47.21	36.15	51.04	24.21	39.98	34.50	22.23	47.21
(2) Speed: slow	57.87	97.22	90.90	87.35	89.04	47.07	55.09	46.30	57.10	86.88
(3) Extent: small	70.04	84.83	80.64	63.63	82.37	35.64	54.13	36.38	47.60	65.11
Validation performance for f_{PRF}=12.2 Hz, HOG=[16,16]										
kNN=5 (euclidean)	91.89	95.92	97.60	92.84	92.67	53.80	59.96	72.43	76.85	87.36
SVM linear	97.26	97.88	97.09	94.02	91.50	63.37	61.41	78.58	88.54	88.37
SVM Gauss	97.04	98.27	97.43	95.53	92.95	64.09	64.32	79.53	88.37	90.27
Test performance for SVM Gaussian classifier, f_{PRF}=12.2 Hz, HOG=[16,16]										
(1) Aspect angle: 45°	32.09	38.18	34.64	27.77	34.25	30.32	33.07	33.37	18.65	36.61
(2) Speed: slow	41.24	52.12	51.27	52.40	40.54	18.93	37.01	25.85	32.35	48.16
(3) Extent: small	55.08	52.46	51.86	48.34	43.62	24.12	32.76	31.76	44.12	41.91



(a) Performance with f_{PRF}=122 Hz, HOG=[16,16]

(b) Performance with f_{PRF}=12.2 Hz, HOG=[16,16]

Figure 3.3: Validation and test performance of gesture experiments for the different f_{PRF} related to Table 3.1. Red, yellow, green data labels for testing performance match in color those in Table 3.1.

to provide the histograms and the feature vector, while the orientation is discretized in the angular bins of the histogram of 20° from 0° to 160°. It is noted that vectorial gradients are proportionally split into the histogram bins if their orientation value is between the bins' nominal values. Furthermore, HOG only considers gradients from 0° to 180° since a contour in an image is non-directional. Furthermore, the histogram bin of 180° does not exist since it is equivalent to the 0° bin. This processed feature vector is then used for classification.

We tested our proposed method with a few of the most common yet effective supervised learning classifiers, the Nearest Neighbor (NN), the Linear Support Vector Machine (linear SVM), and the Gaussian SVM classifier. Specifically, the NN classifier was used with a number of neighbors of five with a Euclidean distance computation. For the linear SVM and the Gaussian SVM classifier, we apply the multi-class setting one-versus-one.

The yellow square box in Fig. 3.1 shows that we also apply feature fusion for classification. For this, individual feature vectors from individual radar data domains are joined into a concatenated feature vector, e.g., $\kappa_{\mathrm{Fu_{all}}}$, which is expressed as $\kappa_{\mathrm{Fu_{all}}}=[\kappa^{\mathrm{T}}_{\mathrm{RM\text{-}PC}}, \kappa^{\mathrm{T}}_{\mathrm{RM\text{-}PO}}, ..., \kappa^{\mathrm{T}}_{\mu\mathrm{D\text{-}PC}}]^{\mathrm{T}}$. We show that concatenating all possible feature vectors from different radar domains does not lead to the best classification results. In this regard, the best accuracy was achieved by using a subset of features, namely, the cropped phase of the RM (*RM-PC*), the cropped RM (*RM-C*), and the μD spectrogram (*μD-O*), so that the concatenated feature vector is formed as $\kappa_{\mathrm{Fu_{best}}}=[\kappa^{\mathrm{T}}_{\mathrm{RM\text{-}PC}}, \kappa^{\mathrm{T}}_{\mathrm{RM\text{-}C}}, \kappa^{\mathrm{T}}_{\mu\mathrm{D\text{-}O}}]^{\mathrm{T}}$.

On the other hand, very promising results were achieved by excluding the μD spectrogram and focusing deliberately on the RM which is directly provided by the radar, so that an additional STFT calculation or even more complex time-frequency distributions can be omitted. Computing the μD spectrogram requires some computational resources and time. Considering the STFT as the simplest approach to calculating spectrograms via time-frequency analysis, a measure of its complexity as the number of floating-point operations (FLOPS) can be computed as,

$$k \cdot log_2(n^n) \quad \text{with:} \quad k = \frac{N-L}{n-L} \tag{3.2}$$

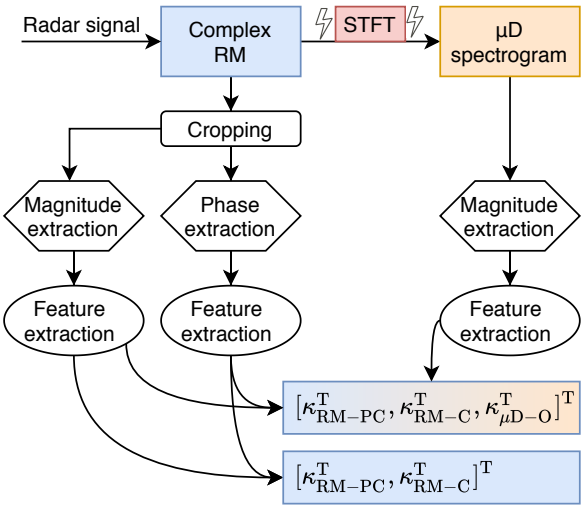


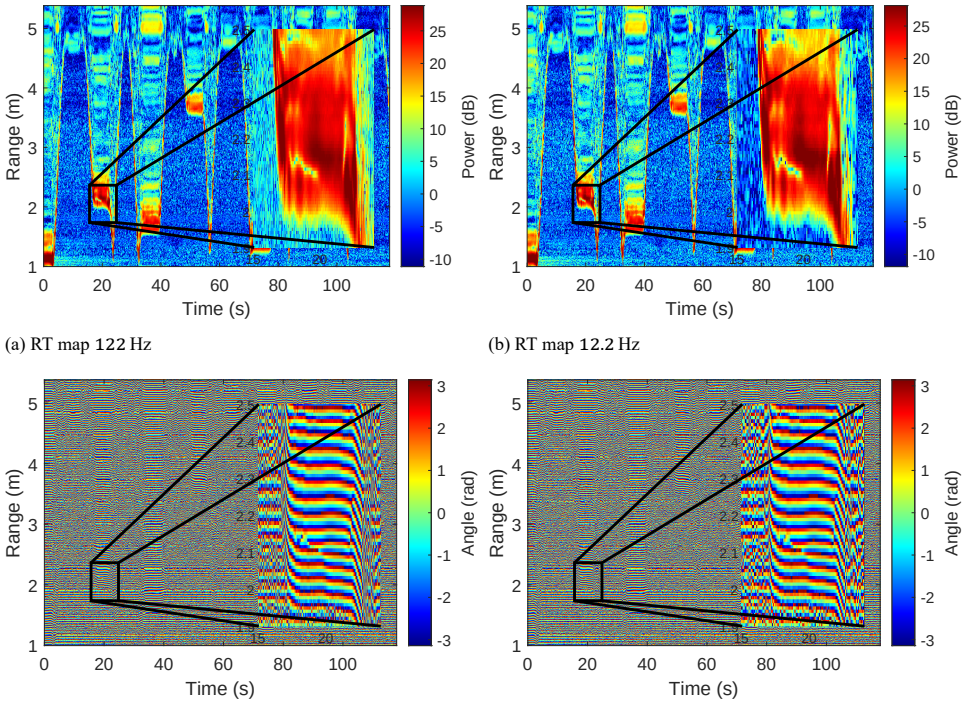
Figure 3.4: The revealing scheme to extract the features from the range only as, $\kappa_{\mathrm{Fu}_{\mathrm{range}}}$, and with containing the μD spectrogram feature as, $\kappa_{\mathrm{Fu}_{\mathrm{best}}}$, which implies the STFT.

with n the STFT window function length, L the overlap length, and N the signal length. By using the STFT with the discrete Fourier transform (DFT), the FLOPS from Eq. (3.2) changes to kn^2 [113, 114]. It is noted, the simplified FLOP calculation does not consider additional smoothing window multiplication, i.e., the Hamming window. In this respect, avoiding the STFT leads to the concatenated feature vector consisting of $\kappa_{\mathrm{Fu}_{\mathrm{range}}}=[\kappa^{\mathrm{T}}_{\mathrm{RM\text{-}PC}}, \kappa^{\mathrm{T}}_{\mathrm{RM\text{-}C}}]^{\mathrm{T}}$, and includes only the cropped RM and the cropped phase of the RM. The scheme is illustrated in Fig. 3.4.

3.3. Experimental Results

Two datasets were collected in the radar laboratory at the Delft University of Technology (TU Delft), consisting of a comprehensive number of classes for the gesture and the gross-motor experiments that are presented in this section. Both sets were recorded with Humatics P410 pulsed radar systems. Four participants were involved in the experimental data collection, with a height between 1.65 m and 1.84 m, and a weight between 65 kg and 86 kg. The number of data samples for the training sets is 280, 282, and 300 samples per class for the gross-motor activities with a f_{PRF} of 12.2 Hz, and gesture activities with a f_{PRF} of 12.2 Hz and 122 Hz, respectively.

In the analytical experiment presented in Fig. 3.5, phase maps are derived from a sequence that shows walking and stopping activities. These maps utilize a PRF of 122 Hz and a synthetically reduced PRF of 12.2 Hz to provide an example. The magnified inserted regions visually capture the actions of an individual walking, stopping, and then resuming the walk. This is demonstrated for both the initial PRF (Figs. 3.5a and 3.5c), and the 10-times reduced PRF (Figs. 3.5b and 3.5d) over a data collection span of 2 min. The findings indicate that despite a 10-times reduction in the sampling rate of the data over



(a) RT map 122 Hz

(b) RT map 12.2 Hz

(c) Phase map 122 Hz

(d) Phase map 12.2 Hz

Figure 3.5: The figures display the phase map derived from a range time (RT) profile of a human walking sequence, collected at PRFs of 122 Hz (RT map: Fig. 3.5a, phase map: Fig. 3.5c) and 12.2 Hz (RT map: Fig. 3.5b, phase map: Fig. 3.5d)

slow-time, there is no noticeable aliasing or degradation in these data domains. This minimal aliasing effect is further corroborated by the classification results obtained.

The test set for gross-motor activities includes 120 samples by considering an aspect angle of 45°. The number of test data for gesture activities for both f_{PRF} amounts to 94, 78, and 84 samples for the aspect angle of 45°, the slow speed, and the small extent, respectively.

Each of the considered radar data domain (e.g. *RM-O*, *RM-PC*, or *μD-O*) was resized to a matrix size of 128x128, from which further processing extracts the HOG feature vectors (κ_{i}), with, i, the seven different domains listed in Section 3.2.2. Collecting training/validation data for a 20-fold cross-validation was performed under controlled aspect angle, spatial extent, and speed of movement. For further evaluation, we collected a separate test set for the gesture activities with (1) an aspect angle (AA) of about 45°, (2) a slower speed, and (3) a smaller extent, as shown in Table 3.1.

Training/validation data for a 20-fold cross-validation of gross motor activities was collected with the same conditions, while the test set (hold-out) contains only data with an aspect angle of 45°, since a slower speed or a smaller extent can be difficult for some gross-motor activities, e.g., falling. It should be noted that the training/validation set does not include data samples that reflect the conditions of the test set (AA, slower speed or

Figure 3.6: Pictures of the activities and related cropped phase matrices: six gestures (top, green-shaded), and eight gross-motor activities performed bidirectionally (bottom, yellow-shaded).

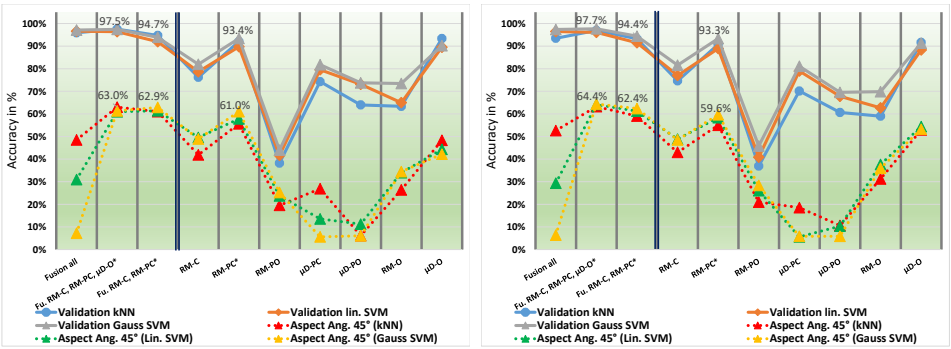
smaller extent). In fact, the classifier is requested to classify data without being explicitly trained in those conditions. In this work, the training/validation was performed using 20-fold cross-validation, and it refers to [115]; essentially:

- Training set size = $\frac{19}{20}$(or 95 %) of the dataset.
- Validation set size = $\frac{1}{20}$(or 95 %) (or 5 %) of the dataset.

3.3.1. Arm Gesture Results

The first dataset consists of six gesture activities, namely, (a) push and pull arms, (b) close arms, (c) open arms, (d) rolling arms, (e) stop sign, and (f) clap hands. The individual gesture activities and the related *RM-PC* are shown in the green-shaded box in Fig. 3.6. The movements were performed facing the radar with a distance of 3.20 m from the radar. For performance comparison, we collected two subsets with f_{PRF} of 122 Hz and 12.2 Hz.

The validation and test performance of the gesture activities can be found in Table 3.1. Specifically, the test dataset includes three different cases, namely (1) a higher aspect angle (AA), (2) a slower speed, and (3) a smaller spatial extent compared to the training/validation data. The results are visualized in Fig. 3.3 related to Table 3.1. In the table, we show the test performance of the gesture activities only for the Gaussian SVM classifier, which gives



(a) Performance with f_{PRF}=122 Hz, HOG=[16,16]

(b) Performance with f_{PRF}=122 Hz, HOG=[32,32]

Figure 3.7: Validation and test performance of gross-motor experiments for two different HOG sizes related to Table 3.2. Red, yellow, green data labels for testing performance match in color those in Table 3.2.

the best results and outperformed the kNN and the linear SVM classifier.

Remarkably, we illustrate that our proposed method of the phase of the RM (*RM-PC)* and HOG features is still capable of classifying gestures with sufficient accuracy when the unambiguous Doppler frequency of the µD spectrogram decreases by a factor of 10x using the lower f_{PRF}=12.2 Hz instead of the f_{PRF}=122 Hz. As expected, due to the resulting ambiguity in micro-Doppler, the classification accuracy decreases when using the *µD-O* domain by more than 9 % with the best performing classifier, Gaussian SVM (99.76 % → 90.27 %). The RM and especially the *RM-PC* is almost unaffected by lowering the radar f_{PRF} by such a significant amount. Regarding the extraction of HOG features, the highest classification results were obtained using a HOG cell size of [16,16] samples. Other HOG cell sizes of [8,8] or [32,32] have also been tested, but have provided lower classification results.

3.3.2. Gross-Motor Activities Results

The second dataset contains an even larger number of classes, while the activities were performed away from the radar in addition to facing the radar. The activities are (g) bending from standing, (h) bending from sitting, (i) kneeling down, (j) kneeling up, (k) sitting down, (l) standing up, (m) falling, and (n) standing up from falling. Taking into account the bidirectional orientation, 16 classes were collected and considered for classification. The activities can be seen in Fig. 3.6 in the yellow-shaded box, together with the relevant bidirectional *RM-PC* plots.

We show the average classification results for gross-motor activities in Table 3.2 and visualize the results in Fig. 3.7. In this case, the dataset was collected with only a f_{PRF} of 122 Hz, but two HOG cell sizes of [16,16] and [32,32] are compared. These are shown in Table 3.2 by the gray-shaded and yellow-shaded boxes, respectively. The results show that doubling the HOG cell size does not lead to a drastic classification performance reduction and both results are rather comparable. In the HOG descriptor, three parameters are pivotal: cell size, block size (typically by default 2x2 cells), and block overlap (by default 50 % of the block size). By increasing the cell size, while keeping the other parameters constant, the

Table 3.2: Average performance of gross-motor experiments for two different HOG cell sizes for the feature fusion case (3 left side columns) and individual domains (7 right side columns). Average test performance presented for 45° AA for the three different classifiers kNN, linear SVM, and Gaussian SVM.

Features / Classifier	Fusion all	Fu. RM-C, RM-PC, µD-O	Fu. RM-C, RM-PC	RM-C	RM-PC	RM-PO	µD-PC	µD-PO	RM-O	µD-O
Validation performance for f_{PRF}=122 Hz, HOG=[16,16]										
kNN=5 (euclidean)	95.89	97.53	94.72	76.21	92.04	38.24	74.30	63.98	63.35	93.37
SVM linear	96.61	96.38	91.95	78.67	89.70	41.66	79.54	73.11	64.97	89.50
SVM Gauss	97.12	97.42	93.91	81.97	93.39	44.00	81.79	73.74	73.45	90.20
Test performance for aspect angle (AA): 45° f_{PRF}=122 Hz, HOG=[16,16]										
AA: 45° (kNN)	48.51	63.01	61.01	41.89	55.63	19.56	26.90	6.18	26.41	48.33
AA: 45° (lin. SVM)	30.95	60.96	61.41	49.58	57.80	23.74	13.56	11.25	34.02	44.33
AA: 45° (Gauss SVM)	7.20	61.36	62.87	48.96	61.01	25.26	5.60	6.09	34.46	42.20
Validation performance for f_{PRF}=122 Hz, HOG=[32,32]										
kNN=5 (euclidean)	93.46	96.94	93.23	74.62	90.81	36.96	70.14	60.66	58.99	91.61
SVM linear	96.45	96.07	91.48	77.02	88.92	40.87	79.00	67.78	62.68	88.29
SVM Gauss	97.37	97.66	94.40	81.57	93.32	45.82	81.09	69.54	69.85	91.30
Test performance for aspect angle (AA): 45° f_{PRF}=122 Hz, HOG=[32,32]										
AA: 45° (kNN)	52.65	63.23	59.05	43.00	55.00	20.99	18.54	10.63	31.26	52.91
AA: 45° (lin. SVM)	29.35	64.16	61.23	48.87	58.34	26.06	5.60	10.36	37.66	54.38
AA: 45° (Gauss SVM)	6.40	64.38	62.38	48.51	59.63	28.37	5.87	5.87	36.02	53.18

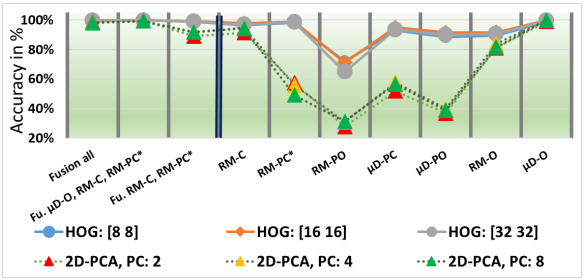


Figure 3.8: Evaluating the HOG versus the 2D-PCA feature extraction methods, paired with the Gaussian SVM classifier. This comparison considers HOG cell sizes of 8, 16, and 32, and principal component vectors (PC) of 2, 4, and 8.

computational load decreases. This is due to the reduced size of the resulting input feature vector, κ_i [116].

For training/validation accuracy, it can be seen that the Gaussian SVM classifier gives slightly higher accuracy for the proposed method based on the *RM-PC* data domain. When feature fusion is considered, the advantage of using the Gaussian SVM classifier with the HOG cell size of [32,32] is less dominant compared to the NN or the linear SVM classifier. Additionally, for classification based on a single data domain, the proposed method of *RM-PC* yields a performance improvement of +1.71 % compared to the best *µD-O* classification (*µD-O* [kNN] 91.61 % → *RM-PC* [Gauss SVM] 93.32 %). Although small in absolute terms, this improvement can be considered significant due to the large number of classes (16).

The test performance in Table 3.2 shows the results for the AA of 45° when using the three different classifiers of kNN, linear SVM, and Gaussian SVM. This is different from Section 3.3.1 where gesture activities were performed at slower speed and with a smaller spatial extent. Specifically, the Gaussian SVM classifier outperforms the other tested classifiers in almost every category, except for the *Fusion all* case, which will be discussed in the next section.

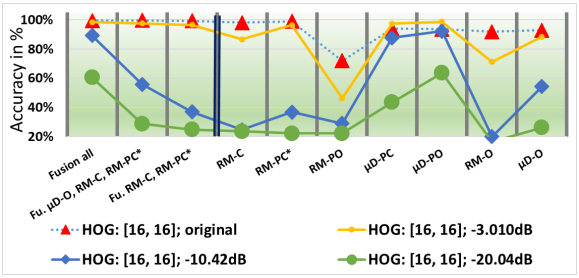
3.3.3. Comparative and Noise Analysis

In this section, a comparison between the investigated method of the HOG classification and the 2D-PCA classification is demonstrated, as well as a noise performance analysis on the gesture dataset. Regarding Table 3.3 and Fig. 3.8, it can be seen that the phase domains, such as RM-PC, RM-PO, µD-PC, µD-PO, as well as their fusion are better classified by the HOG algorithm. The typical µD-spectrogram (µD-O) classification performs almost equal with both methods. This leads to the conclusion that 2D-PCA cannot capture phase-related patterns as well as is done by the HOG algorithm.

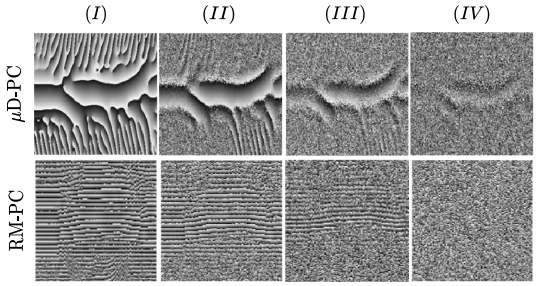
Fig. 3.9 illustrates the impact of reducing the SNR by −3.010 dB, −10.42 dB, and −20.04 dB using additive white Gaussian noise. The alteration in the phase pattern is evident in Fig. 3.9b for the *push and pull arms*. Decreasing the SNR by −3.010 dB in the original signal yields classification results that are nearly comparable, especially for the proposed method (RM-PC). A more significant reduction in SNR results in an anticipated decline in classification for most feature domains, as detailed in Table 3.3.

Table 3.3: Average performance of gesture experiments by comparing the HOG algorithm (first sub-table) and 2D-PCA (second sub-table) for different cell sizes and by using different principal component (PC) vectors, respectively. The last sub-table shows the performance by adding different noise levels. The 3 left columns are the feature fusion accuracy values and the 7 right columns are the individual domains.

Features / Method	Fusion all	Fu. RM-C, RM-PC, µD-O	Fu. RM-C, RM-PC	RM-C	RM-PC	RM-PO	µD-PC	µD-PO	RM-O	µD-O
Validation performance using HOG, f_{PRF}=122 Hz										
HOG: [8 8]	99.12	99.94	98.81	96.55	98.33	71.79	93.04	88.51	89.64	98.99
HOG: [16 16]	99.70	99.94	99.29	97.74	98.99	71.25	95.06	91.55	91.73	99.76
HOG: [32 32]	99.82	99.88	99.17	96.85	98.57	65.06	93.51	90.00	90.60	99.58
Validation performance using 2D-PCA, f_{PRF}=122 Hz										
2D-PCA, PC: 2	98.69	99.41	89.05	91.43	57.50	27.92	51.91	36.85	81.07	99.05
2D-PCA, PC: 4	98.27	99.29	91.73	94.64	55.83	31.01	57.74	39.94	83.93	99.82
2D-PCA, PC: 8	98.10	99.29	91.85	94.35	49.11	31.31	56.61	38.81	81.73	99.88
Validation performance using HOG [16 16] and added noise, f_{PRF}=122 Hz										
HOG, -3.01 dB	98.41	97.55	96.50	86.69	96.28	46.41	97.34	98.58	71.35	88.39
HOG, -10.42 dB	89.23	55.89	36.96	24.70	36.85	29.05	87.74	92.14	20.00	54.41
HOG, -20.04 dB	60.48	28.81	24.94	23.69	22.26	22.38	43.51	63.63	16.67	26.25



(a) Performance by adding noise for the HOG algorithm.



(b) Visualization by of the cropped phase-based RM (RM-PC) and µD-spectrogram (µD-RC), respectively, for push and pull arms.

Figure 3.9: Gesture validation performance at a SNR decrease of (*II*) -3.010 dB, (*III*) -10.42 dB, and (*IV*) -20.04 dB using additive white Gaussian noise. Data were collected at $f_{\mathrm{PRF}} = 122$ Hz. See Table 3.3 for results.

3.3.4. Discussion on the Results

The introduced method of classifying the *RM-PC* showed very promising results, which can be further improved by using feature fusion (early fusion) along with the *RM-C*, or *RM-C* together with *µD-O*. However, we do not suggest using all available domains, which can be seen for the *Fusion all* case, as the classification accuracy can drastically decrease due to overfitting. Nevertheless, the use of our phase-based classification framework is suitable to avoid the calculation of the STFT completely, and omit any form of µD computation and related feature extraction, and instead focus on the range only. This can be a potentially interesting advantage to reduce complexity in view of real-time applications and computationally constrained environments, as demonstrated in Eq. (3.2) together with Fig. 3.4.

The *RM-PC* plots in Fig. 3.6 (yellow-shaded box) for facing the radar can be compared with those recorded for movements away from the radar, and also with gesture activities (green-shaded box in the same figure). Here, the changes in contour patterns of the phase plots are dominant and still visible, although the differences are less clear and intuitive than those typically visible in spectrograms. However, the HOG algorithm can capture well and distinguish changes in contour patterns in the angular direction and intensity. For comparison, Fig. 3.2 shows the HOG feature change for (*push and pull arms*) but performed at different speeds and spatial extents. Specifically, the red-framed HOG features represent the torso movement and are more dominant for a larger spatial extent or

faster speed. Also, the green-framed HOG features are distinguishable in strength and orientation, which represents mainly the arm movement. Here a difference can easily be seen between small and large extents.

In the analysis, it was observed that the use of HOG features is sensitive to noise. Existing literature indicates that as noise levels escalate, the effectiveness of HOG decreases. This underscores the importance of a thorough investigation of the effects of noise on both the HOG and PCA approaches [117, 118]. The analysis is pertinent to the conventional range time map (*RM-C*) and the magnitude of the µD spectrogram (*µD-O*) representations. It does not pertain to phase maps derived from the full range time map (*RM-PO*), the cropped range time map (*RM-PC*), including the phase of the µD spectrogram (*µD-PO*) or its cropped version (*µD-PC*). For these domains, it should be noted that PCA was not able to extract fine line and contour patterns. Moreover, when using features directly derived from the phase information, it is crucial to employ in-phase and quadrature balancing techniques either during data collection or in post-processing [6].

3.4. Conclusion

We propose a novel approach for classification of human gross-motor activities and arm gestures based on the phase information directly extracted from high resolution range-time maps (RM). This approach is an alternative compared to the more conventional use of the magnitude of the micro-Doppler (µD) spectrograms for classification. We investigated the wrapped phase of RM and µD spectrograms, whereas the phase-based RM provides superior results over the phase-based µD spectrograms. However, the unique shape of those wrapped phases in terms of intensity and complexity of the line patterns requires a suitable feature extraction algorithm to capture the relevant information, different from the features typically used on a conventional µD spectrogram. For this, we exploited the histogram of oriented gradients (HOG) algorithm to capture suitable features toward a phase-based classification by using three commonly known classifiers, namely the Nearest Neighbor (NN), the linear Support Vector Machine (SVM), and the Gaussian SVM.

We demonstrate this approach on two experimental datasets, namely one for gross-motor activities (e.g. sitting, standing, bending, kneeling, etc.), and the second for arm gestures (e.g. pushing and pulling arms, waving hands, or pointing, etc.). The latter dataset is recorded with two different Pulse Repetition Frequencies (f_{PRF}). We have shown that the proposed method can be applied to arm gesture recognition measured with a 10-times lower f_{PRF} – which can be beneficial by using low-cost hardware – without any noticeable decrease of performance, while a conventional µD-based approach suffers with such data due to Doppler ambiguities. The method has shown to be robust with respect to the test scenario variables: e.g., the aspect angle to the radar line of sight, the velocity, and the extent of arm movements are also characterized.

Promising validation results consistently above 92 % are demonstrated for arm gestures and gross-motor activities using HOG features extracted from the proposed phase-based RM. These results based on phase domain classification can even be improved by fusing features from different radar data domains, such as the original RM and/or the µD spectrograms, which show a more robust performance under different operational conditions (e.g., different aspect angles, extent, or movement velocity). Superior performance was attained by fusing the proposed phase-based RM together with

the magnitude of the RM. Regarding the fusion of the RM domain mentioned previously for the slow f_{PRF} of 12.2 Hz of the gesture dataset, a validation accuracy improvement of greater 10 % compared to the conventional µD spectrogram classification was achieved. Furthermore, the proposed method may suit radar systems that provide complex in-phase and quadrature signal components directly without resorting to the Hilbert transform, as well as for different operational frequencies and bandwidths.

4

Data Fusion in Distributed Radar Systems for HAR

Continuous human activity recognition (HAR) is investigated in arbitrary directions using a network of five spatially distributed pulsed Ultra-Wideband (UWB) radars. While activities performed continuously and in uncontrolled trajectories, which provides a more realistic and natural scenario for HAR, the network of radar sensors is proposed to address the issue of unfavorable or occluded perspectives when using only a single sensor. Different techniques to combine the relevant information from the multiple radars in the network are investigated, focusing on signal level fusion directly applied on range-time maps (RT), and the selection of radar nodes based on location and velocity of the target derived from multilateration processing and tracking. Recurrent neural networks (RNNs) with and without bidirectionality are used to classify activities based on micro-Doppler spectrograms (μD) obtained for sensor fusion techniques. To assess classification performances, novel evaluation metrics accounting for the continuous nature of the sequence of activities and inherent imbalances in the dataset are proposed and compared with existing metrics. It is shown that the conventional accuracy metric may not capture all the important aspects for continuous HAR, and the proposed metrics can be considered for a more comprehensive evaluation.

Parts of this chapter have been published in:

R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Distributed radar fusion and recurrent networks for classification of continuous human activities", in *IET Radar, Sonar and Navigation*, 2022.
R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Evaluation metrics for continuous human activity classification using distributed radar networks", in *IEEE Radar Conference 2022 (RadarConf22)*, 2022.

