Radar-based Classification of Continuous Sequences of Human Activities

Kruse, N.C.

DOI
10.4233/uuid:31afbccf-a0f8-4bf9-ae23-967d48e935d2

Publication date
2025

Document Version
Final published version

Citation (APA)
Kruse, N. C. (2025). *Radar-based Classification of Continuous Sequences of Human Activities*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:31afbccf-a0f8-4bf9-ae23-967d48e935d2

Radar-based Classification of Continuous Sequences of Human Activities

Radar-based Classification of Continuous Sequences of Human Activities

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
Wednesday 26 March 2025 at 17:30

by

Nicolas Christian KRUSE

Master of Science in Physics,
Rijksuniversiteit Groningen, Groningen, The Netherlands
born in Leiden, The Netherlands

This dissertation has been approved by the promotors:

Prof. dr. A. Yarovoy	Delft University of Technology, promotor
Dr. F. Fioranelli	Delft University of Technology, promotor

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. 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
Prof. dr. ir. F.P. Widdershoven	Delft University of Technology
Dr. M.A. Zuñiga Zamalloa	Delft University of Technology
Dr. M. Ritchie	University College London, United Kingdom
Reserve member:	
Dr. ir. J.H.G. Dauwels	Delft University of Technology

This research was partially funded by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) through the ”Radar-aware Activity Recognition with Innovative Temporal Networks” project (RAD-ART).

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

Keywords: Human Activity Classification, Machine Learning, Radar, Extended Target, Sensor Fusion, Radar Network.

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

Front & Back: Design by Nicolas Kruse.

ISBN/EAN 978-94-6384-756-8 (Paperback)
ISBN/EAN 978-94-6518-033-5 (E-Book/PDF)

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

Author e-mail: nicolaskruse@hotmail.com

Aan mijn reisgenoten.

Contents

List of Acronyms

ADL	Activities of Daily Living
CNN	Convolutional Neural Network
CPI	Coherent Processing Interval
CSI	Channel State Information
DFT	Discrete Fourier Transform
DL	Deep Learning
DoA	Direction of Arrival
FFT	Fast Fourier Transform
GAN	Generative Adversarial Network
(Bi)GRU	(Bidirectional) Gated Recurrent Unit
GT	Ground Truth
HAR	Human Activity Recognition
IMU	Inertial Measurement Unit
kNN	k-Nearest-Neighbours
L1PO	Leave-One-Person-Out
LSE	Least Square Error
(Bi)LSTM	(Bidirectional) Long Short Term Memory
MIMO	Multiple Input Multiple Output
ML	Machine Learning
PC	Point Cloud
PCA	Principal Component Analysis
PRF	Pulse Repetition Frequency
PRI	Pulse Repetition Interval
PT	Point Transformer
RCS	Radar Cross Section
RD	Range Doppler
RNN	Recurrent Neural Network
SISO	Single Input Single Output
STA/LTA	Short Term Average over Long Term Average
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
UWB	Ultra Wideband

SUMMARY

Radar sensors are an emerging technology in the context of non-contact monitoring of vulnerable individuals. Radar-based solutions ensure end-user privacy, whilst providing medical professionals and caregivers with key information concerning the subject's well-being. This thesis proposes novel methods for the classification of sequential human activities using a network of radar sensors. Accurate classification of Activities of Daily Life (ADL) can enable for instance the detection of falls and wandering amongst elderly individuals, and can be employed for the recognition of aggressive or otherwise anomalous behaviour for those receiving mental health care.

This thesis introduces a generic signal model for the used radar signals and data as well as notation conventions in Chapter 2. The methods that are developed for continuous classification of human activities in this study are benchmarked on a common dataset for the purpose of comparison with existing solutions in the literature. Notably, the described signal model is tailored to the radar system that has been used for the collection of this dataset.

In Chapter 3, a novel classification method is proposed for continuous sequences of human activities. The proposed method processes data from Single Input Single Output (SISO) radar sensors and adopts a non-conventional Point Cloud (PC) representation for classification. Specifically, reflection intensity is represented in a range-Doppler-time vector space, in contrast with typical x-y-z coordinate space. The method is essentially one of dimensionality reduction, and can be utilised in the absence of Direction-of-Arrival (DoA) information. Classification of the PCs is achieved by means of a Point Transformer (PT) neural network. Due to the reduced data size of the PCs when compared to more conventional 2D matrix representations such as range-Doppler maps or spectrograms, the PT network is able to effectively utilise the input at full resolution without becoming computationally unwieldy. Furthermore, the proposed method is applied and experimentally verified with a network consisting of multiple cooperating radar sensors, for which different sensor fusion techniques are implemented and demonstrated to increase the overall classification performance.

The PC-based approaches are further developed in Chapter 4, where a novel method is proposed that consists of three main components: a segmentation algorithm, a segment processing algorithm, and a classification network. The segmentation algorithm is aimed at dividing an input sequence of activities into a time-ordered set of single activity segments. This approach incorporates the benefits of utilising the PC processing method, whilst mitigating the problems of using inflexible fixed windows for classification. Segmentation is performed by monitoring a quantity computed from the micro-Doppler spectrogram, namely the Rényi entropy. This quantity is indicative of the type of activity performed, and segmentation is performed when significant fluctuations occur. The proposed method, as well as two alternative segmentation methods, is validated on the same publicly available experimental dataset used throughout this thesis. Notably, it comprises a variety of sequences of nine human activities as observed by a network of five radars.

A novel sensor fusion method is proposed in Chapter 5 that processes raw data from a network of radar sensors and yields three-dimensional representations of both reflection intensity and velocity distribution. Specifically, data from a network of distributed monostatic radars are processed into two 3D fields defined in Cartesian coordinates. The first of these fields contains the reconstructed reflection intensity at each point in a 3D spatial grid; the second is a 3D vector field of reconstructed velocities. In the context of human activity classification, the reconstructed velocities can be related to the motion of the different body parts, characterised in more detail compared to simply using spectrograms or range-Doppler representations with only the radial velocity components. The efficacy of the method is evaluated through two case studies. The first case study entails classification of human activities utilising the proposed method to process 2D data from the same publicly available experimental dataset used throughout this thesis, followed by classification by a CNN-BiLSTM (Convolutional Neural Network - Bidirectional Long Short Term Memory) architecture. The second study demonstrates the feasibility of 3D intensity and velocity reconstruction by processing dedicated data captured specifically for this study.

Chapter 6 finally presents conclusions pertaining to the research performed for this thesis, as well as recommendations for future research. The contributions of the studies are summarised first, and are followed by a section detailing the recommendations. Improvements and refinements of the proposed methods are suggested, and remaining challenges are outlined.

Samenvatting

Radarsensoren zijn een opkomende technologie voor het contactloos monitoren van kwetsbare individuen. Radargebaseerde methoden waarborgen de privacy van de gebruiker en voorzien medische hulpverleners van essentiële informatie met betrekking tot het welzijn van de persoon. In dit proefschrift worden methoden voorgesteld om opeenvolgende menselijke activiteiten te classificeren met behulp van radarsensoren. Nauwkeurige classificatie van Activiteiten van het Dagelijks Leven (ADL) maakt het mogelijk om vallen en dwaalgedrag bij ouderen tijdig te detecteren en kan worden ingezet om aggressief of anderszins afwijkend gedrag te herkennen bij patienten in de geestelijke gezondheidszorg.

Deze scriptie introduceert allereerst een generiek signaalmodel voor de gebruikte radarsignalen en notatieconventies in Hoofdstuk 2. De methoden die in deze studie zijn ontwikkeld voor de continue classificatie van menselijke activiteiten worden geijkt op een dataset met het oog op vergelijking met bestaande oplossingen in de literatuur. Het beschreven signaalmodel is relevant voor het radarsysteem dat is gebruikt voor het verzamelen van de dataset.

In Hoofdstuk 3 wordt een nieuwe classificatiemethode voorgesteld voor continue reeksen van activiteiten. De voorgestelde methode verwerkt gegevens van Single Input Single Output (SISO) radarsensoren en introduceert een atypische puntwolk oftewel Point Cloud (PC)-representatie voor classificatie. Specifiek wordt de reflectie-intensiteit weergegeven in een afstand-Doppler-tijd vectorruimte, in tegenstelling tot de typische x-y-z coördinatenruimte. De methode is in essentie een dimensionaliteitsreductiemethode en kan worden gebruikt zonder Direction-of-Arrival (DoA) informatie. Classificatie van de PC's wordt uitgevoerd met behulp van een Point Transformer (PT) netwerk. Vanwege de gereduceerde datastructuur van de PC's in vergelijking met meer conventionele matrixrepresentaties zoals afstand-Dopplerrepresentaties en spectrogrammen, kan het PT-netwerk de input effectief gebruiken op hoge resoluties zonder dat het computationeel onhandelbaar wordt. Verder wordt de voorgestelde methode toegepast en experimenteel geverifieerd met een netwerk bestaande uit meerdere radarsensoren. Hierbij worden verschillende sensorfusietechnieken geïmplementeerd en wordt aangetoond dat zij de algemene classificatieprestaties verbeteren.

De op PC gebaseerde benaderingen worden verder ontwikkeld in Hoofdstuk 4, waar een methode wordt voorgesteld die uit drie hoofdcomponenten bestaat: een segmentatiealgoritme, een segmentverwerkingsalgoritme en een classificatienetwerk. Het segmentatiealgoritme is gericht op het verdelen van een ononderbroken reeks van activiteiten in een tijdsgeordende reeks van segmenten met slechts één activiteit. Deze aanpak maakt gebruik van de voordelen van de PC-verwerkingsmethode, terwijl de problemen van het gebruik van vaste vensters voor classificatie worden beperkt. Segmentatie wordt uitgevoerd door een grootheid te volgen die is berekend uit het micro-Doppler-spectrogram, de Rényi-entropie. Deze grootheid geeft een indicatie van het type uitgevoerde activiteit, en segmentatie wordt uitgevoerd wanneer significante schommelingen optreden. De voorgestelde methode, evenals twee alternatieve segmentatiemethoden, wordt gevalideerd op dezelfde openbaar beschikbare experimentele dataset die bestaat uit een verscheidenheid aan reeksen van negen menselijke activiteiten, geobserveerd door een netwerk van vijf radars.

In Hoofdstuk 5 wordt een nieuwe sensorfusiemethode voorgesteld die ruwe data van een netwerk van radarsensoren verwerkt en driedimensionale representaties oplevert van zowel reflectie-intensiteit als snelheidsverdeling. Gegevens van een netwerk van gedistribueerde monostatische radars worden verwerkt tot twee 3D-velden, gedefinieerd in cartesiaanse coördinaten. Het eerste veld bevat de gereconstrueerde reflectie-intensiteit op elk punt in een 3D-ruimteraster; het tweede is een 3D-vectorveld van gereconstrueerde snelheden. In de context van menselijke activiteitclassificatie kunnen de gereconstrueerde snelheden worden gerelateerd aan de beweging van de verschillende lichaamsdelen. De effectiviteit van de methode wordt geëvalueerd aan de hand van twee studies. De eerste studie betreft de classificatie van menselijke activiteiten waarbij de voorgestelde methode wordt gebruikt om 2D-gegevens uit een openbaar beschikbare dataset te verwerken, gevolgd door classificatie met een CNN-BiLSTM (Convolutional Neural Network - Bidirectional Long Short Term Memory) architectuur. De tweede studie toont de haalbaarheid aan van 3D-intensiteit- en snelheidsreconstructie door gegevens te verwerken uit een dataset die voor deze studie is vastgelegd.

Hoofdstuk 6 presenteert de conclusies met betrekking tot het onderzoek dat voor deze scriptie is uitgevoerd, evenals aanbevelingen voor toekomstig onderzoek. De bijdragen van de studies worden eerst samengevat, gevolgd door een sectie waarin de aanbevelingen worden beschreven. Verbeteringen en verfijningen van de voorgestelde methoden worden gesuggereerd, en de resterende uitdagingen worden geschetst.

Preface

Four years have passed in record time. As I commenced my PhD programme during the global pandemic, I found myself wondering what I had gotten myself into. Coming from the field of high energy physics, I had to change gears in order to immerse myself in this new topic. Any doubts I had quickly disappeared over the days and weeks as I started meeting more and more of my wonderful colleagues. I greatly valued the discussions with them, and the advice they could give.

Over the years that followed, I have done what I love the most: I learned. I was able to sink my teeth into new material, acquire new skills, and recently managed to print something in colour at the first try. This dissertation is my labour of love and exasperation, it is my profound wish that it can play a role in unburdening the lives of others.

Nicolas Kruse
Delft, February 2025

1

Introduction

Radar can be used to aid medical professionals and caregivers by providing a privacy-preserving way of monitoring vulnerable individuals. In this introductory chapter I explain why and how radar can be used for this purpose. I discuss existing literature and current challenges in the field, before stating the objectives and contributions of this PhD research.

1.1. Radar-based Human Activity Classification

In the general field of healthcare, radar has emerged as a valuable non-contact sensing technology with distinct advantages over alternatives like cameras and wearable sensors [1–6]. Specifically, radar-based systems offer activity monitoring without requiring individuals to remember to wear devices or interact with them, and maintain effectiveness in low-light or glaring conditions as well as in obstructed environments. Importantly, radar's inability to capture visual imagery can help preserve personal privacy.

Monitoring human Activities of Daily Living (ADL), provides essential insights for medical professionals, enabling timely interventions for events such as falls, wandering, and self-harm attempts. Following estimates of the United Nations, the percentage of global population aged 65 and above is expected to nearly double from 9.1% in 2019 to 15.9% in 2050 [7]. This growth comes with an associated increase in the need for healthcare solutions to address age-related risks like falls and decline of cognitive abilities. The US Center for Disease Control and Prevention concluded that approximately 27 000 adults of age 65 and up died due to fall related injuries in 2014 [8]. The same source reports that 28.7% of older adults reported falling at least once, with 37.5% of these falls resulting in injury requiring medical treatment. Another leading cause of death and injury is suicide, which was the tenth most prevalent cause of death in the US in 2013 [9]. Many of these suicides occur among mental health patients, with the American Psychiatric Association reporting a third of the fatalities under patients who are under a 15-minute checking schedule [10]. These figures highlight the demand for reliable, non-intrusive monitoring technologies that can assist in timely medical response.

Low-cost sensing solutions are desirable, as monitoring in an in-home setting is in many cases preferable. Extended hospitalisation and admittance to clinics are costly procedures and put an increased strain on medical staff and other support personnel. From a psychological standpoint, it is generally preferable for patients to remain in their home environment as much as possible for their own well-being. In-home monitoring offers the possibility for long-term observation of a patient, which offers opportunities for preventative care in addition to reactive care. Preventative care can take the form of e.g., the monitoring of wandering among patients with dementia [11]. Furthermore, analyses of step-time variability can be used to study those who are at an elevated fall risk due to e.g., Parkinson's disease or other gait impairments [12]. In the case of mental health care, accurate classification of ADL can be beneficial for the recognition of potentially aggressive or otherwise anomalous behaviour, such as restlessness and attempts at self-harm or suicide.

In this broader context, radar technology has been proposed for a variety of the aforementioned healthcare applications, including fall detection [13, 14], gait analysis [15–19], and vital sign monitoring [20–22]. By leveraging advances in signal processing and machine learning, radar-based Human Activity Recognition (HAR) systems can classify activities continuously, moving beyond traditional "snapshot-style" classification where data to be classified consists of only a single, well-defined activity. Instead, continuous classification methods process extended activity sequences, aligning more closely with real-life behavioural patterns where multiple activities are performed one after the other and there are not necessarily clear and neat transitions in-between.

1.2. Review of Existing Approaches to Radar-based Human Activity Classification

This section presents a brief overview of existing approaches to perform radar-based human activity classification. The individual chapters of this thesis will each include a more in-depth review of the literature pertaining to the specific topics discussed in the respective chapters.

Existing solutions to radar-based classification of ADL can be broadly categorised in two distinct approaches, based on the nature of the data under consideration. One of these two branches is concerned with the classification of data samples where it is a priori known that only a single activity is present. This approach is sometimes referred to as 'snapshot(-style)' classification [12, 23]. Typically, these approaches follow a four-stage pipeline in terms of applied radar signal processing, which is comprised of data acquisition, data processing into various data domains, feature extraction, and classification.

Data acquisition is often achieved through experimental means, where human activities are captured in e.g. a laboratory space [12] or a more realistic living space [24]. Due to the required investments of time and resources into creating large annotated datasets, other means of data acquisition or augmentation are explored as well. As an example, Generative Adversarial Networks (GAN) are explored to produce radar data that is similar, but not identical, to a comparatively small set of real data [25, 26], which can lead to improved classification capabilities. Additional forms of data augmentation include the synthesis of radar-like data from alternative sensors that are more prevalent, such as camera [27].

The specific *processing and representations of radar data* that are optimal for classification is an ongoing topic of study. Due to the ability of most radar sensors to directly measure target velocity accurately, the spectrogram representation is a common choice in the literature [13, 23–25, 28–31]. In spectrograms, range information is discarded in favour of velocity information, which is strongly tied to the subject kinematics as the velocity components from each body part will be superimposed onto the overall spectrogram. This leads to a 2D representation of velocity (or Doppler frequency) as a function of time. Aside from other 2D representations such as range-Doppler maps [32, 33], an increasing number of classification methods include Point Cloud representations, where unordered lists of points are stored rather than matrices with dimensions of range (or spatial coordinates), time, Doppler/velocity, and intensity or Radar Cross Section (RCS). Point cloud representations are often based on detected points in 3D space [3, 34–36], but can also be constructed in different vector spaces, such as range-Doppler-time [37]. Finally, fusion of information from a larger set of domains or representations into a condensed, intermediate domain is increasingly explored [38–40]. Machine learning approaches have proven to be valuable in the creation of fused data representations that, while often effective and computationally efficient, are no longer easily human-interpretable in terms of their physical meaning.

After the determination of an appropriate representation for the radar data, the process of *feature extraction* follows. The aim of this process is to identify and extract quantities of interest from the chosen data domains, primarily based on discrimination power. Feature extraction can be performed manually, yielding so-called 'handcrafted' features, or by means of Machine Learning methods. Handcrafted features often involve statistical quantities derived from the radar data representations, e.g., Doppler spectrum width or average Doppler frequency [41–45]. Some derived quantities can more easily be tied to physical characteristics of the motion being performed, such as for example the period of limb motion in [46]. Approaches to feature extraction that involve Machine Learning are diverse but follow a general trend of a dimensionality reduction that is optimised with respect to a certain cost function, or loss function. Input data from a chosen domain enters a neural network, and subsequent layers of neurons represent linear combinations of the input data. The weights of the linear combinations are continuously adjusted in the training of the network until a desired network is achieved. Between handcrafted features and Machine Learning approaches to feature extraction are methods such as Principal Component Analysis (PCA) [28, 47], that are not trained on any data, but require little user input.

The final stage in the pipeline is *classification* based on the extracted features. Similarly to feature extraction approaches, recent years have seen a transition from algorithmically simpler methods such as k-Nearest-Neighbours (kNN) [28, 32, 42] and Support Vector Machines (SVM) [44, 48, 49], to Deep Learning (DL) methods. The most prevalent of the DL methods employed is the Convolutional Neural Network (CNN). Due to CNNs effectiveness in image processing and classification, the application to radar data which is often represented in 2D matrices is a logical evolution. Many works on radar-based classification of ADL maintain the 2D representations common to image processing, using e.g., domains such as range-Doppler maps or spectrograms [31, 50–53]. Other research efforts are directed towards the utilisation of higher-dimensional inputs, allowing for example range-Doppler-time tensors to be classified in their entirety [40].