4.1. Introduction to Data Fusion

Monitoring activities of daily living (ADL) by radar has gained attention for safe and independent aging-in-place of older and vulnerable subjects. This includes recordings of critical events such as falls, monitoring abnormalities in movements and activities, and in general, providing an assessment of well-being in terms of cognitive and physical state [4, 49, 88].

As human activity recognition (HAR) by radar typically exploits micro-Doppler (μD) signatures of human movements, distributed networks with multiple cooperating radars have attracted significant interest to address the issue of reduced μD signatures recorded at unfavorable aspect angles [38, 46, 119–122]. However, the focus has often been on the classifier's architectures, i.e., neural networks [4], including in our work [123]. To the best of our knowledge, there are only limited experimental studies on the most suitable number and topology of the different radar sensors for HAR, and on the most effective fusion techniques to combine their data [46, 124]. Therefore, such questions remain widely open in HAR.

Griffiths et al. [125, 126] investigated drone payload classification using 3 multistatic radar nodes and achieved superior results by voting-based decision fusion among independent classifiers (named as 'binary voting' and 'threshold voting'). The same authors [36, 127] also investigated the usage of handcrafted features of μD spectrograms (e.g., their centroid and bandwidth) for other classification tasks based on the same multistatic radar network of 3 nodes. These tasks included gait analysis of individuals alone and in pairs to identify whether they were armed or unarmed in outdoor surveillance scenarios. Even in this case, decision fusion appeared to provide the best classification results, but little investigation was devoted to lower-level signal fusion approaches.

Unlike the aforementioned studies, this work investigates novel lower-level fusion approaches applied to a network of five spatially distributed monostatic radars simultaneously observing a surveillance area. The proposed fusion schemes aim to combine data from a selection of the different radar nodes prior to the generation of the μD spectrogram used for classification. Specifically, a fusion method based on *incoherent fusion* of the range-time (RT) domain data from each radar node is investigated. Although rather simple in terms of computation, this approach proves to be the most effective in terms of overall classification results, with the following potential advantages given:

- Minimizing the number of classifiers to be defined and trained in the pipeline to one.
- Avoiding complex methods to combine the partial decisions from separate classifiers.
- Using one single μD spectrogram representation for the entire network, containing information from all nodes with computational simplicity.

In addition, two approaches are evaluated to select a subset of the available radar nodes for the subsequent fusion process. The first approach is the *orthogonal radar selection*, whereby two radar nodes with orthogonal lines of sight are selected to capture the signatures of D μalong the radial and tangential directions. While this selection is static and done once and for all, with the second approach, a dynamic selection of the radar nodes is performed. In this *weighted radar selection*, the location, velocity, and heading of the target are first estimated by multilateration processing [128] combined with a simple

alpha-beta-(gamma) – α, β, (γ) tracking filter. Then, a weighting function is implemented to select at any given time the most suitable radar for classification, i.e., the closest radar to the target (thus assuming the highest SNR), or the one radar with most favorable aspect angle to the trajectory (thus assuming the richest micro-Doppler information).

The different radar data fusion methods mentioned above are explored in the context of classifying continuous sequences of human activities, as opposed to the more conventional classification of artificially separated activities [46, 52, 129]. These continuous activities represent more realistic and natural scenarios for evaluating radar-based HAR algorithms, where the transitions between different activities can happen at any time and in undefined instants. If performed along unconstrained trajectories and directions, the classification of such activities can clearly benefit from the multi-perspective views of distributed radar networks.

However, as discussed in our preliminary results in [130], HAR applied on continuous sequences requires alternative performance evaluation metrics beyond simple accuracy or quantities directly extracted from confusion matrices, regardless of the nature of the radar used for recording, i.e., monostatic or distributed/multistatic. Specifically, four aspects of *Continuity*, *Misalignments*, *Interruptions*, and *Imbalance:* for continuous HAR data must be considered.

Continuity: The activities are performed in a natural way with continuous sequence recordings, where transitions between activities occur at arbitrary times and with various activities in the dataset that occupy different times by nature. For example, 'falling' is typically a short-term action, whereas 'walking' can be perfromed over an extended time. Therefore, it is difficult to precisely pinpoint the time instant where one activity ends and the following activity starts, even in the ground truth. *Misalignments*: As a consequence of the difficulty in accurately estimating the time instant of activity transitions, misalignments between ground truth and prediction labels can occur, i.e., time offsets between ground truth and predictions. Depending on the overall goal of the HAR system, one needs to establish the importance of such misalignments in terms of the performance evaluation for classification algorithms. *Interruptions*: As an activity will occupy an extended number of slow-time bins, a dedicated classifier (predictor) can provide a prediction output where a few samples are associated with the wrong classes and are, in fact, false predictions. Such "short-term" false predictions or temporal fluctuations of the predicted class (i.e., outliers) occur only in a few, isolated slow-time bins. Such "jumps" in the prediction label vector are often not encountered when classifying human activities as artificially separated "snapshots", and not captured when using conventional evaluation metrics for continuous-time sequences. In summary, conventional metrics hardly account for temporal fluctuations in a continuous prediction vector. *Imbalance:* Finally, when evaluating realistic sequences of activities, *imbalances* in the dataset will naturally appear due to different occurrences and time spans of activities during an observation period. A typical example can be the prevalence of the 'walking' class while participants move about in the room to perform single instances of other in-place activities. Furthermore, another typical situation is the small amount of available samples for critical activities, such as 'falling'.

To account for the continuous nature of human activities and provide a more insightful performance analysis based on the aforementioned aspects, this work discusses a collection of 10 possible evaluation metrics with their advantages and disadvantages. These metrics

are used to evaluate the different radar data fusion methods on an experimental dataset with five radar nodes, 15 participants, and 9 labeled activity classes. The presented methods are validated for single human target cases. Although more complex scenarios exist, such as multiple individuals in the field of view or individuals with domestic animals, these are left for future research work.
To summarize, the main contributions are:

- Novel fusion methods to combine data from radars in a spatially distributed network are investigated, focusing on signal level fusion and selection of a subset of nodes to improve HAR performances.
- The original classification performance metrics are shown to account for the continuous nature of human activities. These metrics are used to evaluate an experimental dataset containing data from five radar nodes, with 15 participants performing continuous and unconstrained sequences of 9 activities.
- Four recurrent neural networks (RNNs) are used as classifiers for HAR, namely gated recurrent unit (GRU) and long short-term memory (LSTM), including their bidirectional implementation. These architectures are considered very suitable for HAR based on continuous sequences, as they can directly take the μD sequence as input.

The rest of the chapter is organized as follows. Section 4.2 describes the experimental setup, the collected dataset, and the class set distribution plus their separability, followed by Section 4.3 showing the multilateration tracking approach. Section 4.4 presents the proposed fusion schemes for distributed radar sensors. The evaluation metrics are introduced in Section 4.5 and afterward the hyperparameter tuning for the used RNN in Section 4.6. The experimental results are presented in Section 4.7, and the final remarks are given in Section 4.8.

4.2. Experimental Setup and Dataset

The section introduces the experimental dataset containing the distributed radar node layout with the statistical characteristic of the unbalanced class set.

4.2.1. Experimental Setup

Five ultra-wide band (UWB) radar nodes by Humatics P410 (former PulsON) are simultaneously employed with coded waveform capabilities minimizing interference between nodes. The radar nodes are deployed in a circular baseline, with a spacing of approximately 45° between them, and cover a surveillance area of about 4.39 m as shown in Fig. 4.1.

4.2.2. Dataset and Class Distribution

Continuous sequences of activities are recorded with 15 participants available in a public dataset[1] [69]. The test and training data procedure is performed by excluding one

[1] Dataset DOI: https://doi.org/10.4121/16691500

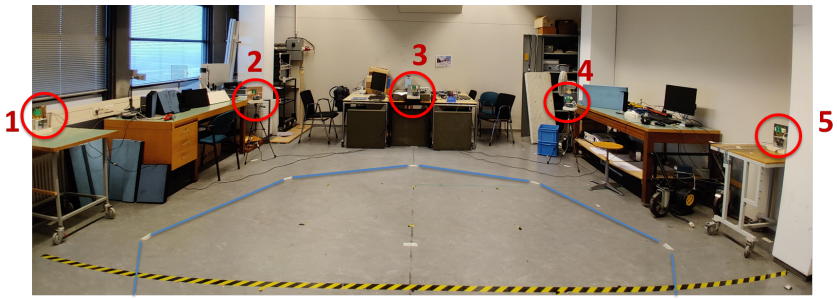


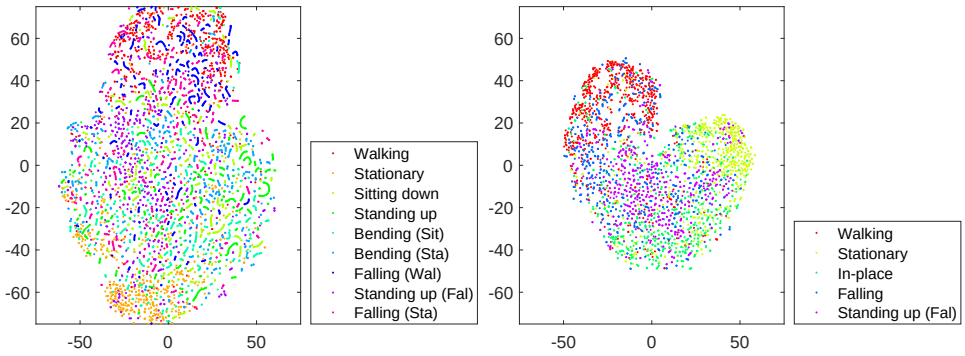
Figure 4.1: Distributed radar network with five radars observing a surveillance area of about 4.39 m in diameter at the MS3 laboratory, TU Delft.

participant from the training data for testing. The procedure is well known as leave-one-person-out (LOPO). For all participants, each collected recording has a total duration of 2 min, and all activities were carried out at predefined locations (sequence type A), and at freely chosen locations within the surveillance area (sequence type B). It should be noted that the participants were free to move in unconstrained directions between performing each activity and to face random directions in terms of aspect angles to the five radars.

Although nine activities were recorded, these were grouped into 5 classes, namely: (I) 'translation' activities (i.e., essentially walking); (II) 'stationary' activity (i.e., essentially the position of standing between two other activities without performing any specific movement); (III) 'in-place' activities (namely, sitting down, standing up from sitting, bending while sitting and standing); (IV) 'falling' (including both falling from standing or walking); and (V) 'standing up from falling'.

This decision in terms of grouping activities together into 'macro-classes' came from limited practical benefits gained by too fine-grained separation of very similar in-place classes such as 'bending from walking' and 'bending from standing', or by distinguishing 'falling from walking' and 'falling from standing'. This also reduces in part the imbalance or skewness of the dataset by grouping together minority classes of activities for which few samples were available, especially the 'in-place' activities. The t-distributed Stochastic Neighbor Embedding (tSNE) representation for the features extracted from the μD domain of the original 9 classes and for the grouped 5 'macro-classes' are presented in Fig. 4.2 after feature scaling is performed.

Furthermore, the pie diagram in Fig. 4.3 illustrates the sample distribution for the collected dataset after the class grouping. As mentioned in the introduction, continuous activities recorded in semi-natural conditions may lead to imbalanced datasets where walking is predominant compared to in-place activities or critical activities such as fall events. This is visualized with the 'walking' class being about 50 % of the dataset. Without resorting to generate synthetic data to correct for the imbalance [131], its effect in performance assessment is addressed by the parameter optimization process shown in Section 4.6, where the dispersion using the standard deviation of the F_1 score across classes has been included in the metrics evaluation process by minimizing the F_1 score fluctuation as shown in Eq. (4.31).



(a) Before class combination (b) After class combination

Figure 4.2: The figure shows the t-SNE distribution of 9 initial classes and the t-SNE distribution after class merging. Therefore, the *in-place* label was assigned to: [sitting down, standing up, and bending] and the *falling* label to: [falling from walking, and falling from standing].

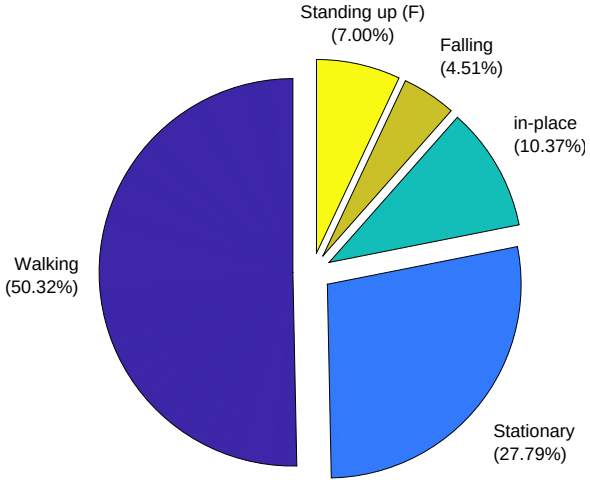


Figure 4.3: The pie diagram shows the imbalance of classes with the minority and majority classes of 'falling' and 'walking' below 5 % and above 50 %, respectively.

4.3. Multilateration and Tracking Approach

This section introduces multilateration-based localization to determine the target's position and is followed by an alpha-beta-(gamma) – α, β, (γ) filter that is applied to estimate the smoothed target's location, velocity, and acceleration. These are then used in the selection of the radar nodes that will be involved in the data fusion process prior to classification. The multilateration processing and the tracking approach with the α, β, (γ) filter can be considered as subsequent processing 'blocks' one after the other. Specifically, while the target's location can be estimated directly from the multilateration processing, the tracking filter is used for smoothing the location and additionally extracting the single target's velocity and acceleration. Although more advanced methods such as various implementations of the Kalman filter may be applied, in this work the simpler α, β, (γ) filter provides satisfactory results for the next stage of selection of radar nodes based on the target's state, which is described in Section 4.4.3.

4.3.1. Multilateration Positioning

Multilateration processing is applied to estimate the target's location in a multi-sensor system. The system is an over-determined equation system, as can be seen in Fig. 4.4. The peak power of the range pulse as in [68] provides the radial range, r_{nx} of the n-th radar with known radar node location (x_{nx}, y_{nx}). The target's estimated position $(\tilde{x}, \tilde{y})$ can be expressed as follows,

$$\begin{cases} (\tilde{x}-x_1)^2+(\tilde{y}-y_1)^2 & = r_1^2 \\ (\tilde{x}-x_2)^2+(\tilde{y}-y_2)^2 & = r_2^2 \\ \vdots \quad\quad \vdots \quad\quad \vdots & \quad\quad \vdots \\ (\tilde{x}-x_{Nx})^2+(\tilde{y}-y_{Nx})^2 & = r_{Nx}^2 \end{cases} \tag{4.1}$$

Eq. (4.1) is linearized for five radars such as,

$$\begin{cases} (\tilde{x}-x_1)^2+(\tilde{y}-y_1)^2-(\tilde{x}-x_5)^2-(\tilde{y}-y_5)^2 & = r_1^2-r_5^2 \\ (\tilde{x}-x_2)^2+(\tilde{y}-y_2)^2-(\tilde{x}-x_5)^2-(\tilde{y}-y_5)^2 & = r_2^2-r_5^2 \\ \vdots \quad\quad\quad \vdots \quad\quad\quad \vdots & \quad\quad \vdots \\ (\tilde{x}-x_4)^2+(\tilde{y}-y_4)^2-(\tilde{x}-x_5)^2-(\tilde{y}-y_5)^2 & = r_4^2-r_5^2 \end{cases} \tag{4.2}$$

Then, Eq. (4.2) can then be rewritten in matrix notation as follows:

$$\mathrm{A} = \begin{bmatrix} 2(x_1-x_5) & 2(y_1-y_5) \\ \vdots & \vdots \\ 2(x_4-x_5) & 2(y_4-y_5) \end{bmatrix} \tag{4.3}$$

$$\mathrm{b} = \begin{bmatrix} x_1^2-x_5^2+y_1^2-y_5^2+r_5^2-r_1^2 \\ \vdots \quad\quad\quad \vdots \quad\quad\quad \vdots \\ x_4^2-x_5^2+y_4^2-y_5^2+r_5^2-r_4^2 \end{bmatrix} \tag{4.4}$$

The over-determined equation system is solved by an ordinary least squares (OLS) estimation that minimizes the error. Thus, the target coordinates can be calculated as follows,

$$\tilde{\mathrm{x}}_n = \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \left(\mathrm{A}^T\mathrm{A}\right)^{-1}\mathrm{A}^T\mathrm{b} \tag{4.5}$$

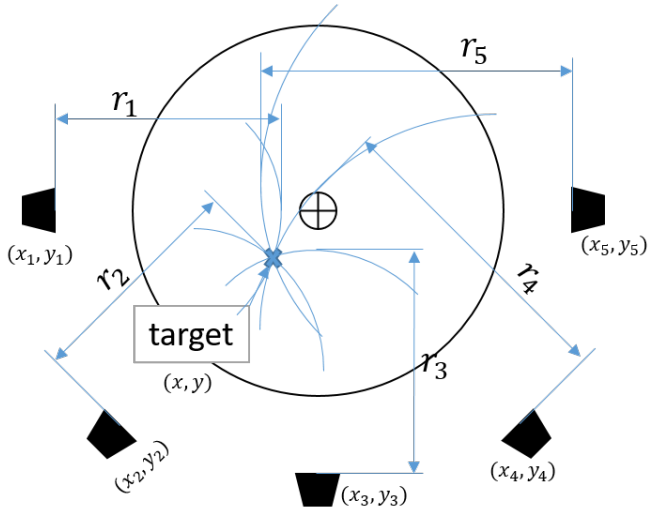


Figure 4.4: The distributed layout for the radar network uses five nodes. The target's location is estimated by the intersections of the determined ranges of each node using multilateration processing.

Table 4.1: The notation of Section 4.3.2 Alpha Beta (Gamma) Filter equations [134].

Δt	Time between instantaneous measurements (referring to PRI)
x_n	Ground truth value
$\tilde{\mathrm{x}}_n$	Measured value at time n
$\hat{\mathrm{x}}_{n,n}$, $\hat{\dot{\mathrm{x}}}_{n,n}$, $\hat{\ddot{\mathrm{x}}}_{n,n}$	Estimate of x, ẋ, ẍ at time n
$\hat{\mathrm{x}}_{n+1,n}$, $\hat{\dot{\mathrm{x}}}_{n+1,n}$, $\hat{\ddot{\mathrm{x}}}_{n+1,n}$	Estimate made at time n of the future state $(n+1)$ of x, ẋ, ẍ
$\hat{\mathrm{x}}_{n,n-1}$, $\hat{\dot{\mathrm{x}}}_{n,n-1}$, $\hat{\ddot{\mathrm{x}}}_{n,n-1}$	Prior prediction at time n, i.e., pred. made at time $(n-1)$ of x, ẋ, ẍ

where $\left(\mathrm{A}^T\mathrm{A}\right)^{-1}\mathrm{A}^T$ is the Moore–Penrose inverse and $(\cdot)_n$ is the discrete-time instant index [128, 132, 133].

4.3.2. Alpha Beta (Gamma) Filter

The output of α, β, (γ) filter is used to estimate location, velocity and acceleration of the single target in the scene, respectively. The tracking filter is jointly used for track smoothing, essential for slow performing movements as those in HAR.

The estimated and unfiltered location is indicated as $\tilde{\mathrm{x}}_n=[\tilde{x}, \tilde{y}]^T$ obtained from Eq. (4.5) in Section 4.3.1. Additional technical insight about the α-β-γ filter is provided by Becker in [134], with notation in Table 4.1.

The filter model is given by,

$$\mathrm{X}_{n+1,n} = \mathrm{F}\mathrm{X}_{n,n} \tag{4.6}$$

with the prediction state $\mathrm{X}_{\{\cdot\},\{\cdot\}}$, while F defines the target kinematic. The kinematic

dependencies are expressed such as,

$$\hat{\mathrm{x}}_{n+1,n} = \hat{\mathrm{x}}_{n,n} + \hat{\dot{\mathrm{x}}}_{n,n}\Delta t + \hat{\ddot{\mathrm{x}}}_{n,n}\frac{\Delta t^2}{2} \tag{4.7a}$$

$$\hat{\dot{\mathrm{x}}}_{n+1,n} = \hat{\dot{\mathrm{x}}}_{\mathrm{n,n}} + \hat{\ddot{\mathrm{x}}}_{n,n}\Delta t \tag{4.7b}$$

$$\hat{\ddot{\mathrm{x}}}_{n+1,n} = \hat{\ddot{\mathrm{x}}}_{n,n} \tag{4.7c}$$

with $\hat{\dot{\mathrm{x}}}$ and $\hat{\ddot{\mathrm{x}}}$ the velocity and acceleration, respectively. The target states in Eq. (4.7) can be formulated in matrix notation such as,

$$\mathrm{X}_{n+1,n} = \begin{bmatrix} \hat{\mathrm{x}}_{n+1,n} \\ \hat{\dot{\mathrm{x}}}_{n+1,n} \\ \hat{\ddot{\mathrm{x}}}_{n+1,n} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathrm{x}}_{n,n} \\ \hat{\dot{\mathrm{x}}}_{n,n} \\ \hat{\ddot{\mathrm{x}}}_{n,n} \end{bmatrix} = \mathrm{F}\mathrm{X}_{n,n} \tag{4.8}$$

The extended version of Eq. (4.8) can be rewritten as,

$$\mathrm{X}_{n+1,n} = \begin{bmatrix} \hat{x}_{n+1,n} \\ \hat{y}_{n+1,n} \\ \hat{\dot{x}}_{n+1,n} \\ \hat{\dot{y}}_{n+1,n} \\ \hat{\ddot{x}}_{n+1,n} \\ \hat{\ddot{y}}_{n+1,n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 & \frac{\Delta t^2}{2} & 0 \\ 0 & 1 & 0 & \Delta t & 0 & \frac{\Delta t^2}{2} \\ 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_{n,n} \\ \hat{y}_{n,n} \\ \hat{\dot{x}}_{n,n} \\ \hat{\dot{y}}_{n,n} \\ \hat{\ddot{x}}_{n,n} \\ \hat{\ddot{y}}_{n,n} \end{bmatrix} \tag{4.9}$$

where the variable $\hat{\mathrm{x}}_{\{\cdot\},\{\cdot\}}$ is the $2x1$ vector $[\hat{x}_{\{\cdot\},\{\cdot\}}, \hat{y}_{\{\cdot\},\{\cdot\}}]^T$, and this applies equivalently for $\hat{\dot{\mathrm{x}}}_{\{\cdot\},\{\cdot\}}$, $\hat{\ddot{\mathrm{x}}}_{\{\cdot\},\{\cdot\}}$. The track update equations enclosing the present measurement, $\tilde{\mathrm{x}}_n$, are given by,

$$\hat{\mathrm{x}}_{n,n} = \hat{\mathrm{x}}_{n,n-1} + \alpha\left(\tilde{\mathrm{x}}_n - \hat{\mathrm{x}}_{n,n-1}\right) \tag{4.10a}$$

$$\hat{\dot{\mathrm{x}}}_{n,n} = \hat{\dot{\mathrm{x}}}_{n,n-1} + \beta\left(\frac{\tilde{\mathrm{x}}_n - \hat{\mathrm{x}}_{n,n-1}}{\Delta t}\right) \tag{4.10b}$$

$$\hat{\ddot{\mathrm{x}}}_{n,n} = \hat{\ddot{\mathrm{x}}}_{n,n-1} + \gamma\left(\frac{\tilde{\mathrm{x}}_n - \hat{\mathrm{x}}_{n,n-1}}{\frac{1}{2}\Delta t^2}\right) \tag{4.10c}$$

with the (residual) error included as $\tilde{\mathrm{x}}_n - \hat{\mathrm{x}}_{n,n-1}$. Δt is the time between instantaneous measurements and refers to the PRI. The location, velocity and acceleration parameters of the tracking filter, α, β, and γ, respectively, are empirically found [135]. The prediction states of the tracking filter are used for the selection of radar nodes and subsequent data fusion, as described in Section 4.4.3.

4.4. Proposed Radar Data Fusion Approaches

This section presents the proposed signal and feature fusion approaches for the network of five UWB radars for HAR. The notation in this section is defined as in Table 4.2, and the length of the feature vector obtained from one μD spectrogram is defined by k with

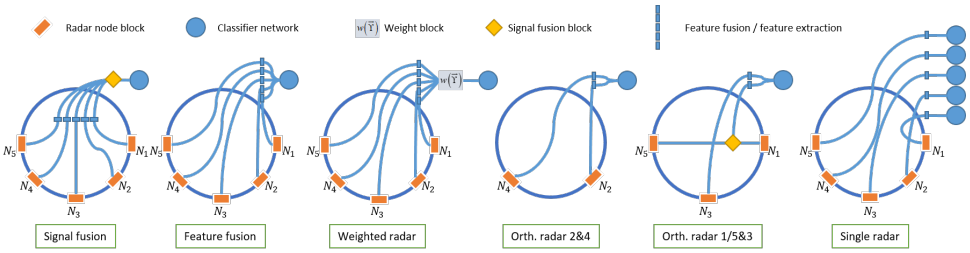


Figure 4.5: Sketches of the proposed fusion approaches of 'signal fusion', 'feature fusion', 'weighted radar selection', and 'orthogonal radar selection' (in two forms) compared with 'single radar' classification applied for HAR in a distributed radar network.

4

Table 4.2: Notation for radar data fusion approaches in Section 4.4.

k / K	individual / total feature vector length
nx / Nx	individual / total radar nodes with {#1, … , #5}
χ_{nx} / X	individual / combined range-time (RT) map
ψ_{nx} / Ψ	individual / combined µD spectrogram feature map
m / t	RT map fast-time index, and slow-time index
m' / t'	µD spectrogram frequency bin –, and the slow-time index

feature scaling performed. Furthermore, Fig. 4.5 shows schematically the proposed fusion approaches presented in this work and compared to the case where a single radar sensor is used.

4.4.1. Signal Level Fusion

The received radar echoes in fast-time provide the target's radial range. The main lobe is typically associated with the target's position and the sidelobes defining the noise floor, assuming sufficient SNR conditions. The PulsON P410 radar nodes are incoherently synchronized, referring to the *Class II* of the *Classification Scheme for Distributed Radar Systems* by Gottinger et al. [70], with additional information provided in Section 2.2. Consequently, summing up the range-time (RT) matrices of all radar nodes yields an incoherent signal level fusion method as,

$$X(m,t) = \frac{1}{Nx} \sum_{nx=\#1}^{Nx} \chi(m,t)^{(nx)} \tag{4.11}$$

The obtained RT matrix, $X(m,t)$, contains information from all radar nodes, as shown in the pipeline in Fig. 4.6. This resulting matrix is further used to calculate a µD spectrogram to be fed as input to the classifier of choice after feature scaling.

The STFT is applied on the RT, X(m,t), as in [82]. The resulting µD spectrogram, Ψ(m',t'), contains the Doppler/velocity information of the target from all nodes, where m' refers to the frequency bins of the µD spectrogram and t' indicates the slow-time bins, respectively. A variety of short-time Fourier transform (STFT) window sizes and overlap

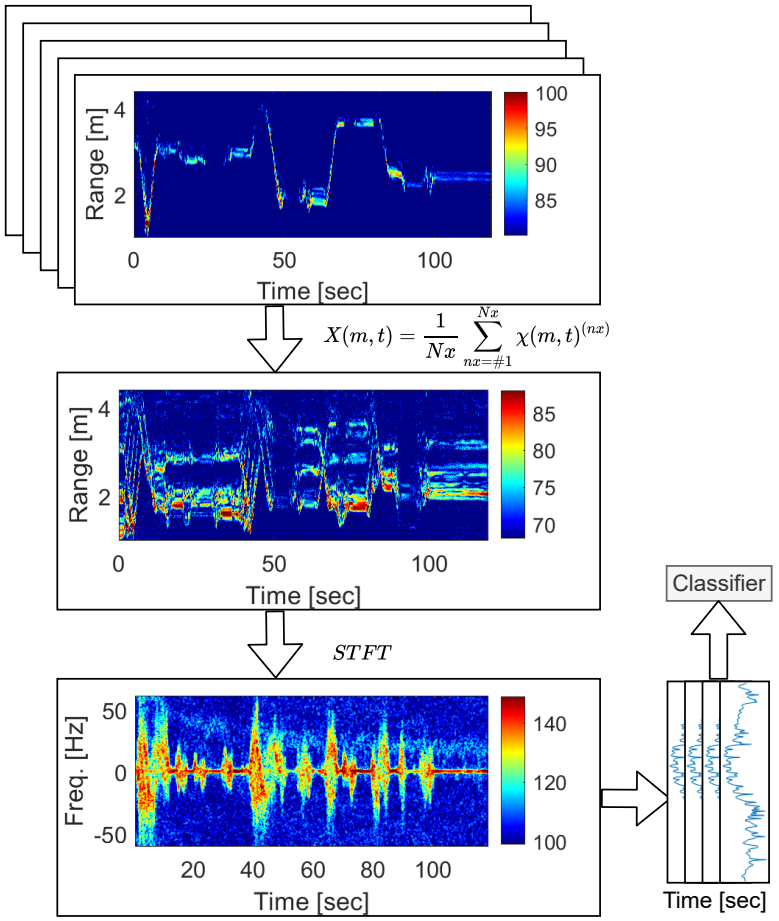


Figure 4.6: Pipeline of incoherent signal fusion, from individual range-time (RT) maps to one µD spectrogram used for the recurrent neural network (RNN) classifier.

values were tested for the best performance between clutter suppression and clarity of limb motions. Clutter cancelation is performed by subtracting the average Doppler frequencies from the µD spectrogram, with a satisfying classification achieved with the STFT overlap of 10 samples (82 ms → t'), and a Hanning window size of 150 samples (1.23 s) [68, 136].

Afterwards, the proposed method uses directly the slow-time bins of the resulting µD spectrogram as feature vectors for classification based on recurring neural networks (RNN) [51], rather than using sliding windows or other techniques to segment the flow of continuous activities.