Moving beyond snapshot-style or non-continuous classification requires the processing of temporally extended sequences of sequential human activities. The amount of activities in a sequence is unknown, as is the onset and duration of each activity. This complication requires alterations to the four-stage classification discussed before, where the expected output is only a single activity label. Three overarching approaches are considered here based on the surveyed literature: sliding window methods, Recurrent Neural Network (RNN)-type classifiers, and segmentation-based methods.

Sliding window methods employ the well-established methods from non-continuous literature. Feature extraction is performed on a window that is of short duration comparative to the full sequence. Classification results in an activity label, and the window is moved forward in time until the entire sequence has been processed. Sliding window methods can be found in conjunction with e.g., SVMs [53], and CNNs [54]. Challenges that accompany sliding window methods include the selection of an appropriate window size, or adaptive window approach. Longer windows generally include more information on the activity being performed, facilitating classification. However, the chance of including multiple activities, and thus introducing ambiguity, increases with longer windows.

RNN-type methods process sequences at some fundamental, often rather short, interval of the measurement setup, for example the radar sensor Pulse Repetition Interval (PRI). Feature extraction is performed on the basis of these individual time steps, where the data is often represented as a 1D vector. Notably, networks in the RNN family of

DL architectures are able to correlate vectors from different time steps, essentially learning from past or future feature vectors. Among networks in the RNN family are (Bidirectional) Gated Recurrent Units (GRUs) [38, 55], and (Bidirectional) Long Short Term Memory ((Bi)LSTM) networks [33, 56]. Recently, Transformer-type networks with attention mechanisms have seen success in classification of extended sequences, despite the challenge in training them effectively with the often scarce radar data available [57].

Finally, classification methods based on sequence segmentation aim to adopt the methods from non-continuous literature, whilst addressing some of the challenges with sliding windows. Activity sequences are segmented into variable-size segments that contain a single activity only. This facilitates subsequent classification, but introduces the problem of finding the transitions between activities and defining metrics to perform this. Most approaches in literature aim to simply distinguish between the absence and presence of motion, often accomplished by evaluating target Doppler frequencies and by applying simple thresholds on the power levels in the signature [58–61].

1.3. Open Challenges

Despite the multitude of proposed solutions to classification of ADL with radar, several key challenges currently remain. Being a classification problem, the primary goal in HAR is to perfectly distinguish a given set of activities. With Machine Learning methods becoming ubiquitous in the literature beyond radar, improvements in classification performance are often achieved with more data, and larger, deeper classification networks. However, the collection and annotation of radar data is typically more challenging than using other sensors, and this can create a barrier to the development of models with increasing depth. Moreover, this trend cannot continue indefinitely, and more research is being performed to reduce network sizes, for example with the aim of efficient implementation on edge devices [62]. Hence, improving classification performance with reduced computational requirements is an open challenge.

Aside from improving classification accuracy on a given set of classes, the amount of classes under consideration is another topic under investigation. Including a larger variety of activity classes generally improves the versatility of the proposed method, but simultaneously increases the complexity of the classification task. Furthermore, the increased variety in activities requires an accompanying increase in the size of the datasets that are to be used for training purposes. Especially datasets that have been captured in realistic scenarios often have a strong imbalance in samples for different classes. In particular, critical classes such as fall events are rare, and their respective sample support being low hinders effective training of a classifier tasked with recognising these key events.

Related to the above challenge, open-set approaches are currently minimally explored [63] but are inherently designed to address this problem. Open-set methods offer a capability of either rejecting previously unseen activity classes, or including them as new activity types in an unsupervised manner. It is expected that realistic scenarios will see many of these unseen activity classes due to the variety in human motions, and the current closed-set, fully supervised approaches offer no effective solution to this problem.

More generally, the creation of large datasets of human activities is important for the development of classification algorithms. However, there is currently no generally accepted primary benchmarking dataset in the community [64]. Such a dataset should ideally include a large variety of annotated activities, with variations in scene, observed participant, variations in the number of people present, and potentially radar sensors. Radar sensors are varied in capabilities and data outputs when compared to e.g. cameras, further complicating the task of establishing a reference dataset that can be utilised for the variety of methods that have been proposed in the literature.

In the development of solutions to HAR for radar, inspiration is taken from other fields, such as the classification of images, video footage, and audio processing. Radar data is however intrinsically different from these domains, and the methods which perform well for other data have to be adapted appropriately. As described in the previous section, CNNs have been demonstrated to be extremely effective at classification of images, and an extension to 2D radar data domains is straightforward. However, certain properties of CNNs, such as translation invariance, are undesirable for e.g., range-Doppler domains, where the coordinates within the 2D representation have strong physical relevance and link to the kinematics of the observed targets.

Finally, the topic of continuous classification of ADL is not solved. The limitations of sliding window approaches have been outlined in the previous sections, and RNN-type classification features drawbacks such as required model complexity and interpretability. Especially transition points between subsequent activities are challenging with regards to classification, primarily due to the rigid approaches most often taken that feature a closed set of activity classes. Classifier predictions are typically single-label only, meaning that the predictions around a transition between activities will abruptly change from one label to the next. Real activities typically flow more smoothly into each other, and exact transition points are often subjective in nature.

In the following chapters of this thesis, more detailed, topic-specific challenges will be outlined and discussed together with the proposed solutions.

1.4. RESEARCH OBJECTIVES

The primary research objective of this thesis is the improvement of classification of human activities, with emphasis on tackling continuous data sequences recorded by a network of cooperating radar nodes. In continuous sequences, constituent activities are of unknown, arbitrary duration, and smoothly transition into each other. With the goal of classification improvements, several methods are proposed in this research to engage with different facets of this overarching challenge.

In broad terms, the problem of continuity in activity sequences is approached from a perspective of segmentation and classification. Rather than processing lengthy portions of radar data with potentially numerous activities, a method is formulated to divide the sequence into homogenous segments that ideally contain only a single activity. Classification can subsequently be performed through the use of neural networks that are only tasked with identifying this single activity.

The classification of extended activity sequences typically requires neural networks capable of managing both the comparatively long duration of the sequence, as well as the complexity of the multiple activities therein. Consequently, full resolution input data such as complete range-Doppler matrices or 4D radar data hypercubes, can generally not be employed due to computational constraints. To address this issue, a method is proposed to process raw radar data into a point cloud data structure, even if the original radar sensor was a SISO (Single Input Single Output) radar without inherent capability to measure the angular position of targets. The proposed processing approach allows utilisation of full resolution data but captures only the salient features in terms of distance, velocity, and temporal distribution of the scattering points of an observed target, in this case a person. This resulting point cloud is then used with a suitable classifier designed for point clouds to demonstrate the method effectiveness, specifically a Point Transformer network.

The individual motions of body parts are hypothesised to be critical to the correct evaluation of the activity that is being performed, as demonstrated in [65]. Thus, a method is proposed to utilise a distributed network of radar sensors to reconstruct the location, shape, and velocity distribution of a human target in 3D. Unlike previous works where the velocity signatures of the body parts are only measured via the line-of-sight projections towards each radar, the proposed approach aims to reconstruct a full 3D profile of the involved velocities. The method is evaluated on experimental data to demonstrate the increased classification performance that is achieved through the application of the method.

Notably, all methods proposed in this research are evaluated on a benchmark radar data set co-developed as part of this thesis [66]. This ensures they can be compared to each other, and additionally to reference works utilising the same data set. The dataset consists of 840 min of radar data of human activities. These data are captured by a distributed network of five monostatic sensors, making the dataset unique in terms of opportunities for sensor fusion techniques. All activities are performed in 2 min sequences, with 14 participants performing 30 sequences each. The sequences are composed of 9 different activity classes, and vary from fall events in different locations to mixed sequences of all 9 activities. In contrast to many available datasets, the activities performed in the sequences are completely unconstrained in duration and location, which makes the dataset challenging from a classification perspective. As part of the PhD research efforts, a new data-loader has been developed to enable seamless integration of the dataset into Python-based methods. This addition increases the utility of the dataset for a wider audience of prospective users.

All methods developed for this research have yielded classification performance metrics that match or outperform reference works from the literature, demonstrating the effectiveness of the proposed methods for continuous human activity classification. These improvements in classification accuracy make radar-based HAR a more suitable solution for human monitoring, which in turn can contribute to improved quality of life for vulnerable individuals.

1.5. Main Contributions

The main contributions of this PhD research can be briefly summarised as follows.

- A novel approach for continuous activity classification is proposed based on the utilisation of Point Cloud (PC) radar data structures and a Point Transformer (PT) network inspired by [67]. In contrast to conventional PC-based approaches, the proposed method adopts range, Doppler, time and reflection intensity as PC coordinates. The result is an approach that combines computational efficiency and classification performance.

- A method for segmentation and classification of continuous activity sequences is proposed based on Rényi entropy [68] rather than simpler power-based thresholds. Classification of the discrete segments allows for the utilisation of a powerful Point Transformer classifier, compared to reference methods from the literature. This in turn increases classification performance at no increase in computational complexity of the classifiers utilised.
- A novel sensor fusion method that processes raw data from a network of radar sensors and yields three-dimensional representations of both reflection intensity and velocity distribution. This method is the first to offer these reconstructions in intensity and velocity of extended human targets, and the efficacy of the method is demonstrated through classification case studies. The added information on limb velocities aids in classification, yielding improved performance metrics.
- Demonstration of the efficacy of the proposed methods through experimental comparison with reference works utilising a challenging dataset including 2 minutes sequences composed of 9 activities performed by 14 participants.

1.6. Thesis Outline

The remainder of this thesis is organised as follows: Chapter 2 introduces the dataset used for benchmarking purposes in terms of sensor description, signal model, and details about the general preprocessing steps applied to the data. Chapter 3 presents the approach to continuous activity classification utilising the proposed point cloud data format and a Point Transformer network for effective classification. In Chapter 4, a segmentation method based on Rényi entropy is introduced to divide continuous activity sequences into discrete segments suitable for classification. Chapter 5 addresses the reconstruction of extended target location, shape, and velocity distribution, using data from a network of radar sensors. Finally, Chapter 6 presents the conclusions and suggests directions for future work.

2

Radar Dataset and Signal Model

Radar data in the vast majority of cases bears no visual resemblance to the person being observed. Whilst this is crucial for the preservation of user privacy, it also means that ample thought must be given to the problem of extracting the necessary information from the radar data. In this chapter I describe the basics of the radar system that I've used for experiments throughout this research, and go over processing steps that yield information on the person's location and movements. Within the research group we have collected a large dataset of human activities, measured by radar. I go over details of this dataset, which we have used extensively for benchmarking the performance of our algorithms and for training machine learning models.

2.1. Pulsed UWB Radar System Characteristics

In order to gauge the effectiveness of the methods proposed in this thesis, an experimental benchmark dataset has been co-developed and utilised in order to compare methods with each other and with reference methods from the literature on radar-based HAR. The radar systems used in the collection of the data are Humatics PulsON P410 pulsed Ultra Wideband (UWB) sensors [69], shown in Figure 2.1. The principle of operation of these sensors is the emission and reception of very short pulses of electromagnetic radiation. The modulated pulses propagate through space and scatter off objects, with a portion of the pulse energy being reflected back towards the receiving element. The time between pulse emission and reception is directly related to the distance between the scatterer and the sensor antenna. Aside from target range, radial velocity of the scatterer can additionally be determined, as will be explained in Section 2.3. The P410 sensors operate at a centre frequency of 4.3 GHz and feature a bandwidth of 2.2 GHz, resulting in a range resolution of approximately 6.8 cm. The sensors are Single Input, Single Output (SISO), which implies that, unless a directional antenna is used, no information can be acquired on the angle of arrival of reflections from scatterers at different azimuth or elevation angles.

2.2. Range-Time Representation

The signal model in this section is in part adapted from [70]. Consider a series of M coherent pulses of a transmitted signal, spaced at integer multiples of Pulse Repetition Interval (PRI) T_{Pulse} as:

$$S_{Tx}(t) = \sum_{m=1}^{M} a(t - mT_{pulse}) \exp(i2\pi f_c t). \tag{2.1}$$

$a(t)$ is a complex-valued modulation term incorporating phase and amplitude modulation of the carrier signal with frequency f_c. The received signal at time t is expressed in three components as:

$$S_{Rx}(t) = S_{Tgt}(t) + S_{Clutter}(t) + S_{Noise}(t). \tag{2.2}$$

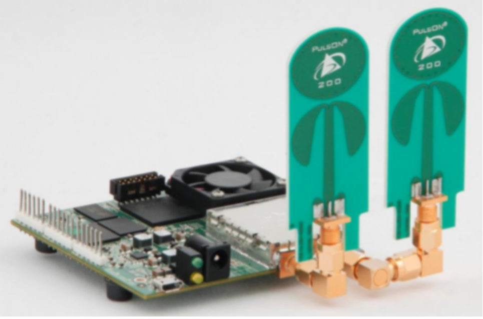


Figure 2.1: A single Humatics PulsON P410 radar sensor with two antennae, one for transmission and one for reception.

The first term $S_{Tgt}(t)$ incorporates reflections from extended targets, modelled here as sets of point scatterers that reflect the incident radiation, giving an attenuated and time-delayed copy of the transmitted signal at the receiver:

$$S_{Tgt}(t) = \sum_{n} \alpha_n S_{Tx}(t - \tau_n(t)), \tag{2.3}$$

where the sum is over all point scatterers, indexed by n, and α_n is the complex-valued attenuation term for each scatterer. $\tau_n(t)$ is the two-way delay time for the scatterer n at time t:

$$\tau_n(t) = \frac{2r_n(t)}{c} = \frac{2(r_n(0) + \Delta r_n)}{c}, \tag{2.4}$$

$$\Delta r_n = \int_0^t \vec{v}_n(t') \cdot \widehat{r}_n(t') dt'. \tag{2.5}$$

Here, $\vec{v}_n(t)$ is the instantaneous velocity of the scatterer n at time t and the adopted convention is that the radial velocity component away from the transmitter is positive. The range to the scatterer n is indicated with $r_n(t)$, where the hat symbol $\widehat{}$ is used to indicate a unit vector. Assuming that the velocity in a single coherent processing interval (CPI) is approximately constant, the two-way round-trip delay is written as:

$$\tau_n(t) = \frac{2(r_n(0) + v_n t)}{c}, \tag{2.6}$$

with v_n the radial component of the velocity of the scatterer n away from the transmitter.

Substituting equations 2.1 and 2.6 into 2.3 gives:

$$\begin{aligned} &S_{Tgt}(t) = \\ &\sum_{n} \alpha_n \sum_{m=1}^{M} a(t - \frac{2r_n(0)}{c} - mT_{pulse}) \ldots \\ &\exp(2\pi i f_c t(1 - \frac{2v_n}{c})). \end{aligned} \tag{2.7}$$

Here, the term $\frac{2r_n(0)}{c}$ in the exponent is absorbed in the attenuation term α_n. The displacement $v_n t$ over a single CPI is assumed to be much smaller than the resolution in range, and is thus omitted from the modulation term a. Finally, equation (2.7) is rewritten as a function of discrete coordinates t' (fast-time) and m (slow-time). $t \equiv t'$ mod mT_{pulse}, which for clarity will be denoted by $t' = t - mT_{Pulse}$, with $t' < T_{Pulse}$ implicit. The resulting equation is:

$$\begin{aligned} &S_{Tgt}[t', m] = \\ &\sum_{n} \alpha_n a(t' - \frac{2r_n(0)}{c}) \exp(2\pi i f_c(t' + mT_{Pulse})(1 - \frac{2v_n}{c})). \end{aligned} \tag{2.8}$$

In order to model the clutter term in equation 2.2, equation 2.8 with the target model is used, but under the assumption that the clutter is primarily static, which is reasonable for the context of this work on indoor monitoring. This means that $v_n \approx 0$, giving in discrete notation:

$$S_{Clutter}[t', m] = \sum_n \alpha_n a(t' - \frac{2r_n(0)}{c}) \exp(2\pi i f_c(t' + mT_{Pulse})). \tag{2.9}$$

The noise term $S_{Noise}(t)$ is assumed to be thermal in nature and is thus included as Gaussian white noise.

The ensemble of the aforementioned three components is finally expressed for one CPI as $S_{Rx}[t', m]$:

$$S_{Rx}[t', m] = S_{Tgt}[t', m] + S_{clutter}[t', m] + S_{Noise}[t', m]. \tag{2.10}$$

The discrete signal $S_{Rx}[t', m]$ collected over multiple pulses is reshaped as a 2D matrix with dimensions fast-time and slow-time, respectively representing range (i.e., the physical distance of scatterers from the radar) and time (i.e., the time sequence of radar pulses one after the other). This 2D range-time matrix is denoted by: $\mathscr{R}_{r,t}$. This complex-valued range-time representation is used throughout this thesis work as a starting point for subsequent processing steps.

2.3. VELOCITY PROCESSING

From the range-time matrices, two representations pertaining to target velocity distributions are computed. The first is the range-Doppler matrix, obtained through the application of a discrete Fourier transform over the slow-time dimension m in a CPI. Using the following property and denoting the Fourier transform over variable m with $\mathscr{F}_m$:

$$\mathscr{F}_m[ce^{2\pi iam}] = c\delta(k - a); \quad a, m, k \in \mathbb{R} \tag{2.11}$$

the target term in (2.8) after the transform becomes:

$$\mathscr{F}[S_{Tgt}[t', m]] = \sum_n \alpha_n a(t' - \frac{2r_n(0)}{c}) \exp(2\pi i f_c(1 - \frac{2v_n}{c})t')\delta(k - f_c(1 - \frac{2v_n}{c})T_{Pulse}), \tag{2.12}$$

which is 0 except where:

$$k = f_c(1 - \frac{2v_n}{c})T_{Pulse}. \tag{2.13}$$

For the clutter term $S_{Noise}[t', m]$, as v_n is assumed to be 0, the transformed term is 0 except where:

$$k = f_c T_{Pulse}. \tag{2.14}$$

The Gaussian noise term $S_{Noise}[t', m]$ transforms to a scaled Gaussian. Summing and reshaping the transformed terms yields a complex matrix with dimensions of fast-time t' and frequency k, the latter of which is directly related to the radial velocities v_n of the scatterers through equations (2.13) and (2.14). The complex matrix is thus denoted with range and velocity variables respectively as $\mathscr{RD}_{r,v}$.