4.4.2. Feature Level Fusion

Feature level fusion is applied on the individual μD spectrograms, indicated as ψ_{nx}(m',t'). Subsequently, the STFT is applied on the individual RT map of each radar node, $\chi_{nx}(m,t)$, as computed in [82], where $(\cdot)_{nx}$ indicates the radar. Feature fusion concatenates the individual μD spectrograms from each radar node as,

$$\Psi(\text{m'},\text{t'}) = \left[\psi_1(\text{m'},\text{t'})^T, \psi_{nx}(\text{m'},\text{t'})^T, \dots, \psi_{Nx}(\text{m'},\text{t'})^T\right]^T \tag{4.12}$$

with, t', the slow time bins of the μD spectrogram. It should be noted that the slow-time sampling t of the RT map, $\chi_{nx}(m,t)$, depends on the PRF of the radars, whereas the slow-time sample t' of the μD spectrogram, ψ_{nx}(m',t'), depends on the STFT window and overlap parameters.

Both the proposed feature fusion and the signal level fusion presented in Section 4.4.1 use information from all the radar nodes in the network. However, with feature fusion, the length of the resulting feature vector is $Nx \times k$, thus Nx times compared to signal level fusion (where in our case Nx is equal to five, the total number of radar nodes in the network). Hence, an advantage of signal level fusion may be the dimensionality reduction of the resulting feature vector.

4.4.3. Dynamic Weighted Radar Selection

As discussed in the previous section, feature fusion increases the dimensionality of the feature vector by the number of radar nodes used. Hence, there could be an advantage in reducing the number of radars used in the fusion process by only keeping those providing relevant information for classification. In the proposed weighted radar selection, this process is not done once and for all but dynamically adjusted on the basis of the target behavior. The position and movement of the target are used to select the most suitable radar with respect to (w.r.t.) the proposed weight function in Eqs. (4.13) to (4.15).

First, the prediction states of the alpha beta (gamma) – $\alpha, \beta, (\gamma)$ filter from Section 4.3.2 provides the location, velocity, and acceleration of the target $[\hat{\text{x}}_{n+1,n}, \hat{\dot{\text{x}}}_{n+1,n}, \hat{\ddot{\text{x}}}_{n+1,n}]$. As the position of the radar nodes $\text{x}_{nx}=[x_{nx}, y_{nx}]^T$ is assumed to be known, the Euclidean distance computes the length of the vector differences and defines the target's distance to each node, such as $\|\vec{\eta}_{nx}\|_{l_2}=\|\text{x}_{nx}-\hat{\text{x}}_{n+1,n}\|_{l_2}$. It should be noted that the discrete time index, $(\cdot)_{n+1,n}$, will be neglected for convenience to improve the readability.

Then, a weighting function can be computed as,

$$\text{w}(\eta) = \frac{\frac{1}{\|\vec{\eta}_{nx}\|_{l_2}^4}}{\sum_{nx=\#1}^{Nx} \frac{1}{\|\vec{\eta}_{nx}\|_{l_2}^4}} = \frac{1}{\sum_{nx=\#1}^{Nx} \|\vec{\eta}_{nx}\|_{l_2}^{-4}} \begin{bmatrix} \frac{1}{\|\vec{\eta}_1\|_{l_2}^4} \\ \frac{1}{\|\vec{\eta}_2\|_{l_2}^4} \\ \vdots \\ \frac{1}{\|\vec{\eta}_{Nx}\|_{l_2}^4} \end{bmatrix} \tag{4.13}$$

where the term $1/(\cdot)^4$ is related to the SNR distance-power relationship inspired by the radar equation [5]. The denominator, $\sum_{nx=\#1}^{Nx} \|\vec{\eta}_{nx}\|_{l_2}^{-4}$, is a normalization term such that $\sum_{nx=\#1}^{Nx} w_{nx}(\eta)=1$, for $\text{w}(\eta)=[w_1(\eta), w_{nx}(\eta), \dots, w_{Nx}(\eta)]^T$. The radar associated with

the maximum value of the weighting function at a given time is the one selected for the subsequent classification process, typically the one physically closest to the target and typically providing the highest SNR. Only its feature vector is forwarded to the RNN classifier as sketched in Fig. 4.5.

Two further weight functions are formulated that account for the target's aspect angle $cos(\phi_{nx})$, and the radial velocity Υ_{nx}, respectively. As previously shown, the first partial derivative of the predicted state, $\hat{x}_{n+1,n}$ is the velocity. The projection of the velocity onto the line of sight vector, $\vec{\eta}_{nx}$, is the radial velocity to each radar node such as, $\Upsilon_{nx} = \frac{\hat{\dot{x}} \cdot \vec{\eta}_{nx}}{\|\vec{\eta}_{nx}\|_{l_2}^2} \vec{\eta}_{nx}$. The weight function of the person's velocity to the radar nodes is computed similarly to Eq. (4.13) such as,

$$w(\Upsilon) = \frac{\left\|\vec{\Upsilon}_{nx}\right\|_{l_2}}{\sum_{nx=\#1}^{Nx} \left\|\vec{\Upsilon}_{nx}\right\|_{l_2}} \tag{4.14}$$

with $\|\cdot\|_{l_2}$ (the l2-norm) as the absolute velocity w.r.t. the radar nodes.

The target's aspect angle to the radar nodes is determined by the angle between the velocity vector, $\hat{x}_{n+1,n}$, and the line of sight vector, $\vec{\eta}_{nx}$ such as, $cos(\phi_{nx}) = \frac{\hat{\dot{x}} \cdot \vec{\eta}_{nx}}{\|\hat{\dot{x}}_{nx}\|_{l_2} \cdot \|\vec{\eta}_{nx}\|_{l_2}}$. The weight function of the target's aspect angle to the radar nodes is then computed as,

$$w(cos(\phi_{nx})) = \frac{|cos(\phi_{nx})|}{\sum_{nx=\#1}^{Nx} |cos(\phi_{nx})|} \tag{4.15}$$

with the angle term $cos(\phi_{nx})$ ranging between [-1,1] and 0 indicating a tangential movement to the radar node ($|\cdot|$ denotes the absolute value). It should be noted that the aspect angle of facing the radar, $|cos(0)|$=1, as well as facing away, $|cos(2\pi)|$=1, are expected to provide the highest Doppler response and RCS for HAR (i.e., directly facing back or torso of the person). The radar node associated with the maximum value of the weighting function w(Υ) or $w(cos(\phi_{nx}))$ at a given time is the one selected for the subsequent classification process, essentially the one perceiving the highest radial velocity or located at the most superior aspect angle, respectively.

During this research, tests were conducted that combine the aforementioned weight functions as follows: $w = w(\eta) \cdot w(\Upsilon) \cdot w(cos(\phi_{nx}))$. However, these tests resulted in limited performance improvement; thus, the results are not reported in this chapter.

4.4.4. Static Orthogonal Radar Data Fusion

It has been found that classification of unconstrained HAR using a single radar suffers from the lack of recognition capabilities of orthogonal movements, i.e., tangential to the line of sight. Hence, with orthogonal radar fusion a pair of 2 nodes are deliberately chosen in the network with lines of sight along radial and tangential directions. As shown in Fig. 4.5, this orthogonal pair of radars is separated by a quarter circle (90°). The choice of this pair of radars is static, that is, it does not change during the measurement. An example of a

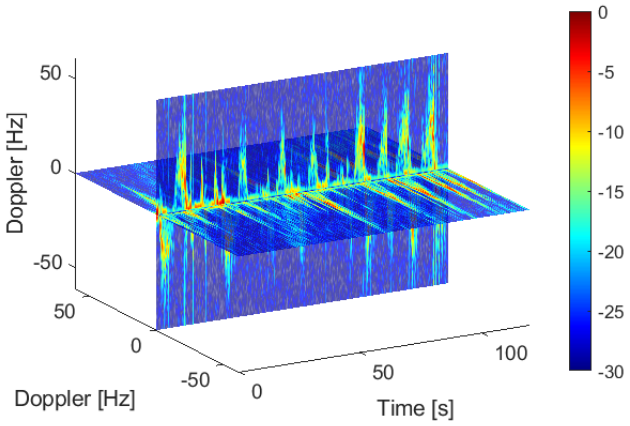


Figure 4.7: The 2D orthogonal spectrogram spanning an x-and y-plane over time domain generating the 2D-µD spectrogram.

resulting 2D-µD spectrogram is shown in Fig. 4.7, with velocity (Doppler) components in principle covering all standard movement directions needed for unconstrained HAR.
Two possible selections are considered:

Orthogonal Radars 2&4 The chosen subset of radar nodes nx={#2, #4} with the RT maps $\chi_{\#2}(m,t)$, and $\chi_{\#4}(m,t)$ are used to extract the µD spectrograms by applying the STFT. The resulting µD spectrograms $\psi_{\#2}$(m',t'), and $\psi_{\#4}$(m',t') are combined through feature fusion as Ψ(m',t')= $\left[\psi_{\#2}(\text{m',t'})^T, \psi_{\#4}(\text{m',t'})^T\right]^T$ and then used for classification. The feature vectors double in size compared to using a single radar or signal level fusion in Section 4.4.1.

Orthogonal Radars 1/5&3 Here the orthogonal radar nodes #1 and #5 are first incoherently fused at the signal level as in Section 4.4.1, as, $X_{\#1,\#5}(m,t) \leftarrow 1/N \sum_{nx=\#1}^{\#5} \chi_{nx}(m,t)$. Then the STFT applied to the concatenated RT domain $\left(X_{\#1,\#5}(m,t)\right)$ obtains the spectrogram as, $\psi_{\#1,\#5}$(m',t') $\leftarrow$ STFT $\left(X_{\#1,\#5}(m,t)\right)$. Finally, feature fusion is performed together with node #3 needed to generate the 2D-µD spectrogram of the target of interest such as $\Psi_{\#1,\#5,\#3}$(2m',t')= $\left[\psi_{\#1,\#5}(\text{m',t'})^T, \psi_{\#3}(\text{m',t'})^T\right]^T$. Thus, the concatenated features are used for subsequent classification with a feature vector length of 2×k:=2×len(m').

4.4.5. Summary of Radar Data Fusion Approaches

This section introduced different fusion methods for combining data collected by distributed radar nodes. Low-level *signal fusion* combines directly the RT data of different radars and generates a single µD spectrogram for classification, hence a feature vector of size equivalent to only using one radar. *Feature fusion* calculates and concatenates the spectrograms from different nodes, thus enlarging the size of the feature vector. Rather

Table 4.3: The overview of the fusion methods for distributed radar the relative feature vector length together with the author's objective notes.

Fusion method	Feature size	Notes
Signal fusion	k	Best case reference
Feature fusion	5×k	Longest feature vector
Weighted radar fusion	k	Tracker needed
Orth. fusion {#2,#4}	2×k	Not all nodes included
Orth. fusion {#1,#5,#3}	2×k	Not all nodes included
Single radar	k	Not suitable for unconstrained HAR

than selecting all nodes at the same time for feature fusion, only pairs of *orthogonal radars* have been considered. Furthermore, a weighted radar selection over time has been formulated, whereby the most suitable radar node is selected at any given time based on a criterion mapped to a weighting function (i.e. the closest radar to the target, the one with the most favorable aspect angle, or velocity presented). The proposed methods are summarized in Table 4.3, and their performances will be compared with the use of a single radar for classification, as if there is no distributed radar network.

4.5. Performance Metrics

Table 4.4: Notation for metric definitions of Section 4.5.

y, A / $\hat{y}$, $\hat{A}$	ground truth / predicted label / area
$\overline{y}$ / $\overline{\hat{y}}$	mean ground truth / mean predicted samples
s / $\hat{s}$	ground truth / predicted block
$(\cdot)_p$, P	sample, set of samples
$(\cdot)^{(c)}$	class index (later neglected for readability)
$tp; tn; fp; fn$	true/false positive/negative rate

In this section, evaluation metrics for continuous and unconstrained HAR introduced in our earlier work [130] are described in detail using the notation in Table 4.4. These metrics will be used in the following sections for performance assessment.

4.5.1. Accuracy

To compare predictions and ground truth labels the identity function for each class c $I^{(c)}(\hat{y}_p, y_p)$ is introduced to measure incorrect predictions:

$$I^{(c)}(\hat{y_p}, y_p) = \begin{cases} 0 \leftarrow \hat{y}_p = y_p \\ 1 \leftarrow \hat{y}_p \neq y_p \end{cases} \tag{4.16}$$

The number of misclassifications is provided by,

$$M^{(c)} = \frac{1}{P}\sum_{p=1}^{P} I^{(c)}(\hat{y}_p, y_p) \tag{4.17}$$

with the resulting accuracy being equal to:

$$A^{(c)} = 1 - M^{(c)} \tag{4.18}$$

When evaluating classification performances, the accuracy metric does not capture inequalities of false negative (fn) and false positive (fp), and does not account for imbalanced datasets. This may lead to overlooked performance drops [115].

4.5.2. F-score with Precision, Recall and Specificity

The F_β score (F_1) provides a more concise metric accounting for fn and fp imbalances, and consists of a combination of precision and recall. Together with precision, recall, and F_β score (F_1), the specificity can be also computed as,

$$precision = \frac{tp}{tp + fp} \tag{4.19a}$$

$$TPR = sensitivity = recall = \frac{tp}{tp + fn} \tag{4.19b}$$

$$TNR = \text{specificity} = \frac{tn}{tn + fp} \tag{4.19c}$$

$$F_\beta{}^{(c)} = (1 + \beta^2) \times \frac{precision \times recall}{\beta^2 \cdot precision + recall} \tag{4.19d}$$

In the formula for the F_β score, precision and recall are evenly treated if $\beta = 1$, known as the F_1 score. Otherwise, the formula favors precision if $\beta > 1$ [137].

4.5.3. Dice Index

The Dice similarity index (also named the Sørensen-Dice coefficient) normalizes the length of the vector labels $\hat{y}$ and ground truth y, and divides them by the total number of non-zero entries. Multiplication by a factor of 2 scales the measurement range between $[0, 1]$ and 1 which means label vectors identical to the ground truth [138]. It is expressed as,

$$Dice^{(c)} = 2 \times \frac{|\hat{A} \cap A|}{|\hat{A}| + |A|} = \frac{2tp}{2tp + fp + fn} \tag{4.20}$$

4.5.4. Jaccard Index

The Jaccard index or the Tanimoto coefficient defines the intersection divided by the union of two label vectors.

$$Jac^{(c)} = \frac{|\hat{A} \cap A|}{|\hat{A}| + |A| - |\hat{A} \cap A|} = \frac{tp}{tp + fp + fn} \tag{4.21}$$

It should be noted that the denominator denotes the union as $|\hat{A}| + |A| - |\hat{A} \cap A| = |\hat{A} \cup A|$.

4.5.5. Consecutive Block Detection (CBD)

This proposed metric considers and penalizes interruptions and misalignments in the sequence of predicted samples, $\hat{y}_p$, with respect to the corresponding ground truth labels, y_p. To the best of our knowledge, this aspect is not always well considered in the literature when evaluating radar-based HAR for continuous activities.

Unweighted Consecutive Block Detection (CBD)

Firstly, the individual ground truth blocks and the prediction blocks are counted as shown for the ground truth in Eq. (4.22) and the predictions in Eq. (4.23), respectively, as,

$$s\left(y_p\right) = \frac{1}{2}\sum_{p=2}^{P-1}\sqrt{\left(y_p - y_{p-1}\right)^2} \tag{4.22}$$

and

$$\hat{s}\left(\hat{y}_p\right) = \frac{1}{2}\sum_{p=2}^{P-1}\sqrt{\left(\hat{y}\left(\hat{y}|y_p = 1\right)_p - \hat{y}\left(\hat{y}|y_p = 1\right)_{p-1}\right)^2} \tag{4.23}$$

with the counter index, $(\cdot)_p$, in the sequence of a total length, P. The ratio of blocks, as shown in Fig. 4.8, is computed as,

$$Ed^{(c)} = \frac{s\left(y\right)}{\hat{s}\left(\hat{y}\right)} \tag{4.24}$$

with the range between $[0, 1]$, where 1 indicates the same number of blocks found within the ground truth sequence of a class and the prediction. It should be noted that block length differences are not considered in Eqs. (4.22) to (4.24), and this can affect the result with a solution provided in the following Section 4.5.5.

Weighted Consecutive Block Detection (CBD)

Due to the aforementioned concern of the block length differences, a corresponding penalty factor can be computed as,

$$w = \sqrt{\frac{\left|\hat{A} \cap A\right|}{|A|}} \tag{4.25}$$

with the numerator indicating the intersection between the ground truth and the prediction, $\left|\hat{A} \cap A\right|$, over the ground truth, $|A|$. The non-linearity impact of the weight, $w = \sqrt{(\cdot)}$, is introduced to minimize penalization on small misalignments. The weighted Consecutive Block Detection (CBD) is then computed by combining Eqs. (4.24) and (4.25):

$$Ed_w{}^{(c)} = Ed \cdot w = \frac{s\left(y\right)}{\hat{s}\left(\hat{y}\right)} \cdot \sqrt{\frac{\left|\hat{A} \cap A\right|}{|A|}} \tag{4.26}$$

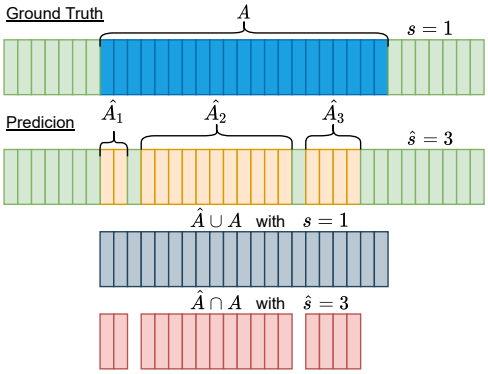


Figure 4.8: The intersection and union sequences are demonstrated and are used for CBD and IoU, as well as, the penalization term of IoU as, $H(\hat{s}, s) = 3/4$, for this example.

4.5.6. Intersection-Over-Union (IoU)

IoU is another metric that penalizes interruptions and misalignments in the prediction sequences to the ground truth sequence. It is a known technique for evaluating camera-based object detection algorithms and is, under certain conditions, equivalent to the Jaccard index. It defines the similarity of bounding boxes [139], which are generally uninterrupted entities in vision-based detection methods. A modified expression can account for interruptions in vectors of labels as,

$$H(\hat{s}, s) = 1 - \left(\frac{2 \cdot \hat{s}}{\hat{s} + s} - 1 \right)^2 \tag{4.27}$$

with s and $\hat{s}$ the concatenated sequence blocks for ground truth and predictions, respectively:

$$\begin{aligned} \mathrm{IoU}^{(\mathrm{c})} &= Jac \cdot H(\hat{s}, s) \\ &= \left(\frac{\hat{A} \cap A}{\hat{A} \cup A} \right) \cdot \left(1 - \left(\frac{2 \cdot \hat{s}}{\hat{s} + s} - 1 \right)^2 \right) \end{aligned} \tag{4.28}$$

Eq. (4.28) penalizes interrupted sequences even if the predictions are broadly correct and aligned with the ground truth [50].

4.5.7. Correlation Index or Matthews Correlation Coefficient (MCC)

The correlation index or Matthews correlation coefficient (MCC) is less commonly used for classification performance assessment. It is based on the Pearson correlation coefficient, typically used to find linear similarities between vectors. This can also be used for sequence

classification as,

$$R^{(c)}(\hat{y},y)=\frac{\sum_{p=1}^{P}\left(y_p-\bar{y}\right)\left(\hat{y}_p-\overline{\hat{y}}\right)}{\sqrt{\sum_{p=1}^{P}\left(y_p-\bar{y}\right)^2\sum_{p=1}^{P}\left(\hat{y}_p-\overline{\hat{y}}\right)^2}},\ R\in\mathbb{R};[-1,1] \tag{4.29}$$

with $\overline{\hat{y}}$ and $\overline{y}$ the means of the ground truth and prediction vector, respectively. Alternatively, the equation can be expressed as,

$$R^{(c)}=\frac{tp\cdot tn+fp\cdot fn}{\sqrt{(tp+fp)\cdot(tp+fn)\cdot(tn+fp)\cdot(tn+fn)}} \tag{4.30}$$

and is known as Matthews Correlation Coefficient (MCC) [140]. It should be noted that $R(\hat{y},y)=-1$ is equivalent to perfectly misclassified sequence and to a perfect classified sequence with $R(\hat{y},y)=1$, respectively, while $R(\hat{y},y)=0$ is the expected value of an unbiased "coin tossing classifier" for a balanced dataset.

4.6. RNN and Hyperparameters Optimization

Recurrent neural networks (RNNs), characterized by their recurrent layers comprising neurons with connections that retain information about past states, are particularly advantageous in capturing the temporal dynamics of input sequences. In various real-life scenarios involving audio, speech, language, or videos with temporal contexts, the expectation is to generate output values that are correlated with the history of corresponding inputs. However, traditional classifiers (e.g., SVM, KNN) and several deep learning networks (e.g., AlexNet, ResNet, MLP) lack the ability to incorporate temporal information as they operate in a stateless manner. Consequently, researchers have turned their attention to RNNs, specifically the widely recognized architectures such as long short-term memory (LSTM) or gated recurrent unit (GRU), to address this limitation [141, 142]. Gradually, these architectures have gained interest in applications like continuous HAR, particularly in the context of continuously recorded radar data [52, 92]. However, the application of RNNs in distributed radar networks of the scale presented in our work has been scarcely explored in the existing literature.

Therefore, with the aim of accounting for temporal relations in the continuous data sequence structure, the following RNNs were chosen in this work:

- GRU Gated Recurrent Unit
- Bi-GRU Bidirectional Gated Recurrent Unit
- LSTM Long Short-Term Memory
- Bi-LSTM Bidirectional Long Short-Term Memory

Bi-RNN (Bi-LSTM or Bi-GRU) networks have received attention due to their enhanced capabilities over standard RNN (LSTM/GRU) networks. One of the primary distinctions of a Bi-RNN is its ability to process sequence data bidirectionally, assimilating the context from both preceding and subsequent data points in a sequence. This characteristic often results in greater accuracy and increased robustness in tasks such

Table 4.5: Range of hyperparameters and their best performing values after Bayesian optimization

Descr.	Range	Bi-LSTM	LSTM	Bi-GRU	GRU	Default
optimizer	ADAM RMSprop SGDM	best performing: ADAM				–/–
epochs	$[10, 50]$	flatten growth after 45				–/–
hidden unit	$[3, 200]$	168	60	199	189	–/–
Mini batch	$[32, 128]$	converging →128				128
initial LR	$[10^{-4}, 0.1]$	0.0076	0.0029	0.0015	0.0025	0.0010
LR drop fac.	$[0.1, 0.5]$	0.2484	0.1361	0.1778	0.1039	0.1000
L2 reg.	$[10^{-7}, 1]$	0.0001	0.0001	0.0003	0.0039	0.0001

as sequence data processing and classification. Moreover, the bidirectional nature equips Bi-RNN with the ability to capture asymmetric dependencies in sequences and address long-term dependencies more effectively than standard RNNs. Additionally, by processing data from both directions, Bi-RNNs can extract a more comprehensive set of features, thereby constructing a more robust model. In the context of continuous text classification, a Bi-RNN was exploited by Zhang et al. [143]. Their findings underscored the superiority of Bi-LSTM over conventional LSTM algorithms in specific text classification scenarios.

Their key hyperparameters are optimized using the experiment manager of MATLAB with Bayesian optimization, as this is the key to improving performances in classification tasks [144]. The following hyperparameters are considered: number of epochs, hidden units, mini batch size, initial Learning Rate (LR), LR drop factor, and the L2 regularization factor [145]. The optimization ranges are mentioned, and the optimal parameters are logged for each network and presented in Table 4.5. The epochs of the optimization algorithm have been set to 50 for each RNN network. Hyperparameter optimization is evaluated using the following metric:

$$\max_{\forall \text{Opti. Parameter}} \left(\frac{F_{1\,macro}}{\left(1 + std(F_{1\,macro})\right)^2} \right) \tag{4.31}$$

with $F_{1\,macro}$ being the macro F_1 score, that is the mean of the F_1 score for the individual classes. The denominator, $(1 + std(F_{1\,macro}))^2$, ensures that the fluctuation across the classes is minimized, while the numerator strives towards a high macro F_1 score. Furthermore, the 3 optimizers stochastic gradient descent momentum (SGDM), root mean square propagation (RMSProp), and adaptive moment estimation (ADAM) were tested, with the best results achieved using ADAM across all networks. Specifically, the output function called stopIfAccuracyNotImproving() [146] was applied, which stops the training of the network if the classification accuracy of the validation data (or other metrics, i.e., those presented in Eq. (4.31)) stagnates for a defined period of epochs. The stopping function in combination with suitable hyperparameter tuning within the parameters' limits, shown in Table 4.5, is meant to optimize the network and avoid a high variance (overfitting) of the network.

Table 4.6: Classification performance metrics for the proposed radar data fusion methods using a Bi-LSTM classifier with the default hyperparameters from Table 4.5. For comparison, the performance for average classification using only one single radar is shown. Note: The bold methods, 'signal-', 'feature-', 'weighted radar s.t. w (η)', both 'orthogonal fusion' concepts, and the 'single radar average' are used to compute the mean and the standard deviation shown in the last tow rows of this table.

Method / Metric	**Accuracy**	**F1 score**	**TPR**	**TNR**	**Dice**	**Jaccard**	**CBD**	**Weight CBD**	**IoU**	**Corr.**
Signal fusion Bi-LSTM	0.933	0.84	0.693	0.95	0.722	0.583	0.829	0.686	0.566	0.68
Feature fusion Bi-LSTM	0.924	0.814	0.63	0.941	0.669	0.526	0.774	0.615	0.5	0.625
Weighted radar Bi-LSTM s.t. w (η)	0.92	0.803	0.623	0.94	0.654	0.51	0.725	0.573	0.463	0.604
Weighted radar Bi-LSTM s.t. w (Υ)	0.916	0.793	0.61	0.938	0.634	0.493	0.746	0.589	0.477	0.584
Weighted radar Bi-LSTM s.t. w $(cos\,(\phi_{nx}))$	0.887	0.719	0.5	0.914	0.514	0.383	0.582	0.42	0.293	0.436
Orthogonal radar 1/5&3 Bi-LSTM	0.925	0.815	0.625	0.942	0.665	0.523	0.782	0.618	0.5	0.626
Orthogonal radar 2&4 Bi-LSTM	0.919	0.796	0.603	0.939	0.635	0.495	0.729	0.569	0.463	0.589
Single radar average Bi-LSTM	0.909	0.773	0.566	0.93	0.599	0.456	0.687	0.521	0.436	0.543
Mean across all fusion methods	0.922	0.807	0.623	0.94	0.657	0.516	0.754	0.597	0.488	0.611
Standard dev. across all fusion methods	0.008	0.023	0.042	0.006	0.041	0.042	0.051	0.056	0.045	0.045

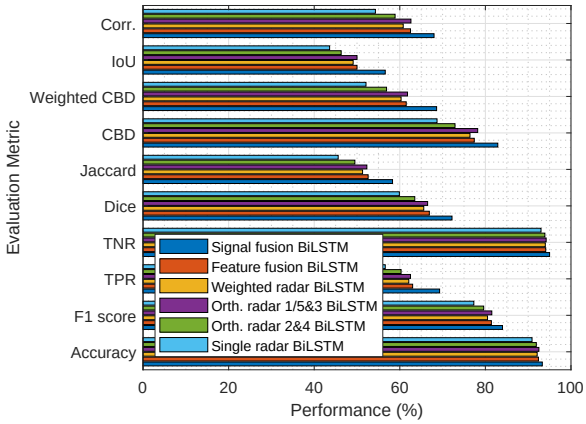


Figure 4.9: Overview of classification performance metrics for all investigated radar data fusion methods in leave-one-person-out (LOPO) test with Bi-LSTM network with default hyperparameters as classifier.

4.7. Experimental Results

In this section, the classification results for the four selected recurrent neural networks and proposed data fusion methods for multiple radar nodes are reported.

4.7.1. RNN Results Utilizing Sensor Fusion Approaches

The section presents the results achieved by using the previously discussed radar fusion methods, namely: signal fusion, feature fusion with all nodes, weighted radar selection, orthogonal radar fusion of node {#2,#4} and of node {#1,#5,#3}. These results are compared with those obtained when using only one single radar classification with a Bi-LSTM classifier, implemented with the default parameters presented in Table 4.5. The results are reported in Table 4.6 for the LOPO test and visualized in Fig. 4.9. Using the LOPO test data exposes the classifier to unseen data from a different participant. This individual participant may exhibit a distinct activity pattern and motion profile compared to what the classifier was originally trained on, i.e., using a 70%/30% training and validation split.

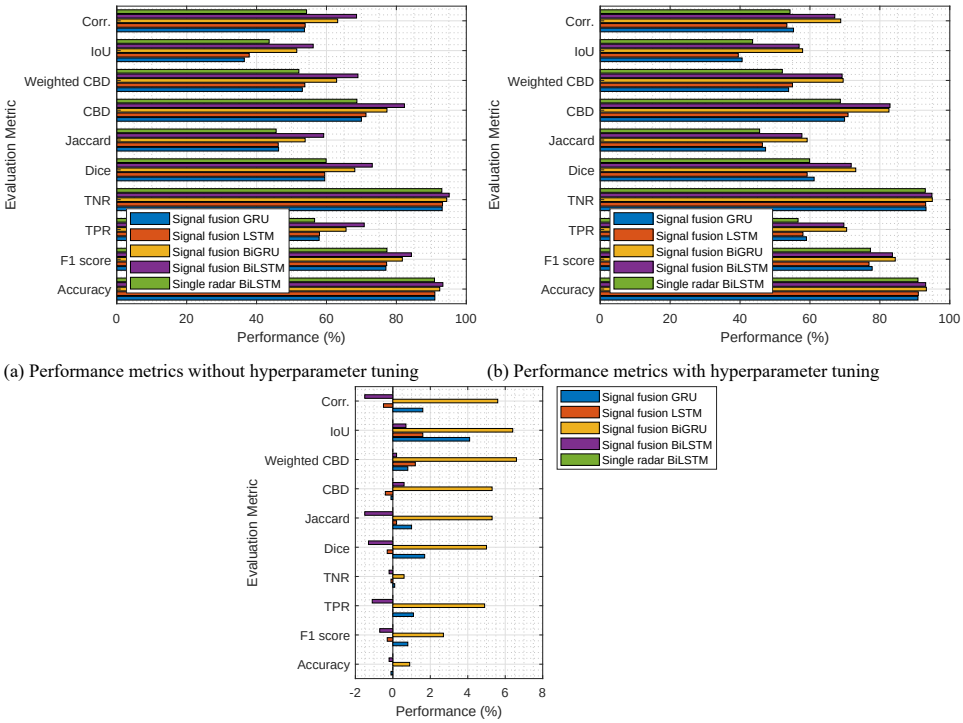
The appropriateness of performance metrics for the shown fusion methods, highlighting the performances of one method over others can be analyzed by evaluating the standard deviation across the tested methods. In this regard, the results show higher dispersion (standard deviation) for metrics other than the more conventional accuracy or F_1 score, such as the Dice, Jaccard, and Correlation (MCC) indexes as block-based metrics, such as CBD and IoU.

Regardless of these metrics, the best suitable fusion method is incoherent signal fusion, followed by feature fusion and orthogonal radar fusion of nodes {#1,#5,#3}, which show almost equivalent performance. Weighted radar selection and orthogonal radar fusion of nodes {#2,#4} perform slightly worse than the other fusion methods. All presented radar data fusion approaches improve their performance in contrast to single radar classification and can be inspected in Fig. 4.9.

Focusing on block-based evaluation metrics, such as the CBD, weighted CBD, and the

IoU, the low-level signal fusion approach appears to be the best performing approach with a performance gain of more than +7 % over the mean across all methods by analyzing IoU and even more than +8 % by using the weighted CBD. Performance-wise, feature fusion approaches follow with about +1 % better performance compared to the mean across all fusion methods. Specifically, feature fusion, weighted radar selection, and orthogonal radar fusion of nodes {#1,#5,#3} achieve an average IoU of approximately 0.5 (50 %), whereas single radar classification achieves 0.436 only.

4.7.2. RNN Results with Hyperparameter Tuning



(a) Performance metrics without hyperparameter tuning

(b) Performance metrics with hyperparameter tuning

(c) Performance improvement (positive values refer to performance improvements after hyperparameter tuning)

Figure 4.10: Overview of performance classification metrics for different RNNs without and with hyperparameter tuning (from Table 4.7). Signal level fusion is considered in leave-one-person-out (LOPO) test.

The section presents performance results achieved with different RNNs as classifiers when incoherent signal fusion is applied, as this was shown to be the best fusion approach. Table 4.7 shows the results using default parameters in the rows indicated as $[\cdot]^*$, whereas the bold rows marked as $[\cdot]^\#$ show the obtained results after hyperparameter tuning. In all cases, the LOPO approach meant that one test sequence was excluded while training the classifier, and this is repeated and averaged across all participants. Moreover, these results are also summarized in the bar graphs in Fig. 4.10, focusing on performance improvements due to hyperparameter tuning, as shown in Fig. 4.10c.

Table 4.7: Classification performance metrics for different considered RNN classifiers using signal level fusion. Each network results are compared before $[\cdot]^*$, and after $[\cdot]^\#$ hyperparameter tuning. For comparison, the performance for average classification using only one single radar is shown. The mean and standard deviation are shown in the last four rows of this table across all architectures.

Method / Metric	**Accuracy**	**F1 score**	**TPR**	**TNR**	**Dice**	**Jaccard**	**CBD**	**Weight CBD**	**IoU**	**Corr.**
Signal fusion GRU *	(0.91)	(0.77)	(0.579)	(0.931)	(0.595)	(0.463)	(0.7)	(0.531)	(0.365)	(0.537)
Signal fusion GRU#	**0.909**	**0.778**	**0.59**	**0.932**	**0.612**	**0.473**	**0.699**	**0.539**	**0.406**	**0.553**
Signal fusion LSTM*	(0.91)	(0.772)	(0.58)	(0.932)	(0.595)	(0.462)	(0.713)	(0.538)	(0.379)	(0.539)
Signal fusion LSTM#	**0.91**	**0.769**	**0.58**	**0.931**	**0.592**	**0.464**	**0.709**	**0.55**	**0.395**	**0.534**
Signal fusion Bi-GRU*	(0.924)	(0.817)	(0.656)	(0.944)	(0.681)	(0.539)	(0.773)	(0.629)	(0.515)	(0.632)
Signal fusion Bi-GRU#	**0.933**	**0.844**	**0.705**	**0.95**	**0.731**	**0.592**	**0.826**	**0.695**	**0.579**	**0.688**
Signal fusion Bi-LSTM*	(0.933)	(0.843)	(0.708)	(0.951)	(0.731)	(0.592)	(0.823)	(0.69)	(0.562)	(0.686)
Signal fusion Bi-LSTM#	**0.931**	**0.836**	**0.697**	**0.949**	**0.718**	**0.577**	**0.829**	**0.692**	**0.569**	**0.671**
Single radar average Bi-LSTM*	(0.909)	(0.773)	(0.566)	(0.93)	(0.599)	(0.456)	(0.687)	(0.521)	(0.436)	(0.543)
Single radar average Bi-LSTM#	**0.909**	**0.773**	**0.566**	**0.93**	**0.599**	**0.456**	**0.687**	**0.521**	**0.436**	**0.543**
Mean across all*	(0.917)	(0.795)	(0.618)	(0.938)	(0.64)	(0.502)	(0.739)	(0.582)	(0.451)	(0.588)
Mean across all#	**0.92**	**0.806**	**0.638**	**0.94**	**0.661**	**0.522**	**0.76**	**0.611**	**0.488**	**0.609**
Standard deviation across all*	(0.011)	(0.033)	(0.062)	(0.009)	(0.063)	(0.061)	(0.057)	(0.074)	(0.085)	(0.068)
Standard deviation across all#	**0.012**	**0.036**	**0.066**	**0.01**	**0.066**	**0.064**	**0.068**	**0.083**	**0.085**	**0.073**

The accuracy metric provides an inconclusive evaluation of the macro results across the tested network architectures, indicated by a low standard deviation of around 0.012. Thus, the performance improvement after hyperparameter tuning is not too evident when using standard metrics. Slightly better is the F_1 score with about 0.036 (3 %), which shows the gain in performance for the Bi-GRU network with the optimized hyperparameters of Table 4.5. The next metrics, the Dice and Jaccard index, provide even better evidence that the hyperparameter tuning of the RNN yields an improvement in performance of approximately 5 % for the Bi-GRU, whereas the Bi-LSTM dropped slightly. It should be noted that the slight decrease in performance is only given for the sample-based evaluation metrics, such as accuracy, F_1 score, Jaccard, Dice, or Matthews correlation coefficient (MCC), and can be neglected. Instead, block-based evaluation metrics, such as CBD, weighted CBD, and IoU, provide a more reliable evaluation, since interruptions and outliers toward other classes will be taken into account. The results achieved after hyperparameter tuning have increased compared to the default values for all network architectures tested, with the most significant improvement for the Bi-GRU. Nevertheless, Bi-GRU and Bi-LSTM provide nearly comparable results, as shown for the sequence evaluation metrics, i.e., IoU or both CBD metrics. In addition, the weighted CBD and IoU provide the highest standard deviation, indicating superior evaluation capabilities compared to sample-based metrics, i.e., simply accuracy evaluation.

4.7.3. Class Evaluation using Spider Diagram

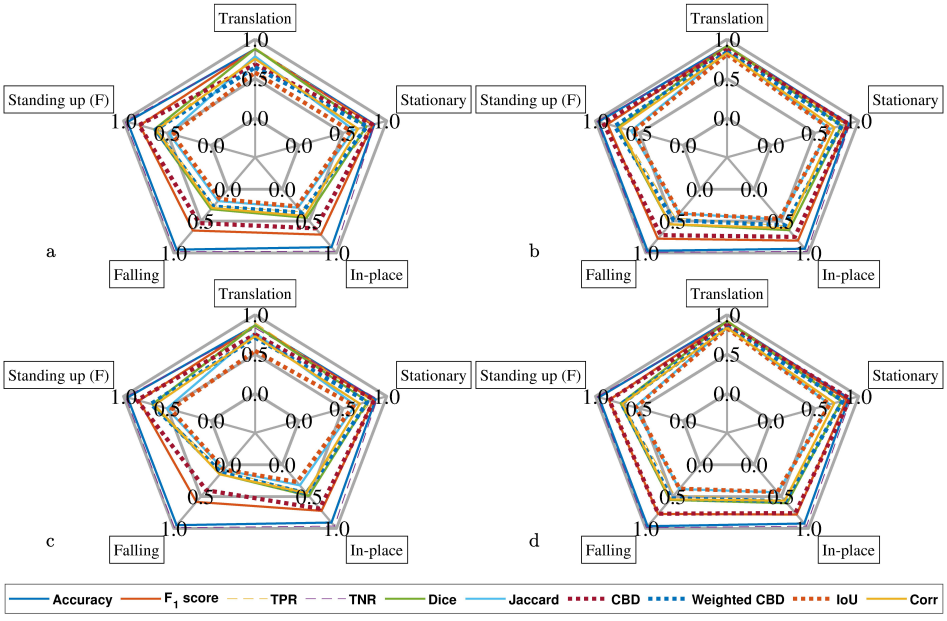


Figure 4.11: Spider graphs with performance metrics addressing the continuous human activity sequence evaluation per class. (a) GRU, (b) Bi-GRU, (c) LSTM, (d) Bi-LSTM

The previous Sections 4.7.1 and 4.7.2 discussed the most suitable multi-sensor fusion

methods and performance improvements by using various RNNs with optimized hyperparameters. However, the discussion and conclusions were drawn on the 'macro' results across all classes of interest. In this section, class specific results ('micro' results) for the individual classes are discussed for incoherent signal fusion using unidirectional and bidirectional RNN architectures, namely, the GRU, LSTM, Bi-GRU, and Bi-LSTM. These results are reported in Fig. 4.11 and discussed in this section for each considered performance metric.

Accuracy

Accuracy appears to be very high for all the classifiers considered (macro accuracy $> 90\,\%$). However, it should be noted that accuracy as a single evaluation does not capture the performances of datasets with class imbalances (e.g., fewer samples of 'falling' compared to other classes). By visual inspection of the other metrics, in addition to accuracy, the GRU classifier in Fig. 4.11a suffers in detecting the falling class in the lower left corner of each spider diagram, whereas both bidirectional classifiers, Bi-GRU (Fig. 4.11b) and Bi-LSTM (Fig. 4.11d), outperform their unidirectional counterparts. This will often remain unnoticed when using the accuracy metric only.

F_1 score, TPR, TNR

Evaluating TPR (sensitivity or recall) and TNR (specificity) on their own is less effective than using the F_1 score, as this can provide a better global view of performance for each specific class. An average of the F_1 score across all classes, macro F_1 score, is also possible. For this case study, the performance differences between individual classes increase to approximately $12\,\%$ for signal fusion using Bi-LSTM, specifically referring to the 'translation' ($91.6\,\%$) and 'standing up from falling' ($78.9\,\%$) activity, as shown in Fig. 4.11d. However, the drastic difference can be seen for unidirectional RNNs, e.g., the LSTM where the differences between 'translation' and 'standing up from falling' rises to more than $20\,\%$.

Dice Index

The Dice index is a more rigorous metric than the accuracy metrics or the F_1 score shown previously. Here, for example, 'standing up from falling' degrades to $58.8\,\%$ (F_1 score: $78.9\,\%$) for the Bi-LSTM. Nevertheless, the bidirectional classifiers perform almost equally high across the individual classes, whereas, i.e., LSTM (Fig. 4.11c) shows a break down by detecting 'falling', which, in general, will be overseen by using the previous more conventional metrics.

Jaccard Index

The Jaccard index is related to the Dice index, with performance always lower than the Dice index, except at their extrema [130]. In fact, this metric will report even lower performances, i.e., 'standing up from falling' degrades to $41.6\,\%$ (Dice index: $58.8\,\%$). In fact, the metric is the most rigorous evaluation method for individual sample detection, apart from block sequence based detection, such as the IoU.

Consecutive Block Detection (CBD)

The CBD operates differently than the metrics shown above. Here, the interruption ratios of the prediction label blocks to the ground truth label blocks have an impact, see Eq. (4.24). In a particular example drawn, the activities 'stationary' and 'in-place' provide the best and worst classification results, respectively, for incoherent fusion using the Bi-LSTM classifier (see Fig. 4.11d) with 88.0 % and 78.3 %. Furthermore, both bidirectional RNNs (Bi-GRU and Bi-LSTM) show good results that are equally distributed across all classes.

However, the simple CBD defined in Section 4.5.5 accounts for the number of detected blocks only, while differences in block length are neglected. Instead, the weighted CBD, as defined in Section 4.5.5, considers the detection length differences of the predictions versus the ground truth labels. Specifically (see Fig. 4.11d for Bi-LSTM), the best and worst classes become the activity 'translation' and 'falling' with 83.5 % and 56.6 %, respectively. Correspondingly, bidirectional RNNs outperform unidirectional RNNs with significant improvement provided for minority classes as 'falling'.

Intersection-Over-Union (IoU)

The IoU metric is an alternative metric that accounts for the detected block ratio and the differences in block length. The IoU is the most extreme evaluation metric for our dataset since it is a product of the Jaccard index (a hard metric on its own) multiplied with a block detection term [50]. Thus, the activity 'standing up from falling' degrades to 41.1 % (Jaccard index: 41.6 %), and the 'translation' activity to 82.4 % (Jaccard index: 84.7 %) as the best class. Likewise, using unidirectional classifiers (LSTM and GRU) has adverse effects specifically on minority classes such as 'standing up from falling', 'falling', and 'in-place' activities.

Correlation Index or Matthews Correlation Coefficient (MCC)

The MCC is rather challenging to compare with the previously introduced metrics due to its diverse definition. An advantage of this metric (MCC) is a distinct indication when a classifier provides an output that results in $R(\hat{y}, y) < 0$. Such results immediately indicate a mismatch between the ground truth and prediction samples. As a side note, it will be mentioned that a coin tossing classifier would converge towards the limit $R(\hat{y}, y) = 0$ for a balanced dataset. For the presented dataset, the activities 'falling' and 'translation' provide the worst and best results with 57.7 % and 83.1 %, respectively. Similarly to the metrics introduced before, the unidirectional classifiers (LSTM and GRU) suffer for minority classes as previously inspected, which can be seen for the macro results in Fig. 4.10 and for the micro results in Fig. 4.11.

4.7.4. Performance w.r.t. Radar Nodes

A Monte Carlo simulation was set up to investigate the effect on the proposed classification performance metrics of selecting a different number of radar nodes. The Monte Carlo approach selects different nodes out of the possible combinations of 2, 3, and 4 nodes, respectively, but in all cases the verification is performed with the L1P0 approach. As shown in Fig. 4.12, the lowest performance for each metric is achieved by using one single node only. This performance degradation is caused by the unfavorable aspect angle when using one single node, and is compensated by considering multiple nodes, which leads to a better performance.

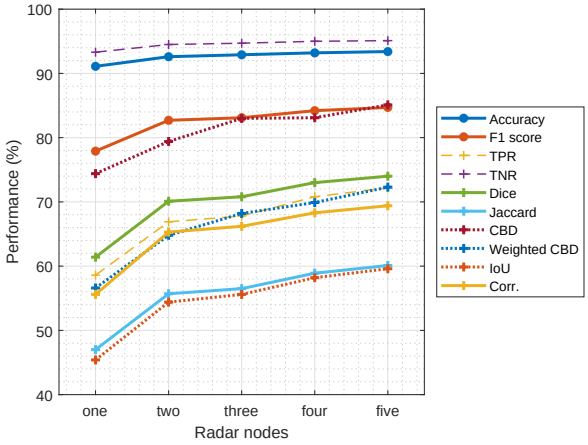


Figure 4.12: Performance metrics as a function of selected number of radar nodes (note that incoherent signal fusion is used when two or more nodes are selected). Training/testing is performed using LOPO and Bi-LSTM classifier.

It is important to note that this performance increase as a function of the number of radar nodes is less pronounced for sample-based metrics like the accuracy metric, or the F_1 score metric (visualized by a flattening curve in Fig. 4.12), but more evident when using block sequence-based metrics such as CBD, weighted CBD, or IoU are used. In the latter case, the increase in performance with the number of radar nodes is more noticeable. From this analysis, it appears that a higher number of radar nodes achieves better performances when using incoherent signal fusion and the Bi-LSTM classifier, specifically for the metrics proposed for continuous sequential activities.

4.7.5. Discussion on Performance Metrics

In the analysis of the results, the pros and cons of each evaluation metric were discussed when applied to the different RNN architectures and radar data fusion methods. Some considerations from this initial analysis follow:

- *For balanced (equally-distributed) data evaluation*: The conventional accuracy metric provides satisfactory results, even if it does not describe where mistakes (e.g., missed detections or false alarms) occur for a given class. For that, precision/recall or their combination into F_1 score is more suitable.

- *For unbalanced (skewed) data evaluation*: The F_1 score becomes a more sophisticated metric than plain accuracy and is widely used. Its importance becomes crucial as the accuracy metric can overestimate the performance of a network, as seen in our case study. The same applies to the Dice and Jaccard indices. Both, Dice and Jaccard, are the most drastic sample-based metrics, as seen in the results from the considered dataset. In addition, the correlation index or Matthews correlation coefficient (MCC) accounts for imbalances in the dataset and it is widely used in the medical domain [50].

- *For the evaluation of continuous sequences of activities*: The previous metrics suffer from evaluating continuous sequences of activities with unconstrained and seamless transitions. Therefore, the proposed CBD is preferable for such cases with its modification such as the weighted CBD and the IoU. These metrics can account for outliers (i.e., misalignments and interruptions) in the prediction label vector and are well suited for HAR based on continuous sequences. When used at the output of RNN classifiers, they can directly assess their sequential output predictions and penalize instabilities/interruptions that propagate errors within the network's memory cells.

4.8. Conclusion

The chapter proposed a variety of approaches for data fusion in a network of five distributed radar sensors in the context of human activities classification. The implementation of signal level fusion applied on range-time (RT) maps has been researched and is compared to state-of-the-art methods. Within the fusion process, the selection of radar nodes based on a weighting function that accounts for the target location and velocity/trajectory has also been investigated. RNN-based classification algorithms, namely GRU, LSTM, Bi-GRU, and Bi-LSTM, were used to process the resulting μD spectrograms derived from the fusion process.

The proposed techniques are evaluated on an experimental dataset with 15 participants and 9 activities, combined into 5 macro-classes. Notably, the dataset contains continuous sequences of activities performed in random locations and with arbitrary and unconstrained trajectories and unfavorable aspect angles to the radar sensors. New evaluation metrics are proposed and compared to account for the specific nature of continuous activities in radar-based HAR, such as the presence of misalignments and interruptions (outliers).

The results emphasize the necessity for metrics beyond traditional accuracy or precision/recall in evaluating continuous HAR. This is particularly relevant when employing recurrent neural networks for classification, assessing performance over short, continuous time intervals, or segments within the sequences. Specifically, evaluation metrics that account for outliers in the prediction vector (i.e., misalignments, interruptions, and fluctuations), such as the weighted CBD and IoU appear to provide a more comprehensive performance evaluation than simple accuracy. For example, while the IoU shows around 20 % difference between imperfect and reasonably performing classifiers, conventional accuracy evaluation gives only 2 % discrepancy, and hence a too coarse assessment. Classifiers with bidirectional capabilities are shown to provide slightly better performance, especially for classes with few samples in the unbalanced dataset used for evaluation. This imbalance is to some extent typical in continuous, realistic activities. For example, there will be more walking samples than in-place activities, whereas critical activities (e.g., falls) will be generally rare. For the radar data fusion, the incoherent signal level fusion of the RT data from each node appeared to outperform other methods, with the best results provided by using the full set of radars in the used network.

Future work can extend the proposed techniques and assess the evaluation metrics in more realistic home environments and scenarios, such as multiple targets (e.g., multiple people or a person including pets). This could be approached by implementing a multi-target tracker with modified hypotheses for a dynamic radar selection in such scenarios of

multiple targets in home environments. Moreover, the usage of various radar sensors in the network could be considered (e.g., adding V/W-band MIMO nodes).

5

Multipath Effect Exploited for Human Activity Recognition

Radar sensor multipath problems, a common phenomenon for indoor employed radar networks and often an unwanted side effect, have been considered in this study. Limited research, mainly focusing on nonextended targets using single radars, has demonstrated the need to use multipaths and boost a radar's network classification capabilities.
In this chapter, a pipeline is proposed from isolating a target's line of sight (LOS) to determining its position and concluding the higher-order multipaths. Furthermore, edge cases are discussed, such as handling extended targets observed over multiple range bins and various aspect angles, even when utilizing only one radar. The oftentimes unseen advantages of the rendered higher-order multipath components are exploited in addition to the LOS observations to leverage the network's perceptions.
The pipeline has been verified by comparing results to only a single radar and the radar network's LOS. For perception tasks, the multipath components with the LOS components are fused before exploiting a 12-layer convolutional neural network (CNN). Its results were compared to a simple multilayer perceptron (MLP) classifier fed with features extracted using principal component analysis (PCA). A significant test performance improvement in the order of +11 % is demonstrated by using a multi-radar network with its LOS and multipaths for continuous and consecutive recorded human activities.

Parts of this chapter are coming from:

R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Multipath Exploitation for Human Activity Recognition using a Radar Network", in *IEEE Transaction on Geoscience and Remote Sensing [under review]*, 2023.

5.1. Introduction to Multipath Effect

Radar sensor networks can increase the efficacy of perception by leveraging diverse observations from multiple radar nodes. Applications of radar networks include nonintrusive monitoring and activity classification, especially of vulnerable individuals, such as falls or other potentially dangerous events among other daily activities [3, 29].

Existing literature on human activity classification using radar measurements primarily characterizes the received signal with micro-Doppler (μD) or range-Doppler (RD) signatures of a radar's line of sight (LOS) signal. In principle, Doppler modulations from micro-motions were found to have a substantial predictive impact on classification performances, with the benefits of using networks over single radar observations [36, 123, 147, 148]. Some studies have also introduced ensemble learning methods using a sensor network with boosted, bagged, and stacked machine learning models exploited [50].

Recent literature has emphasized that continuous sequences naturally occur in human kinematics. Thus, sophisticated classifiers are needed for it, with benefits from the multi-perspective views intrinsically provided by radar networks [48]. Some work even incorporates in the network RF-based illuminators of opportunities widely available in consumer home environments, even if they may be affected by multipath and provide by themselves lower performances compared to their active counterparts [149, 150]. Nonetheless, examples of high-quality μD signatures using passive WiFi in the context of human activity recognition (HAR) and hand gestures are reported in the literature [151].

In the introduced literature on human activity classification, whether using single, monostatic radars or radar networks, multipath is often viewed as an unwanted phenomenon that degrades the system's overall performance. This perspective is echoed in other chapters of this thesis where multipath was not leveraged. However, in certain scenarios, multipath components can enhance indoor localization, as demonstrated for monostatic setups [152]. Inspired by this idea, in this work a novel processing pipeline is introduced to enable the exploitation of multipath components by a radar network in the overall classification process. The rationale of the proposed approach is that the LOS signal (0^{th} component) reflected by a target and its 1^{st} order and 2^{nd} order multipath components are not only capturing the target's signature from different aspect angles, but the components are also coherent with each other as generated by the same radar. This provides additional sources of information to characterize the target's scattering behavior and movement pattern, as if the multipath components were additional physical radar sensors observing the target from different perspectives. Utilized in the context of a physical radar network with different nodes, the proposed pipeline allows it to potentially 'augment' the network by adding to the classification process related to each physical radar node, which would be normally discarded in conventional processing.

The proposed pipeline consists in short of the following steps, which will be detailed in the following sections:

- LOS determination via a hierarchical clustering algorithm to deal with the extended nature of the human target due to the high spatial resolution of the radars used. This enables one to isolate and track the LOS from the multipath components in the range-time (RT) data.
- Determination of the location of the target via multilateration, as well as the location

of the generated multipath components as the target moves.

- Extraction of features from range-Doppler (RD) images of each 'data domain', i.e., the LOS and the 1st order and the 2nd order multipath components generated for each radar in the network.
- Machine learning classifiers trained and tested on data for verification, which also includes the upsampling of the data using synthetic minority over-sampling technique (SMOTE) [153, 154].
- Rigorous leave-one-person-out (LOPO) test for final verification of the classifiers' performance.

More specifically, the proposed pipeline has been verified with experimental data collected with a network of 3 pulsed radar nodes. The data collection included 6 activities performed by 14 volunteers, used to compare different classification algorithms, namely support vector machine (SVM), multilayer perceptron (MLP), and convolutional neural network (CNN). It is shown that F_1 score classification performances can be increased by +11 % (with respect to using a monostatic radar node) thanks to the inclusion of data and features retrieved from the multipath components using the proposed processing pipeline.

The rest of the chapter is organized as follows. The proposed pipeline is discussed in Section 5.2 with its signal processing architecture, target tracking, and ML models. The dataset with the participant's statistics and the obtained results can be found in Section 5.3. Finally, conclusions and comments for future work are given in Section 5.4.

5.2. Proposed Pipeline for Multipath Exploitation for HAR

Multipath is often considered as an undesired effect to remove in indoor radar applications, including human activity classification. However, research in wireless communication has demonstrated the benefits of using multipath [152, 155] for improved localization. Thus, a novel processing pipeline is proposed to locate multipath components related to human movement in an indoor setting and exploit them in the classification process for human activity recognition. The rationality of using multipath components for classification is that they capture the target's signature from different spatial perspectives (aspect angle), and are coherent with the main LOS signal as generated by the same radar. This allows us to consider the signature in the multipath components as if generated by additional physical radar nodes in a network, thus enabling the exploitation of additional and diverse data for classification. An essential task in the proposed pipeline is the identification of the LOS and multipath components from the range-time (RT) data and their tracking over time while the human target in the scene moves. The proposed processing pipeline is shown in Fig. 5.1 and is described in the rest of this section.

Nevertheless, it presents challenges with, first, an a priori isolation of the LOS signal required to distinguish the target's location with discarded multipath. Separation of the LOS signal by means of multilateration processing, as introduced in the following sections. Furthermore, Fig. 5.1 illustrates the intricacy of the methodologies needed to accurately

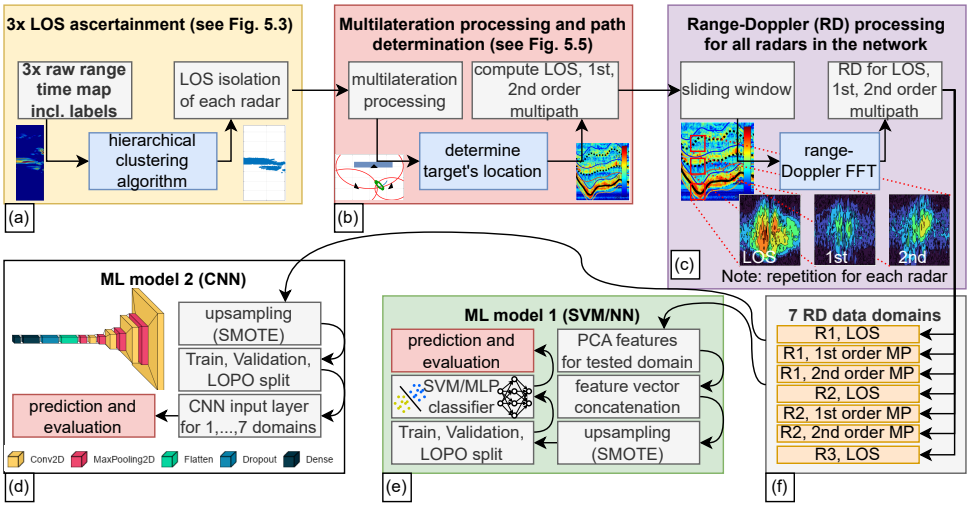


Figure 5.1: Proposed processing pipeline for multipath exploitation in human activity classification with two classifiers tested. The blocks of *LOS ascertainment*, and *Multilateration processing & path determination* are explained in detail in Figs. 5.3 and 5.5, respectively.

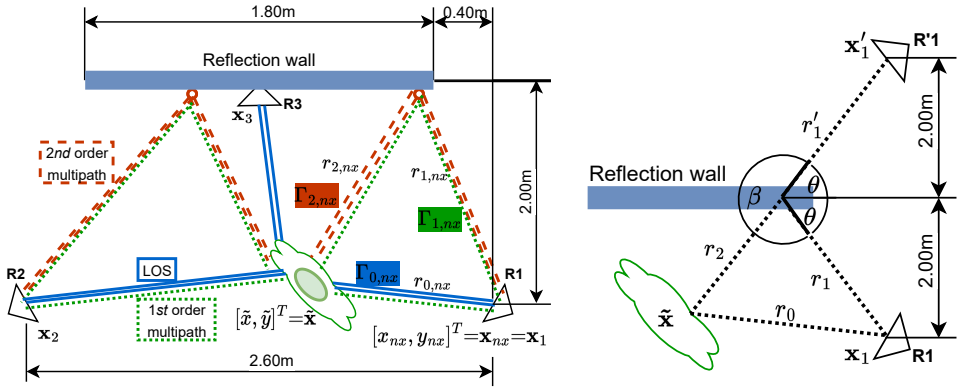
define a target's higher-order multipaths. Given that the full range provided by each radar is typically not utilized for classification, there is no static control over the signal's pathways.

In fact, utilizing the entire range of each radar arbitrarily may lead the network into a high variance situation (i.e., overfitting to those data). A glance at Fig. 5.2b provides insight into this. Hypothetically shifting a target closer to the reflection wall would result in a notable change in its signature, presenting challenges for feature distributions and subsequent classification. Analyzing the variances within the RD map reveals the significant impact of such geometrical changes in the scenario on the expected classification performances, even for identical activities. Hence, the necessity for a distinct pathway separation becomes evident, rather than utilizing the entire range data in the classification process.