A second velocity representation that will be used throughout this thesis is the velocity-time or spectrogram representation $\mathscr{V}_{v,t}$. First, a Fourier transform of the target term $S_{Tgt}[t',m]$ from equation (2.8) is applied in the fast-time dimension as follows:

$$\begin{aligned}\mathscr{F}_{t'}[S_{Tgt}[t',m]] &= \sum_n \alpha_n \mathscr{F}[a(t' - \frac{2r_n(0)}{c})] \circledast \delta(k' - (f_c(1-\frac{2v_n}{c})))\dots\\ &\qquad \exp(2\pi i f_c(1-\frac{2v_n}{c})mT_{Pulse}))\\ \mathscr{F}_{t'}[S_{Tgt}[t',m]] &= \sum_n \alpha_n \exp(\frac{2ik'r_n(0)}{c})\mathscr{F}_{t'}[a(t')] \circledast \delta(k' - (f_c(1-\frac{2v_n}{c})))\dots\\ &\qquad \exp(2\pi i f_c(1-\frac{2v_n}{c})mT_{Pulse})),\end{aligned} \tag{2.15}$$

where the second line follows from the time-shifting property of the Fourier transform $\mathscr{F}_{t'}[f(t'-t'_0)] = \exp(-ik't'_0)\mathscr{F}_{t'}[f(t')]$. Denoting the Fourier transform of the modulation term as $\mathscr{F}_{t'}[a(t')] = A(k')$ and utilising the sifting property of the Dirac delta yields the following:

$$\begin{aligned}\mathscr{F}_{t'}[S_{Tgt}[t',m]] &= \sum_n \alpha_n \exp(\frac{2ir_n(0)f_c(1-\frac{2v_n}{c})}{c})A(f_c(1-\frac{2v_n}{c}))\dots\\ &\qquad \exp(2\pi i f_c(1-\frac{2v_n}{c})mT_{Pulse}))\end{aligned} \tag{2.16}$$

Finally, a second Fourier transform over dimension m yields a fast-time-independent velocity representation:

$$\begin{aligned}\mathscr{F}_m[\mathscr{F}_{t'}[S_{Tgt}[t',m]] &= \sum_n \alpha_n \exp(\frac{2ir_n(0)f_c(1-\frac{2v_n}{c})}{c})A(f_c(1-\frac{2v_n}{c}))\dots\\ &\qquad \delta(k - f_c(1-\frac{2v_n}{c})T_{Pulse}))\end{aligned} \tag{2.17}$$

which as before is non-zero only for $k = f_c(1-\frac{2v_n}{c})T_{Pulse}$. Equation (2.17) is only dependent on the conjugate variable k of the slow-time m, which directly relates to the radial velocities v_n of the scatterers. To obtain the spectrogram representation, (2.17) is applied for an arbitrary number of sequential or optionally overlapping CPIs, forming a new time dimension, and yielding $\mathscr{V}_{v,t}$.

2.4. Dataset of Continuous Human Activities

An extensive dataset with human activities has been collected and made public to benchmark the methods proposed in this thesis [66]. This dataset consists of sequences of two minutes each, captured by a network of five cooperating PulsOn radar sensor nodes. The data comprise sequences from 14 participants of varying age, gender, height and body type. It is important to note that this dataset is unique in its kind and features several key characteristics that make it valuable for research efforts pertaining to the classification of ADL, notably:

- **Dataset size.** Each participant has performed 30 different sequences, each two minutes in duration. In total there is therefore 840 min of data, which is much greater than alternative datasets that are available in the same domain of radar-based HAR, such as 480 min for the OPERAnet dataset [71], 119 min for the dataset of the University of Alabama [30], and 28 min for the dataset of the University of Glasgow [23].

- **Randomisation and variety.** The motions that the participants have performed in the experimental space feature great variety and are randomised in location and orientation. They are furthermore unconstrained in activity duration, and activity onset times are also determined randomly by the participants. This variety and randomisation yields large benefits for the training of classification networks, whilst simultaneously offering a very challenging task with regards to continuous classification of ADL.

- **Sequence types.** The sequences included in the dataset are varied in composition, as further detailed in Section 2.4.2. Among the 30 different types, notably there are realistic mixes of 9 different activities, offering a challenging test bed for classification methods. Critical events such as falls are included and incorporated in the sequence compositions as well, including two different modalities of falling, i.e. from a stationary position or while walking.

- **Distributed Sensor Network**. All activities are recorded simultaneously by five sensor nodes, spaced around the experimental area. Aside from effectively increasing the amount of data five-fold, the simultaneous capture enables new classification approaches to be developed. Access to data captured from multiple orientations yields unparalleled potential for the development of sensor fusion approaches, the benefit of which will be demonstrated in Chapters 3 and 5 of this thesis.

In the rest of this section, the sensor network geometry is discussed in sub-section 2.4.1, with a description of experimental measurement area. The sensor parameters used in the collection are also provided. Subsection 2.4.2 discusses the sequence types and contains examples of radar data in different representations.

2.4.1. SENSOR GEOMETRY AND DATA CAPTURE PARAMETERS

The capabilities of the simultaneous operation of the PulsOn UWB units is a result of pulse integration, and pseudorandom encoding of the Pulse Repetition Interval of each individual sensor. Every sensor can be individually set to a unique code channel that governs minute variations in PRI. Furthermore, each unit can coherently integrate a preset amount of pulses, in order to increase the ratio between consistent targets and clutter, and randomly distributed noise. Combining the PRI coding and the pulse integration allows for each sensor to emit simultaneously, with the PRI variations ensuring that pulses from sensors on different channels will only contribute to an elevated noise floor (i.e., instead of generating more destructive interference) due to the relative shifts in time that are effectively applied.

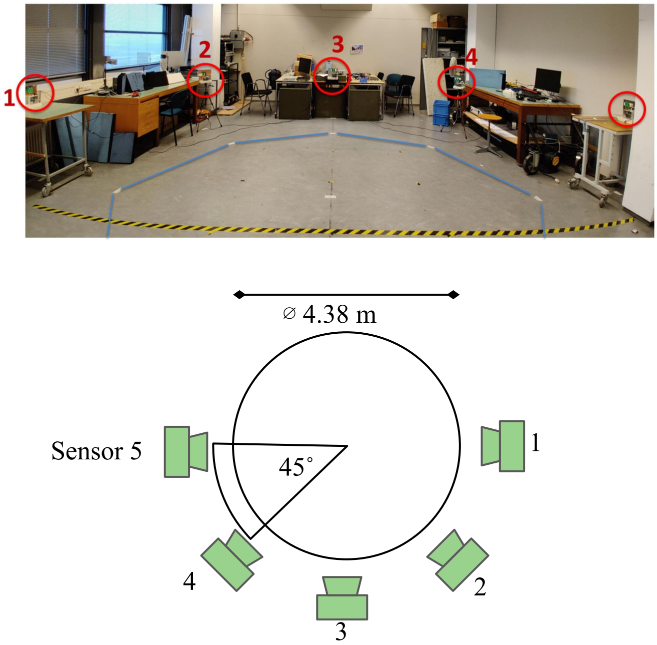


Figure 2.2: Photograph (top) and diagram (bottom) of the experimental measurement area. Five sensors are arranged in a semicircle, spaced at regular 45° intervals with a 6.38 m diameter of the semicircle.

The five radar sensors are arranged in a semicircle, as shown in Figure 2.2. They are spaced at regular 45° intervals with a 6.38 m diameter of the semicircle. The activity area is a circle with diameter 4.38 m, concentric to the semicircle of the sensor network. The sensors are placed approximately 1 m above the floor to point approximately at the torso of the subjects while performing the sequences of activities. Although the observation area is clear of objects, some static clutter is present in the laboratory in the form of e.g., desks and metal shelving that may cause clutter and multipath returns.

The PRF of each sensor node is set to 122 Hz, resulting in a maximum unambiguous velocity of ±2.13 $\mathrm{m\,s^{-1}}$. Each radar sensor is equipped with an antenna with a pattern that is approximately symmetric in azimuth. Combined with the single-channel nature of the sensors, this means that angle-of-arrival estimation can not be performed with a single sensor.

2.4.2. Sequence Characteristics

In total, nine different activities are performed in the activity sequences, listed in Table 2.1. The participants range in age from 20 to 37 years and comprise 10 males and 4 females. The sequence types are diverse in the order and combination of the nine activities, including walking around the measurement area and falling at various locations and orientations. 'Mixed' sequences containing all nine activities performed with random duration and locations are also included to provide a more realistic and challenging

Table 2.1: Distribution of samples among the nine activity classes of the collected dataset.

Activity	**Fraction**
Walk	43.12%
Stationary	11.52%
Sit Down on Chair	4.33%
Stand Up (From sitting)	3.96%
Bending (Sitting)	9.87%
Bending (From Standing)	10.51%
Falling (From Walking)	2.81%
Falling (From Standing)	4.39%
Standing Up (After Fall)	9.48%
Total	100%

set of data. The ground truth labels are created by the participants themselves by means of a handheld remote control signalling transitions from one activity to another.

To demonstrate the various data representations, the processing techniques previously described in this chapter are applied on a two minute sequence captured with a single PulsON P410 UWB sensor. Figure 2.3 presents, from top to bottom, a range-time matrix $\mathscr{R}_{r,t}$, a spectrogram $\mathscr{V}_{v,t}$, and a range-Doppler map $\mathscr{RD}_{r,v}$. For the spectrogram, a window size of 1.05 s is utilised (corresponding to 128 CPI), with an overlap of 0.98 s. For the range-Doppler map, the window size is also 1.05 s.

Figure 2.3 presents a sequence where the participant performs a variety of activities, including walking around, sitting down, and falling twice. To better illustrate the challenges associated with discriminating various activity types, Figure 2.4 displays a collection of spectrograms corresponding to different activities.

Finally, as equations (2.13) and (2.14) imply, the radar sensor is principally only capable of measuring the radial velocities associated with targets. The result is a strong effect on the recorded velocity of the relative location of the sensor and the target. Figure 2.5 displays a fall event in spectrogram representation, but captured by several radar sensors simultaneously, hence from different line of sight positions with the respect to the direction of the fall itself. For the top sensor in particular, most of the motions are in a plane that is orthogonal to the radar line-of-sight. As such, less information can be gathered from this particular spectrogram, as evidenced by the weaker signature.

2.5. Conclusion

This chapter introduces details pertaining to the dataset that has been co-developed and used extensively throughout this thesis to benchmark the proposed activity classification methods. The UWB pulsed radar system is described, along with a related signal model, followed by processing approaches that yield range-time and velocity representations. The dataset itself is described in detail. The dataset is a valuable contribution to the scientific community due to its large size, the variety in activity sequence types, subject locations, and orientations. Unique to this dataset is the distributed network of

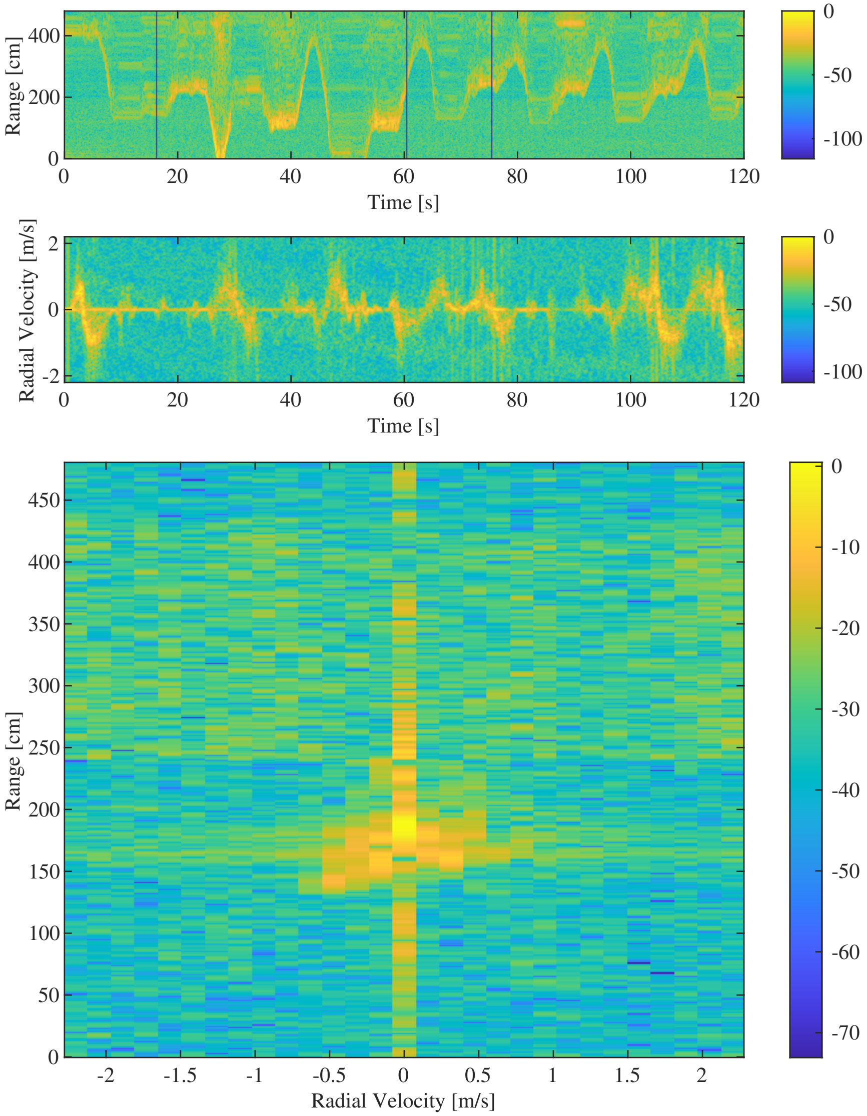


Figure 2.3: Example representations of radar data using a sequence of human activities analysed in this thesis. Signal power is shown in dB, normalised with respect to the highest measured value. *Top*: Range-time matrix containing two minutes of human activities of a single subject. *Middle*: Velocity-time (spectrogram) representation of the same two minute data sequence. *Bottom*: Range-Doppler map of a 1.05 s segment of the same data sequence.

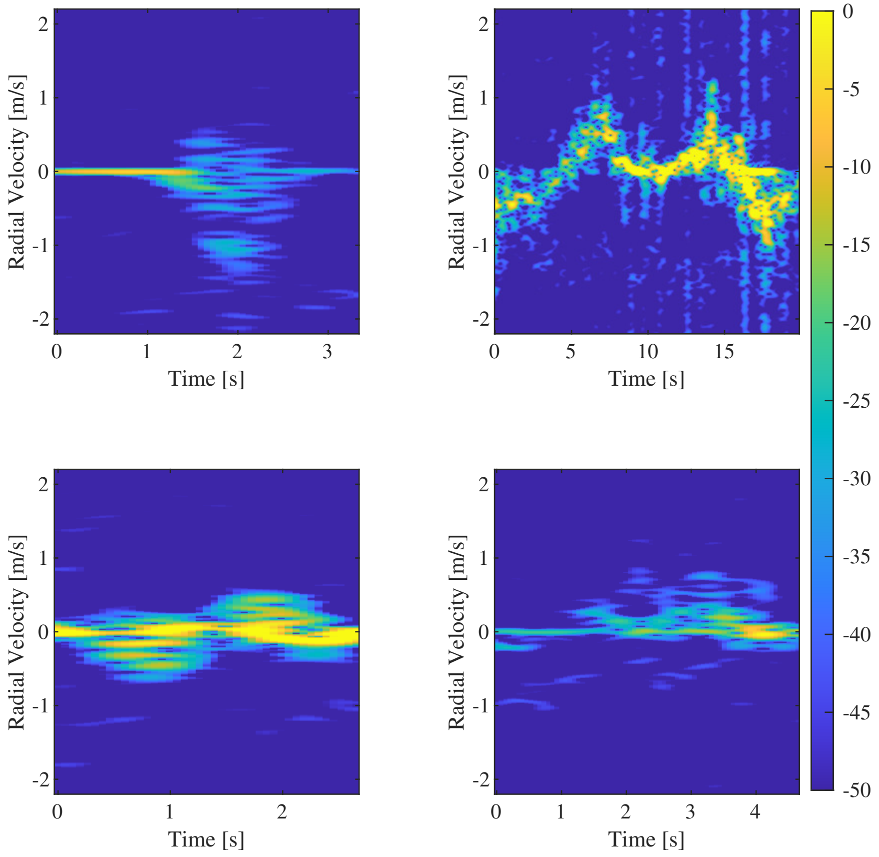


Figure 2.4: Different activities presented in spectrogram representation. Left to right, top to bottom: A fall from a stationary position, walking around, bending down from a seated position, standing up from the ground. Signal power is shown in dB, normalised with respect to the highest measured value.

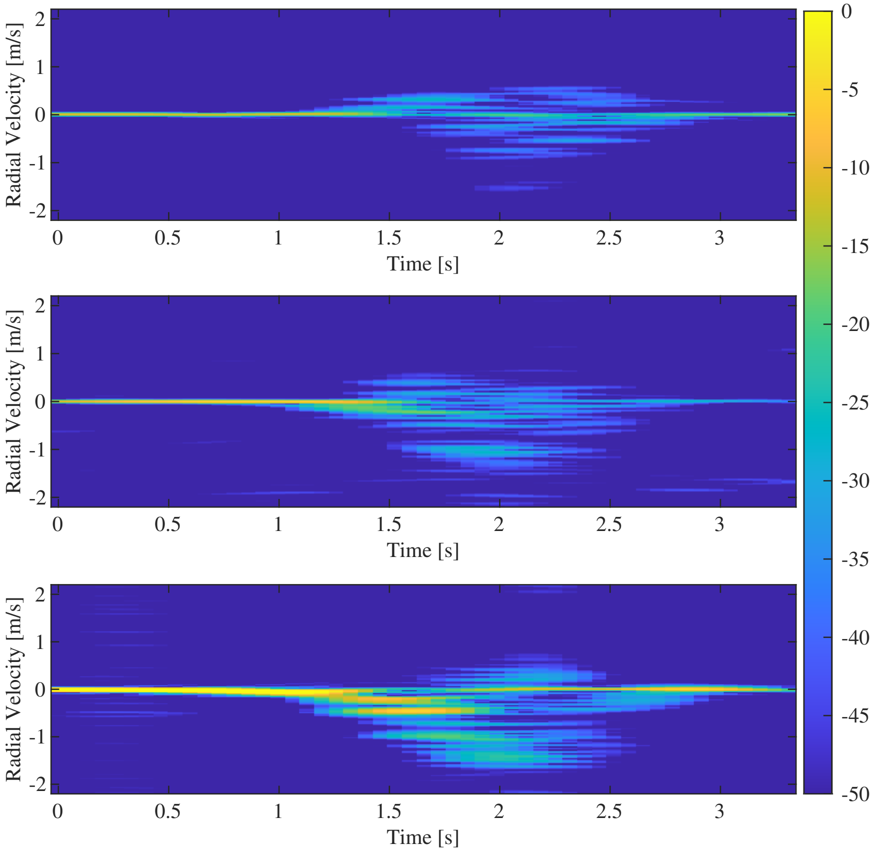


Figure 2.5: Fall event from three different angles of observation in a horizontal plane at waist height. Signal power is shown in dB, normalised with respect to the highest measured value.

cooperating sensors that has been utilised to collect it, allowing for the development of sensor fusion techniques, which will be detailed in the following chapters.

3

Radar Point Cloud Representations for Activity Classification

Typically, radar data is stored for the entirety of the field of view of the radar sensor, including the empty space. This is sometimes referred to as the radar data (hyper)cube. Consequently, a lot of memory is required to store all of this data, and it takes a long time for computer algorithms to go through all of the data. In this chapter I proposed a method of finding and storing only the data that corresponds to a human under observation, thus reducing the size of the data considerably. In other words, the method gets rid of the data that is about the empty space around the human, as if cutting out a person from a photograph. The small data size allows for a strong machine learning classifier to be used, and I demonstrate the effectiveness of this approach experimentally.