5.2.1. LOS Isolation from Multipath Components

In this section, we introduce the proposed method for isolating the LOS from higher-order multipath components, emphasizing that these multipaths typically follow a non-Gaussian distribution.

Guided by the pipeline Fig. 5.1 and the layout Fig. 5.2, the person's precise location must be determined to conclude the pathways. Therefore, a hierarchical clustering algorithm (also: agglomerative clustering) [156] was used to exclude a radar-target's LOS. Numerous alternatives up to a multiple target tracking approach [5] even for distinct propagation paths [157] were considered. However, the applied hierarchical clustering method has shown superior results for an extended target, as in the used case, with no specified number of clusters required and feasible for arbitrary shapes. An optimization method was created with an R^2-score evaluation metric used to minimize the error of the



(a) Layout (b) Fictive mirroring

Figure 5.2: (a) Sketch of the room layout with 3 radars of which two, R1 and R2, can benefit from multipath. R3 is placed at the reflection wall and provides the range to compute the target's location $\tilde{\mathbf{x}}$ using trilateration. (b) Geometrical relationship for computing the travel paths needed for the multipath ranges, r_1, r_1', r_2.

5

hierarchical clustering method by finding the optimal parameters. Alternatively, complex methods, such as HOTA (higher order metric for evaluating multi-object tracking), might equitably assess the algorithm's performance [158].

Mini Optimization Problem Isolating LOS

The data recorded from two radar nodes (Fig. 5.2a) provide the person's LOS and higher-order multipath components. Higher-order components are considered 1^{st}, 2^{nd}, and higher-order multipaths. In turn, multipath is a time-delayed signal return, and a reiterative signal of the same target at a far-off location. In most cases, the multipath signal, which is generally the backscatter from a different target's reflection perspective, captures the target from a different angle of arrival (AoA).

Problematically as seen in the range-time (RT) M map in Fig. 5.3, the target's backscatter does not follow a normal Gaussian distributed $\mathcal{N}(\mu, \sigma^2)$ with mean, μ, and variance, σ^2, because of intentional multipaths. Hence, ordinary filters, such as the standard Kalman filter, failed due to the multiple pathways (in a case study tested as constant-velocity and constant-acceleration filters). Therefore, a problem of isolating the most dominant pathway was formulated. This dominant pathway is, in theory, the closest path and must be the person's LOS path. Then, the target's LOS range of each radar determines the person's location thanks to the synchronized multi-radar setup, while trilateration processing resolves the unknown position.

Hierarchical Clustering Algorithm Hierarchical clustering requires the creation of clusters from the initial RT map, a matrix $\mathrm{M}=\mathrm{X}_{LOS}+\mathrm{X}_{multipaths}+\mathrm{N}$ including the signals X and the noise term N. Therefore as a first step, the k-strongest range bins of the RT map

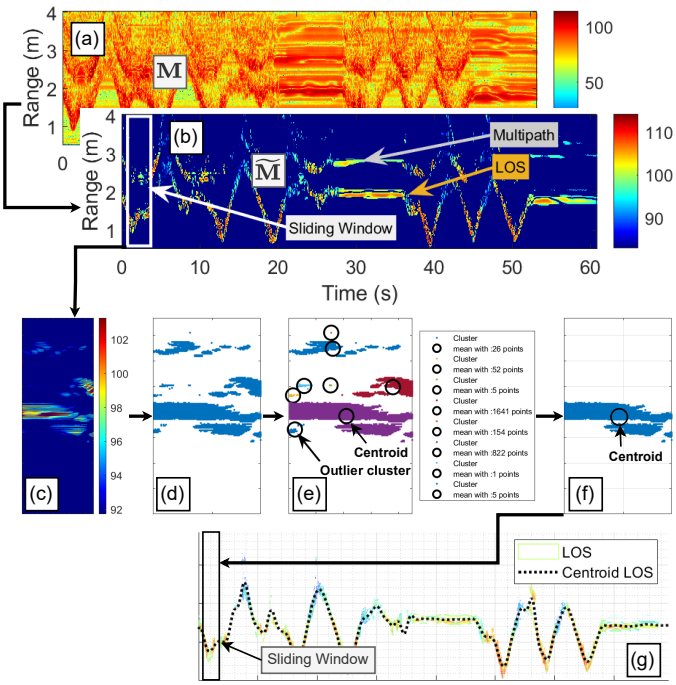


Figure 5.3: The flowgraph with insights from Fig. 5.1a. It is shown in (a) the initial range-time (RT) map including LOS plus multipath components, followed by (b) the selection of the k-strongest range bins, (c-f) the clustering steps to separate the LOS cluster from noise and multipath, and finally (g) the new RT map with the LOS signal including its centroid.

are selected as,

$$\widetilde{\mathrm{M}}_{ij} = \begin{cases} \arg\max\limits_{k \,\forall\, i} \mathrm{M}_{ij}, & k\text{-strongest} \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

with, i and j the row and column elements, respectively, of the matrices M_{ij}, forming the sparse matrix $\widetilde{\mathrm{M}}_{ij}$ with mainly LOS components and some fractions of multipath remaining, as shown in Fig. 5.3.

Sliding slow-time windowing is performed with a window length of $3k$, with k the parameter Eq. (5.1) denoted as k-strongest. It is the empirically optimal sample number mitigating arbitrary connections between LOS and multipath clusters, as in Fig. 5.3 with $3k$ being 90 slow-time sample bins ($\approx$0.74 s).

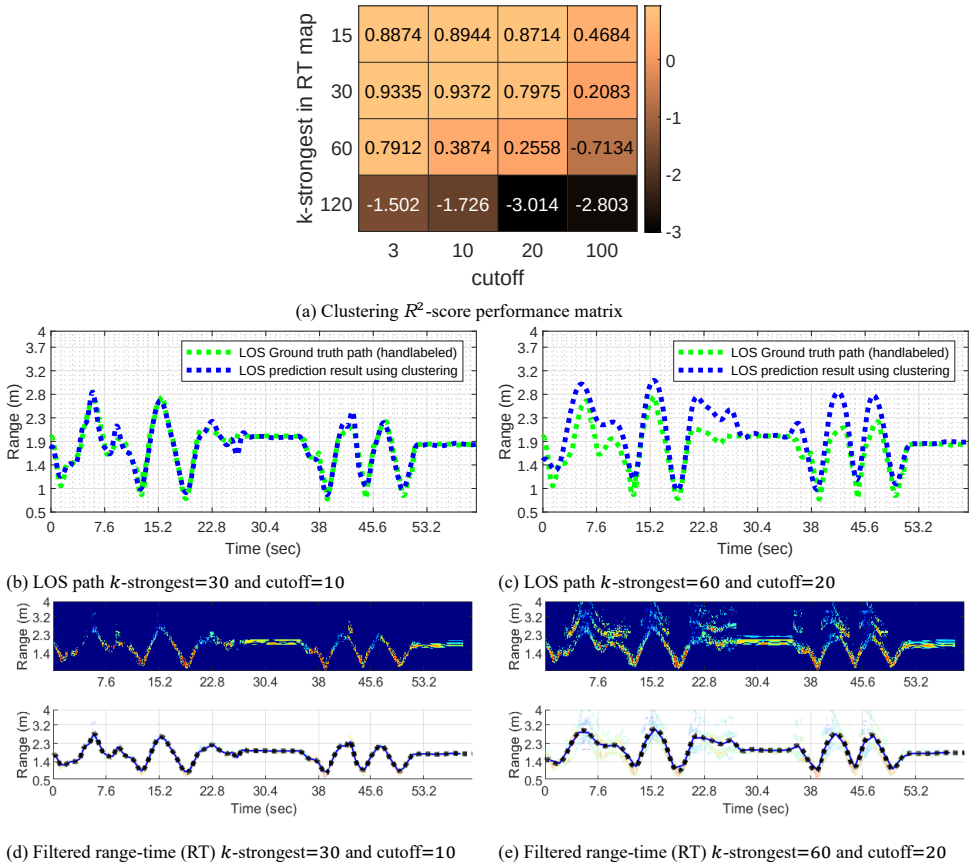
The hierarchical clustering (also: agglomerative clustering) method is applied (Fig. 5.3e) with the following advantages:

- Requires specified number of clusters: No
- Arbitrarily shaped clusters: Yes

The hierarchical structure of the clusters is often visualized using a dendrogram, which specifies the distances between clusters and shows how they merge at each level [159]. The cutoff parameter, one of the two optimization parameters, provides the distances between

the clusters, with the Euclidean distance linkage method chosen by default [156]. It is noted that myriad cluster algorithms are available, and others may compete with the implemented algorithm.

Furthermore, the clusters found are selected w.r.t. to the range in the RT profile, as shown in Fig. 5.3e, denoting that the closest cluster most likely belongs to the LOS path. An empirical condition is specified excluding smaller cluster clouds, as shown in Fig. 5.3e (*Outlier cluster*). If the condition does not meet, the algorithm selects the next farthest cluster and then moves the sliding window in slow-time. The found cluster with its centroid is recorded and provides the RT map with the LOS signal return only (Fig. 5.3g). Throughout the investigation, the two parameters, the *cutoff distance* and *k-strongest samples*, were defined in the mini optimization problem introduced in the next section to meet the system requirements.



(a) Clustering R^2-score performance matrix

(b) LOS path k-strongest=30 and cutoff=10

(c) LOS path k-strongest=60 and cutoff=20

(d) Filtered range-time (RT) k-strongest=30 and cutoff=10

(e) Filtered range-time (RT) k-strongest=60 and cutoff=20

Figure 5.4: The R^2-scores results using the clustering algorithm are provided in (a). (b) and (d) provide the best results obtained using the optimized parameters with k-strongest=30 and cutoff=10 (R^2-score of 93.72 %). (c) and (e) show an example of non-optimal parameters obtained by k-strongest=60 and cutoff=20 (R^2-score of 25.58 %).

Parameter Optimization for Hierarchical Clustering Algorithm The parameters for the hierarchical clustering algorithm are optimized using the following parameter ranges:

- Cutoff distance: $\in [3,\ 10,\ 20,\ 100]$
- k-strongest obtaining $\widetilde{\mathrm{M}}$: $\in [15,\ 30,\ 60,\ 120]$

The performance metrics R^2 was applied, with the property $\mathbb{R} \in (-\infty, 1]$, a metric typically used to evaluate similarities in regression problems. It is used to optimize $\mathtt{argmax}\ R^2$, which is s.t. cutoff, k-strongest. Therefore, a brute force optimization approach is chosen to tune the parameters for each iteration and monitor its results [m,n]:=[k-strongest, cutoff] of R^2, defined as,

$$\begin{aligned} r_i &= y_i - \hat{y}_i, \forall\ i \in (0, N-1) \\ R^2_{m,n} &= 1 - \frac{\sum_i r_i^2}{\sum_i (y_i - \mathbb{E}[y])^2} \end{aligned} \tag{5.2}$$

with r_i the residual between y_i the ground truth, $\hat{y}_i$ the predicted sample, and $\mathbb{E}[y]$ the expected value (mean of the ground truth) [160]. The *LOS ground truth*, as in Fig. 5.4a, was hand-labeled, and finally, the optimum was found for the k-strongest sample number of 30 and the cutoff distance of 10, as shown in Fig. 5.4a. These parameters are used throughout the following analysis.

Furthermore, two examples are provided with an achieved R^2-score of 93.72 % as for the best case using the parameters of k-strongest=30 and cutoff=10, with the LOS centroid and the cleaned RT map obtained, respectively. The generated results are shown in Figs. 5.4b and 5.4d. A contrast example is illustrated in Figs. 5.4c and 5.4e with a slight off LOS path obtained by the parameters k-strongest=60 and cutoff=20, and an evaluation R^2-score of 25.58 % only.

5.2.2. Motivation of the Multipath Model

After the LOS ranges are determined, multilateration (trilateration) processing was used to discriminate a person's location. The motivation for using multipath provides the following advantages:

- Time-of-flight pulses returning from the (extended)-target at different times, resulting to distinct range cells.
- The dominant range cells of the target's LOS (0^{th} order), 1^{st} , and 2^{nd} order multipaths of one radar are representations of the same target (single target case).
- Each radar's LOS, 1^{st} , and 2^{nd} order multipath capture the target from different aspect angles.

It will be pointed out that, w.r.t. to a radar's resolution and the target size, other competitive investigations may consider an additional tracking filter, as proposed in the previous Section 4.3, whereas it was neglected for that project presented in this chapter since each block increases the complexity and is vulnerable to the pipeline's error propagation. This section provides techniques for exploiting the multipath model, as illustrated in Fig. 5.2.

Multipath Model

The model considers a multipath-assisted localization (MAL) from 0^{th} to 2^{nd} order reflections, as in Fig. 5.2a. A priori knowledge of the reflector's and radar's location is assumed. The MAL 0^{th} order is the LOS seen by the radar, indicated by the one-way range r_0. The dominant multipath reflected on the wall is separated in $r_{1,nx}$ and $r_{2,nx}$. By fictive mirroring the radar on the reflector, the symmetrical equivalence of $r_{1,nx}$ is $r'_{1,nx}$ with the radar location x_{nx} becoming $\mathrm{x'}_{nx}$, as in Fig. 5.2b shown for x_1 and $\mathrm{x'}_1$, respectively. According to the geometrical relations, the three predominant pathways can be determined as,

$$\Gamma_{0,nx} = \frac{2 \cdot r_{0,nx}}{2} = r_{0,nx} \tag{5.3a}$$

$$\Gamma_{1,nx} = \frac{r_{0,nx} + r'_{1,nx} + r_{2,nx}}{2} \tag{5.3b}$$

$$\Gamma_{2,nx} = \frac{2 \cdot \left(r'_{1,nx} + r_{2,nx}\right)}{2} = r'_{1,nx} + r_{2,nx} \tag{5.3c}$$

with Γ half of the roundtrip ranges, equivalent to the target's range in a typical RT maps and the indices $\{\cdot\}_{0,\ldots,2,\{\cdot\}}$ for the 0^{th}, 1^{st}, and 2^{nd} pathway and $\{\cdot\}_{\{\cdot\},nx}$ the radar node [54, 55, 152].

Target's Location Finding

Multilateration processing is applied to estimate the target's location in the radar network. The full derivation can be found in Section 4.3, which solves the system of overdetermined equations and includes graphical illustrations. Hence, the found LOS path Section 5.2.1 provides the radar-target's radial range, $r_{0,nx}$ of the radar node with its location $\mathrm{x}_{nx}=[x_{nx}, y_{nx}]^T$. The target's estimated position is given by, $\tilde{\mathrm{x}}=[\tilde{x}, \tilde{y}]^T$, can be formulated as,

$$\begin{cases} (\tilde{x} - x_1)^2 + (\tilde{y} - y_1)^2 & = r_{0,1}^2 \\ (\tilde{x} - x_2)^2 + (\tilde{y} - y_2)^2 & = r_{0,2}^2 \\ \vdots \quad \vdots \quad \vdots & \quad \vdots \\ (\tilde{x} - x_{Nx})^2 + (\tilde{y} - y_{Nx})^2 & = r_{0,Nx}^2, \text{ with } Nx = 3 \end{cases} \tag{5.4}$$

Eq. (5.4) is linearized and optimized for 3 radars (Nx), as in the used case, such as,

$$\begin{cases} (\tilde{x}-x_1)^2+(\tilde{y}-y_1)^2-(\tilde{x}-x_3)^2-(\tilde{y}-y_3)^2 & = r_{0,1}^2 - r_{0,3}^2 \\ (\tilde{x}-x_2)^2+(\tilde{y}-y_2)^2-(\tilde{x}-x_3)^2-(\tilde{y}-y_3)^2 & = r_{0,2}^2 - r_{0,3}^2 \end{cases} \tag{5.5}$$

For simplicity, Eq. (5.5) can then be rewritten in matrix notation as,

$$\mathrm{A} = \begin{bmatrix} 2(x_1 - x_3) & 2(y_1 - y_3) \\ 2(x_2 - x_3) & 2(y_2 - y_3) \end{bmatrix} \tag{5.6}$$

$$\mathrm{b} = \begin{bmatrix} x_1^2 - x_3^2 + y_1^2 - y_3^2 + r_{0,3}^2 - r_{0,1}^2 \\ x_2^2 - x_3^2 + y_2^2 - y_3^2 + r_{0,3}^2 - r_{0,2}^2 \end{bmatrix} \tag{5.7}$$

The equation system is solved by an ordinary least squares (OLS) estimation. Thus, the target coordinates can be computed as,

$$\tilde{\mathrm{x}} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \left(\mathrm{A}^T\mathrm{A}\right)^{-1}\mathrm{A}^T\mathrm{b} \tag{5.8}$$

where $\left(\mathrm{A}^T\mathrm{A}\right)^{-1}\mathrm{A}^T$ is the Moore–Penrose inverse, and with more insights provided by Zhou et al. [128]. Although the processing pipeline is designed not to have excessive complexity, it can extend its capabilities with an additional filter. Specifically, it should be noted that the pipeline can be extended by a tracking filter as presented in the previous chapter, i.e., the Kalman filter or the alpha-beta-(gamma) – α, β, (γ) tracking filter (with further information: in Sections 4.3 and 4.3.2). Such filters can smooth the target's track and may remove outliers. However, should this extension be pursued, the results would necessitate further validation to ensure their accuracy and reliability.

As the radar's location is known, the Euclidean distance between the dynamic target location and the stationary radar location can be recomputed as, $|\mathrm{x}_{nx}-\tilde{\mathrm{x}}|=\hat{r}_{0,nx}$, with $\hat{r}_{0,nx}$ the LOS distance for the node nx. In theory, $|\hat{r}_{0,nx}-r_{0,nx}|=\epsilon\to 0$ indicates the error of the multilateration processing, and it is assumed to be close to 0 with $\hat{r}_{0,nx}\approx r_{0,nx}$. Then, by substituting the range $r_{0,nx}$ in Eq. (5.3a), the round trip range is computed. It is straightforward, without prior knowledge needed of the multipath's angle of incidence and reflection, the radar's location can hypothetically be mirrored on the reflection wall ($y_{wall}=0$). Mathematically, a reflection on the unit vector $\hat{\mathbf{i}}_x=[1,0]^T$ will be computed, since it is assumed that the reflection wall forms the coordinate system's x-axis at ordinate 0. Thus, the angle θ between the radar location and the unit vector is needed for the rotation matrix $\mathbf{R}(\beta)$ to compute the mirrored radar location x'_{nx} as,

$$cos(\theta) = \frac{\mathrm{x}_{nx}\hat{\mathbf{i}}_x}{|\mathrm{x}_{nx}||\hat{\mathbf{i}}_x|} \to \theta = cos^{-1}\left(\frac{\mathrm{x}_{nx}\hat{\mathbf{i}}_x}{|\mathrm{x}_{nx}||\hat{\mathbf{i}}_x|}\right) \tag{5.9a}$$

$$\beta = 2\,(\pi-\theta) \tag{5.9b}$$

$$\mathbf{R}(\beta) = \begin{bmatrix} cos\beta & -sin\beta \\ sin\beta & cos\beta \end{bmatrix} \tag{5.9c}$$

$$\mathrm{x}'_{nx} = \mathbf{R}(\beta)\mathrm{x}_{nx} = \begin{bmatrix} cos(\beta)x_{nx} - sin(\beta)y_{nx} \\ sin(\beta)x_{nx} + cos(\beta)y_{nx} \end{bmatrix} \tag{5.9d}$$

such as the hypothetical radar location $\mathrm{x}'_{nx}=[x'_{nx}, y'_{nx}]^T$ is obtained. The sum of r'_1+r_2 is then simply the Euclidean distance of the $|\mathrm{x}'_{nx}-\tilde{\mathrm{x}}|=(r'_1+r_2)_{nx}$, and can be substituted in Eqs. (5.3b) and (5.3c), with a graphical illustration shown in Fig. 5.2b. Finally, the LOS, 1st, and 2nd order multipaths of the two radars that observe the reflection wall are determined.

Furthermore, due to the known scattering effects and the extended target concerns of a human individual, a window of 50 cm width–the empirically defined shoulder dimension of a person–was placed around the found centroid ranges and can be observed by the upper and lower bounds of the Figs. 5.5 c, e, and f.

Path Discrimination for Further Processing

The previous Section 5.2.2 introduces the variables $\Gamma_{0,nx}$, $\Gamma_{1,nx}$, and $\Gamma_{2,nx}$ that provide the ranges associated to the 0th, 1st, and 2nd order pathways, respectively. It can be seen in

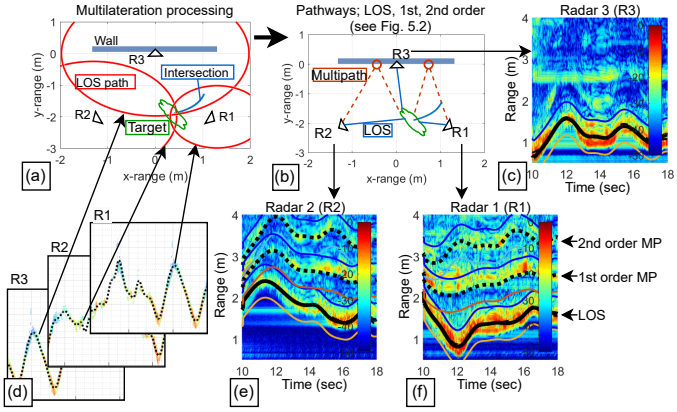


Figure 5.5: Flowgraph with detailed insights on the multilateration approach to identify the location of the target and related multipath components. It shows in (d) the target's LOS from each radar needed for (a) multilateration processing with its (b) LOS and multipaths extracted in the RT images (c,e,f).

5

Fig. 5.5 c, e, and f, the ranges of the pathways change over time, which is the result of the movement of an individual. Therefore, the variable Γ must be computed dynamically for each slow-time hop.

A sliding window with a height of 0.50 cm and a width of 0.82 sec (100 samples) in slow-time is placed at the centroid range determined by Γ, as in Fig. 5.1c shown. Over its window, the 1D-FFT in slow-time dimension computes the range-Doppler (RD) map for seven domains. Finally, all domains are obtained that consider the 0^{th} (LOS), 1^{st} , and 2^{nd} order pathways of radar R1 and R2, respectively, and the LOS domain for R3.

5.2.3. Machine Learning and Preprocessing

The measured radial velocities of the propagating paths are extracted using the RD FFT [5], with seven domains obtained in total. Subsequently, two machine learning (ML) models are used after converting the complex RD maps to image-like data and z-score normalization [161]. The two compared models differ in their structure, with ML model 2 being a 12-layer convolutional neural network (CNN) proposed by Vanschoren [162], which extracts features due to its convolutional layers. ML model 1 requires prior feature extraction, where a simple support vector machine (SVM) classifier with radial basis function (RBF) kernel or a three hidden unit multilayer perceptron (MLP) classifier was tested on the principal component analysis (PCA) feature, with the PCA being a powerful and nonparametric tool used to discard irrelevant dimensions and keep the salient features [163].

After training the models, rigorous evaluation utilizing a leave-one-person-out (LOPO) test assesses the classifier's performance on unseen data. Throughout the project, two types of classification pipelines are considered suitable, illustrated in Fig. 5.1e and 5.1d as ML model 1, and ML model 2, respectively. The ML model 1 pipeline extracts the RD features using PCA, followed by an SVM or MLP classifier. In contrast, the ML Model 2 pipeline operates on the RD domains directly without separate feature extraction required due to its use of a convolutional neural network (CNN) while the network learns its feature pattern.

Data Preprocessing and Upsampling

Regardless of the applied machine learning pipeline, the RD data domains are preprocessed before applying PCA feature extraction or forwarding the images to a CNN. Firstly, the complex RD maps are converted to a logarithmic magnitude scale before resizing each RD domain to a 64x64 image size. Furthermore, image normalization is applied to convert the dynamic range to $\mathbb{R} \in [0, 1]$.

Secondly, it should pointed out that the collected dataset exhibits class imbalances, with the majority class 1–*Walking without an object* and the minority class 4–*Standing up from sitting* comprising 42.4 % and 3.4 % respectively, as depicted in Fig. 5.8b. After examining the method, synthetic minority over-sampling technique (SMOTE) was selected [153]. SMOTE was applied to the feature vectors for ML model 1 and directly to the RD images when using ML model 2. Although, the latter is not the typical scope of the SMOTE algorithm, this proves to be feasible, as demonstrated by Bhattacharya et al. [154] and Reza et al. [164] for CNN architectures. An alternative way can be the use of a generative adversarial network (GAN), as demonstrated on µD spectrogram data [131].

ML Model 1

ML Model 1 consists of the following block operating on the preprocessed RD images:

- PCA feature extraction
- Feature vector concatenation
- SMOTE upsampling (if applied)
- Training, validation, LOPO test split
- Classifier (SVM, or MLP) → prediction, result evaluation

with the flowchart shown in Fig. 5.1e.

Therefore, a prominent feature extraction tool, the principal component analysis (PCA)–a powerful, nonparametric tool used to discard irrelevant dimensions and keep the salient features [163]–was used to obtain human activity patterns from all RD data, domain-independently. It means, PCA is used on the 0^{th} , 1^{st} , and 2^{nd} order domains of all the providing radars. Thus, the covariance matrix is computed as,

$$\mathrm{H} = \frac{1}{I} \sum_{i=0}^{I-1} \left(\mathrm{X}_i - \bar{\mathrm{X}}_i\right)^T \cdot \left(\mathrm{X}_i - \bar{\mathrm{X}}_i\right) \tag{5.10}$$

with X, and $\bar{\mathrm{X}}$ the images and the mean image, respectively, of the dataset, $\in \mathbb{R}^{\eta \times \eta}$ of the dataset size I. Then the eigendecomposition of the covariance matrix H is used to compute the eigenvalues λ_i and the eigenvectors contained in $\Phi = \left[\nu_1, \nu_2, \ldots, \nu_k, \nu_{k+1}, \ldots \nu_\eta\right] = \left[\Phi_k, \nu_{k+1}, \ldots \nu_\eta\right]$. Then the five eigenvectors (k = 5) associated with the highest ordered eigenvalues are selected before projecting the RD images X on Φ_k to compute the feature matrix, $\upsilon \in \mathbb{R}^{\eta \times k}$ used for classification as, $\upsilon = \mathrm{X}\Phi_k$ [37]. The feature matrix υ of each domain is flattened to obtain the vector $\tilde{\upsilon} \in \mathbb{R}^{(\eta k \times 1)}$. Succeeding, the PCA feature extraction is used on the seven domains to obtain the total feature vector as,

$$\Upsilon = \left[\tilde{\upsilon}^T_{R1|0^{\text{th}}}, \tilde{\upsilon}^T_{R1|1^{\text{st}}}, \tilde{\upsilon}^T_{R1|2^{\text{nd}}}, \tilde{\upsilon}^T_{R2|0^{\text{th}}}, \tilde{\upsilon}^T_{R2|1^{\text{st}}}, \tilde{\upsilon}^T_{R2|2^{\text{nd}}}, \tilde{\upsilon}^T_{R3|0^{\text{th}}}\right]^T \tag{5.11}$$

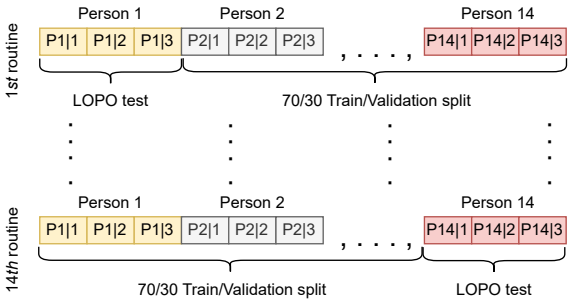


Figure 5.6: Details on the training and validation split after excluding one person's data (3 recorded sequences per person) for the LOPO test. This is repeated 14 times, i.e., for every participant in the dataset, with an average metric calculated across the repetitions.

with $_{R\{\cdot\}|\{Domain\}}$ the indicated radar and domain. The feature vector has the length of $\mathbb{R}^{7\eta k\times 1}$. It will be noted, the feature vector Υ shortens accordingly, if domain selection is used.

After obtaining the feature vector, the SMOTE upsampling technique can be applied using the feature vector, Υ, with its associated label [153]. Regardless of SMOTE, the subsequent step involves the selection of data from one of the 14 participants for the LOPO test. The data of the remaining 13 participants is used for a 70/30 training/validation split. The process is repeated for each participant, as shown in Fig. 5.6.

The training/validation data are forwarded to a SVM classifier with RBF kernel, after training, the LOPO sequence is tested. Alternatively, by using the same data, a simple MLP classifier was trained with 3 hidden units and 8 nodes each using ReLu activation functions and an ADAM optimizer [165]. Both–the SVM and the MLP–are implemented in Python using the scikit-learn environment [166] before the weighted average F_1 score and the accuracy for the validation data and LOPO test are computed.

ML Model 2

The pipeline of ML Model 2 consists of fewer blocks than ML model 1 Section 5.2.3 with:

- SMOTE upsampling (if applied)
- Training, validation, LOPO test split
- CNN classifier → prediction, result evaluation

as the flowchart shows in Fig. 5.1d. Further, the network operates on the RD image data X, and therefore no separate feature extraction is required. Next, a 3*rd*-order tensor is created as the network's input data using X of up to 7 testable domains. As for the tensor, note that its composition can vary between $\in \mathbb{R}^{\eta\times\eta\times\{1,\ldots,7\}}$.

The aim was to use image-like RD domains directly fed into a CNN with few networks tested, and the best results were achieved by a 12-layer network (input layer uncounted) proposed by Vanschoren [162]. The Keras network [167] was modified to suit the input data size with its architecture shown in Table 5.1 and Fig. 5.7.

Simple hyperparameter tuning is performed using ADAM optimizer [165] with a learning rate found to be 0.001 (limits: [0.0001, 0.01]), an early stopping with patience cycle of 10 epochs using the validation loss and a maximum of 250 epochs, whereas the

Table 5.1: Modified CNN architecture proposed by Vanschoren [162] for a 6 class problem with the graphical structure shown in Fig. 5.7. Notes: *Up to seven channels are utilized, contingent upon the number of input data domains available.

Layer type	Input	Output
InputLayer	(64, 64, 7)*	(64, 64, 7)*
Conv2D	(64, 64, 7)	(62, 62, 32)
MaxPooling2D	(62, 62, 32)	(31, 31, 32)
Conv2D	(31, 31, 32)	(29, 29, 64)
MaxPooling2D	(29, 29, 64)	(14, 14, 64)
Conv2D	(14, 14, 64)	(12, 12, 128)
MaxPooling2D	(12, 12, 128)	(6, 6, 128)
Conv2D	(6, 6, 128)	(4, 4, 128)
MaxPooling2D	(4, 4, 128)	(2, 2, 128)
Flatten	(2, 2, 128)	(512)
Dropout	(512)	(512)
Dense	(512)	(512)
Dense	(512)	(6)

network stopped training at around 100 epochs. Additionally, the Dropout layer was set to 0.5 (steps: [0.2, 0.35, 0.5]), with 0.5 as the general maximum probability for a typical network [168].

5.3. Dataset Description and Results

In the subsequent sections, a dataset comprising recordings from 14 participants, obtained through ultra-wideband (UWB) radar nodes, is detailed, emphasizing the dataset's distribution and inherent class imbalances. Methodologies employed are then discussed, with particular attention to the application of upsampling techniques and the innovative utilization of radar's multipath propagation for enhanced classification performance.

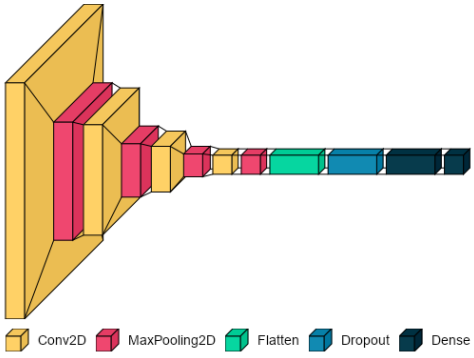
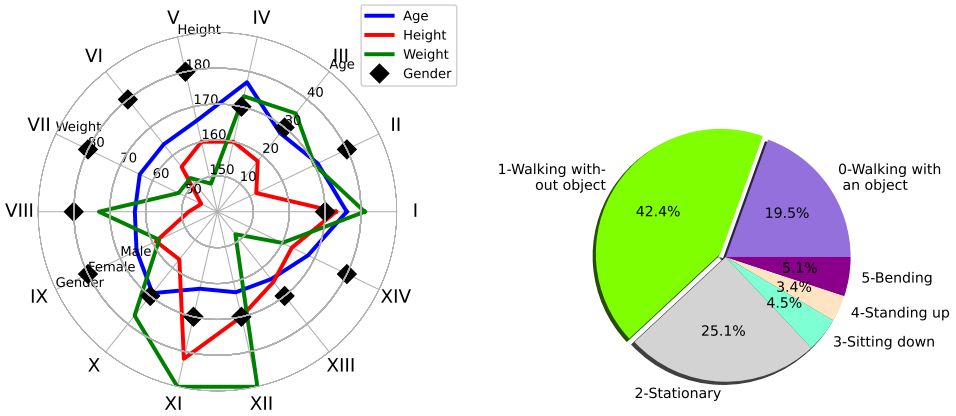


Figure 5.7: Graphical structure for the network proposed by Vanschoren [162].



(a) Metrics for participant I,II,...,XIV

(b) Class distribution within the dataset

Figure 5.8: The metrics for each participant are shown in Fig. 5.8a, and Fig. 5.8b provides the class distribution within the dataset, with the majority class 1–*Walking without an object* and the minority class 4–*Standing up from sitting*.

5

5.3.1. Dataset

Three continuous data sequences with all classes, including of 14 participants, were recorded. The recordings have a length of 1 min, with a gender-balanced occurrence of 7 female and 7 male participants with an average of age: 27.1 years (std: 4.7 years), height: 171 cm (std: 10 cm), weight: 67.6 kg (std: 14.7 kg), with the individual's statistics shown in Fig. 5.8a.

As natural, the dataset is initially unbalanced with the majority class 1–*Walking without an object* and the minority class 4–*Standing up from sitting*, as illustrated in Fig. 5.8b. The data collection was carried out using 3 Humatics P410 UWB radar nodes, which are capable of recording the participants' actions simultaneously. The radars were deployed in a triangle formation as shown in Fig. 5.2a, with more details about the UWB radar provided by Guendel et al. [87] and He [59].

5.3.2. Proposed Methodology and Results

The section discusses, firstly, the impact of upsampling a dataset, a common practice among others addressing unbalanced datasets and its impact on the classification. Secondly, the more novel investigation includes the use of multipath propagations to boost classification performances, with results provided on single radar classification (sR) and using the entire radar network (aR) employing the classification pipelines *Model 1* and *Model 2*.

Upsampling Results

Table 5.2 presents the results using the unbalanced dataset as it originates from the data collection with the majority classes 0–*Walking with an object*, 1–*Walking without an object*, 2–*Stationary condition*, with a graphical illustration about the sample distribution provided in Fig. 5.8b and of similar data from a different data collection in Fig. 4.3.

The investigation showed that unbalanced classes lead to poor prediction, as can be

Table 5.2: The validation accuracy & F_1 score together with LOPO results across 14 participants. The two top rows show results for the initial unbalanced dataset (*Upsampling: None*), and the two bottom rows for a balanced training/validation set using *SMOTE*. The MLP and SVM classifiers were tested using PCA features from all radars and all possible domains, including their multipaths.

Upsampling	Classif.	Domains	Valid. Acc.	LOPO Acc.	Valid. F1-sc.	LOPO F1-sc.
None	MLP	All domains	71 %	52 %	70 %	51 %
None	SVM	All domains	84 %	61 %	85 %	60 %
SMOTE	MLP	All domains	84 %	58 %	84 %	59 %
SMOTE	SVM	All domains	**89** %	**63** %	**89** %	**64** %

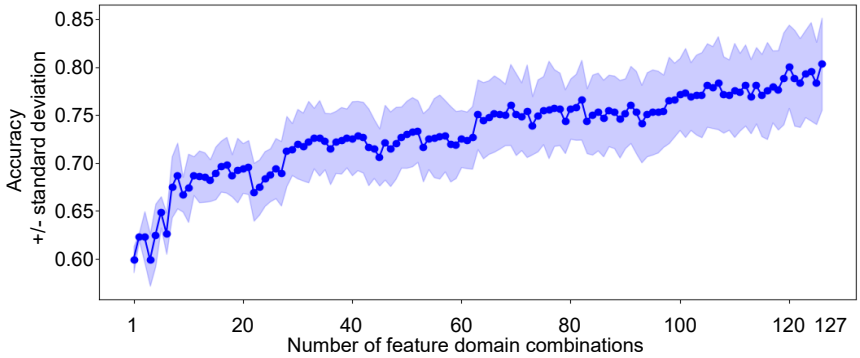


Figure 5.9: Accuracy and standard deviation results from exhaustive feature selection, testing the performance from using each individual domain to combining all 7 domains.

observed from the first rows of Table 5.2. In this regard, Haibo et al. [169] introduced techniques to handle unbalanced classes and analyzes the proposed techniques. Thereafter, throughout the project, the synthetic minority oversampling technique (SMOTE) was used.

The improvement of a balanced training/validation dataset is conclusive, as observed in the last rows of Table 5.2 with +8 % using MLP and +4 % using SVM on the LOPO test with the best results achieved of 64 % using a simple yet effective SVM classifier.

Domain Investigation

Both *Model 1* (PCA+MLP/SVM) pipeline and *Model 2* (CNN) pipeline were tested to exhibit the multipath's impact, with an exhaustive feature selection (EFS) performed using the less computationally complex *Model 1*. As presented, the radar layout can benefit from up to seven domains, consisting of 3xLOS ($3x0^{th}$), $2x1^{st}$ and $2x2^{nd}$ order multipath domains, respectively. Therefore, 127 domain combinations can be selected for classification, with the sum of the binomial coefficient from one domain up to seven domains ($m = 7$), computed using the following equation, $127 = \sum_{i=1}^{m} \binom{m}{i}$.

Table 5.3: Performance metrics using different domains, namely 3 LOS domains and 4 multipath, and their combinations. Performances are compared with using only LOS of single radars, as well as single radars with only their own multipath components. *Classification with a single radar has a high training/validation variance leading to low performances with unseen data, as for LOPO F_1 score.

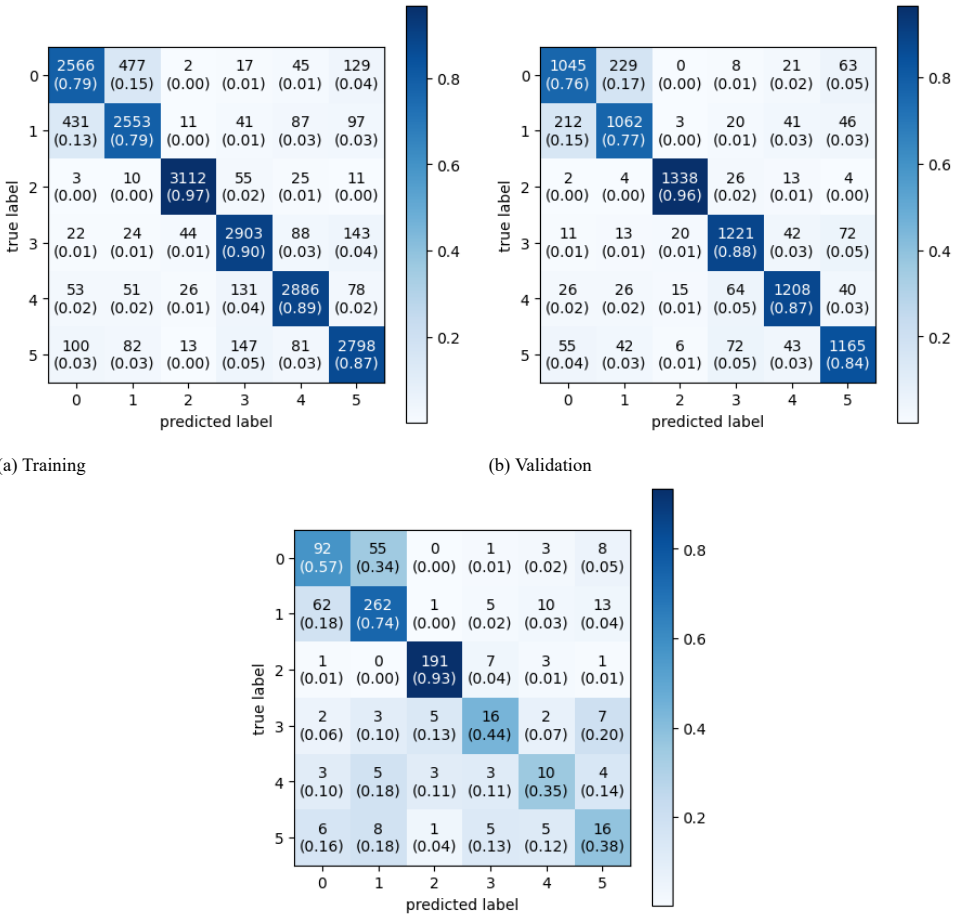
Classifier: CNN	**Upsampling: SMOTE**				**Gain**
Domains	**Valid. Acc.**	**LOPO Acc.**	**Valid. F1-sc.**	**LOPO F1-sc.**	**LOPO F1-sc.**
All domains (aR)	83 %	**83** %	83 %	**70** %	+11 %
All LOS (all radars (aR))	81 %	81 %	81 %	68 %	+9 %
LOS+1^{st}+2^{nd} (sR)	82 %	64 %	82 %	65 %	+6 %
LOS+2^{nd} (sR)	85 %	63 %	85 %	63 %	+4 %
LOS+1^{st} (sR)	87 %	64 %	87 %	64 %	+5 %
LOS (Single radar (sR))	*__90__ %	59 %	*__90__ %	59 %	0 %

Performance gain using Model 1 Exhaustive feature selection (EFS) was carried out using the MLxtend [170] package. The MLP classifier was then applied to the principal component vectors of each domain. By evaluating the accuracy results, the classifier is trained on every domain combination with a 5-fold cross-validation split. The average score with its upper and lower standard deviation is shown in Fig. 5.9, starting from 60 % accuracy for one domain up to 83 % by using all seven domains. In addition, sequential feature selection (SFS) was investigated, and conclusions were drawn equivalently with the best performance achieved by using the full set of domains. It should be noted that, due to the use of the *MLxtend* package [170], an indication of the specific details of each tested instance out of the large combinatorics of the domains observed on the x-axis in Fig. 5.9 cannot be provided. However, it has been shown empirically that the line of sight provides the most relevant information.

Performance gain using Model 2 As introduced in Section 5.2.3, *Model 2* (CNN model) uses the domain inputs without prior feature extraction of up to seven domains and utilizes the RD maps as images. Then, the input size of the network was adjusted according to its number of domains.

Only one input domain is employed to evaluate the system's performance on a single radar (sR) using only the LOS domains with the results provided in Table 5.3 in the last row. It will be noted, the model overfits and is rather incapable of classifying unseen data, as the LOPO F_1 score of 59 % shows. Fortunately, observing the target from different directions using just an sR LOS combined with its 1st order multipath has yielded an LOPO F_1 score of 64 %, a gain of 5 % compared to using only the sR's LOS domain. Using an sR LOS with its 2nd order multipath component results in a score of 63 % and a gain of 4 %. In fact, it is shown that the SISO radar can therefore observe the target from various AoA and boost the classification.

Feeding all domains into the classifier yields the best LOPO test results, averaging 70 % across the 14 tested individuals. The training, validation and LOPO test classification matrices can be seen in Fig. 5.10. Again, the results of the training and



(a) Training

(b) Validation

(c) LOPO Testing

Figure 5.10: The classification metrics of the best performing CNN classifier listed in Table 5.3 training/validation data (Figs. 5.10a and 5.10b) using SMOTE and the LOPO test result (Fig. 5.10c). The average for 14 participants is computed. The listed classes are: 0–*Walking with an object*, 1–*Walking without an object*, 2–*Stationary condition*, 3–*Sitting down*, 4–*Standing up from sitting*, and 5–*Bending from standing*.

validation data are balanced, as observed in the classification matrices Figs. 5.10a and 5.10b, while the result of the LOPO test is Fig. 5.10c unbalanced. Furthermore, the metrics provide the average computed across all participants, with 2–*Stationary condition* the best class detected with low precision and recall. On the contrary, natural confusion is given for 0–*Walking with an object* versus 1–*Walking without an object* by reason of the similarity of its feature distribution. This was observed by analyzing its t-distributed stochastic neighbor embedding (t-SNE) feature distribution [161].

The classes, including 3–*Sitting down*, 4–*Standing up from sitting*, and 5–*Bending from standing*, represent minority classes in the dataset. They exhibit roughly balanced

distributed outliers. This distribution underscores the inherent challenges of many HAR datasets, which often skew towards walking and stationary activities rather than transient actions or short-term actions such as sitting, standing up, or bending.

5.4. Conclusion

A seven-domain classification problem was proposed by utilizing only three radars and deliberately exploiting the radars' multipath. Correspondingly, the performance of including the radars' multipath was evaluated on a single node with +6 % F_1 score improvement for a leave-one-person-out test (LOPO) and even +11 % by testing the entire radar network. Novel investigations show the separation of the line of sight (LOS) return, with a complex clustering pipeline proposed and final multilateration processing performed. The primary aim of this pipeline is to compute the person's location and to exclude the target's multipaths from the LOS for further classification.

Two machine learning models are proposed, the first operating on prior principal component analysis (PCA) feature vectors before a support vector machine (SVM) or multilayer perceptron (MLP) is employed. The second pipeline uses a 12-layer convolutional neural network (CNN) architecture, outperforming the previous pipeline, without separate feature extraction required.

A gender-balanced dataset of 14 participants was recorded with the networks trained with data from 13 participants. One person at the time was excluded for the LOPO test, with the classifier being evaluated on unseen data, and the routine was repeated for each person in the dataset. The data is continuously recorded with the person's free movement directions with no aspect angle limitations during the recordings. Notably, the classification results benefit from balancing the dataset, due to a natural majority and minority occurrence of classes of 42.4 % for walking and 3.4 % for standing up, respectively. Synthetic minority over-sampling technique (SMOTE) was employed, which favors the class samples towards the amount of the majority class.

The realization of using multipaths employing a synchronized radar network with continuously recorded data has provided innovative advantages beyond state-of-the-art methods that have ever touched any form of intended multipath investigation.

Future work may benefit from using more directive antennae that allow enlarging the experimental area with an increased reflection wall, boosting the SNR and permitting a more comprehensive dataset with perhaps the falling class included. Second, the proposed methods might be used with MIMO radar nodes, whereas the angular detection capability using multipath components needs to be investigated.

6

Multi People Monitoring using Regression

Radar-based human activity recognition in crowded environments is addressed using regression approaches. Whereas previous analysis has focused on single activities and subjects, this problem of continuous activity recognition involving up to five individuals moving in arbitrary directions in an indoor area is introduced. To treat the problem, a regression-based approach is used, which offers innovative insights into creating robust and accurate systems for monitoring human activities.

Novel approaches utilizing long short-term memory (LSTM) or convolutional neural network (CNN) regression techniques with a linear regression (LR) and a support vector machine (SVM) regressor are compared on extracted features from radar data through the histogram of oriented gradients and principal component analysis. These approaches are rigorously evaluated by a leave-one-group-out (LOGO) method, with performance assessed using common regression metrics such as the root mean square error (RMSE). The most promising outcomes were observed for crowds of three and five individuals, with respective RMSE of approximately 0.4 and 0.6. These results were achieved primarily by using the micro-Doppler (μD) spectrogram or range-Doppler data domain.

Parts of this chapter have been published in: R. G. Guendel, I. Ullmann, F. Fioranelli, and A. Yarovoy, "Continuous People Crowd Monitoring defined as a Regression Problem using Radar Networks", in *2023 20th European Radar Conference (EuRAD), Berlin, Germany*, 2023.

6.1. Introduction to Multi People Regression Monitoring

Human activity recognition (HAR) has emerged as a crucial research area, not only for enabling vulnerable individuals to maintain an independent lifestyle, but also for ensuring safety in self-determined living environments. A range of technologies, from contactless sensors such as radio frequency (RF) based products to wearable sensors in the form of smartwatches and other devices, have shown the capability to measure various vital metrics, including location, pulse rate, body temperature, blood pressure, and motion characteristics [9]. However, when it comes to monitoring multiple individuals simultaneously, wearable sensors and contactless video-based approaches, such as cameras and lidar sensors, have limitations in terms of usability and privacy concerns. As a result, radar has arisen as a promising alternative due to its ability to overcome these restrictions [171].

The treatment of crowd monitoring as a discretized classification problem has been discussed in the literature, and Bendali-Braham et al. [172] have provided further insight into the complexities of crowd monitoring. For instance, a slight deviation in the classifier's prediction may result in significant classification accuracy errors, as illustrated in the case where the ground truth provides 10 walking individuals, but a classifier predicts only 9. In such cases, the accuracy becomes 0 %, similar to the accuracy obtained if the classifier predicts no walking. Furthermore, this issue is compounded by the possibility of people starting or stopping walking within the sliding window used for classification, leading to similar errors. Therefore, to address these limitations, this study aims to treat the problem of predicting the number of people walking versus idling people as a regression problem, rather than a classification problem. The contributions can be summarized as follows:

- A regression problem was defined to predict the number of people walking in the scene instead of a more conventional discretized classification problem.
- A variety of regressors including deep learning methods, such as the long-short-term memory (LSTM) network, were applied on features extracted from continuous radar recordings, and their results were evaluated using relevant metrics such as the root mean squared error (RMSE). Other regressors included a convolutional neural network (CNN) operating directly on the image domain.
- The proposed approach was validated with data collected with a five-node radar network, synthetically combining the signatures of up to five people walking and stopping to simulate crowd movements in an indoor area.

The rest of this chapter is organized as follows. Section 6.2 presents the data collection, the dataset, and the proposed regression approach, with experimental results presented in Section 6.3. Finally, Section 6.4 concludes the chapter.

6.2. Proposed Regression-based Method for Multi-Person Monitoring

Accurate monitoring of multiple individuals in shared spaces challenges traditional HAR methods. Recognizing the limitations of classification with a discrete number of classes,

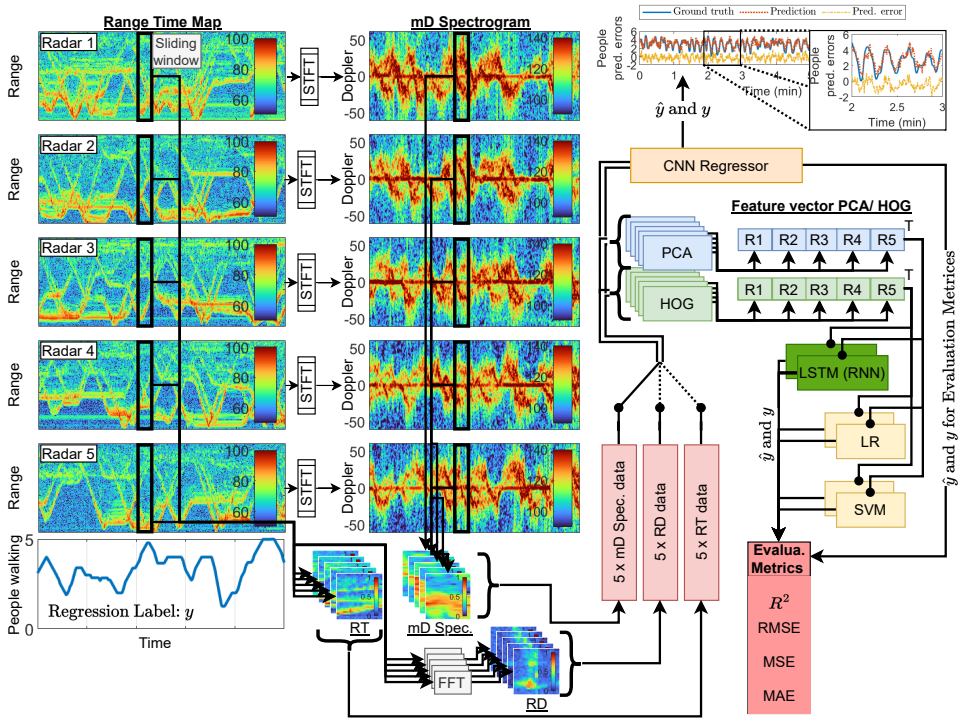
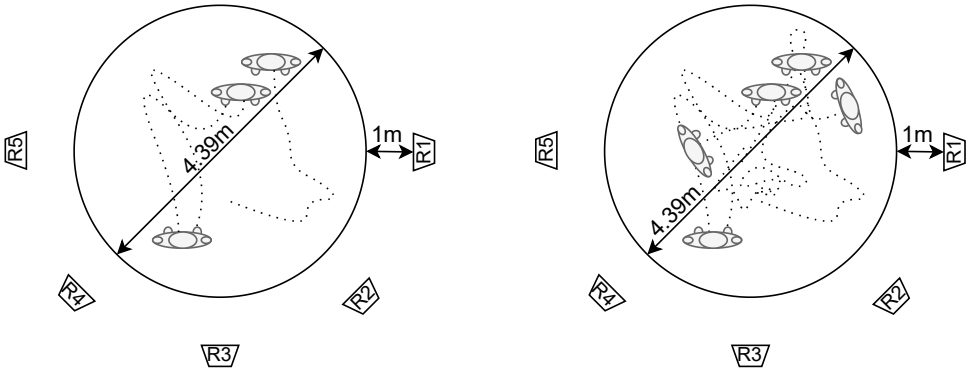


Figure 6.1: The flowchart of the proposed approach shows the synthetically merged data providing the RT maps for a group of five people recorded with a radar network consisting of five nodes. In terms of signal processing, three data domains are extracted, namely the RT–, RD–, and μD spectrogram domain, followed by the feature extraction chains of PCA and HOG, and the ML/DL-based regressors. Lastly, a leave-one-group-out (LOGO) test is shown with its prediction error plot (top-right corner) and the four applied evaluation metrics (bottom-right corner).

this research adopts a regression-based approach for the continuous estimation of the number of walking individuals in a scene. A typical example of this approach can be the estimation of crowd activities using distributed radars. The methodology integrates traditional regressors like linear regression (LR) and support vector machine (SVM) with advanced models such as long short-term memory (LSTM) networks and convolutional neural networks (CNN). These models are trained on features from continuous radar recordings, capturing crowd movement dynamics. A key method is the synthetic fusion of data, combining individual recordings to simulate real-world crowd scenarios. This approach enriches the dataset, ensuring diverse training scenarios for the regressors.

Then, this section outlines the novel proposed methodology, emphasizing regression techniques and synthetic data fusion for enhanced HAR in multi-person environments. It delves into the radar data and examines the extracted domains, including feature fusion. Additionally, the regression approaches of both conventional and deep learning regressors are detailed, along with the typical evaluation metrics suitable for regression.



(a) Group of 3 people

(b) Group of 5 people

Figure 6.2: The sketches depict the layout with (a) three and (b) five people, respectively, walking and remaining idle facing arbitrary directions. The dotted line represents a generic random walking pathway, and the layout represents a five-radar network indicated as R1 to R5.

6.2.1. Radar Data and Domains

The publicly available data [69] contains single human subjects; hence, multiple recordings of different individuals were coherently fused to generate synthetic data with multiple people, such that: $\bar{R} = \frac{1}{K}\sum_{i=1}^{K} R_i$, with K the maximum number of people, R a complex range-time (RT) sequence and $\bar{R}$ the fused result. The generated output in a logarithmic scale can be visually examined in the RT maps on the left side of Fig. 6.1, with five subjects performing walking and standing in an unconstrained trajectory. Additionally, a layout sketch depicting three and five people in a scene is presented in Fig. 6.2. The dotted lines are intended to represent the walking motion in arbitrary directions of some individuals within the measurement area, while others remain standing in an idle position. More details and technical insights on the radar nodes can be found in [68, 87].

Three radar data domains, the range-time (RT) map, range-Doppler (RD) map, and micro-Doppler (µD) spectrogram, are obtained using a sliding window on continuously recorded data, as illustrated in Fig. 6.1. As seen, the network consists of five radar nodes, enabling unconstrained human activity recognition (HAR). A sliding window of 1 sec with a hop size of 0.25 sec is applied. Furthermore, the row-wise fast Fourier transform (FFT) was used over the same window to obtain the RD map.

Before using the same sliding window approach on the µD spectrogram, the short-time Fourier transform (STFT) was applied to the RT signals, with a window size and hop-size of 64 and 63 samples, respectively. Finally, downsampling to 28x28 pixels was applied on all gathered domains of five nodes, with a few examples shown at the bottom of Fig. 6.1.

6.2.2. Feature Extraction

On the generated images for the three data domains, principal component analysis (PCA) was applied, selecting the five strongest principal components associated with the five strongest singular values; this resulted in a feature vector size of 140 samples for the given

input images of 28x28 pixels. Similarly, as a comparative method to PCA, the histogram of oriented gradients (HOG) is used with a cell size of 8x8, a block size of 2x2, 9 bins, and 50 % block overlap, resulting in a feature vector size of 144 samples.

Before forwarding the extracted features to the regression algorithm, feature fusion was employed to combine the information seen by all five radar nodes resulting in a total feature vector length of 700 and 720 for PCA and HOG, respectively.

6.2.3. Regression Approach

This section describes the conventional regressors and the deep learning networks used in this research.

Conventional Regressors

Several regression methods were tested with the best compromise between computational load and performance provided by the LR and the SVM regression. These two regressors are used throughout this study on the concatenated features described in Section 6.2.2.

Deep Learning Regressors

Furthermore, the following two deep learning based regressors were used. First, a modified convolutional neural network (CNN) for regression, as proposed in [173], which operates on image tensors of dimension 28x28x5 directly, and no feature fusion via PCA or HOG required. Then, the recurrent neural network (RNN), proposed in [174] with its version of the long-short-term memory (LSTM) network was modified, simply changing its last layers (*Softmax Layer*, the *Classification Layer*) to a *Regression Layer*. Furthermore, we reduced the network's depth to 400 hidden units instead of the originally proposed 1500.

The ADAM optimizer was used to train both deep learning (DL) networks with 50 epochs and an initial learning rate of 10^{-3}. It should be noted that additional hyperparameter and network tuning may further improve performance, but this is left for future work beyond the scope of this chapter.

6.2.4. Evaluation Metrics

The main evaluation metric throughout this work is the root mean squared error (RMSE), by default the most popular metric when evaluating regression problems and defined as, $\epsilon_{RMSE}=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}$ [161], with $\hat{y}_i$, y_i the prediction and the ground truth, respectively, and m the samples in the leave-one-group-out (LOGO) test set. The results are reported in Table 6.1 and Fig. 6.3, with additional metrics such as the mean squared error (MSE), the mean absolute error (MAE), and the R^2 score.

6.3. Experimental Results

The reported results are for the DL models, namely a CNN and a LSTM, the latter using PCA and HOG feature domains, respectively. Additionally, those features were tested by the following conventional regressors, the LR and the SVM regression model. The CNN regression model, by nature an image processing regression model [173], was modified for the input data size of 28x28x5, where five represents the number of radars in the network; thus, no prior feature extraction is required. An example of the attained performance can be

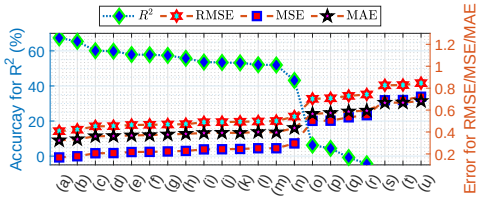
Table 6.1: Leave three and five people out test results, known as leave-one-group-out (LOGO), using a CNN, LSTM, and conventional regressors applied on the µD spectrogram, the RD map, and the RT map, respectively. The RMSE results using a CNN are also shown in Fig. 6.3.