Parts of this chapter have been published as:

Kruse, N. C., Fioranelli, F., & Yarovoy, A. (2023). *Radar Point Cloud Processing Methods for Human Activity Classification With Point Transformer Networks.* IEEE Transactions on Radar Systems, 2, 1–12.

Kruse, N. C., Fioranelli, F., & Yarovoy, A. (2023). *Continuous Human Activity Classification with Radar Point Clouds and Point Transformer Networks.* 2023 20th European Radar Conference (EuRAD), September, 302–305.

3.1. Introduction

The need for human activity classification arises significantly in the field of healthcare, where observations of patient behaviours and activities can give medical professionals important information necessary for personalised and timely care, thus reducing the need for hospitalisation and intensive intervention. These observations include fall detection [13, 14], gait analysis [15], and general classification of the activities performed by a patient [57, 72–74]. Radar is an advantageous sensor for such monitoring tasks, as no recognisable imagery is captured of the subject, thus preserving privacy. Additionally, radar sensors function in low-light conditions, do not require the patients to wear, carry, or interact with any device, and have some through-wall capabilities [74].

Many earlier works studying the classification of Activities of Daily Living (ADL) incorporated forms of feature extraction that were either manually constructed, or based on e.g., Principal Component Analysis (PCA) [17, 28, 32, 42, 75]. Later research more often utilised machine learning approaches to perform automatic feature extraction, notably Convolutional Neural Networks (CNN) architectures applied on radar data treated as images [76, 77]. The introduction of classifiers based on Recurrent Neural Networks (RNN) led to reported increases in classification performance [23, 78, 79], and to a focus shift from 'snapshot-style' classification, where activities are recorded and classified in isolation, to classification of continuous sequences where the different activities are performed one after the other in a more realistic manner [23, 29, 79–82].

More recently, driven by the progress in novel architectures of neural networks, Transformers have been proposed to overcome the limitations of RNNs when dealing with long sequences of data, and their difficulty in learning patterns across segments of the data that are far from each other [83]. Also for Human Activity Recognition (HAR), the utilisation of transformer-type architectures has been seen in various forms. For example, in [84–87], image-based approaches are taken through the implementation of Vision Transformer networks. The radar data is represented in a velocity-time (spectrogram) domain, and the role of the Vision Transformer is that of an image recognition network. An alternative approach is taken in [88], where a mm-Wave radar sensor outputs a 3D point cloud (PC) of the subject, which is converted into voxels. A CNN extracts features from the voxelised representation, with a Transformer architecture performing classification based on these vectors. Finally, in [57], feature extraction is performed on several 2D representations of human motion by means of an auto-encoder, and classification is again achieved by means of a Transformer model operating on the output of the auto-encoders.

In this chapter, a novel method is proposed to perform activity classification on continuous sequences of human activities of unconstrained duration. This is realised by processing the output of a Single Input Single Output (SISO) radar sensor into a multi-dimensional point cloud representation of reflection intensity in a range-Doppler-time space. From this data representation, classification is subsequently performed by means of a Point Transformer (PT) model [67]. It should be noted that, because the SISO sensors are unable to provide Direction-of-Arrival information individually, the processed point clouds contain no 3D spatial coordinates aside from the range to the capturing radar sensor, and are thus atypical compared to point cloud representations derived from mm-wave Multiple Input Multiple Output (MIMO) radars often encountered in lit-

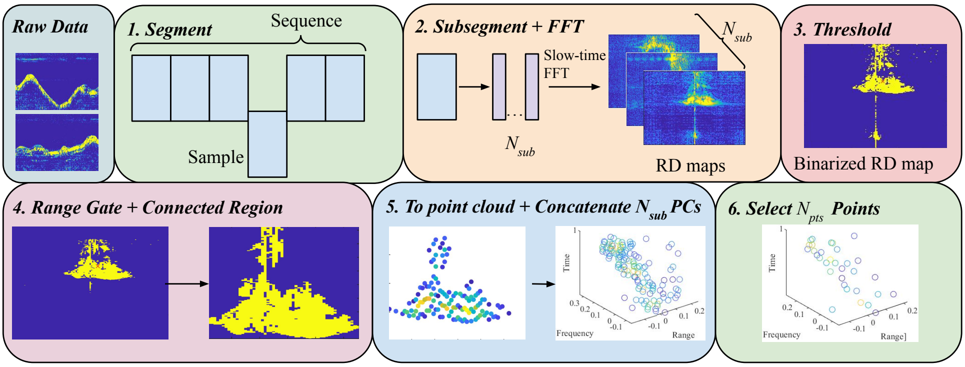


Figure 3.1: Proposed processing pipeline for the generation of point cloud samples suitable for Point Transformer networks from continuous sequences of human activities recorded by a SISO radar.

erature [36, 85, 89]. In this regard, the proposed processing method can be considered similar to a dimensionality reduction method, which preserves the salient features necessary for effective classification and encodes them into a point cloud format suitable to leverage the classification capabilities of PT networks. The point cloud format enables the storage of only those points that are associated with the extended target or targets, and not the remaining values of the often sparse input domains. The reduction in sample size becomes more apparent as input dimensions are added, since the point cloud format size scales linearly with input dimension, whereas conventional grid-based formats scale exponentially. Furthermore, the proposed method is applied and experimentally verified with a network consisting of multiple radar sensors, for which different sensor fusion techniques are implemented and demonstrated to increase the overall classification performance. The contributions of this chapter can be summarised as:

- A novel approach for radar-based continuous activity classification utilizing Point Transformer networks.
- Utilisation of sensor fusion techniques to enhance classification performance in scenarios with networks of multiple radars.
- Demonstration of the method efficacy through experimental comparison with reference works utilizing the same challenging dataset including 2 minutes sequences composed of 9 activities performed by 14 participants.

The remainder of this chapter is organised as follows. In Section 3.2, the proposed point cloud processing methods are discussed, as well as the sensor fusion techniques, and the Point Transformer model. Results are displayed in Section 3.3 and discussed further in Section 3.4. Section 3.5 finally contains the conclusion to this chapter.

3.2. Proposed Methodology

This section contains a description of the proposed processing method for generating samples in point cloud representation, suited for classification with the Point Transformer architecture. Also included is a description of the Point Transformer architecture itself. The sample generation method is visually summarised in Figure 3.1 from raw range-time radar data to point cloud samples to be used an input to the Point Transformer.

3.2.1. Signal Model and Dataset

The signal model used in this chapter is described in Chapter 2. The transmitted signal is again modeled as a coherent pulse train of M pulses, spaced at integer multiples of Pulse Repetition Interval (PRI) T_{Pulse} as:

$$S_{Tx}(t) = \sum_{m=1}^{M} a(t - mT_{pulse}) \exp\left(i2\pi f_c t\right). \tag{3.1}$$

$a(t)$ is a complex-valued modulation term incorporating phase and amplitude modulation of the carrier signal with frequency f_c. The processing steps in Chapter 2 are followed to obtain the fast-time/slow-time representation from equation (2.10), which is rewritten as a complex range-time matrix $\mathscr{R}_{r,t}$.

The dataset utilised for experimental validation is the same as used throughout this thesis work [66]. A full description of this dataset is available in Chapter 2. Each individual radar node outputs a real-valued vector representing the backscattered signal. A Hilbert transform is applied to obtain the quadrature component of the signal, and a fast-time/slow-time matrix is then constructed. Thus, for each sequence the resultant complex-valued range-time matrix forms the starting point for further processing and is denoted by *Raw Data* in Figure 3.1.

3.2.2. Processing steps

Sequence Segmentation

Since the raw data consist of continuous sequences of activities, an approach has to be selected on how to perform classification in a continuous manner via a form of segmentation; this is the first step in the processing pipeline (step **1.** in Figure 3.1). In broad terms, these approaches can be grouped into two categories, i.e., Time step-based Classification and Window-based Classification. For *time step-based classification*, individual samples are short in duration and often correspond to some fundamental measuring scale of the sensing system, such as the PRF or the individual time window over which spectrograms have been calculated. This approach is suited for classification with RNN-type networks and has been used for continuous activity classification [19, 29, 82]. Downsides to this approach are firstly the ambiguity that arises when evaluating an activity over such a short duration, especially during activity transitions, and secondly the low information content of a single sample, necessitating the use of a classifier that evaluates multiple samples jointly. In *window-based classification*, the sample durations are longer and aimed at capturing either an entire activity, or a fraction encapsulating enough of the activity to perform classification [12]. Windows can be fixed in duration, or adapted in real time to better suit the duration of the activity that is occurring. For this approach, window duration selection is critical. Too short, and the information content of the sample is insufficient to properly evaluate the activity, too long and the window may include multiple activities, leading to ambiguity issues in defining the sample label.

For this chapter, a fixed-window approach is adopted, comparable to sliding-window methods in the literature [14, 32], but with no overlap between windows. This approach is taken in order to best utilise the Point Transformer architecture, which is in principle unable to link multiple samples in a time sequence, but which performs well at the task of feature extraction and classification from the point cloud samples [67]. Based on previous work [90], a segment length of $256 \cdot \mathrm{PRI} = 2099\,\mathrm{ms}$ (hereafter denoted as 2 s for the sake of brevity) is chosen to maximise sample information content and minimise the classification performance impact of sample label ambiguity. Each sample is assigned a single activity class label based on the sequence ground truth. If multiple ground truth labels occur in the segment, for example at the transition between two different activities, the final label is determined by a majority-rule, i.e. by assigning the label of the activity that occupies the majority of the segment.

Point Cloud Generation

After the initial sequence segmentation step, each data sample is further split into N_{sub} sub-segments of equal duration (step **2.** in Figure 3.1). This step is added so that within a sample there is an evolution over time, important for distinguishing complementary activities such as e.g. standing up from sitting down. A Fast Fourier Transform (FFT) is performed over the slow-time dimension of all N_{sub} sub-segments as described in Section 2.3, resulting in N_{sub} time-ordered Range-Doppler (RD) maps $\mathscr{RD}_{r,v}$. In previous work, the optimum value for N_{sub} was determined to be 6 [90] for the dataset used, balancing Doppler and time resolution for the highly dynamic nature of human motions. Combined with the 2 s sample duration, this provides a Doppler resolution of $10.5\,\mathrm{cm\,s^{-1}}$.

An initial noise-rejection step (step **3.** in Figure 3.1) is then performed through the application of a static threshold filter on the magnitude of each of the N_{sub} RD maps, as:

$$R_{i,j}^{bin} = \begin{cases} 1, & \text{if } |R_{i,j}| > \alpha \max_{i,j} |R_{i,j}| \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

with α the threshold level between $[0,1]$, $|R_{i,j}|$ the magnitude of an element of the RD map in dB scale, and $R_{i,j}^{bin}$ a binary element of the filtered RD map. The optimal value for α is determined experimentally to be 0.8 [90].

To further select useful information on the moving human, a range gate is applied (step **4.** in Figure 3.1), centered on the centroid of the binarised RD maps, whose range-Doppler coordinates are calculated by:

$$\vec{R}_c = \begin{bmatrix} \frac{\sum_i \sum_j i \cdot R^{bin}(i,j)}{\sum_i \sum_j} \\ \frac{\sum_i \sum_j j \cdot R^{bin}(i,j)}{\sum_i \sum_j}. \end{bmatrix} \tag{3.3}$$

where the indices i and j represent range and Doppler bins respectively. The range gate extends for 2 m in length, which is deemed enough to fully cover an adult laying down in case of a fall.

Human limb motions are primarily restricted to rotations at the joints. For this reason, a range-Doppler representation of a human will tend to occupy a connected region within a RD map, barring occlusions. This property is exploited to achieve additional noise/clutter suppression, by selecting only the largest connected regions in the binary RD maps (step **4.** in Figure 3.1). This is realised by evaluating the regions with adjacent nonzero pixels, and selecting those three regions that feature the largest area. The choice for three regions is empirical: in general it is found that the largest region is orders of magnitude larger than the second largest and thus selecting three regions ensures that the area of interest is captured without the inclusion of further noisy regions. Every remaining pixel from the connected regions is now stored as a point in a list and assigned four basic variables: range, Doppler, intensity of the corresponding pixel in the RD map, and time expressed as a fraction of N_{sub}.

The point lists for all N_{sub} sub-segments can then be concatenated (step **5.** in Figure 3.1), resulting in a point cloud with variables range, Doppler/velocity, time (expressed as the sub-segment number within the segment), and intensity.

Afterwards, a uniform amount of N_{pts} points per segment is selected to ensure consistency with the Point Transformer network input size (step **6.** in Figure 3.1). To this end, the points are sorted by intensity, and the highest N_{pts} points are kept. If initially there are fewer than N_{pts} points, randomly sampled points are duplicated until the required number is met. The duplicate points do not alter the physical characteristics of the sample and are the first to be removed in the down-sampling layers of the Point Transformer network. The resulting cloud consists of N_{pts} points, with the variables range, Doppler/velocity, time, and intensity, and represents a 2 s observation of human motion.

Fusion approaches

To more effectively utilise the simultaneous data capture from five radar nodes in a network, fusion techniques are implemented in the form of:

- Feature fusion during segment processing
- Decision fusion during classification

In the case of feature fusion, the methodology from Figure 3.1 is followed until step **5.** where, prior to the point up/down-sampling operation, the points from the nodes to be fused are pooled into a single point cloud. Afterwards, the up/down-sampling operation is applied to obtain a point cloud of the required number of points. This approach, hereafter referred to as *Simple Fusion*, does not transform any of the variables, nor are there any new variables added to the resulting fused point cloud.

For decision fusion, each data segment is processed using only data from a single node, and segments from multiple nodes are processed by multiple, independent Point Transformer networks in parallel. Each network outputs a vector $\vec{y}_c^n$ of prediction confidences for the activity classes under consideration, where $n \in N \subseteq \{1,2,3,4,5\}$ is the node number and c indicates the respective class label. Two decision fusion methods are utilised in this chapter as follows:

Softmax Summation The output vectors for all Point Transformer networks are summed, and the class with the highest resulting prediction confidence is selected for the segment label $\tilde{y}$:

$$\tilde{y} = \arg\max_{c} \sum_{n} \vec{y}_c^n \tag{3.4}$$

Majority Voting The segment label is determined through the statistical mode of the set of individual predictions $\tilde{y}^n$, which are in turn acquired by selecting the class with the highest prediction confidence in the output vector from each network:

$$\tilde{y} = \quad \mathrm{mod}\ \{\tilde{y}^n\}, \quad \tilde{y}^n = \arg\max_{c} \vec{y}_c^n \tag{3.5}$$

Position and Angle Estimation

When a network of multiple radar nodes with known locations is available, as in the considered dataset, additional features for the point cloud can be extracted from the data. Utilizing a subset $N \subseteq \{1,2,3,4,5\}$ of the five nodes, position estimation can be for instance performed through multilateration, given that $|N| \geq 3$. For this, it is necessary that the extended target be simplified as a point target. First the approximate target range r_n is determined for each node by convolving the range profile of the node with a triangular function with a base of $2a = 45\,\mathrm{cm}$

$$f(r) = \begin{cases} 1 - |\frac{r}{a}|, & |r| < a \\ 0, & \text{otherwise,} \end{cases} \tag{3.6}$$

which is determined empirically and reduces the influence of noise/clutter on the target range estimate. The range at which the convolution reaches the maximum is selected as the range to the point target. The target location is then estimated following [91], and expressed in terms of the target ranges as:

$$r_n^2 = (\tilde{x} - x_n)^2 + \left(\tilde{y} - y_n\right)^2, \quad n \in N \subseteq \{1,2,3,4,5\}, \tag{3.7}$$

where r_n denotes the range to the node with index n, $(\tilde{x}, \tilde{y})$ the estimated point target coordinates, and (x_n, y_n) the coordinates of the node with index n. Expanding the squares of equation 3.7 and subtracting one of the $|N|$ equations from the others (in this case the last, with index $n = |N|$) gives:

$$\begin{aligned} \left(r_n^2 - r_{|N|}^2\right) - \left(x_{|N|}^2 - x_n^2\right) - \left(y_{|N|}^2 - y_n^2\right) &= \ldots \\ 2\left(x_{|N|} - x_n\right)\tilde{x} + 2\left(y_{|N|} - y_n\right)\tilde{y} & \\ = 2\left[x_{|N|} - x_n \quad y_{|N|} - y_n\right] \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix}. & \end{aligned} \tag{3.8}$$

For $|N| \geq 3$ this system of equations is overdetermined, and a solution can be approximated by minimizing the mean square error (MSE) $\arg\min_{\tilde{x}} ||\mathbf{A}\vec{x} - \vec{b}||^2$, where $\mathbf{A}\vec{x}$ corresponds to the right hand side of (3.8), and $\vec{b}$ to the left hand side. Minimizing the MSE corresponds to finding the null space of the gradient of the MSE expression: $\mathbf{A}^T\mathbf{A}\vec{x} - \mathbf{A}^T\vec{b} = \vec{0}$, which leads to the following expression for the point target location estimate:

$$\vec{x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\vec{b} \tag{3.9}$$

With the location estimates calculated, target tracking is performed by means of the ($\alpha\beta$)-filter which assumes a system model comprising two state variables, one of which is determined by integrating the other. For the current case of point target tracking this assumption is assumed to hold. The input of the filter is the location estimates $\vec{x}$, and the output consists of location and velocity variables $\vec{\tilde{x}}$ and $\vec{v}$.

From the velocity output $\vec{v}$, the approximate target *heading* is acquired with respect to an axis $(\vec{C}-\vec{x}_3)$, i.e., an axis crossing through node 3 and the experimental area centre point, as shown in Figure 3.2. To improve the reliability of the heading estimate, an absolute velocity cut-off is implemented at $||\vec{v}||^2 < 1000\,\mathrm{cm^2\,s^{-2}}$. Below this velocity, the heading variable is set to a dummy value to allow for location jitter in the tracker output. Thus, the computation of the heading ϕ_H is expressed as:

$$\cos(\theta_H) = \begin{cases} \frac{(\vec{C}-\vec{x}_3)\cdot\vec{v}}{||(\vec{C}-\vec{x}_3)||\;||\vec{v}||}, & ||\vec{v}||^2 > 1000\,\mathrm{cm^2\,s^{-2}} \\ c_{dummy}, & \text{otherwise.} \end{cases} \tag{3.10}$$

Aside from the target heading, two other angular variables are computed. Firstly, the *Angle off-Boresight* represents the angle between an axis through a node centre-line $(\vec{C}-\vec{x}_n)$ and the line between the target and the observing radar node $(\vec{\tilde{x}}-\vec{x}_n)$, shown in Figure 3.2 and expressed as follows:

$$\cos(\theta_{AoB,n}) = \frac{(\vec{C}-\vec{x}_n)\cdot(\vec{\tilde{x}}-\vec{x}_n)}{||(\vec{C}-\vec{x}_n)||\;||(\vec{\tilde{x}}-\vec{x}_n)||}. \tag{3.11}$$

Additionally, if heading information is available, the *aspect angle* can be computed, which is the angle between the direction of motion of the target and the line between the target and the observing radar node:

$$\cos(\theta_{AA,n}) = \begin{cases} \frac{(\vec{\tilde{x}}-\vec{x}_n)\cdot\vec{v}}{||(\vec{\tilde{x}}-\vec{x}_n)||\;||\vec{v}||}, & ||\vec{v}||^2 > 1000\,\mathrm{cm^2\,s^{-2}} \\ c_{dummy}, & \text{otherwise.} \end{cases} \tag{3.12}$$

These angles can be added to the features of the point cloud when feature fusion is used, i.e., at the aforementioned step when point cloud samples from different radar nodes are combined together. It should be noted that, for a given node, *heading* and *aspect angle* are dependent variables.

3.2.3. Point Transformer Network as Classifier

The classification network architecture utilised in this chapter is the Point Transformer proposed in [67], which is a point cloud-based variant in the transformer family of architectures [92]. At the heart of a transformer architecture lies the so-called ‘attention mechanism’ [83]. In a network layer, the attention mechanism allows for the value of an input element to influence the weights of other elements on the output dynamically. These dependencies can be learned from the data in contrast to conventional neural networks where weights are input-independent and only change during training.

The Point Transformer model proposed in [67] consists of four sequential stages, which in turn are composed of a Point Transformer block and a point down-sampling

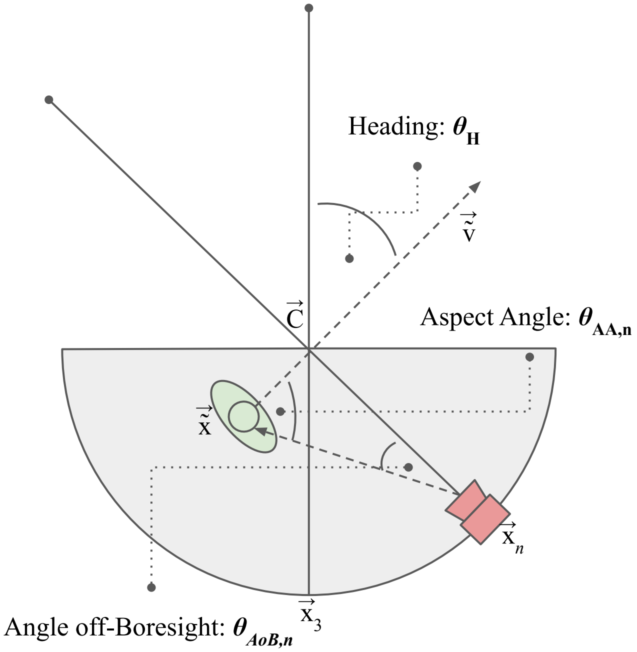


Figure 3.2: Representation of the three angles computed as part of the feature fusion method. Only a single radar node is shown out of 5 nodes for simplicity. $\vec{x}_n$ denotes the location of the node with index n. *Heading angle* is the direction of motion of the target with respect to the axis through node 3 and the experimental area centre point $\vec{C}$. *Aspect angle* is the angle between the direction of motion of the target $\vec{\tilde{v}}$ and the line between the target $\vec{\tilde{x}}$ and the observing node. *Angle off-Boresight* is the angle from the node centre-line to the estimated target location $\vec{\tilde{x}}$.

block. The former block features a Point Transformer layer which is where the attention mechanism is utilised to exploit correlations between points in an input cloud. The Point Transformer code is adapted from [89].

3.3. EXPERIMENTAL RESULTS

In this section results are presented for various experiments, intending to gauge the classification performance of the proposed method with a Point Transformer classifier. Firstly, results are discussed pertaining to single-radar performance, followed by results where feature fusion has been applied to more effectively utilise multiple radar nodes. Finally, experimental results are discussed where decision fusion has been performed.

In this section, all results shown are based on a sample holdout scheme where 80 % of samples are used for training and validating the classifier, and 20 % of samples are used for testing. Later in Section 3.4 a comparison to reference works is made, where a more challenging Leave-One-Person-Out (L1PO) testing scheme is employed. For all results, two performance metrics will be displayed: the test accuracy on the unseen test dataset, and the macro F1-score which gives clearer insight in the performance on the

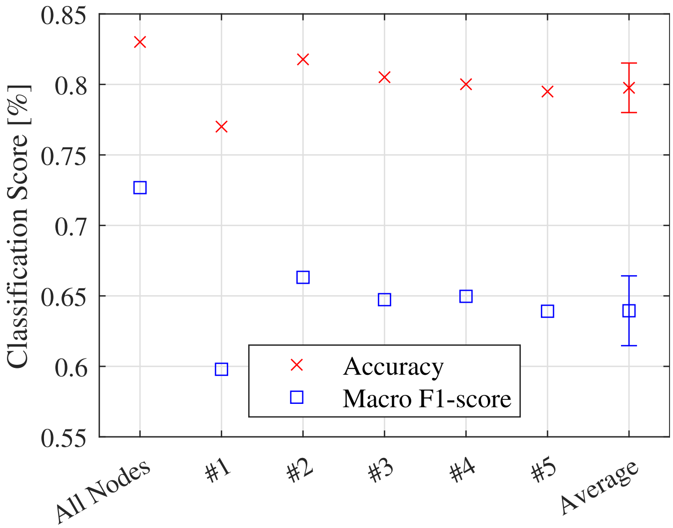


Figure 3.3: Performance metrics for point cloud samples generated with individual radar nodes. For the *All Nodes* result, individual samples are also generated with individual nodes, but all samples are then pooled into a single training and testing dataset. The remaining numbered results and the *Average* result are for the respective nodes and their average. See Figure 2.2 for the relative locations of the nodes.

imbalanced dataset.

3.3.1. Baseline results

Classification metrics for Point Transformers utilizing point cloud samples generated from radar nodes used in isolation, i.e. with no fusion method to combine data from multiple nodes into a single PC, are displayed in Figure 3.3. The deviating performance of node #1 is most likely attributable to strong clutter (large metal shelving) in the vicinity of this node. The comparable results of the other four nodes are interpreted as indicating that the activities in the sequences are diverse in location and orientation, thus making none of the five line of sight directions of the nodes as the dominant, most advantageous one. For the experiment labeled as *All Nodes*, the samples originating from all radar nodes are pooled into a single training and testing dataset. The increased performance reveals that the sample support from a single node is insufficient to maximise performance with the chosen network, regardless of the aforementioned diversity in location and orientation. Reliance on large datasets is indeed a known characteristic of Transformer architectures [93].

3.3.2. Fusion results

Feature Fusion

The results for experiments pertaining to feature fusion approaches are discussed in this subsection and displayed in Figure 3.4. As a reference result for the various fusion methods, the first column *No Fusion* corresponds to the *All Nodes* result from Figure 3.3. For

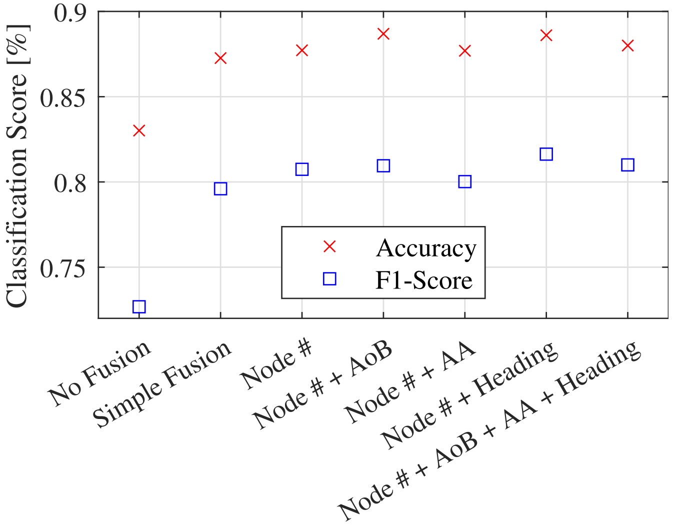


Figure 3.4: Performance metrics for various feature fusion approaches. For all results, the generated point clouds comprise at least the variables range, Doppler, time, and intensity. *No Fusion* involves samples generated with single nodes only and corresponds to the *All Nodes* result from Figure 3.3. The samples for the *Simple Fusion* result are constructed using all nodes, as described in Section 3.2.2.
The remaining results are denoted by the variables that are added to the point clouds thanks to the combination of data from multiple radar nodes. *Node #*: Extra PC variable containing the originating node for each point. *AoB*: Angle off-Boresight, angle between axis through node centerline and estimated target location. *AA*: Aspect Angle, angle between axis through node centerline and estimated target heading. *Heading*: Estimated target heading relative to fixed axis through experiment area centre and node #3.

all results, the generated point clouds comprise at least the variables range, Doppler, time (relative to start of the segment), and intensity. The *Simple Fusion* column displays the performance for samples generated utilizing data from all five nodes, as described in section 3.2.2, but without the addition of any new variables to the point clouds, i.e., each point has 4 variables. The notable performance increase between *No Fusion* and *Simple Fusion* reveals that, even without information on the originating node of a point in the point cloud, the Point Transformer can make stronger inferences of the target class when it has access to observations from multiple directions. This contrast is even stronger when considering the fact that the sample support for *No Fusion*, i.e., the amount of data for training the network, is effectively five times larger than for the fused case. Furthermore, adding a variable to label every point in the point cloud with its originating node does not markedly improve performance as can be seen in the *Node #* column. This last result is surprising, as intuitively the originating node would be instrumental information in the effective fusion of the information from various nodes. It is possible that a learned clustering approach is embedded in the trained Point Transformer, based on distinctly grouped range and Doppler values in the multi-node point cloud. This learned clustering would allow the implicit assignment of originating node number by the network, and thus explain the low performance gained by explicitly including the node numbers as variables.

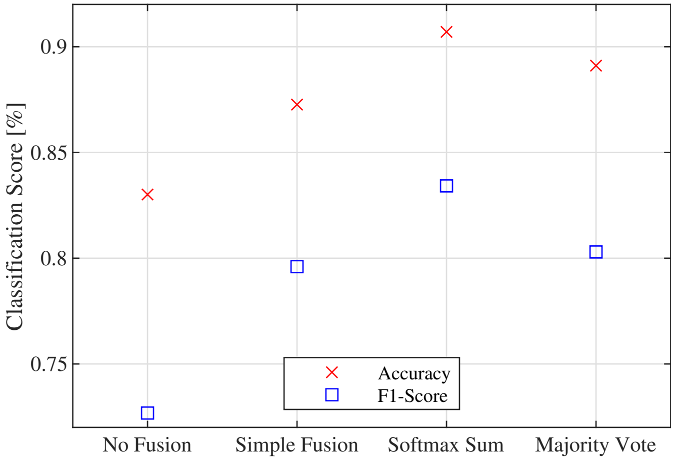


Figure 3.5: Performance metrics for various decision fusion approaches. *No Fusion* again corresponds to the *All Nodes* result from Figure 3.3, and *Simple Fusion* corresponds to *Simple Fusion* in Figure 3.4. For the *Softmax Sum* result, the class prediction vectors for all five nodes are summed as described in Section 3.2.2 and the maximum value is selected as the predicted class. For *Majority Vote*, the statistical mode of the class predictions of all five nodes is taken to be the decision-fused class prediction.

Three additional experiments are performed where different observation conditions are incorporated in the point clouds in the form of the additional angular variables defined in the previous section. The intended outcome is the inclusion of information in the point cloud on the observation quality of a particular node, for example the approximate component of a motion taking place perpendicular to the radar line-of-sight. The computation of the estimates of *Angle off-Boresight*, *Aspect Angle*, and *Heading* is described in Section 3.2.2. However, it is noted that no appreciable performance increase is attained in any of these experiments, indicating that no salient information appears to be utilised from these variables.

Decision Fusion

Two methods of decision fusion are implemented and evaluated for this chapter, with results displayed in Figure 3.5 and specifics detailed in Section 3.2.2. For reference, results for the case of *No Fusion* and *Simple Fusion* are shown as before. It should be noted that *Simple Fusion* and the two decision fusion methods *Softmax Sum* and *Majority Vote* are mutually exclusive procedures. The results reveal a significant difference in performance between the majority voting method and the softmax summation method, from which it is concluded that correct label predictions are dominated by a small fraction of comparatively confident node predictions. Inspection of the data shows no bias of these high confidence values to any specific node, implying that other factors such as potentially observation angle may be responsible.

To further explore this result, Figure 3.6 displays classification performance after decision fusion with a variable amount of nodes, i.e., by performing an ablation study where specific nodes are left out from the decision fusion process. Firstly it can be seen

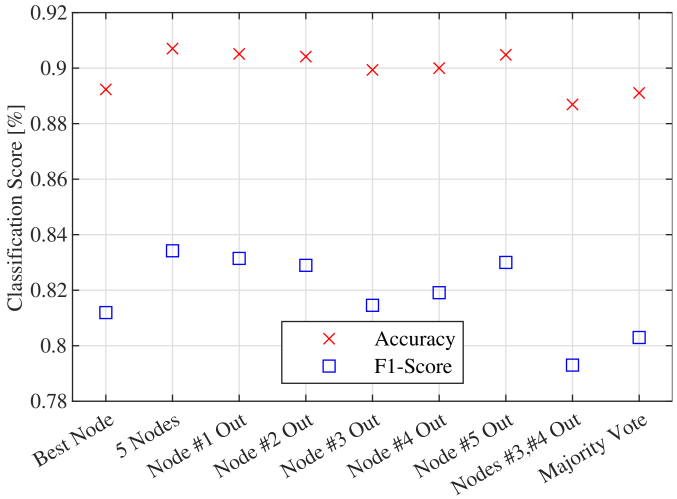


Figure 3.6: Performance metrics for the node ablation study. *Best Node* indicates the result when only the most confident node is selected for prediction. *5 Nodes* indicates the performance of the *Softmax Sum* decision fusion scheme applied to all five nodes. *Node # out* represents the result when a specific node is left out of the *Softmax Sum* decision fusion process, and *Majority Vote* indicates the result for the majority voting scheme applied to all five nodes.

in the *Best Node* column that using only the most confident node for prediction (i.e. the highest softmax value) gives worse performance than fusing any combination of four nodes, indicating that the fusion process improves prediction capabilities. It is also apparent that none of the nodes have a particularly large impact if left out, and only when the two most influential nodes (3&4) are left out does the performance drop below the value achieved with *Majority Vote*.

Finally, a study is performed on the average performance gained by increasing the amount of sensors used in the *Softmax Sum* decision fusion scheme. The results are shown in Figure 3.7, with the amount of nodes fused on the horizontal axis. For this experiment, all possible combinations of e.g., 2, 3, and 4 nodes are fused, and the average performance is shown in the figure.

3.4. DISCUSSION

This section contains a discussion on the results obtained, mostly focusing on the computational requirements of the method, a comparison to reference results using the same dataset, as well as an investigation into the most prevalent classification errors.

3.4.1. COMPUTATIONAL REQUIREMENTS

The computational requirements of the proposed method are summarised here. In Table 3.1 the proposed method is compared to two alternative methods in terms of sample size and processing time for a full two-minute sequence. The *RD Sequence* row indicates the generation of RD maps at the same resolution and interval as in the first steps of the

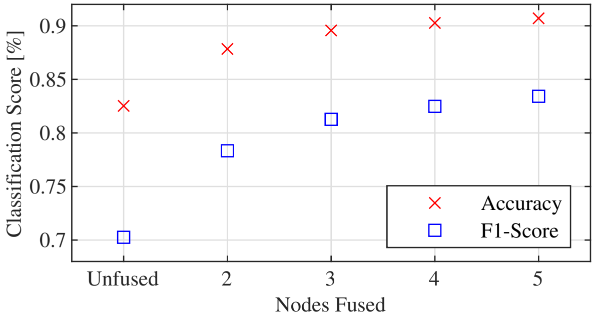


Figure 3.7: Performance metrics for a study on the performance gained by the addition of nodes in the *Softmax Sum* decision fusion scheme. For each amount of nodes on the horizontal axis, all possible combinations of nodes are tested and the average performance is reported.

Table 3.1: Comparison of computational requirements for the proposed method and 2 alternative methods, for the two-minute sequences used in this chapter. *Sample size* refers to the size on disk of the sample(s) constituting one full sequence; *Processing Time* denotes the time required to process a full sequence into sample(s) using a 3.40 GHz i5-7500 CPU. *RD Sequence* denotes the processing of a sequence into a series of RD maps using the same parameters as for the proposed method.

Method	**Sample Size [MB]**	**Processing Time [s]**
Proposed Method	0.74	4.36±0.51
RD Sequence	3.68	1.28±0.04
Spectrogram[38]	1.68	1.27±0.03

proposed pipeline. The *Spectrogram* row is based on the generation of spectrograms and follows the approach in [38]. The longer processing time for the proposed method compared to solely the RD map generation is predominantly due to the steps of range-gating operation and the evaluation of the connected components of the binarised RD maps.

With regards to the Point Transformer model, using an Nvidia RTX 3090 board, a training time of 18 s per epoch is observed, with an inference time of approximately 38 ms per full two-minute sequence. A typical model size is 17 MB.

Lastly, a study is performed to determine the optimal number of points per sample, balancing classification performance and computational requirements. Figure 3.8 displays the results of Point Transformers trained on an increasing amount of points per sample. It can be seen that performance gain is saturated beyond 1024 points.

3.4.2. Performance Comparison

In order to compare classification performance of the method to alternative approaches in literature, two earlier reference works are selected that report method validation on the same experimental dataset [33, 38]. In these works, results are presented for a Leave-One-Person-Out (L1PO) validation scheme, where data from each individual participant are used in turn as the unseen testing set for the classification. As such, the proposed

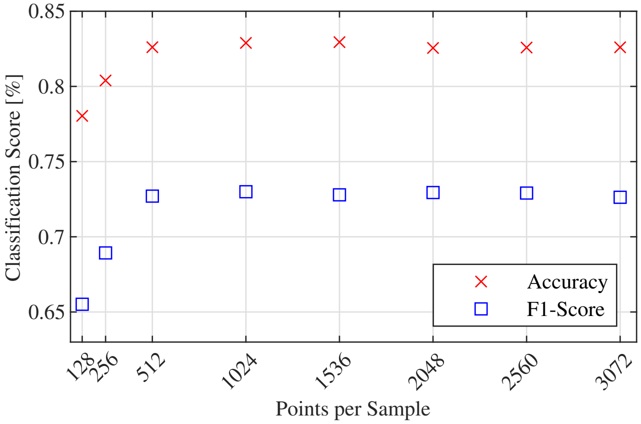


Figure 3.8: Performance metrics for point cloud samples generated with an increasing amounts of points per sample. Performance saturation is observed for networks trained on samples containing more than 1024 points.