Regression model	Group size Classif./Ev. Metrics	3 People RMSE	5 People RMSE
DL	CNN µD spec.	0.408	0.633
DL	CNN RT map	0.703	1.035
DL	CNN RD map	0.421	0.606
DL	HOG LSTM µD spec.	0.471	0.642
DL	HOG LSTM RT map	0.496	0.645
DL	HOG LSTM RD map	0.469	0.634
Conventional	HOG LR µD spec.	0.453	0.653
Conventional	HOG SVM µD spec.	0.452	0.636
Conventional	HOG LR RT map	0.490	0.671
Conventional	HOG SVM RT map	0.489	0.651
Conventional	HOG LR RD map	0.466	0.635
Conventional	HOG SVM RD map	0.464	0.616
DL	PCA LSTM µD spec.	0.541	0.795
DL	PCA LSTM RT map	0.826	1.074
DL	PCA LSTM RD map	0.708	0.960
Conventional	PCA LR µD spec.	0.494	0.654
Conventional	PCA SVM µD spec.	0.500	0.663
Conventional	PCA LR RT map	0.828	1.111
Conventional	PCA SVM RT map	0.846	1.128
Conventional	PCA LR RD map	0.729	0.949
Conventional	PCA SVM RD map	0.742	0.958

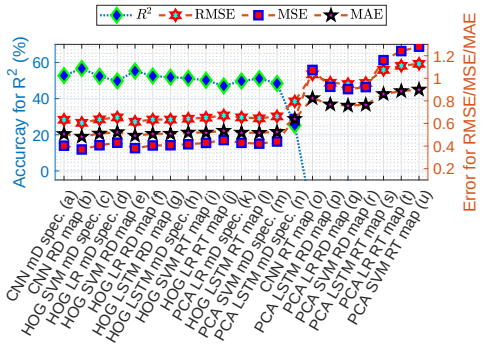
seen in Fig. 6.1 (top right corner) by comparing the *Ground Truth* (blue) with the *Prediction* curve (brown), and the *Prediction error* curve (yellow), here demonstrated for a group of five people.

6.3.1. Group of 3 People

The associated results for a group of three people in the scene are shown in Fig. 6.3a with ordered performance declining from left to right, as well as in the third column of Table 6.1. The best overall performance for the leave-one-group-out (LOGO) test was achieved using the CNN regressor applied on the µD spectrogram with an RMSE of 0.4, closely followed by the CNN architecture operating on RD maps. Subsequently, conventional regression models such as the LR and the SVM were applied, achieving an RMSE of approximately 0.45. The extraction of PCA features, when applied to the µD spectrogram with five principal component vectors selected, yields satisfactory results. In that regard, the regression models LR and SVM both achieve an RMSE of 0.50, while the LSTM model has an RMSE of 0.54. Regardless of the chosen regression model, a drastic performance drop can be observed using the RT map in combination with features extracted from the PCA.



(a) Group of people is 3



(b) Group of people is 5

Figure 6.3: Regression results shown for different regression models and for three– and five-people groups in (a) and (b), respectively. The considered evaluation metrics are the R^2 score, RMSE, MSE, and MAE.

6.3.2. Group of 5 People

The performance results for groups of five individuals are presented in Fig. 6.3b and detailed in the fourth column of Table 6.1. The observed regression performance trends align with those discussed in Section 6.3.1. Notably, the µD spectrogram and the RD domain consistently exhibit superior performance. The CNN regressor, when applied to the RD map, achieves the best result with an RMSE of 0.6. This is closely followed by the LSTM, SVM, and LR models using HOG-extracted features. Conversely, the RT map combined with PCA yields less satisfactory outcomes.

6.3.3. Discussion

The poor regression performance using PCA-based features from 28x28 images of the RT map may be due to its rotation-invariant feature extraction. For RT maps, the feature vector may not convey information about the slope of the dominant signature of a walking human, and the RT domain may not provide direct crucial velocity information, compared to the µD spectrogram or the RD domain. Similarly, the CNN applied to the RT map also provided poor performance, possibly due to similar concerns. It is also notable that, although the HOG feature descriptor is not inherently rotation invariant, it has shown superior performance in detecting the slopes of people within an RT map [109]. Finally, while such a small image size is perhaps also not favorable, this choice was made to limit the computational runtime and burden for the regression models.

These initial results suggest that domains incorporating velocity information are essential, or alternatively, an effective feature extraction method like HOG should be

employed. Although the LSTM does not compete performance-wise with the CNN, it should be noted that the LSTM has more freedom for hyperparameter and network tuning, and with finer tuning, it may outperform other models. However, it is worth noting that the LSTM and conventional regressors (LR, SVM) operate on the HOG or PCA feature vectors, whereas the CNN operates directly on the sample images, making a fair comparison arguable.

6.4. Conclusion

In this chapter, the challenge of estimating crowd activities using distributed radars is addressed by adopting regression approaches, with the objective of accurately predicting the ratio of walking to standing individuals in various orientations. Conventional regression models, specifically the linear regression (LR) and the support vector machine (SVM) regressor, are compared with more advanced deep learning models.

Experimental data were gathered from 15 individual subjects, and synthetic datasets for groups of three and five individuals were generated using synthetic signal fusion. This synthetic data was segmented into training and a leave-one-group-out (LOGO) test set, aiming to evaluate the performance of the trained regressor on unseen data. Three different radar data domains were tested, namely the micro-Doppler spectrogram, the range-time domain, and the range-Doppler domain. Both the principal component analysis (PCA) and histogram of oriented gradient (HOG) feature extraction methods were applied to these domains. Subsequently, LR, SVM, and long short-term memory (LSTM) regressors were tested. Additionally, the convolutional neural network (CNN) regressor was examined, with no prior feature extraction needed.

The results confirm that the CNN regressor, when applied to the micro-Doppler spectrogram domain, achieved superior performance with an RMSE of 0.4 for crowds of three people. Similarly, for crowd groups of five, the range-Doppler map yielded an RMSE of 0.6 with the CNN operating directly on the image domain. On the other hand, the LSTM, LR, and SVM regressors, when applied on HOG features, provided nearly equivalent performance, registering an RMSE of 0.45 (for crowds of three) and 0.65 (for crowds of five). In contrast, PCA often underperformed compared to its aforementioned counterparts, especially in the range-time domain.

7

Conclusions and Further Work

This chapter of the thesis summarizes the key research findings that have been discussed throughout the previous chapters and proposes possible future research directions.

7.1. Major Results, Novel Contributions, and Added Values

- *Novel Data Fusion Approaches for Distributed Radar Nodes (Chapter 4)*

  In order to improve the classification performance of human activity recognition (HAR), a distributed radar network has been used in this research. Due to the diversity of aspect angles for human observations, spatially distributed radars gather more information on human activities, and different approaches were exploited to fuse this information. In addition to feature and decision fusion methods previously considered in the literature, a novel low-level signal fusion concept was proposed combining data before extracting its features. Such a novel signal-level fusion method applied on RT maps has been implemented and compared to state-of-the-art methods, with benefits shown in terms of performance and robustness. With this approach, the subsequent STFT applied on the combined RT map generates one micro-Doppler (μD) spectrogram that contains the information from all radars in the network, with the best results provided by using the full set of radars in the used network.

  Other novel fusion concepts were proposed that account for a person's prior information, such as the location or the velocity/trajectory. For this reason, multilateration processing was exploited on each radar's range pulses to determine a person's location. This was followed by the implementation of an alpha-beta (gamma)–filter to estimate the target's location, velocity, and acceleration direction. With this information, a weight function computes the best suitable radar within a network with respect to different criteria, such as the target's location, steering direction, and steering velocity.

  With the proposed signal fusion method, an F_1 score of 84 % could be achieved, that is an improvement of +6.7 % compared to using a random single radar yielding 77.3 %. With the proposed signal fusion method when using only three radars, the F_1 score performance was 81.5 %, marginally better than using conventional feature fusion methods on all radar nodes, yielding 81.4 %. In fact, two nodes were discarded, and the same performance was achieved using our proposed fusion. The dynamic radar selection using one radar with respect to the target-radar distance achieved 80.3 % and 3 % better than choosing a random single radar for HAR.

- *Novel method for Human Movement Detection and Classification (Chapter 3)*

  For the first time, the complex phase information extracted directly from high-resolution range-time (RT) maps has been used for human movement detection and classification. This approach is complementary to the conventional method, which typically uses the μD spectrogram for classification. When extracting features from the (wrapped) phase map for classification, traditional methods that work on a conventional μD spectrogram are not suitable. The histogram of oriented gradients (HOG) algorithm is proposed to capture the line pattern of a phase map and provide features that can be classified. The proposed phase-based classification approach has been validated by using three state-of-the-art classifiers, namely the nearest neighbor (NN), linear support vector machine (SVM), and the Gaussian SVM.

It has been shown that the proposed method can be applied to the problem of arm gesture recognition measured with a 10-times lower pulse repetition frequency (PRF)–which can be beneficial by using simpler hardware and lower data throughput–without any noticeable decrease in performance, while a conventional μD-based approach may suffer from Doppler ambiguities. Furthermore, the robustness of the proposed approach with respect to the test scenarios was investigated, e.g., the aspect angle to the radar line of sight, the velocity, and the extent of arm movements were also characterized.

To summarize, these findings show a superior performance by fusing the proposed phase-based range-time (RT) map with the magnitude of the RT. This approach is beneficial for simpler, low-cost radars operating at a low PRF where the target velocity exceeds the radar's unambiguous velocity, with performances above conventional μD spectrogram classification. The proposed method is suitable for radars providing the complex I and Q signal components, and can leverage limited hardware with finite resources, while additional Fourier transforms, e.g., for obtaining the μD spectrogram, become no longer necessary.

■ *Novel Advanced Metrics for Continuously Recorded Human Activities (Chapter 4)*

Continuously recorded human activities require advanced performance metrics. Without such metrics, fluctuations in a classifier's prediction can be overlooked when evaluating human activities, e.g., by using artificially separated *snapshot* images for continuous-time sequences. Furthermore, data imbalances may complicate the evaluation of realistic sequences, specifically because such prior-named imbalances will naturally appear due to different occurrences and time spans of activities during an observation period. Typical examples can be the dominance of a *walking* class while participants can move in a room to perform single instances of other in-place activities. In contrast, a typical situation of a critical activity is *falling*, which occurs almost never and is, in fact, a short-duration action compared to *walking* or other typical *in-place* activities.

Therefore, new approaches have been studied to evaluate classifiers in the circumstances mentioned above and new evaluation metrics were proposed. These are compared to the state-of-the-art metrics to account for the specific nature of continuous activities in radar-based HAR, such as the presence of misalignments and interruptions (e.g., outliers) in the classifier's prediction.

These novel metrics, such as the weighted consecutive block detection (CBD) and the intersection over union (IoU), provided a more comprehensive performance evaluation than simply accuracy or F_1 score and were proposed for HAR tasks that account for outliers in the prediction vector, such as misalignments, interruptions, and fluctuations. Notably, the advantage of IoU has been demonstrated with 20 % difference between imperfect and reasonable performing classifiers, while traditional accuracy evaluation gives only 2 % discrepancy, hence a too coarse assessment.

- *Novel Approach in Human Activities Classification Accounting for Temporal Relation (Chapter 4)*

  Throughout the investigation, it has been found that classifiers accounting for temporal relations can provide superior performance over non-temporal classifiers operating on *snapshot* samples. Therefore, recurrent neural network (RNN)-based classification algorithms, namely GRU, LSTM, Bi-GRU, and Bi-LSTM, were investigated to process the resulting low-level μD spectrograms derived from the aforementioned fusion processes. To the best of the author's knowledge, such families of networks remain largely unexplored for HAR using distributed radar datasets, which address continuous activities and unpredictable transitions. This investigation is distinct from conventional approaches, where convolutional neural networks (CNN) are used on data with only one activity per sample or, at best, a few predefined transitions. Furthermore, it is demonstrated that classifiers with bidirectional capabilities (Bi-GRU and Bi-LSTM) exhibit superior F_1 score performance. By utilizing Bi-GRU, the F_1 score improves by +7.5 % compared to unidirectional GRU, achieving 84.4 % and 77.8 %, respectively. Additionally, the intersection over union (IoU), a proposed metric that accounts for interruptions in continuous HAR predictions, provides a more comprehensive assessment. The results show a score of 57.9 % for Bi-GRU and 40.4 % for GRU, highlighting the clear difference in their performance.

- *Novel Approach to Generate Additional Virtual Radar Nodes by Means of Multipath (Chapter 5)*

  Rather than canceling multipath as conventionally performed for indoor applications, it has been proposed for the first time to track the LOS and multipath component of a target, as these multipath components are equivalent to virtual radars observing the target from other aspect angles. With the knowledge of the target location, its higher-order multipath components can be determined by knowing the reflector and radar locations a priori. Therefore, information from the multipath components (e.g., μD spectra) is extracted by leveraging on spatial diversity and aspect angles in observations of the same target, rather than by using only the LOS component. Importantly, as both the LOS and multipath target observations are generated from the same physical radar node, the target's multipath signatures are coherent with the LOS signature.

  Quantified results are provided in Table 5.3, where an F_1 score for leave-one-person-out (LOPO) of 70 % was achieved by using only 3 physical radars and generating 4 additional virtual radar domains, which led, in summary, to 7 extracted signal domains. This is a LOPO result improvement of +11 % compared to a single radar with 59 % using only its LOS signal. The same has been validated by comparing this single radar's LOS domain result with all its possible three domains of LOS, 1^{st}, and 2^{nd} order multipath. Finally, a LOPO result of 65 % and consequently a performance gain of +6 % was obtained, and the need for utilizing multipath for boosting HAR has been demonstrated.

- *Indoor Multi People Monitoring using Regression (Chapter 6)*

  To the best of the author's knowledge, for the first time a regression approach was developed to predict the number of multiple people *walking* in arbitrary directions versus those *standing* in an indoor environment, exploiting a distributed 5-node radar configuration. Unlike conventional approaches based on classification, regressors can provide an approximate estimate which is often sufficient in scenarios such as the analysis of crowds and occupancy in indoor environments [175]. For this, state-of-the-art conventional regressors have demonstrated favorable performance. Highlighting the importance of appropriate feature extraction methods, the principal component analysis (PCA) shows its strengths when applied to the μD spectrogram for feature extraction. On the other hand, the histogram of oriented gradients (HOG) proves to be beneficial for RD or RT maps, as well as μD spectrogram features, prior to employing linear regression (LR) or support vector machine (SVM) regressors. Furthermore, the validation results were obtained using groups of 3 and 5 individuals, achieving RMSE values of 0.452 (with μD spectrogram features) and 0.616 (with RD map features) when utilizing an SVM regressor with HOG-extracted features. Compared to conventional regressors, the deep learning regressors that were tested marginally outperformed their conventional counterparts. They yielded RMSE values of 0.408 for groups of 3 when using features of the μD spectrogram and 0.606 for groups of 3 when using features of the RD map. Importantly, due to their inherent convolutional structure of CNNs, there is no need to extract features using HOG or PCA prior to their use. These findings highlight the benefits of regression tasks in HAR, particularly for providing approximate estimations of crowd behavior and activities.

7.2. Recommendations for Future Work

Possible research directions can be thought of as a continuation of the work presented in this thesis and are outlined below.

- *Proposal: Bistatic/Multistatic Radar Setup*

  Bistatic or multistatic radar setups have already demonstrated their significance in observing a target from various aspect angles and capturing hidden features that monostatic radars may fail to detect. These beneficial features can arise from bistatic and forward scatter mechanisms, which cannot be captured by simple monostatic nodes.

  However, bistatic/multistatic nodes may experience lower pulse repetition frequencies (PRFs) compared to their monostatic counterparts. This limitation arises due to challenges associated with over-the-air clock synchronization between nodes, among other issues. In essence, to utilize low PRF data, the proposed solution involves processing the complex-valued data of a bistatic/multistatic network using a phase map applied to a low-resolution range-time (RT) map. This approach offers a potential solution for addressing ambiguities that may arise in Doppler measurements.

  Furthermore, bistatic/multistatic nodes sometimes have the flexibility to operate as

monostatic nodes as well. A potential approach involves concurrently fusing information generated from both bistatic/multistatic and monostatic nodes, allowing for enhanced data integration and analysis.

- *Proposal: Coherent Radar Networks*

  Gottinger et al. [70] proposed a classification scheme for distributed radar systems, ranging from *Class I: Unsynchronized* to *Class IV: Carrier phase and phase noise coherent*. It has been suggested that coherent radar approaches, specifically those operating in Class III and above, offer several advantages. These include a higher integration gain, improved capability for bistatic/multistatic Doppler measurements, enhanced suppression of stationary clutter, and reduced phase noise resulting from range correlation effects. The utilization of such advanced systems is believed to provide additional information on the velocity components of a human when observed by distributed and coherent nodes, as compared to an incoherent setup of distributed, simultaneously operating monostatic nodes as the one used for this thesis.

- *Proposal: Multi-Frequency Radar Network*

  Gurbuz et al. [91, 176] have successfully implemented a multi-frequency radar system. They compared the degradation in multi-frequency training using various approaches and deep neural network (DNN) architectures. For this comparison, they used data from different radars, each operating at distinct frequencies, for training and testing. These tested systems consist of three radars operating at frequencies of 77 GHz, 24 GHz, and 10 GHz, respectively, each employing different frequency modulation schemes. With regard to the fusion of nodes, an improvement in classification performance for ASL sign language recognition and HAR tasks has been observed using such systems.

  It is believed that such multi-frequency radar setups can enhance the performance of perception algorithms in network settings, where diverse radar systems are assigned to specific tasks. For example, lower-frequency radars can provide a coarse return signal, which is beneficial for conventional HAR tasks. On the other hand, higher frequency nodes, even within the mm-wave range, with their finer resolution, are suitable for recognition tasks involving weak Doppler returns or low SNR signals caused by a small RCS, such as heart-rate/breath monitoring or gesture recognition. However, mm-wave radars may have limitations in terms of longer ranges and increased vulnerability to multipath reflections.

  Taking into account these factors, the integration of a multi-frequency radar network is believed to further enhance the permanence and robustness of classification algorithms, as it offers a wider spectrum of applications, ranging from heart-rate monitoring to HAR at the same time.

- *Proposal: Distributed Network Topology*

  Throughout the thesis, the concepts of distributed radar architecture were explored, employing a semi-circular baseline with five radars, or using three radars from which multipath and scatter behaviors were exploited. However, by incorporating previously proposed ideas such as the *Bistatic/Multistatic Radar Setup* or the *Cross-Frequency Radar Network*, the topology of a network becomes crucial in maximizing information acquisition regarding the target's behavior as well as the environmental layout. For instance, when detecting a person's fall, the radar node that captures the highest Doppler response is ideally positioned in line with the individual's falling trajectory, enabling a top-down view. Conversely, for everyday activities and monitoring of vital signs that depend on postures such as standing or lying, a different node placement may be optimal.

  Moreover, irrespective of the type of activities monitored, it is essential to design a network topology that prevents interferences among nodes and mitigates unwanted reflections, such as ground reflections. Research focusing on the adaptability of sensors is exemplified by the utilization of space-time adaptive processing (STAP) algorithms. These algorithms concentrate on events within specific spatial areas and dynamically adjust the antenna position. Considering these aspects, the integration of a distributed radar network topology within the context of HAR and assisted living opens up novel proposals and contributions.

- *Proposal: Complex Phase Maps for Multiple Radar Nodes*

  Throughout the thesis, the complex phase map was investigated for ultra-wideband (UWB) radars, and its necessity has been demonstrated due to its benefits, such as reduced computational steps by eliminating the need for a short time Fourier transform (STFT). The features were extracted using the histogram of oriented gradients (HOG) algorithm.

  As part of further investigation, this method can be applied to data collected by a distributed radar network, where the nodes are spatially distributed to address potential concerns regarding aspect angles. This approach allows participants to perform activities freely in all movement directions. It is believed that such spatial freedom would result in consistent classification performance, even when utilizing features from the proposed phase map only. However, it is necessary to ensure that the participants are observed from all directions using all nodes in order to achieve classification performance that is independent of the individual's facing direction.

- *Proposal: Network of Illuminators of Opportunity*

  Li et al. [177] proposed a method for utilizing any type of signal transmission from standalone WiFi devices and demonstrated their capability for human activity sensing. These WiFi devices are typically referred to as illuminators of opportunity, as they can leverage various signals such as digital radio, television signals, and WiFi signals for sensing purposes. In the context of indoor monitoring, Li et al. successfully extracted Doppler information from a continuous-time sequence of a walking human.

  This concept is suitable for a distributed node setup, wherein nodes may operate simultaneously as both bistatic/multistatic and monostatic nodes. Additionally,

WiFi-based systems offer inherent synchronization capabilities by design, owing to their primary purpose of network communications. As a result, these illuminators of opportunity have the potential to serve as radars while fulfilling their communication functions.

Bibliography

[1] C. Waldschmidt, J. Hasch, and W. Menzel, "Automotive radar – from first efforts to future systems," *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 135–148, 2021.

[2] J. Marimuthu, K. S. Bialkowski, and A. M. Abbosh, "Software-defined radar for medical imaging," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 2, pp. 643–652, 2016.

[3] J. Le Kernec, F. Fioranelli, C. Ding, H. Zhao, L. Sun, H. Hong, J. Lorandel, and O. Romain, "Radar signal processing for sensing in assisted living: The challenges associated with real-time implementation of emerging algorithms," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 29–41, 2019.

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

[5] W. A. Holm, Ed., *Principles of Modern Radar: Basic principles*, ser. Radar, Sonar and Navigation. Institution of Engineering and Technology, 2010.

[6] V. Chen, *The Micro-Doppler Effect in Radar, Second Edition*. Arttech house, 2019.

[7] R. J. Javier and Y. Kim, "Application of linear predictive coding for human activity classification based on micro-Doppler signatures," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1831–1834, 2014.

[8] F. Fioranelli, H. Griffiths, M. Ritchie, and A. Balleri, Eds., *Micro-Doppler Radar and its Applications*. The Institute of Engineering and Technology (IET), 2020.

[9] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 81–94, 2016.

[10] K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andrès, "From fall detection to fall prevention: A generic classification of fall-related systems," *IEEE Sensors Journal*, vol. 17, no. 3, pp. 812–822, 2017.

[11] T. G. Stavropoulos, A. Papastergiou, L. Mpaltadoros, S. Nikolopoulos, and I. Kompatsiaris, "Iot wearable sensors and devices in elderly care: A literature review," 5 2020.

[12] M. Z. Uddin and A. Soylu, "Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning," *Scientific Reports*, vol. 11, 12 2021.

[13] Z. Wang, V. Ramamoorthy, U. Gal, and A. Guez, "Possible life saver: A review on human fall detection technology," *Robotics*, vol. 9, no. 3, 2020. [Online]. Available: https://www.mdpi.com/2218-6581/9/3/55

[14] N. Pannurat, S. Thiemjarus, and E. Nantajeewarawat, "Automatic fall monitoring: A review," pp. 12 900–12 936, 7 2014.

[15] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sensing*, vol. 11, no. 9, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/9/1068

[16] R. Shahbazian and I. Trubitsyna, "Human sensing by using radio frequency signals: A survey on occupancy and activity detection," *IEEE Access*, vol. 11, pp. 40 878–40 904, 2023.

[17] B. Tan, K. Woodbridge, and K. Chetty, "Awireless passive radar system for real-time through-wall movement detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 5, pp. 2596–2603, 2016.

[18] O. C. Ann and L. B. Theng, "Human activity recognition: A review," in *2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014)*, 2014, pp. 389–393.

[19] M. Ehatisham-Ul-Haq, A. Javed, M. A. Azam, H. M. A. Malik, A. Irtaza, I. H. Lee, and M. T. Mahmood, "Robust human activity recognition using multimodal feature-level fusion," *IEEE Access*, vol. 7, pp. 60 736–60 751, 2019.

[20] X. Liang, H. Li, W. Wang, Y. Liu, R. Ghannam, F. Fioranelli, and H. Heidari, "Fusion of wearable and contactless sensors for intelligent gesture recognition," *Advanced Intelligent Systems*, vol. 1, no. 7, p. 1900088, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.201900088

[21] C.-Y. Hsu, Y. Liu, Z. Kabelac, R. Hristov, D. Katabi, and C. Liu, "Extracting gait velocity and stride length from surrounding radio signals," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 2116–2126. [Online]. Available: https://doi.org/10.1145/3025453.3025937

[22] H. Hong, L. Zhang, H. Zhao, H. Chu, C. Gu, M. Brown, X. Zhu, and C. Li, "Microwave sensing and sleep: Noncontact sleep-monitoring technology with microwave biomedical radar," *IEEE Microwave Magazine*, vol. 20, no. 8, pp. 18–29, 2019.

[23] C. Li, J. Cummings, J. Lam, E. Graves, and W. Wu, "Radar remote monitoring of vital signs," *IEEE Microwave Magazine*, vol. 10, no. 1, pp. 47–56, 2009.

[24] W. H. Organization, "WHO global report on falls prevention in older age." *Community Health*, p. 53, 2007. [Online]. Available: https://www.who.int/publications/i/item/9789241563536

[25] J. A. Haagsma *et al.*, "Falls in older aged adults in 22 european countries : incidence, mortality and burden of disease from 1990 to 2017," *BMJ Open Access*, vol. 26, no. Supp 1, pp. 67–74, 2020. [Online]. Available: http://dx.doi.org/10.1136/injuryprev-2019-043347

[26] E. R. Burns, J. A. Stevens, and R. Lee, "The direct costs of fatal and non-fatal falls among older adults - United States." *Journal of safety research*, vol. 58, pp. 99–103, 9 2016.

[27] T. Al-Aama, "Falls in the elderly," *Canadian Family Physician*, vol. 57, no. 7, pp. 771–776, 2011. [Online]. Available: https://www.cfp.ca/content/57/7/771

[28] M. Rantz, M. Skubic, C. Abbott, C. Galambos, M. Popescu, J. Keller, E. Stone, J. Back, S. J. Miller, and G. F. Petroski, "Automated in-home fall risk assessment and detection sensor system for elders," *Gerontologist*, vol. 55, pp. S78–S87, 6 2015.

[29] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71–80, 2016.

[30] E. Cardillo, C. Li, and A. Caddemi, "Heating, ventilation, and air conditioning control by range-Doppler and micro-Doppler radar sensor," in *2021 18th European Radar Conference (EuRAD)*, 2022, pp. 21–24.

[31] Q. Wu, Y. D. Zhang, W. Tao, and M. G. Amin, "Radar-based fall detection based on Doppler time–frequency signatures for assisted living," *IET Radar, Sonar and Navigation*, vol. 9, no. 2, pp. 164–172, 2015.

[32] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, 2009.

[33] M. Mercuri, I. R. Lorato, Y. H. Liu, F. Wieringa, C. V. Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," *Nature Electronics*, vol. 2, no. 6, pp. 252–262, 2019. [Online]. Available: http://dx.doi.org/10.1038/s41928-019-0258-6

[34] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "Human tracking and identification through a millimeter wave radar," *Ad Hoc Networks*, vol. 116, p. 102475, 2021.

[35] T. Sakamoto, T. Sato, P. Aubry, and A. Yarovoy, "Fast imaging method for security systems using ultrawideband radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 2, pp. 658–670, 2016.

[36] F. Fioranelli, M. Ritchie, and H. Griffiths, "Classification of unarmed/armed personnel using the netrad multistatic radar for micro-Doppler and singular value decomposition features," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1933–1937, 2015.

[37] M. G. Amin and R. G. Guendel, "Radar human motion recognition using motion states and two-way classifications," in *2020 IEEE International Radar Conference (RADAR)*, 2020, pp. 1046–1051.

[38] ——, "Radar classifications of consecutive and contiguous human gross-motor activities," *IET Radar, Sonar and Navigation*, vol. 14, no. 9, pp. 1417–1429, 2020. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rsn.2019.0585

[39] M. G. Amin, "Micro-Doppler classification of activities of daily living incorporating human ethogram," *Proc. SPIE 11408, Radar Sensor Technology XXIV*, vol. 1140808, p. 14, 2020.

[40] R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Derivative target line (DTL) for continuous human activity detection and recognition," in *2020 IEEE Radar Conference (RadarConf20)*, 2020, pp. 1–6.

[41] V. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*. Arttech house, 2001.

[42] G. Li, R. Zhang, M. Ritchie, and H. Griffiths, "Sparsity-based dynamic hand gesture recognition using micro-Doppler signatures," in *2017 IEEE Radar Conference (RadarConf)*, 2017, pp. 0928–0931.

[43] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on Doppler radar motion signature," in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, 2011, pp. 222–225.

[44] J. Zhang, "Analysis of human gait radar signal using reassigned WVD," *Physics Procedia*, vol. 24, pp. 1607–1614, 2012, international Conference on Applied Physics and Industrial Engineering 2012.

[45] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 8–12, 2016.

[46] H. Li, A. Mehul, J. Le Kernec, S. Z. Gurbuz, and F. Fioranelli, "Sequential human gait classification with distributed radar sensor fusion," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7590–7603, 2021.

[47] B. K. Chalise, D. M. Wong, M. G. Amin, A. F. Martone, and B. H. Kirk, "Detection, mode selection, and parameter estimation in distributed radar networks: Algorithms and implementation challenges," *IEEE Aerospace and Electronic Systems Magazine*, vol. 37, no. 11, pp. 4–22, 2022.

[48] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "A survey on radar-based continuous human activity recognition," *IEEE Journal of Microwaves*, vol. 3, no. 3, pp. 938–950, 2023.

[49] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *IEEE Access*, vol. 8, pp. 83 791–83 820, 2020.

[50] P. Chen, X. Wang, M. Wang, X. Yang, S. Guo, C. Jiang, G. Cui, and L. Kong, "Multi-view real-time human motion recognition based on ensemble learning," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20 335–20 347, 2021.

[51] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, "Bi-LSTM network for multimodal continuous human activity recognition and fall detection," *IEEE Sensors Journal*, vol. 20, no. 3, pp. 1191–1201, 2020.

[52] M. Wang, Y. D. Zhang, and G. Cui, "Human motion recognition exploiting radar with stacked recurrent neural network," *Digital Signal Processing*, vol. 87, pp. 125–131, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1051200418308571

[53] R. Feng, E. D. Greef, M. Rykunov, H. Sahli, S. Pollin, and A. Bourdoux, "Multipath ghost recognition for indoor MIMO radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2022.

[54] P. Setlur, M. Amin, and F. Ahmad, "Multipath model and exploitation in through-the-wall and urban radar sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 4021–4034, 2011.