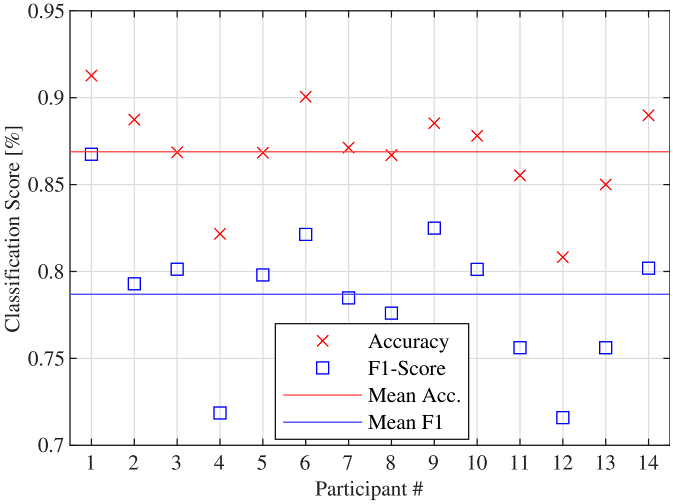


Figure 3.9: Results for the L1PO testing scheme. Averages for test accuracy and macro F1-score across the participants are indicated with horizontal lines.

method in this chapter is also evaluated following this scheme, utilizing point clouds with the variables range, Doppler, time, intensity, node number, and the three additional angular variables discussed in Section 3.2.2. Decision fusion via Softmax Summation is also performed to maximise classification performance.

Figure 3.9 shows macro F1-score and test accuracy for all participants for the nine activity class classification problem, with averages for these metrics across all participants also indicated. Referencing [33], a mean L1PO test accuracy of 85.1 % is reported for

Table 3.2: Merging scheme for the consolidation of the full nine activity classes into a set of five activity classes. Classes are grouped based on similarity.

Constituent Classes	Merged Class
Walking	Walking
Stationary	Stationary
Sitting Down, Standing up (from sitting), Bending (from sitting), Bending (from standing)	In Situ
Falling (from walking), Falling (from stationary)	Falling
Standing up (from ground)	Standing up

Table 3.3: Distribution of Samples Among the Nine Activity Classes for the L1PO Testing Scheme.

Activity	Sample Support	Fraction
Walk	37810	43.12%
Stationary	10100	11.52%
Sit Down	3800	4.33%
Stand Up (Sitting)	3475	3.96%
Bending (Sitting)	8655	9.87%
Bending (Standing)	9220	10.51%
Falling (Walking)	2460	2.81%
Falling (Stationary)	3850	4.39%
Standing Up (Ground)	8315	9.48%
Total	87685	100%

the hybrid CNN-RNN architecture utilised therein, indicating that the proposed method achieves better performance for unseen subjects with an average test accuracy of about 86.9 %.

In [38], multiple classifier models based on RNNs are evaluated on the same dataset, with a focus on performance metrics such as macro F1-score, suited for the class imbalance of the dataset. The respective processed sample supports for the activity classes are shown for the L1PO case in Table 3.3. It should be noted that in [38] a five-class problem was considered, where some of the nine original activity classes are collated in for instance an "in-situ" class. This grouping, shown in Table 3.2, is reproduced here for the comparison with the proposed method, and the results are reported in Table 3.4.

The table reveals that the proposed method achieves increased performance over the bidirectional reference RNN architectures, which themselves offer better classification performance than their unidirectional counterparts. Although there is no temporal information shared between individual 2 s segments in the proposed approach, the point cloud processing method ensures that within a single segment there is information on the activity dynamics through the sub-segmentation procedure. In the *No Fusion* case, the difference in classification performance between the proposed and reference method is greater, indicating efficacy of the fusion methods.

Finally, to complement the results reported in Table 3.4, the method performance

Table 3.4: Macro F1-score results for the proposed classification method and four reference classifiers in [38]. Both single radar performance *No Fusion* and multi radar performance *Fusion* is listed, when available. The results on this table are for a five-class classification problem.

Classifier	No Fusion	Fusion
Proposed	**0.807**	**0.862**
GRU	-	0.778
LSTM	-	0.769
bi-GRU	-	0.844
bi-LSTM	0.773	0.836

is compared to three different implementations of the ResNet-50 [94] CNN architecture on the full nine-class problem. The ResNet-50 model is pretrained on an image dataset from ImageNet containing over a million sample images. Specifically, the three implementations here include:

1. The ResNet-50 network retrained on RD maps computed with 2 s windows.
2. The ResNet-50 network retrained on portions of spectrograms of 2 s duration.
3. The ResNet-50 network retrained on portions of spectrograms of 2 s duration, thresholded at 70% of maximum intensity.

ResNet-50 has been chosen for this comparison as a representative CNN architecture which proved successful for image-based classification tasks, including using radar data. In all three implementations, ResNet stages 1 through 4 are frozen, and the last stage 5 is retrained and fine-tuned using the full radar dataset. The test accuracy achieved using the RD input, the spectrogram input, and the thresholded spectrogram input, are 0.688, 0.655, and 0.618 respectively, whereas the proposed method attains a test accuracy of 0.825 without fusion.

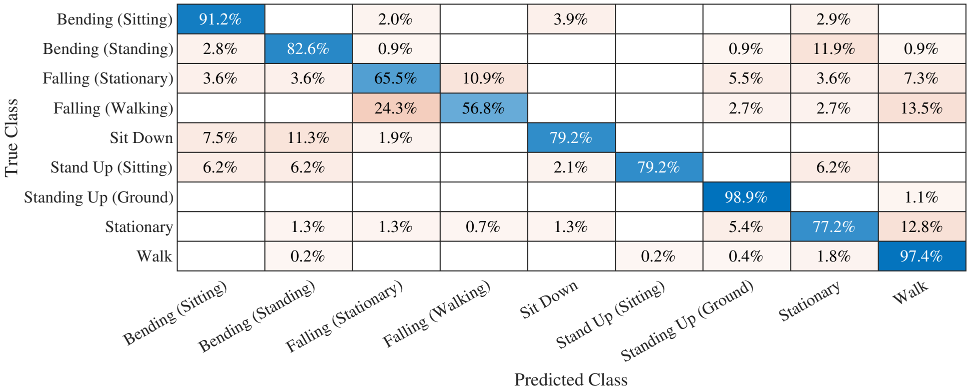
3.4.3. ERROR ANALYSIS

Figure 3.10 displays an example of a confusion matrix for one of the 14 participants used in the L1PO testing scheme. In the matrix, some of the more prevalent errors can be identified such as:

1. Confusion between similar activities (e.g. the two fall types, i.e., one from a stationary position and one from walking, or bending from standing and sitting down)
2. Boundary-type errors (seen in the columns for *Stationary* and *Walk*)

The first of these error types is not only due to the challenging nature of differentiating similar activities, but can also partly be attributed to the segmentation strategy. For example, a 2 s window that does not incorporate the start of a fall will generally not provide enough information to distinguish the type of fall that has occurred.

The second error type is mainly due to the fixed segmentation strategy. Due to the nature of the sequences, many of the activity transitions include either *Walk* or *Stationary*, such as *Walk*→*Falling (Walking)* and *Stationary*→*Bending (Standing)*→*Stationary*.



True Class \ Predicted Class	Bending (Sitting)	Bending (Standing)	Falling (Stationary)	Falling (Walking)	Sit Down	Stand Up (Sitting)	Standing Up (Ground)	Stationary	Walk
Bending (Sitting)	91.2%		2.0%		3.9%			2.9%	
Bending (Standing)	2.8%	82.6%	0.9%				0.9%	11.9%	0.9%
Falling (Stationary)	3.6%	3.6%	65.5%	10.9%			5.5%	3.6%	7.3%
Falling (Walking)			24.3%	56.8%			2.7%	2.7%	13.5%
Sit Down	7.5%	11.3%	1.9%		79.2%				
Stand Up (Sitting)	6.2%	6.2%			2.1%	79.2%		6.2%	
Standing Up (Ground)							98.9%		1.1%
Stationary		1.3%	1.3%	0.7%	1.3%		5.4%	77.2%	12.8%
Walk		0.2%				0.2%	0.4%	1.8%	97.4%

Figure 3.10: Example of a confusion matrix for one of the 14 participants in the L1PO testing scheme. Percentages are row-normalised fractions.

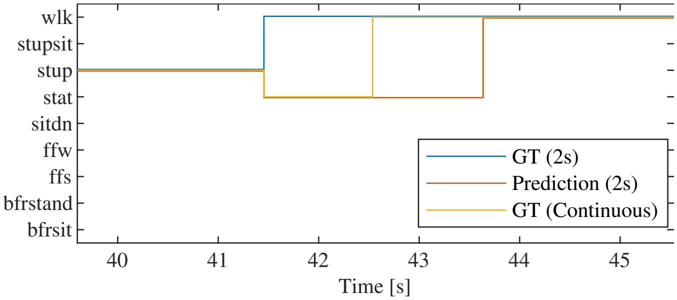


Figure 3.11: Example of a classification error due to sample label ambiguity. Ground truth is indicated in yellow for the original time steps of 1 PRI (8.2 ms), and in light blue for 2 s segments. In this case, the sample label ambiguity is apparent due to the activity transition occurring exactly halfway into the 2 s segment between ~41.5 s and ~43.6 s

When a sample window includes such a transition, ambiguity arises on the activity type, and errors can occur. An example of such an error is given in Figure 3.11, where a 2 s segment is in nearly equal parts *Stationary* and *Walk*, and the resulting ground truth label is *Stationary*. However, the erroneous prediction from the classifier is in this case *Walk*.

The errors associated with the fixed-window segmentation approach can possibly be partially mitigated by adopting a variable-size segmentation approach. These approaches most often entail the monitoring of a quantity that is descriptive of the presence, absence, or change of an activity within the data sequence. This includes for instance spectrogram bandwidth [30], range profile variance [95], and spectrogram entropy [96]. With these approaches, an input sequence is divided into segments of different durations that ideally contain a single activity, which can subsequently be classified per individual segment. Regardless of the segmentation method utilised, ambiguous segments will always occur, in no small part due to the subjective nature of defining where the transition between two distinct activities occur. For this reason, the utilisation of a multi-label classification approach [97] can be also beneficial in detecting and classifying the presence of multiple activities within one segment.

3.5. CONCLUSION

This chapter explores the feasibility of radar-based continuous human activity classification utilizing Point Transformer networks and a novel point cloud processing method. The proposed processing method converts SISO radar output into a point cloud data representation in range-Doppler-time space. This approach contrasts methods in literature where targets are required to be localised in 3D space before employing point cloud-based methods. The method is validated on a publicly available dataset [66], gauging both single-radar performance, and performance after feature and decision fusion across a network of five simultaneously operating radar nodes. For the unconstrained nine activity class problem, a single-radar test accuracy and macro F1-score of 81.0 % and 69.8 % are achieved using L1PO validation. After feature and decision fusion, a test accuracy and macro F1-score 86.9 % and 78.7 % are achieved with the same validation strategy.

In future work, temporal relations between samples in an activity sequence will be exploited using alternative transformer architectures to further improve classification performance.

4

Segmentation of Continuous Sequences of Human Activities

The word 'Continuous' in the title of this dissertation represents a key challenge in this research. Classification of an uninterrupted sequence of activities is an unsolved problem as of yet. For a large part this is due to computer algorithms having trouble with activities transitioning into eachother. In this chapter I propose an approach of 'divide et impera'. The idea is to take the sequence of activities, and to extract a quantity from the data that behaves in a predictable manner when an activity changes into the next. With this information, the sequence can be split into segments that ideally have only a single activity each, which can be classified individually.

Parts of this chapter have been published as:

N. C. Kruse, R. G. Guendel, F. Fioranelli, and A. G. Yarovoy, *Segmentation of Micro-Doppler Signatures of Human Sequential Activities using Rényi Entropy* in International Conference on Radar Systems (RADAR 2022), 2022, pp. 435–440.

and are currently under review in:

N. C. Kruse, A. Daalman, F. Fioranelli, and A. G. Yarovoy, *Radar Point Cloud-based Continuous Human Activity Classification Using Rényi Entropy Segmentation Methods*, IEEE Transactions on Radar Systems.

4.1. Introduction

Automated human monitoring is a beneficial capability for healthcare professionals. Systems that offer these capabilities can monitor e.g., vital signs [21, 22, 98], detect harmful events such as falls [13, 14], and can ensure a timely response of relevant personnel to assist vulnerable people. Wearable sensors, such as Inertial Measurement Units (IMUs) have been utilised for these tasks [99, 100], but are not always feasible, as subjects may forget or object to wear these devices. Camera-based observation is highly dependent on lighting conditions, with darkness and glaring reflections potentially impeding proper operation. Additionally, video-based monitoring comes with elevated privacy concerns. Radar sensing presents a remote monitoring solution that can overcome the limitations of the aforementioned sensor modalities, and offers a promising, versatile platform for human monitoring tasks.

Radar-based classification of Activities of Daily Living (ADL) is therefore an active area of research, with a considerable amount of efforts being directed to the open challenge of continuous classification. Continuous classification here refers to the classification of extended sequences of human activities, with each activity of unknown duration, and frequently with activities smoothly transitioning into each other. Three main approaches can be identified in the current literature on continuous activity classification: sliding window methods, Recurrent Neural Networks (RNN) or similar models to process the entire data sequence, and segmentation-based methods. Sliding window approaches include for instance the work in [53], where a set of features is computed from a window, and used as input to classifiers such as, among others, Support Vector Machines (SVM). In [54], a coarse sliding window (i.e., 30 s with 10 s overlap) is employed in conjunction with a Convolutional Neural Network (CNN). Approaches where activity sequences are processed in their entirety feature such classifiers as (Bidirectional) Recurrent Neural Networks ((Bi)RNNs) [38, 55], hybrid models consisting of CNNs and (Bidirectional) Long Short Term Memory ((Bi)LSTM) networks [33, 56], Gated Recurrent Unit networks [101], and Transformer-based models [57].

The approaches to continuous classification that focus on segmentation of activity sequences generally do so by discriminating periods of motion from those without motion. In [58], the beginning and end of an activity are identified based on fluctuations in wi-fi Channel State Information (CSI) variance. A similar approach is taken in [59], where the amount of detections from the data is used as an indicator of an activity starting and stopping. The STA/LTA (Short-Term Average / Long-Term Average) change detection algorithm is employed in [30, 60] to segment the original sequence into motion-detected intervals, based on the spectrogram envelope. Motion-detected intervals can also be determined through machine learning methods, as demonstrated in [61]. The activity sequence is divided into fixed-size segments, and a CNN-LSTM network is utilised to determine the presence of an activity. A dynamic segment duration algorithm is proposed in [99], where an initial 1 s segment is processed to yield information on the bandwidth. If the 1 s segment appears very dynamic, the segment length is extended.

The reviewed existing approaches in literature for continuous classification of ADL have notable disadvantages. Sliding window methods with a non-zero overlap require input data to be processed multiple times, degrading the computational efficiency of these solutions. Furthermore, window sizes are primarily fixed, which may not necessarily be optimal for the classification of different activity types in realistic sequences. RNNs, transformers, and other neural networks that process complete sequences, are generally less computationally lightweight as their counterparts that are designed to classify comparatively shorter data sequences with single activities, and require a large amount of data for effective training. Among the segmentation methods, the prevalent means of identifying individual activities is the presence of a pause in motion between subsequent activities, revealed by the lack of returned power or detections in the data under test. This pause can be absent for activities that smoothly transition into the succeeding, often without a clear stop of the body movement from one to another.

To address the above issues, in this chapter a continuous ADL classification method is proposed that consists of three main components: a segmentation algorithm, a segment processing algorithm, and a classification network. Segmentation of the input activity sequences is based on a quantity derived from the micro Doppler spectrograms, namely the Rényi entropy [68]. This scalar quantity is constructed to be representative of the distribution of velocities at a given time [102], and monitoring this quantity for fluctuations gives an indication in the transition between activities. Classification of the individual segments extracted in this way is then achieved by first processing the data to a Point Cloud (PC) representation, and then utilising a Point Transformer network inspired by [67]. The proposed method, as well as two alternative segmentation methods, is validated on a publicly available experimental dataset which consists of a variety of sequences of nine human activities. It is demonstrated that favourable performance figures can be achieved on a challenging test dataset, with a test accuracy and macro F1-score of 89.3 % and 82.0 % respectively. These metrics are an improvement on previous approaches utilising the same dataset for testing.

The contributions in this chapter can be summarised as follows:

- A novel approach for classification of continuous sequences of ADL, based on segmentation with Rényi entropy, a quantity describing more complex fluctuations in the data than simpler power-based indicators.

- Experimental validation and performance evaluation of the proposed method with respect to reference methods from the literature, showing that the proposed method outperforms the reference approaches with a Leave-One-Person-Out (L1PO) test accuracy and macro F1-score of respectively 89.3 % and 82.0 %.

- Comparison of the proposed segmentation method with two alternative segmentation approaches. This study reveals that a segmentation approach based on Machine Learning can be situationally more effective, achieving a L1PO macro F1-score of 86.0 %.

The remainder of this chapter is organised as follows: the proposed method, as well as the two reference segmentation methods, are described in Section 4.2. The experimental case study, designed to evaluate the performance of the proposed method with respect to reference works in literature, is outlined in Section 4.3. Results for the case study are presented and discussed in Section 4.4, and conclusions follow in Section 4.5.

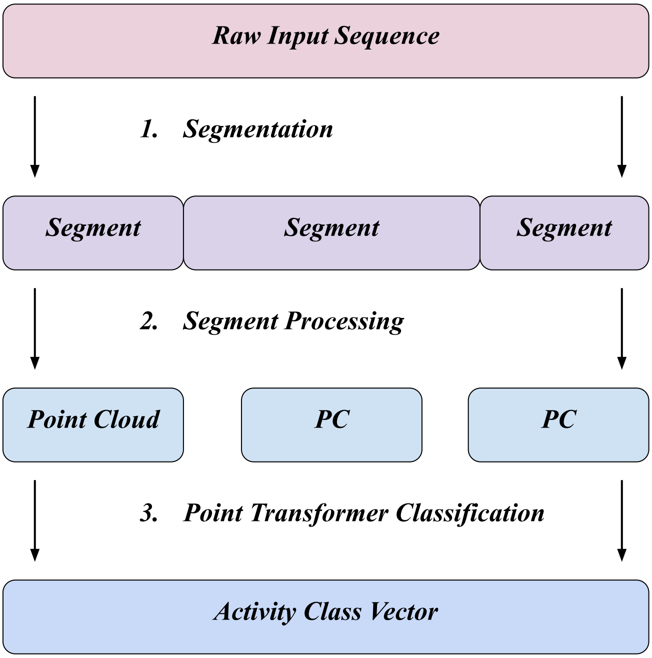


Figure 4.1: The proposed three-stage classification pipeline. A raw input sequence of human activities is segmented into single activities by a segmentation algorithm, the individual segments are processed to generate a point cloud (PC) format, and classification is finally achieved by means of a Point Transformer neural network.

4.2. Proposed Method

The proposed classification method consists of three main elements: a sequence segmentation algorithm, a processing algorithm for the individual segments, and a Point Transformer neural network as classifier. In this section, the three components will be described, as well as relevant preprocessing steps. Additionally, two alternative segmentation methods are outlined in Section 4.2.3.

4.2.1. Signal Model and Notation

Analogous to the analytical computations in Chapter 2, range-time representations of radar data are acquired as follows: real-valued backscattering signal vectors are considered for a set $\mathcal{N}$ of distributed pulsed radar systems. The cardinality of set $\mathcal{N}$ is denoted by N. The quadrature components of the N vectors are obtained through the application of a Hilbert transform, and the resulting complex-valued vectors are reshaped into N complex-valued fast-time slow-time matrices. The fast-time and slow-time dimensions correspond to range and time respectively, and these range-time representations are denoted by $\mathscr{R}_{r,t}$.

Velocity-time (spectrogram) representations are computed from the complex $\mathscr{R}_{r,t}$, following Section 2.3. To this end, a Fast Fourier Transform is first applied in the range dimension, yielding an intermediate matrix equivalent to equation (2.15). Subsequently, a Short-time Fourier transform (STFT) is applied to the frequency bin corresponding to the centre frequency of the system. This operation yields the spectrogram $\mathcal{V}_{v,t}$.

4.2.2. SEGMENTATION

The aim of segmentation is to divide a continuous sequence of various activities into segments that contain only a single activity, simplifying the subsequent task of classification. To this end, a time-dependent quantity is extracted from the spectrogram representation that is indicative of the kinematics of the activity being performed. The quantity that is selected for this purpose is the Rényi Entropy H_α [68], which is defined as:

$$H_\alpha(\mathscr{P}) = \frac{1}{1-\alpha}\log\left(\sum_i p_i^\alpha\right) \tag{4.1}$$

for $0 < \alpha < \infty$ and $\alpha \neq 1$. In the above formula, $\mathscr{P}$ is a discrete probability distribution with p_i a vector of probabilities of the members of the distribution. The parameter α weighs the contribution of individual elements in $\mathscr{P}$ on the overall entropy. For this research, the probability distribution is instead replaced by a velocity distribution. Specifically, at time t_i, a single time bin $\mathcal{V}_{v,t_i}$ of an input spectrogram is normalised by dividing the time bin vector by the sum of its constituent elements, and subsequently utilised for the entropy calculation, as:

$$H_\alpha(t_i) \equiv H_\alpha(\mathcal{V}_{v,t_i}) = \frac{1}{1-\alpha}\log\left(\sum_v\left(\frac{\mathcal{V}_{v,t_i}}{\sum_v \mathcal{V}_{v,t_i}}\right)^\alpha\right). \tag{4.2}$$

Changes in the distribution of velocities of the human target result in corresponding changes in entropy extracted from the spectrogram. Sudden entropy changes are associated with changes in activity, and fluctuations are monitored to indicate these activity transition events. To this end, an inequality is established that serves as an entropy difference threshold:

$$|H_\alpha(t) - H_\alpha(t - T_{lag})| \geq \beta\sigma_H, \tag{4.3}$$

where T_{lag} is a parameter governing the time scale of fluctuations that will trigger an activity transition event. σ_H represents the standard deviation in entropy over a longer interval, for example the full duration of the activity sequence. The parameter β is a constant that determines the required fluctuation magnitude to indicate an activity transition. The utilisation of the Rényi entropy over alternatives such as signal power or spectrogram envelope, such as in [30], is motivated by the invariance of the entropy under a set of key transformations of the distribution of velocities. Specifically, the value of the Rényi entropy will remain unchanged under both a scaling of the input velocity distribution, and a translation. A scaling of the velocity distribution occurs when an otherwise identical motion is performed faster or slower, or when the motion is performed in different orientations, resulting in an altered projection of the target velocity profile onto the radar line-of-sight. In both these cases, it is desirable to have an unchanged entropy, as the nature of the motion remains the same. A translation of the velocity distribution corresponds to an offset in the bulk velocity of the target, implying that the same motion is performed whilst moving. When a human target is walking in various directions, the entropy thus remains constant.

Whenever the threshold value $\beta\sigma_H$ is exceeded, the precise time is recorded, yielding a vector of transition event time stamps. With the acquisition of the vector of transition time stamps, the input sequence can be segmented into a set of range-time matrices of varying durations. It is assumed that there is a minimum duration to human activities of interest, and a minimum segment duration is thus implemented as a parameter. Segments that are shorter than this parameter T_{min} are split evenly across the adjoining two segments.

4.2.3. ALTERNATIVE SEGMENTATION METHODS

Two alternative reference segmentation methods will be employed to gauge the effectiveness of the proposed segmentation approach. They are described in this section.

STA/LTA

Short Term Average over Long Term Average (STA/LTA) is a change detection algorithm based on the ratio between two moving averages of different window size. The short and long term averages of a generic signal $s(t)$ are given by:

$$\mathrm{STA}(s(t)) = \sum_{t'=t-T_s}^{t} \frac{s(t')}{T_s} \tag{4.4}$$

$$\mathrm{LTA}(s(t)) = \sum_{t'=t-T_l}^{t} \frac{s(t')}{T_l}, \tag{4.5}$$

where T_s and T_l indicate the short and long window durations respectively. The initiation and termination of a segment is given by the following two sets of conditions respectively:

$$\frac{\mathrm{STA}(H_\alpha(t))}{\mathrm{LTA}(H_\alpha(t))} > \sigma_2 \ \& \ \mathrm{STA}(H_\alpha(t)) > \sigma_1 \tag{4.6}$$

$$\frac{\mathrm{STA}(H_\alpha(t))}{\mathrm{LTA}(H_\alpha(t))} < \sigma_2 \ \& \ \mathrm{LTA}(H_\alpha(t)) < \sigma_3. \tag{4.7}$$

The entropy $H_\alpha(t)$ is here taken as the specific signal of interest, and $[\sigma_1, \sigma_2, \sigma_3]$ are method parameters. σ_1 and σ_3 are thresholds that govern the required entropy increase and decrease for detecting the onset of a segment. The required ratio between the short and long term averages to indicate the start of a segment is given by the remaining parameter σ_2. Together with the short and long window durations T_s and T_l, a total of five parameters are thus required to configure the STA/LTA algorithm.

For the purpose of the case study, the optimal parameters for the STA/LTA algorithm are determined using the built-in genetic algorithm (GA) optimiser in MATLAB. The objective function is a sum of two terms, both in the range $[0,1]$. The first term expresses the suitability of a segment between two detected transitions in terms of the most occurring activity label. The closer to 1, the better the segment captures a single activity. The second term penalises the difference between the amount of detected transitions and the amount of transitions in the ground truth target vector.

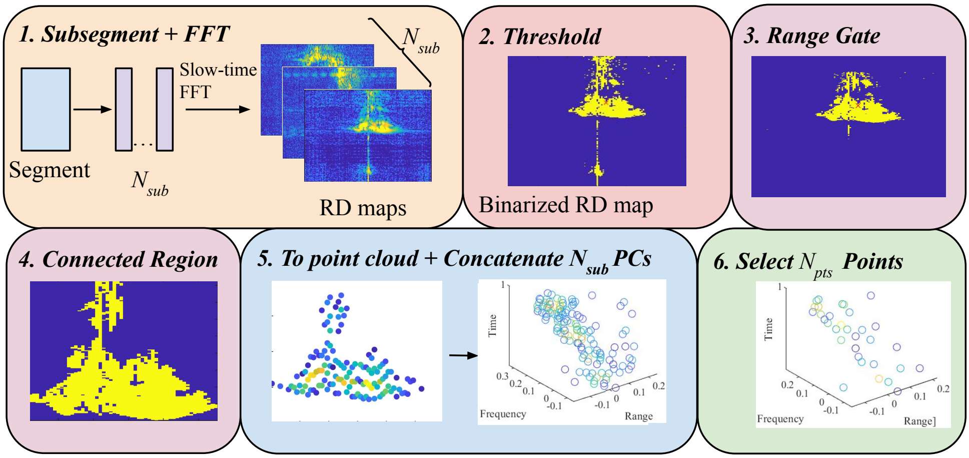


Figure 4.2: Diagram of the proposed segment processing pipeline for the generation of point cloud samples suitable as inputs for Point Transformer networks.

BiLSTM for Segmentation

A machine learning-based approach for segmentation is also investigated. Specifically, a bidirectional Long Short Term Memory (BiLSTM) network is trained to detect transition events from an entropy signal input $H_\alpha(t)$. The target sequence used to train this model is a binary vector of the same length as the entropy input. At every transition between two activities, the value of the target sequence is '1'. It is '0' everywhere else. The target vector is relatively sparse, as the amount of transitions in a sequence is generally orders of magnitude smaller than the duration of the sequence in time steps. This significant sparsity hinders the ability of the model to correctly train and predict transition events. To alleviate this problem, it is here proposed that the target vector is convolved with a rectangular function with a width of approximately 0.33 s. A physical interpretation of this convolution is the non-instantaneous nature of activity transitions and an inevitable degree of subjectivity in defining exactly when they happen. The value of 0.33 s is empirically found to provide good performance without making transition events unrealistically long.

4.2.4. Segment Processing and Classification

Every range-time matrix containing a segment of activities is processed individually, yielding a point cloud (PC) representation. The point cloud representations are computationally more efficient to manipulate than using complete data matrices, and allow for classification by the powerful Point Transformer (PT) family of neural networks [67]. This section describes the point cloud processing and classification. The processing approach is similar to that of Chapter 3, but with the important distinction that segments are no longer fixed in size.

Every segment is first evenly divided into N_{sub} subsegments, as shown in step (1) of Figure 4.2. Each subsegment is thus a complex range time matrix $\mathscr{R}_{r,t}$, and an FFT along the time axis yields a set of N_{sub} range Doppler maps. A threshold function is applied in step (2), yielding a binary map of where the signal power exceeds a fraction γ of the maximum signal power. For the purpose of reduction of noise and clutter contributions, a range gate of 2 m is additionally employed in step (3), centred on the centre of mass of each binary map, with the assumption that the largest detected region correspond to the target of interest. The three largest connected regions are then selected in step (4). Connected region in this case refers to a set of matrix elements that have non-zero neighbours. This step is performed since human anatomy and kinematics dictate a smoothly varying range-Doppler profile. Selecting the large connected regions thus suppresses speckle-type noise. The N_{sub} binary maps are subsequently used in step (5) to select the points of interest from the original N_{sub} range-Doppler maps. These points form a point cloud with dimensions range, Doppler, time, and signal power for each point. For consistency between segments, a fixed number of points is required. Thus, the point cloud is upsampled or downsampled to N_{pts} points in step (6) based on this requirement. Upsampling is achieved by duplication of existing points in the cloud, downsampling by uniform subsampling of the cloud in range-Doppler space.

For classification of each segment, a point transformer neural network is utilised [67]. Point transformer networks fall under the transformer family of deep learning architectures and are suited to tasks including object classification and scene segmentation. Point clouds serve as inputs to these networks and a modified attention mechanism [103] is the means of feature extraction. Details on the network can be found in the original paper [67]. For this chapter, based on previous research in Chapter 3 and inspired by the research in [34], the architecture proposed by [67] is selected. This choice is motivated by the demonstrated classification performance, respective to two alternative architectures that have been considered [104, 105]. Three parameters govern the implementation of the architecture into a specific neural network: the number of transformer blocks, the number of neighbours considered for each point, and the size of each point transformer layer. Based on the research in Chapter 3, they are set to 4, 16, and 128 respectively.

The point cloud corresponding to each individual segment is assigned an activity label based on the primary activity performed during the segment, which is determined through a majority ruling. An activity prediction is the output of the Point Transformer model, which is compared to the ground truth for that particular segment to determine the classification performance of the method.

4.3. Case Study

To gauge suitability of the proposed method for continuous human activity classification, as well as study the effects of alternative segmentation approaches, two case studies are performed. The studies are conducted using the publicly available dataset [66], described in Chapter 2, that contains sequences of activities from 14 participants. Specifically, these case studies include:

1. A comparison of classification performance between the proposed segmentation method and two alternative segmentation methods described in Section 4.2.3.
2. A Leave-One-Person-Out (L1PO) validation of the proposed method and the best performing alternative method to compare to reference classification approaches

Table 4.1: Summary of parameters used in the proposed method. Parameters relating to segmentation are found at the top, those relating to the subsequent processing of the segments are found at the bottom.

Parameter	**Notes**
Segmentation	
α	Influences dependence of entropy on velocity values with strong intensity
T_{lag}	Time scale of monitored entropy fluctuations
β	Required fluctuation magnitude to indicate transition event
Processing	
N_{sub}	Number of subsegments extracted from segment
γ	Threshold value for detection (binarisation) of range Doppler maps, as a fraction of maximum signal power
N_{pts}	Number of points in processed point cloud

in literature. L1PO validation is further detailed in Section 4.3.1, and entails training on data from all but one participant, and testing on the data of omitted participant

A summary of the parameters employed in the proposed method, based on Rényi entropy, is given in Table 4.1.

4.3.1. CLASSIFICATION APPROACH

The approach for classification in the case studies conducted in this chapter is as follows: the activity sequences is first segmented using the method under investigation. The found segments are processed in accordance with Section 4.2.4, and subsequently used as input to the Point Transformer network for classification. Comparing the output of the Point Transformer to the ground truth for the segment yields performance measures, in this research expressed in terms of test accuracy and macro F1-score. Test accuracy is selected due to its prevalence in classification tasks in the literature, and macro F1-score is reported as it can give a good insight in the performance on underrepresented classes.

The validation approaches taken for the case studies fall under two categories in terms of training/testing split.

1. A Leave-one-Person-Out (L1PO) approach, where sequences from 13 of the participants are used to train the Point Transformer, and sequences from the remaining single participant are used to test the model. This process is repeated for each participant and average performance figures are reported. This validation strategy yields the most comprehensive result, as the capability of the model to respond to unseen participants is evaluated. It is however significantly more time-consuming due to the amount of models that have to be trained for each experiment.

2. A sample holdout method, where 80% of the segments are used to train the Point Transformer model, and 20% are used for testing.

Table 4.2: Test accuracy and macro F1-score results for the proposed classification method and two reference segmentation methods, detailed in Section 4.2.3. Results are presented for a 80 %/20 % sample holdout scheme.

Method	**Test Accuracy**	**Macro F1-Score**
Proposed	0.909±0.003	0.853±0.005
STA/LTA	0.903±0.006	0.829±0.006
BiLSTM	0.924±0.010	0.914±0.011

Due to the computational requirements of the L1PO validation strategy, this approach will only be taken when it is necessary to compare the methods in this chapter with reference methods from literature.

4.4. Results and Discussion

4.4.1. Study on Segmentation Methods

Table 4.2 contains the classification results of a comparison between the proposed method and two reference methods. As references, the STA/LTA and BiLSTM segmentation methods from Sections 4.2.3 and 4.2.3 are specifically shown. Each model is trained three times using 80%/20% sample holdout validation to analyze statistical fluctuations. Highest performance is achieved using the BiLSTM segmentation method, as this approach is able to learn more complex patterns in entropy rather than just strong fluctuations. The proposed segmentation method outperforms STA/LTA segmentation in terms of Macro F1-Score. Inspection of the segments created with the STA/LTA algorithm reveals that transitions are generally correctly found when transitioning from stationary to motion and vice versa, but that complex transitions between different types of motion are not detected as effectively.

4.4.2. L1PO Result

Figure 4.3 shows the results of the L1PO validation of the proposed method, indicated with solid markers and lines. Additionally included is the performance for ‘perfect’ segmentation where ground truth labels have been employed to yield segments that contain a single activity only, which is assumed to be ‘optimal’ segmentation. These performance figures are shown with empty markers and dashed lines. It should be noted that this ‘optimal’ result relies on information that is unavailable in a real scenario, and is provided to indicate the effect of segmentation on the final classification effectiveness.

On average, the performance metrics for the segmentation based on ground truth information are higher than for segmentation following the proposed method based on Rényi entropy. Average test accuracy for the former and latter are 90.9 % and 89.3 % respectively. Notable outliers however are participants #3 and #4. An important conclusion that can be drawn is that segmentation into human-interpretable boundaries is not necessarily the best strategy for classification. Single activity ground truth segments may align well with human understanding, but do not necessarily facilitate model-based classification. Breaking down complex or long segments into possibly more homoge-

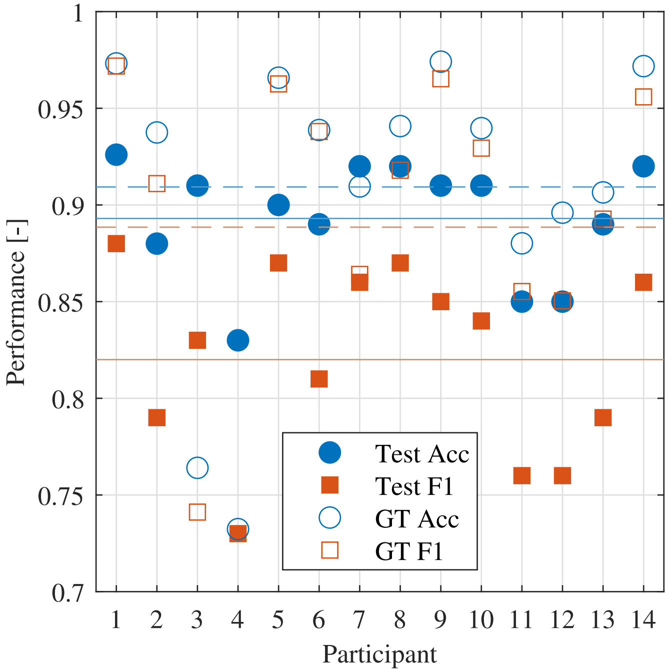


Figure 4.3: Results for the proposed classification method, evaluated using a L1PO testing scheme. Test accuracy and Macro F1-score results are shown for each of the 14 participants. Averages across all participants are indicated with horizontal lines. *GT acc* and *GT F1* refer to classification performance achievable when ground truth information is available about the moments where transitions occur, i.e., segments containing only a single activity. The averages for this 'optimal' GT segmentation are indicated with dashed lines.

neous parts can enhance classification performance, as seen in the cases of participants #3 and #4.

4.4.3. Comparison to Reference Methods in Literature

Table 4.3 presents the L1PO test accuracy and Macro F1-score for the proposed method, the BiLSTM segmentation method, and two reference methods from the literature using the same dataset [33, 37]. Additionally included is the result for classification of segments created using the ground truth labelling information (*GT Segments*). Comparing the proposed method with the Point Transformer operating on fixed, 2 s windows [37] reveals an increase in test accuracy, as well as a +3.3 % increase in Macro F1-Score. This improvement highlights the benefit of using adaptive segments, mitigating the transition errors that occur when using fixed-duration segments. The CNN-BiGRU approach in [33] processes the activity sequences into series of range-Doppler maps, where feature extraction is performed by a Convolutional Neural Network (CNN) on a per-map basis. The extracted features form a time series, which is then used as input to a Bidirectional Gated Recurrent Unit (BiGRU) which performs classification. To keep the size of this hybrid model computationally feasible, the range-Doppler maps are scaled down, possibly explaining the difference in performance with the proposed method. Of note

Table 4.3: Test accuracy and macro F1-score results for the proposed classification method and four alternative methods. *GT Segments* refers to the utilisation of the Point Transformer network, but with 'perfect' segmentation using the ground truth data to locate transitions. The top three rows correspond to different methods discussed in this chapter, the lower two rows to reference works from literature. All results are based on the same dataset [66] and the same L1PO validation scheme. The results in this table are for the full nine-class classification problem. *:[37],§:[33].

Method	**Test Accuracy**	**Macro F1-Score**
Proposed	0.893	0.820
Proposed (GT Segments)	0.909	0.889
BiLSTM Segmentation	0.878	0.860
PT (Fixed Segments)*	0.869	0.787
CNN-BiGRU§	0.851	-

Table 4.4: Merging scheme for the consolidation of the full nine activity classes into a set of five activity classes. Classes are grouped based on similarity.