[55] P. Setlur, G. E. Smith, F. Ahmad, and M. G. Amin, "Target localization with a single sensor via multipath exploitation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 1996–2014, 2012.

[56] R. G. Guendel, N. K. Kruse, F. Fioranelli, and A. Yarovoy, "Multipath exploitation for human activity recognition using a radar network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 1, no. 1, p. 13, 2023, [under review].

[57] M. B. Özcan, S. Z. Gürbüz, A. R. Persico, C. Clemente, and J. Soraghan, "Performance analysis of co-located and distributed MIMO radar for micro-Doppler classification," in *2016 European Radar Conference (EuRAD)*, 2016, pp. 85–88.

[58] R. G. Guendel, I. Ullmann, F. Fioranelli, and A. Yarovoy, "Continuous people crowd monitoring defined as a regression problem using radar networks," in *2023 20th European Radar Conference (EuRAD)*, 2023, pp. 294–297.

[59] Y. He, "Human target tracking in multistatic ultra-wideband radar," Ph.D. dissertation, TU Delft, 2014.

[60] A. Lazaro, M. Lazaro, R. Villarino, and D. Girbau, "Seat-occupancy detection system and breathing rate monitoring based on a low-cost mm-wave radar at 60 GHz," *IEEE Access*, vol. 9, pp. 115 403–115 414, 2021.

[61] Acconeer, "Pulsed Coherent Radar Sensors," https://www.acconeer.com/, 2023, [Accessed 05-10-2023].

[62] NXP Semiconductors, "Radar," https://www.nxp.com/design/designs/radar:RADAR, 2023, [Accessed 05-10-2023].

[63] Infineon Technologies AG, "Radar sensors," https://www.infineon.com/cms/en/product/sensor/radar-sensors/, 2023, [Accessed 05-10-2023].

[64] Texas Instruments Incorporated, "Automotive mmWave radar sensors," https://www.ti.com/sensors/mmwave-radar/automotive/overview.html, 2023, [Accessed 05-10-2023].

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

[66] H. Cui and N. Dahnoun, "High precision human detection and tracking using millimeter-wave radars," *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 1, pp. 22–32, 2021.

[67] H.-U.-R. Khalid, A. Gorji, A. Bourdoux, S. Pollin, and H. Sahli, "Multi-view CNN-LSTM architecture for radar-based human activity recognition," *IEEE Access*, vol. 10, pp. 24 509–24 519, 2022.

[68] R. G. Guendel, M. Unterhorst, E. Gambi, F. Fioranelli, and A. Yarovoy, "Continuous human activity recognition for arbitrary directions with distributed radars," in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–6.

[69] R. G. Guendel, M. Unterhorst, F. Fioranelli, and A. Yarovoy, "Dataset of continuous human activities performed in arbitrary directions collected with a distributed radar network of five nodes," 11 2021. [Online]. Available: https://doi.org/10.4121/16691500

[70] M. Gottinger, M. Hoffmann, M. Christmann, M. Schütz, F. Kirsch, P. Gulden, and M. Vossiek, "Coherent automotive radar networks: The next generation of radar-based imaging and mapping," *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 149–163, 2021.

[71] F. Deudon, S. Bidon, O. Besson, and J.-Y. Tourneret, "Velocity dealiased spectral estimators of range migrating targets using a single low-PRF wideband waveform," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 1, pp. 244–265, 2013.

[72] W.-H. Wang, "Human-motion-analysis," https://github.com/mattwang44/Human-Motion-Analysis-MATLAB/releases/tag/0.0.1, 2023.

[73] J. Geisheimer, W. Marshall, and E. Greneker, "A continuous-wave (CW) radar for gait analysis," in *Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers (Cat.No.01CH37256)*, vol. 1, 2001, pp. 834–838 vol.1.

[74] P. v. Dorp and F. Groen, "Human walking estimation with radar," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 150, pp. 356–365(9), 10 2003.

[75] L. Vignaud, A. Ghaleb, J. L. Kernec, and J.-M. Nicolas, "Radar high resolution range & micro-Doppler analysis of human motions," in *2009 International Radar Conference "Surveillance for a Safer World" (RADAR 2009)*, 2009, pp. 1–6.

[76] R. Boulic, N. Thalmann, and D. Thalmann, "A global human walking model with real-time kinematic personification," *The Visual Computer*, vol. 6, p. 344–358, 1990.

[77] Y. He, P. Molchanov, T. Sakamoto, P. Aubry, F. Le Chevalier, and A. Yarovoy, "Range-Doppler surface: a tool to analyse human target in ultra-wideband radar," *IET Radar, Sonar and Navigation*, vol. 9, no. 9, pp. 1240–1250, 2015. [Online]. Available: https://ietresearch-onlinelibrary-wiley-com.tudelft.idm.oclc.org/doi/abs/10.1049/iet-rsn.2015.0065

[78] B. Erol and S. Z. Gurbuz, "A kinect-based human micro-Doppler simulator," *IEEE Aerospace and Electronic Systems Magazine*, vol. 30, no. 5, pp. 6–17, 2015.

[79] C. M. University, "CMU graphics lab motion capture database," http://mocap.cs.cmu.edu/, 2013.

[80] Y. Zhang, X. Li, G. Ma, J. Ma, M. Man, and S. Liu, "A new model for human running micro-Doppler fmcw radar features," *Applied Sciences*, vol. 13, no. 12, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/12/7190

[81] 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," *Applied Sciences (Switzerland)*, vol. 12, 2 2022.

[82] R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Phase-based classification for arm gesture and gross-motor activities using histogram of oriented gradients," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7918–7927, 2021.

[83] A. Petroff, "Ultra-wideband: A tool for teaching undergraduates," in *2014 IEEE International Conference on Ultra-WideBand (ICUWB)*, 2014, pp. 368–373.

[84] S. A. Rane, A. Gaurav, S. Sarkar, J. C. Clement, and H. K. Sardana, "Clutter suppression techniques to detect behind the wall static human using UWB radar," in *2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT)*, 2016, pp. 1325–1329.

[85] *Data Sheet P410*, Time Domain, 11 2013.

[86] M. E. Liggins, D. L. Hall, and J. Llinas, *Handbook of multisensor data fusion : theory and practice*, 2nd ed., ser. The Electrical Engineering and Applied Signal Processing Series. Taylor and Francis Group, 2008.

[87] R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Distributed radar fusion and recurrent networks for classification of continuous human activities," *IET Radar, Sonar and Navigation*, vol. 16, no. 7, pp. 1144–1161, 2022.

[88] S. A. Shah and F. Fioranelli, "RF sensing technologies for assisted daily living in healthcare: A comprehensive review," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 26–44, 2019.

[89] D. Wu, L. Pigou, P.-J. Kindermans, N. D.-H. Le, L. Shao, J. Dambre, and J.-M. Odobez, "Deep dynamic neural networks for multimodal gesture segmentation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1583–1597, 2016.

[90] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, "Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ser. UIST '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 851–860. [Online]. Available: https://doi.org/10.1145/2984511.2984565

[91] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, T. Macks, and F. Fioranelli, "Cross-frequency training with adversarial learning for radar micro-Doppler signature classification (rising researcher)," in *Radar Sensor Technology XXIV*, K. I. Ranney and A. M. Raynal, Eds., vol. 11408, International Society for Optics and Photonics. SPIE, 2020.

[92] N. Golestani and M. Moghaddam, "Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks," *Nature communications*, vol. 11, no. 1, pp. 1–11, 2020.

[93] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "A multisensory approach for remote health monitoring of older people," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 2, no. 2, pp. 102–108, 2018.

[94] Y. He, P. Molchanov, T. Sakamoto, P. Aubry, F. Le Chevalier, and A. Yarovoy, "Range-Doppler surface: a tool to analyse human target in ultra-wideband radar," *IET Radar, Sonar Navigation*, vol. 9, no. 9, pp. 1240–1250, 2015.

[95] M. G. Amin, A. Ravisankar, and R. G. Guendel, "RF sensing for continuous monitoring of human activities for home consumer applications," in *Big Data: Learning, Analytics, and Applications*, F. Ahmad, Ed., vol. 10989, International Society for Optics and Photonics. SPIE, 2019, p. 1098907. [Online]. Available: https://doi.org/10.1117/12.2519984

[96] H. Li, X. Liang, A. Shrestha, Y. Liu, H. Heidari, J. Le Kernec, and F. Fioranelli, "Hierarchical sensor fusion for micro-gesture recognition with pressure sensor array and radar," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 4, no. 3, pp. 225–232, 2020.

[97] J. Kwon and N. Kwak, "Human detection by neural networks using a low-cost short-range Doppler radar sensor," in *2017 IEEE Radar Conference (RadarConf)*, 2017, pp. 0755–0760.

[98] J. Guo, C. Shu, Y. Zhou, K. Wang, F. Fioranelli, O. Romain, and J. Le Kernec, "Complex field-based fusion network for human activities classification with radar," in *IET International Radar Conference (IET IRC 2020)*, vol. 2020, 2020, pp. 68–73.

[99] S. Yang, J. L. Kernec, F. Fioranelli, and O. Romain, "Human activities classification in a complex space using raw radar data," in *2019 International Radar Conference (RADAR)*, 2019, pp. 1–4.

[100] X. Zhang, Q. H. Abbasi, F. Fioranelli, O. Romain, and J. Le Kernec, "Elderly care - human activity recognition using radar with an open dataset and hybrid maps," in *Body Area Networks. Smart IoT and Big Data for Intelligent Health Management*, M. Ur Rehman and A. Zoha, Eds. Springer International Publishing, 2022, pp. 39–51.

[101] Z. Li, F. Fioranelli, S. Yang, J. Le Kernec, Q. Abbasi, and O. Romain, "Human activity classification with adaptive thresholding using radar micro-Doppler," in *2021 CIE International Conference on Radar (Radar)*, 2021, pp. 1511–1515.

[102] Z. Li, J. L. Kernec, Q. Abbasi, F. Fioranelli, S. Yang, and O. Romain, "Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms," *Scientific Reports*, vol. 13, 12 2023.

[103] W. A. M. Al-Ashwal, "Measurement and modelling of bistatic sea clutter," Ph.D. dissertation, UCL (University College London), 2011.

[104] W. Fischer, *Basic Principles of Digital Modulation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 209–236.

[105] B. R. Mahafza, *Radar Signal Analysis and Processing Using MATLAB*, 1st ed. Chapman and Hall CRC, 2008.

[106] P. H. Quanjer, A. Capderou, M. M. Mazicioglu, A. N. Aggarwal, S. D. Banik, S. Popovic, F. A. Tayie, M. Golshan, M. S. Ip, and M. Zelter, "All-age relationship between arm span and height in different ethnic groups," *European Respiratory Journal*, vol. 44, no. 4, pp. 905–912, 2014.

[107] V. Chen and H. Ling, *Time-frequency Transforms for Radar Imaging and Signal Analysis*, ser. Artech House Mobile Communicat. Artech House, 2002. [Online]. Available: https://books.google.nl/books?id=h-8knwEACAAJ

[108] W. Troy, M. Thompson, and Y. Li, "ISAR imaging of rotating blades with an UWB radar," in *2015 Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*, 2015, pp. 1–4.

[109] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, pp. 886–893 vol. 1.

[110] Y.-F. Pan, X. Hou, and C.-L. Liu, "A robust system to detect and localize texts in natural scene images," in *2008 The Eighth IAPR International Workshop on Document Analysis Systems*, 2008, pp. 35–42.

[111] L. Weng, "Object detection for dummies part 1: Gradient vector, HOG, and SS," *lilianweng.github.io/lil-log*, 2017, accessed July 20, 2020. [Online]. Available: http://lilianweng.github.io/lil-log/2017/10/29/object-recognition-for-dummies-part-1.html

[112] S. J. D. Prince, *Computer Vision: Models, Learning, and Inference*, 1st ed. USA: Cambridge University Press, 2012.

[113] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. USA: Prentice Hall Press, 2009.

[114] M. Alam, I. Islam, and M. R. Amin, "Performance comparison of STFT, WT, LMS and RLS adaptive algorithms in denoising of speech signal," *International Journal of Engineering and Technology*, vol. 3, no. 4, 6 2011.

[115] J. Watt, R. Borhani, and A. K. Katsaggelos, *Machine Learning Refined: Foundations, Algorithms, and Applications*, 2nd ed. Cambridge University Press, 2020.

[116] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, 2006, pp. 1491–1498.

[117] X. Huang, Y. Zhang, J. Zheng, Z. Wang, and X. Cai, "A pedestrian detection method combining i-HOG and sobel-ULBP features," in *Proceedings of the 2020 3rd International Conference on E-Business, Information Management and Computer Science*, ser. EBIMCS '20. New York, NY, USA: Association for Computing Machinery, 2021, p. 587–591. [Online]. Available: https://doi.org/10.1145/3453187.3453399

[118] S. Routray, A. K. Ray, and C. Mishra, "Analysis of various image feature extraction methods against noisy image: SIFT, SURF and HOG," in *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2017, pp. 1–5.

[119] A. Gorji, T. Gielen, M. Bauduin, H. Sahli, and A. Bourdoux, "A multi-radar architecture for human activity recognition in indoor kitchen environments," in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–6.

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

[121] U. M. Khan, Z. Kabir, S. A. Hassan, and S. H. Ahmed, "A deep learning framework using passive WiFi sensing for respiration monitoring," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[122] I. Alnujaim, S. S. Ram, D. Oh, and Y. Kim, "Synthesis of micro-Doppler signatures of human activities from different aspect angles using generative adversarial networks," *IEEE Access*, vol. 9, pp. 46 422–46 429, 2021.

[123] S. Zhu, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Continuous human activity recognition with distributed radar sensor networks and CNN–RNN architectures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[124] M. Pham, D. Yang, and W. Sheng, "A sensor fusion approach to indoor human localization based on environmental and wearable sensors," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 1, pp. 339–350, 2019.

[125] M. Ritchie, F. Fioranelli, H. Borrion, and H. Griffiths, "Multistatic micro-Doppler radar feature extraction for classification of unloaded/loaded micro-drones," *IET Radar, Sonar and Navigation*, vol. 11, no. 1, pp. 116–124, 1 2017.

[126] C. Clemente, F. Fioranelli, F. Colone, and G. Li, Eds., *Radar Countermeasures for Unmanned Aerial Vehicles*, ser. Radar, Sonar and Navigation. Institution of Engineering and Technology, 2021.

[127] F. Fioranelli, M. Ritchie, and H. Griffiths, "Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-Doppler," *IET Radar, Sonar and Navigation*, vol. 10, pp. 1702–1710(8), 12 2016.

[128] J. Zhou and J. Shi, "RFID localization algorithms and applications-a review," *Journal of Intelligent Manufacturing*, vol. 20, pp. 695–707, 2009.

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

[130] R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Evaluation metrics for continuous human activity classification using distributed radar networks," in *2022 IEEE Radar Conference (RadarConf22)*, 2022, pp. 1–6.

[131] B. Erol, S. Z. Gurbuz, and M. G. Amin, "Motion classification using kinematically sifted ACGAN-synthesized radar micro-Doppler signatures," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 4, pp. 3197–3213, 2020.

[132] Z. Belhadi and L. Fergani, "Fingerprinting methods for RFID tag indoor localization," in *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, 2014, pp. 717–722.

[133] V. Tran-Quang, T. Ngo-Quynh, and M. Jo, "A lateration-localizing algorithm for energy-efficient target tracking in wireless sensor networks," *Ad-Hoc and Sensor Wireless Networks*, vol. 34, no. 1-4, pp. 191–220, 2016.

[134] A. Becker, *Kalman Filter from the Ground Up, Second Edition*. kalmanfilter.net, 10 2023, [Accessed 12-10-2023]. [Online]. Available: https://www.kalmanfilter.net/alphabeta.html

[135] E. Brookner, *Tracking and Kalman Filtering Made Easy*. John Wiley and Sons, Ltd, 1998. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471224197

[136] A. Petroff, "A practical, high performance ultra-wideband radar platform," in *2012 IEEE Radar Conference*, 2012, pp. 880–884.

[137] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation," in *AI 2006: Advances in Artificial Intelligence*, A. Sattar and B.-h. Kang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1015–1021.

[138] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press, 1999.

[139] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6 2019.

[140] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020.

[141] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167278919305974

[142] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[143] Y. Zhang and Z. Rao, "n-biLSTM: BiLSTM with n-gram features for text classification," in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 2020, pp. 1056–1059.

[144] T. Agrawal, *Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient*. Apress, 2020.

[145] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012, p. 12.

[146] S. B. Buthelezi, "Learning rate optimisation of an image processing deep convolutional neural network," Ph.D. dissertation, DURBAN UNIVERSITY OF TECHNOLOGY, 2020. [Online]. Available: https://hdl.handle.net/10321/3750

[147] J. Liang and Q. Liang, "Design and analysis of distributed radar sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 11, pp. 1926–1933, 2011.

[148] F. Luo, E. Bodanese, S. Khan, and K. Wu, "Spectro-temporal modelling for human activity recognition using a radar sensor network," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–1, 2023.

[149] A. Ledergerber and R. D'Andrea, "A multi-static radar network with ultra-wideband radio-equipped devices," *Sensors*, vol. 20, no. 6, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/6/1599

[150] I. Milani, C. Bongioanni, F. Colone, and P. Lombardo, "Fusing measurements from Wi-Fi emission-based and passive radar sensors for short-range surveillance," *Remote Sensing*, vol. 13, no. 18, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/18/3556

[151] B. Tan, K. Woodbridge, and K. Chetty, "A real-time high resolution passive WiFi Doppler-radar and its applications," in *2014 International Radar Conference*, 2014, pp. 1–6.

[152] Z. Hao, H. Yan, X. Dang, Z. Ma, P. Jin, and W. Ke, "Millimeter-wave radar localization using indoor multipath effect," *Sensors*, vol. 22, no. 15, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/15/5671

[153] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[154] A. Bhattacharya, "How to use SMOTE for dealing with imbalanced image dataset for solving classification problems," 8 2022. [Online]. Available: https://medium.com

[155] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. USA: Prentice Hall PTR, 2001.

[156] Y. Lu and Y. Wan, "PHA: A fast potential-based hierarchical agglomerative clustering method," *Pattern Recognition*, vol. 46, no. 5, pp. 1227–1239, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320312005018

[157] M. Subramaniam, R. Tharmarasa, M. Pelletier, and T. Kirubarajan, "Multipath-assisted multitarget tracking using multiframe assignment," in *Signal and Data Processing of Small Targets 2009*, O. E. Drummond and R. D. Teichgraeber, Eds., vol. 7445, International Society for Optics and Photonics. SPIE, 2009, p. 74450Z. [Online]. Available: https://doi.org/10.1117/12.830470

[158] J. Luiten, A. Osep, P. Dendorfer, P. Torr, A. Geiger, L. Leal-Taixe, and B. Leibe, "HOTA: A higher order metric for evaluating multi-object tracking," *International Journal of Computer Vision*, vol. 129, pp. 548–578, 2021. [Online]. Available: https://doi.org/10.1007/s11263-020-01375-2

[159] A. Distante and C. Distante, *Handbook of Image Processing and Computer Vision: Volume 3: From Pattern to Object*. Springer Cham, 2020.

[160] G. Bonaccorso, *Machine learning algorithms : Popular algorithms for data science and machine learning, 2nd edition*, 2nd ed. Birmingham : Packt Publishing Ltd, 2018.

[161] A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. O'Reilly Media, Inc., 2019.

[162] J. Vanschoren, "Machine learning for engineers, lecture 9: Convolutional neural networks," accessed: 2023-05-01. [Online]. Available: https://ml-course.github.io/

[163] B. G. Mobasseri and M. G. Amin, "A time-frequency classifier for human gait recognition," in *Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI*, B. V. Kumar, S. Prabhakar, A. A. Ross, C. S. Halvorson, and Šárka O. Southern, Eds., vol. 7306, International Society for Optics and Photonics. SPIE, 2009, p. 730628. [Online]. Available: https://doi.org/10.1117/12.819060

[164] M. S. Reza and J. Ma, "Imbalanced histopathological breast cancer image classification with convolutional neural network," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, 2018, pp. 619–624.

[165] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[166] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[167] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[168] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[169] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[170] S. Raschka, "Mlxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack," *The Journal of Open Source Software*, vol. 3, no. 24, 4 2018. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.00638

[171] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 3, pp. 1629–1645, 2020.

[172] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, and P.-A. Muller, "Recent trends in crowd analysis: A review," *Machine Learning with Applications*, vol. 4, p. 100023, 2021.

[173] The MathWorks, Inc. (2021, 10) Train convolutional neural network for regression. [Online]. Available: https://www.mathworks.com/help/deeplearning/ug/train-a-convolutional-neural-network-for-regression.html

[174] K. Itakura. (2021, 10) Video classification using LSTM with MATLAB. [Online]. Available: https://github.com/KentaItakura

[175] L. Ren, A. Yarovoy, and F. Fioranelli, "Grouped people counting using mm-wave FMCW MIMO radar," *IEEE Internet of Things Journal*, p. 13, 2023.

[176] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, E. Malaia, A. C. Gurbuz, D. J. Griffin, and C. Crawford, "Multi-frequency RF sensor fusion for word-level fluent ASL recognition," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 11 373–11 381, 2022.

[177] W. Li, R. J. Piechocki, K. Woodbridge, C. Tang, and K. Chetty, "Passive WiFi radar for human sensing using a stand-alone access point," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1986–1998, 2021.

[178] B. Nijholt, "thesis-cover, parametrically designing the cover of my PhD thesis," https://github.com/basnijholt/thesis-cover, 2020, [Accessed 22-10-2023].

Acknowledgements

My profound gratitude goes to Prof. DSc. Olexander Yarovyi, my promoter and committee chair. His guidance went beyond mere patience and feedback. His comprehensive knowledge of radar signal processing and antenna theory was invaluable as we navigated through the complexities of exploring and understanding the distributed radar system. Prof. Yarovyi's proactive advice, especially in areas like advanced radar data domains, significantly shaped the direction and depth of my work.

Similarly, my sincerest thanks go to Dr. Francesco Fioranelli, my second promoter and daily supervisor. The depth and frequency of our discussions, backed by his expert guidance and consistent encouragement, were foundational in defining the trajectory of my research. Their collective insights not only enriched my research questions and methodology, ensuring its rigorousness and relevance, but also tremendously aided in sharpening the novelties of our numerous published papers.

To my doctoral committee members: Prof. Dr. Wouter A. Serdijn, Dr. Justin Dauwels, Dr. Marco Antonio Zúñiga Zamalloa, Dr. Raf Roovers, Dr. Julien Le Kernec, and Prof. Dr. Mark Neerincx, I offer my sincere thanks. Their thorough evaluation and constructive feedback were invaluable in elevating the quality of my dissertation. The dedication and effort they invested, especially those who undertook international journeys amidst their demanding schedules, have not gone unnoticed.

The support and assistance from the Microwave Sensing, Signals and Systems (MS3) group's academic staff, as well as the secretarial and support teams, were pivotal during my years of research. I am particularly grateful to Esther de Klerk and Minke van der Put for their meticulous handling of administrative challenges and consistent support. My acknowledgment extends to Fred van der Zwan for his indispensable guidance on computational matters, and to Yun Lu and Pascal Aubry, whose expertise proved invaluable during our data measurement campaigns.

Special commendation goes to Dr.-Ing. Ingrid Ullmann from the Friedrich-Alexander-University Erlangen-Nuremberg. As a visiting researcher, her collaboration was instrumental in yielding joint publications. Her methodical and rational approach to resolving challenges greatly influenced our work together.

Supervising master students Max Cortes Peralta and Zhongyuan Guo has been both enlightening and enriching. Particularly with Zhongyuan Guo, the collective expertise culminated in a novel publication. Furthermore, I am grateful for the enriching collaborations with other master students, notably Ximei Yang and Yubin Zhao. Although they were not under my direct supervision, our joint efforts in various projects were invaluable and led to publications. A special acknowledgment goes to Simin Zhu, whose substantial contributions during his master studies were marked by a joint publication.

Deep gratitude goes to Matteo Unterhorst and his thesis advisor, Prof. Dr. Ennio Gambi from Università Politecnica delle Marche. Matteo's pivotal role during his Erasmus internship and the consequential published open dataset has provided a foundation for

other researchers. The culmination of his tenure at TU Delft was marked by a joint publication that stands as a beacon for continuous activity recognition using distributed radars.

I extend my warm thanks to all my Ph.D. colleagues and postdoctoral researchers. Their diverse expertise and insights have been invaluable in broadening my perspective and enriching my research journey. Among them, I would like to mention Nicolas Christian Kruse. Our collaborative efforts and in-depth discussions have been integral to my academic progress. Furthermore, I am especially grateful for his commendable effort in translating the dissertation's summary into Dutch.

I thank Carla Feijen for her guidance with Adobe InDesign and Emile Velthuis for refining my thesis cover. Additionally, credit goes to the creators of the thesis-cover design [178] for the impactful background image.

Finally, my academic path would have been vastly different without the influence of Dr. Bijan G. Mobasseri, Dr. Ahmad Hoorfar, Dr. Mojtaba Vaezi and Dr. Moeness G. Amin from the Center for Advanced Communications (CAC). Their guidance and support during my master's at Villanova University were foundational in steering me towards a Ph.D. at TU Delft.

About the Author

Ronny Gerhard Gündel received his Diploma-Engineering degree (Dipl.-Ing. (FH)) in Automotive Electronics from the University of Applied Sciences Zwickau in Germany in 2017. He worked at the Fraunhofer Institute for Machine Tools and Forming Technology (IWU). In 2017/18, he joined Villanova University (USA) as a Fulbright Scholar in Electrical Engineering. During the summer of 2018, he worked on vehicular wireless communications at the Vodafone Chair Dresden, guided by Prof. Gerhard Fettweis. In 2018/2019, the Center for Advanced Communications of Villanova University offered him a research assistantship under Dr. Moeness G. Amin, focusing on continuous activity classification. He graduated with a Master of Science in Signal Processing and Communications in 2019. From January 2020 to December 2023, he pursued his Ph.D. at Delft University of Technology. Ronny Gerhard Gündel is going to defend his dissertation and obtain his Ph.D. from Delft University of Technology in November 2023. His research interest includes monitoring continuous human activities in range and time beyond micro-Doppler using distributed radar sensor networks and classification for human signature analysis applied to healthcare and security.

List of Publications

Journal papers

5. **R. G. Guendel**, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Multipath Exploitation for Human Activity Recognition using a Radar Network", in *IEEE Transaction on Geoscience and Remote Sensing [under review]*, 2023.

4. I. Ullmann, **R. G. Guendel**, N. C. Kruse, F. Fioranelli and A. Yarovoy, "A Survey on Radar-based Continuous Human Activity Recognition", in *IEEE Journal of Microwaves*, 2023.

3. S. Zhu, **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Continuous Human Activity Recognition with Distributed Radar Sensor Networks and CNN–RNN Architectures", in *IEEE Transaction on Geoscience and Remote Sensing*, 2022.

2. **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Distributed Radar Fusion and Recurrent Networks for Classification of Continuous Human Activities", in *IET Radar, Sonar and Navigation*, 2022.

1. **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Phase-based Classification for Arm Gesture and Gross-Motor Activities using Histogram of Oriented Gradients", in *IEEE Sensors*, 2020.

Conference proceedings

12. **R. G. Guendel**, I. Ullmann, F. Fioranelli, and A. Yarovoy, "Continuous People Crowd Monitoring defined as a Regression Problem using Radar Networks", in *2023 20th European Radar Conference (EuRAD)*, 2023.

11. I. Ullmann, **R. G. Guendel**, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Radar-based Continuous Human Activity Recognition with Multi-Label Classification", in *2023 SENSORS Conference (Vienna, Austria) [accepted]*, 2023.

10. F. Fioranelli, **R. G. Guendel**, N. C. Kruse, and A. Yarovoy, "Radar Sensing in Healthcare: Challenges and Achievements in Human Activity Classification & Vital Signs Monitoring", in *IWBBIO 2023: Bioinformatics and Biomedical Engineering*, 2023.

9. Z. Guo, **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Point Transformer-based Human Activity Recognition using High-Dimensional Radar Point Clouds", in *2023 IEEE Radar Conference (RadarConf23)*, 2023.

8. **R. G. Guendel**, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Exploiting Radar Data Domains for Classification with Spatially Distributed Nodes", in *SET-312 Research Specialist Meeting Distributed Multi-Spectral/Statics Sensing, Bled, Slovenia.*, 2022.

7. X. Yang, **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Radar-based Human Activities Classification with Complex-valued Neural Networks", in *2022 IEEE Radar Conference (RadarConf22)*, 2022.

6. **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Evaluation Metrics for Continuous Human Activity Classification using Distributed Radar Networks", in *2022 IEEE Radar Conference (RadarConf22)*, 2022.

5. N. C. Kruse, **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Segmentation of Micro-Doppler Signatures of Human Sequential Activities using Renyi Entropy", in *International Conference on Radar Systems (RADAR 2022)*, 2022.

4. Y. Zhao, **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Distributed Radar-based Human Activity Recognition using Vision Transformer and CNNs", in *2021 18th European Radar Conference (EuRAD)*, 2021.

3. **R. G. Guendel**, M. Unterhorst, E. Gambi, F. Fioranelli, and A. Yarovoy, "Continuous Human Activity Recognition for Arbitrary Directions with Distributed Radars", in *2021 IEEE Radar Conference (RadarConf21)*, 2021.

2. **R. G. Guendel**, "Further Investigation of Two-way Classification for Activities of Daily Living", in *2020 17th European Radar Conference (EuRAD)*, 2020.

1. **R. G. Guendel**, F. Fioranelli, and A. Yarovoy, "Derivative Target Line (DTL) for Continuous Human Activity Detection and Recognition", in *2020 IEEE Radar Conference (RadarConf20)*, 2020.

Datasets

1. **R. G. Guendel**, M. Unterhorst, F. Fioranelli, and A. Yarovoy, "Dataset of Continuous Human Activities Performed in Arbitrary Directions Collected with a Distributed Radar Network of Five Nodes", 2021 [Online] Available: https://doi.org/10.4121/16691500.