Constituent Classes	**Merged Class**
Walking	Walking
Stationary	Stationary
Sitting Down, Standing up (from sitting), Bending (from sitting), Bending (from standing)	In Situ
Falling (from walking), Falling (from stationary)	Falling
Standing up (from ground)	Standing up

is the test accuracy of the BiLSTM segmentation method, which is lower than that of the proposed method. This contrasts with the results achieved under the sample hold-out validation scheme in Table 4.2, where the BiLSTM-based method outperforms the proposed method both in terms of test accuracy and macro F1-score. This result highlights the problem of overfitting in machine learning approaches, where a neural network trained on a set of data can have issues with generalisation capabilities outside of the training set. In this case, the test data of the unseen participant proves challenging to the BiLSTM segmentation algorithm.

Some of the reference works with methods benchmarked on the same dataset utilise an aggregated set of five activity classes. In this reduced set, displayed in Table 4.4, activities such as *Falling from walking* and *Falling whilst standing still* are joined into a singular *Falling* class. Table 4.5 shows the results for this five-class problem. The top three rows again correspond to methods discussed and proposed in this chapter, the lower rows are performance metrics reported in reference methods from literature [37, 38]. The methods in [38] involve the computation of a spectrogram representation of the activity sequences, which is subsequently used as input to various Recurrent Neural Network (RNN) architectures. The performance of the proposed method, as well as the segmentation approach utilising a BiLSTM, both surpass that of the reference method [38]. Notably, the combination of a BiLSTM with the Point Transformer model demonstrates

Table 4.5: Test accuracy and macro F1-score results for the proposed classification method and several alternative methods. *GT Segments* refers to the utilisation of the Point Transformer network, but with ‘perfect’ segmentation using the ground truth data to locate transitions. The top three rows correspond to methods discussed in this chapter, the lower six rows to reference works from literature. All results are based on the same dataset [66] and the same L1PO validation scheme. The results in this table are for a five-class classification problem. *:[37],†: Signal Fusion [38], ‡: Feature Fusion [38].

Method	**Test Accuracy**	**Macro F1-Score**
Proposed	0.928	0.880
Proposed (GT Segments)	0.947	0.943
BiLSTM Segmentation	0.913	0.900
PT (Fixed Segments)*	-	0.862
GRU†	0.909	0.778
LSTM†	0.910	0.769
bi-GRU†	0.933	0.844
BiLSTM†	0.931	0.836
BiLSTM‡	0.924	0.840

superior results compared to the BiLSTM operating solely on spectrogram data. This improvement can be attributed to the more efficient utilisation of the BiLSTM network in the combined approach. Specifically, the BiLSTM in the hybrid model processes only the Rényi entropy, rather than the entire spectrogram, and is tasked with predicting transitions rather than classifying specific activity types. Consequently, the combined model remains more compact at 19 MB, compared to the stand-alone BiLSTM of approximately 25 MB.

The L1PO experiments demonstrate promising results for the proposed method, with the BiLSTM-based segmentation delivering the highest performance in terms of macro F1-score. However, the proposed method based on Rényi entropy offers several key advantages in certain scenarios:

- If minimising model size is a significant priority, the proposed method is advantageous, as it relies solely on the Point Transformer model with minimal segmentation processing.

- The segmentation approach used in the proposed method offers superior interpretability compared to the more complex BiLSTM-based segmentation.

- The Point Transformer model takes inputs such as range, velocity, and time, which are consistent across experimental scenes and radar systems. In contrast, BiLSTM segmentation may need retraining across different scenarios, as the Rényi entropy patterns could vary. This argument is further strengthened by the disparity in performance of the BiLSTM-based method between the sample holdout experiment and the L1PO experiment, representing poor generalisation capabilities.

Table 4.6: Test accuracy and macro F1-score results for three windowing approaches. *Segmentation* is the method proposed in this chapter, *Fixed Window* refers to the 2 s window method from Chapter 3, *Dual Fixed Window* refers to an additional experiment that has been performed to test two fixed windows of different sizes in parallel. The results presented are for a L1PO validation scheme on a subset of 4 randomly selected participants.

Classifier	Test Accuracy	Macro {F1-score}
Segmentation	**0.911**	**0.859**
Fixed Window	0.864	0.797
Dual Fixed Window	0.888	0.825

4.4.4. Dual Window Comparison

A final experiment is performed to gauge the effectiveness of a hybrid approach between that presented in Chapter 3, and the segmentation-based approach proposed in this chapter. Rather than using a single fixed, 2 s window as in Chapter 3, every sequence in the hybrid approach is classified by two separate fixed window classifiers, operating in parallel. At every time step, the output prediction vectors of the two classifiers are compared, and the activity class with the highest total confidence level is selected. The two window choices for the dual classifiers are based on the statistical mode and the median of the segment sizes that are found with the segmentation algorithm, rounded for convenience. They are 1.31 s and 2.30 s respectively. All other processing is identical to this chapter.

Table 4.6 shows the results for the experimental comparison. For validation, a L1PO scheme is used. Due to time constraints, only a subset of 4 randomly selected participants is evaluated out of the total of 14. It is noted that the dual classifier approach is superior to the single window approach, but does not reach the performance of the adaptive segmentation approach proposed in this chapter.

4.5. Conclusion

This chapter proposes a novel method for classification of continuous sequences of human activities. The proposed method consists of three main elements: segmentation of the sequences, processing of the segments, and classification of the segments. As a key step of the proposed processing, segmentation is achieved through the monitoring of fluctuations in Rényi entropy, a scalar quantity computed from micro Doppler spectrograms. The proposed method offers a solution to the problem of continuous activity classification that is more reliable in terms of classification performance compared to reference methods from the literature. Additionally, it is also computationally efficient due to the effectiveness of the Rényi entropy as an indicator of activity changes.

The proposed method is experimentally validated on a publicly available dataset. A Leave-One-Person-Out test accuracy of 89.3 %, and a macro F1-score of 82.0 % are achieved on the dataset which consists of a variety of sequences of nine human activities. Alternative segmentation methods are also investigated. These include the STA/LTA change detection algorithm, and a BiLSTM network taking Rényi entropy as input. Between the proposed method and the BiLSTM segmentation method, the highest classification performance metrics are attained by the BiLSTM. The proposed method is however preferable in terms of computational efficiency and interpretability, and outperforms reference methods from literature on the dataset.

An idealised form of segmentation is also studied, where ground truth labels are used to create segments that contain a single activity only. Classification performance in this case is generally higher than on segments produced by the various segmentation algorithms. In some cases however, the opposite is true. It is concluded that segmenting to the ground truth of an activity sequence is not necessarily optimal, if at all possible in a realistic scenario, and segmentation into classifier-interpretable, homogeneous segments might be preferable.

Further refining the method proposed in this chapter could involve the inclusion of a multi-label classification method, such as that presented in [50], to further mitigate the problem of multiple activities in a segment. The proposed segmentation algorithm in this chapter relies on a single parameter governing the time scale that is considered when searching for transitions between activities. In reality, activity transitions can vary in duration. A single time scale has been utilised here as an initial approach, but this can be expanded to multiple values to allow more flexibility in the detection of activity transitions. Finally, alternative segmentation methods are to be investigated that strike a balance between interpretability, computational efficiency, and classification performance.

5

Reconstruction of Extended Human Target Intensity and Velocity Distributions

Radar sensors are typically best suited at measurements in the radial direction, meaning away from and towards the antenna. For a human under observation this means that the radar only sees the part of the movements that are directly toward or away from the radar's antenna. For quite some time we discussed the possibility of using multiple radars in different locations in order to retrieve the movements of the body in all directions, but in most cases we hit theoretical barriers. In this chapter I propose a method that uses information from a group of radar sensors, and fuses it together to yield a recreation of the movements of a human subject in three dimensions. Through experimental work I then show that the recreation is accurate enough to classify what the person was doing at a given moment in time.

Parts of this chapter have been published as:

Kruse, N. C., Guendel, R. G., Fioranelli, F., & Yarovoy, A. (2025). *Reconstruction of Extended Target Intensity Maps and Velocity Distribution for Human Activity Classification.* In IEEE Transactions on Radar Systems, 3, 14-25.

Kruse, N. C., Guendel, R., Fioranelli, F., & Yarovoy, A. (2024). *Distributed Radar Fusion for Extended Target Location and Velocity Reconstruction.* 2024 IEEE Radar Conference (RadarConf24), 1–6.

5.1. Introduction

Within the context of healthcare, radar is considered a promising sensor modality for the monitoring of patients and vulnerable individuals in their home environments. The monitoring capabilities of radar include vital sign estimation [21, 106], gait analysis [15, 107], fall detection [13, 14], gesture recognition for interaction with smart devices and automatic sign language interpretation [30, 108–110], and activity classification [12, 37, 55, 111]. The non-contact nature of radar sensing allows for monitoring in situations where wearable sensors would prove disadvantageous, such as in cases where subjects may forget or object to wearing a sensor. Additionally, radar is an active sensor and functions in complete darkness or glaring lights, with no hindrance to the user. Finally, no visual images are captured, which can be beneficial in terms of perceived privacy from the side of the end-users.

To aid in activity classification, sensor fusion has been utilised to great effect in literature. The combination of data from multiple heterogeneous sensors [71, 112], or from multiple radar sensors operating in a network [19, 38, 113] allows them to complement each other and compensate for weaknesses of individual sensors. For the case of radar in particular, the most precise position and velocity measurements are generally in the radial direction. This suggests that a network of radars can improve the perception of the location and velocity distribution of the different body parts of the observed subject compared to a single radar, provided that the network is sufficiently spatially diverse.

With the proliferation of machine learning (ML) techniques, sensor fusion approaches often rely on a fused representation that constitutes a latent space in a ML model [19, 33, 112, 113]. In other words, the feature space after fusion is no longer easily human-interpretable. These fused feature spaces are optimised for a specific task and mostly do not generalise well to other applications. Other common fusion types are decision-based, where predictions from multiple classifiers are merged on the basis of e.g., prediction confidence [19, 37]. A disadvantage of these approaches is that the individual classifiers do not have access to a representation with information from all sensors, thus potentially limiting performance. Finally, simple fusion approaches are utilised where the fused representation is a concatenation of radar data domains from multiple sensors [38, 111, 113]. In these cases, no data association between sensors is performed.

Various fusion approaches for radar networks in the literature for human activity classification have been investigated. To the best of this author's knowledge there is no method that aims to explicitly model and estimate both reflection intensity and velocity distribution of the target in 3D. This capability would be important to combine interpretability and task versatility. Interpretability comes from a shared data representation from the different nodes that is based on intensity and Doppler/velocity, both of which are quantities that are easily understandable and related to the kinematics of the observed activities. Task versatility is here defined as the potential to utilise the fused representations for multiple applications, for example target tracking or activity classification.

In this chapter, a novel sensor fusion method is proposed that processes raw data from a network of radar sensors and yields three-dimensional representations of both reflection intensity and velocity distribution. Specifically, range-angle-Doppler representations of data from a network of distributed monostatic radar are processed into two 3D fields defined in cartesian coordinates. The first of these fields contains the reconstructed reflection intensity at each point in a 3D spatial grid; the second is a vector field of reconstructed velocities, in the context of human activity classification related to the combined movement of the different body parts. The efficacy of the proposed fusion method is evaluated through a classification case study. A challenging, publicly available dataset of continuous human activities is processed using the proposed method. The fused intensity maps and velocity fields are then used as input to a CNN-BiLSTM (Convolutional Neural Network - Bidirectional Long Short Term Memory) architecture, tasked with discriminating nine different human activities. Additionally, an experimental feasibility study is performed to demonstrate the ability of the proposed method to yield 3D representations of extended target shape and velocity distribution.

The remainder of this chapter is organised as follows. In Section 5.2, the proposed sensor fusion method for reconstruction of reflection intensity and velocity profile of human movements is explained, followed by a description of the 2D case study used to validate the method in Section 5.3. Results of this case study are presented and discussed in detail in Section 5.4. The 3D feasibility study is presented in Section 5.5, and conclusions follow in Section 5.7.

5.2. Proposed Method

The fusion method proposed in this chapter comprises two main elements: a voxelisation of the observed human target into a 3D spatial grid, and a reconstruction technique of the measured reflection intensity and dominant velocity in each voxel. The method is agnostic to the angular capabilities of the radar sensors utilised, and functions with any number of sensors, provided that this exceeds the amount of spatial dimensions considered for the problem. At least four spatially distributed radar nodes are needed to reconstruct intensity and velocity distribution in a 3D grid, in the case that angular information is not available.

Inputs to the proposed fusion method are N radar data tensors from a network of N radar sensors. The tensors represent measured signal amplitude and have dimensions of range, Doppler (radial velocity), and azimuth and elevation if the radar sensors are capable of measuring them. The outputs of the method are a spatial distribution of reflection intensity as well as a vector field of reconstructed velocities.

It is assumed that the human body is a non-rigid, extended target. For the purposes of activity classification, the simplification to a point target with a single velocity is too limiting. Although the distribution of radial velocities of the human target can be measured by multiple radar systems, an association problem prevents the determination of a 3D velocity profile through direct sensor fusion. The human body is however subject to kinematic constraints. The allowed joint movements are principally rotational in nature, which implies that the velocity profile along the limbs will vary smoothly. Given this implication, it follows that a volume element that is small in comparison to the human target can be assumed to contain a dominant velocity. This assumption underlies the velocity reconstruction aspect of the proposed method. It should be noted that the assumption does not hold in the cases where e.g. two limbs pass in close proximity relative to the size of the volume element. In these cases, volume elements may contain multiple dominant velocities.

The proposed fusion method takes radar data tensors $\mathscr{D}(r,\theta,\phi,v)$ from a network of radar sensors as input, where azimuth θ and elevation ϕ are optional, depending on the front-end architecture of the considered radar. Dimensions r and v represent range and Doppler respectively. As this information can be provided by a variety of radar systems with varying processing approaches, no specific architecture or type of radar is assumed, provided that Doppler shifts and thus radial velocity components can be measured. In the specific case of a pulsed single-channel system, with a complex valued fast-time/slow-time matrix as in (2.12) in Chapter 2, the radar data tensor is a range-Doppler map acquired through application of a Discrete Fourier Transform (DFT) along the slow-time dimension:

$$\mathscr{D}(r,v) = \mathscr{F}_{t_s}[\mathscr{S}(t_f,t_s)]. \tag{5.1}$$

Here, $\mathscr{F}_{t_s}$ is the DFT operation along the slow-time dimension t_s. The notation $\mathscr{D}$ is maintained for the radar data tensor.

The set of radar sensors will be indicated by $\mathscr{N}$, and the cardinality of this set is denoted by $N = |\mathscr{N}|$. In the remainder of this section, vectors and unit vectors will be indicated with arrows ($\vec{p}$) and hat operators ($\hat{p}$) respectively. Thus, $\vec{p} = p\hat{p}$ and $p = \vec{p}\cdot\hat{p}$. An estimated vector is indicated with a tilde ($\tilde{p}$).

5.2.1. Intensity Maps

As a first step to the proposed method, the 3D space in the field of view of the radar network is uniformly voxelised in Cartesian coordinates, resulting in a grid with a spacing that can be adjusted based on e.g., the resolution characteristics of the radar sensors utilised. For each volume element in the grid the range to each sensor is computed, as well as the azimuth and elevation angles if available for the radar sensors used. These computations are achieved through a coordinate transfer mapping for each individual sensor.

Consider the centre of a volume element in Cartesian coordinates $\vec{x} = [x,y,z]$. The vector is first translated to place the sensor n in the origin, and is rotated such that the sensor boresight is in the positive x-direction, as:

$$\vec{x}'_n = R_n^{-1}(\vec{x} - \vec{x}_n), \tag{5.2}$$

where $\vec{x}_n$ is the position of sensor n and

$$R_n = R(\theta_{Az})R(\phi_{El}) = \begin{bmatrix} \cos\theta_{Az}\cos\phi_{El} & -\sin\theta_{Az} & \cos\theta_{Az}\sin\phi_{El} \\ \sin\theta_{Az}\cos\phi_{El} & \cos\theta_{Az} & \sin\theta_{Az}\sin\phi_{El} \\ -\sin\phi_{El} & 0 & \cos\phi_{El} \end{bmatrix}. \tag{5.3}$$

Here, the rotation matrix R_n for the sensor n is defined as a function of the boresight elevation ϕ_{El} and azimuth angle θ_{Az} with respect to the positive x-axis. Subsequently, a mapping $f:\mathbb{R}^3 \to \mathbb{R}^3$ is employed to transform to a spherical coordinate system: $f(\vec{x}) =$

$\vec{r} = [r, \theta, \phi]$. The transformation is given by:

$$r = ||\vec{x}||, \tag{5.4}$$

$$\theta = \arctan \frac{y}{x}, \quad x > 0, \tag{5.5}$$

$$\phi = \arcsin \frac{z}{||\vec{x}||}, \quad x > 0. \tag{5.6}$$

Note that the angles ϕ and θ are limited to the hemisphere in the positive x-direction.

After transforming the grid to the frames of the individual sensors as in (5.2) and applying the aforementioned mapping into spherical coordinates, the respective radar data tensors can be evaluated to yield the measured signal amplitude at each volume element. As the transformed grid may not align with the range-angle map of a sensor, the signal amplitude of a volume element is taken to be that of the closest data point in the sensors' range-angle map. The individual tensors are of dimension $\mathscr{D}(r, \theta, \phi, v)$, and reflection intensity for each volume element $\vec{x}$ is determined by summing received signal amplitude over the radial velocity dimension v. Contributions from the set of radar sensors $\mathscr{N}$ are then summed up, as:

$$I(\vec{x}) = \sum_{n \in \mathscr{N}} \sum_{v} \mathscr{D}\left(f\left(\vec{x}'_n\right), v\right). \tag{5.7}$$

By repeating this process for each volume element in the grid, an intensity map I can be constructed.

5.2.2. Velocity Field Reconstruction

The reconstruction of the velocity field for the extended target starts with the same grid definition and transformation as outlined in the previous subsection. In contrast to the intensity map computation, the dominant Doppler/velocity component is recorded for each voxel in the grid, and for each radar sensor in the network. For this chapter, the dominant Doppler component is considered to be the Doppler index with the highest amplitude, as:

$$v_n = v_n(r, \theta, \phi) = \arg\max_v \left(\mathscr{D}\left(r, \theta, \phi, v\right)\right). \tag{5.8}$$

The resultant set of N Doppler components from the set of radar sensors is assumed to originate from a dominant velocity in the volume element. In order to reconstruct this full 3D velocity from the N projections, i.e., the projections on the line of sight of each radar, the process of orthogonal projection is inverted by means of a minimisation problem.

First, consider the generic plane defined by the vectors $\vec{p}$ for which the following holds:

$$\vec{p} \cdot \hat{n} + c_i = 0. \tag{5.9}$$

Here, $\hat{n}$ is a unit vector orthogonal to the plane and c_i a constant. For the specific case of a Doppler/velocity projection seen by a radar sensor, the true target velocity $\vec{u}$ projects onto the sensor's line of sight $\hat{x}'_n$, yielding $\vec{v}_n$ as shown in Figure 5.1. It follows that all possible $\vec{u}$ are given by the vectors $\vec{u} - \vec{v}_n$ that are in the plane that is orthogonal to $\hat{x}'_n$.

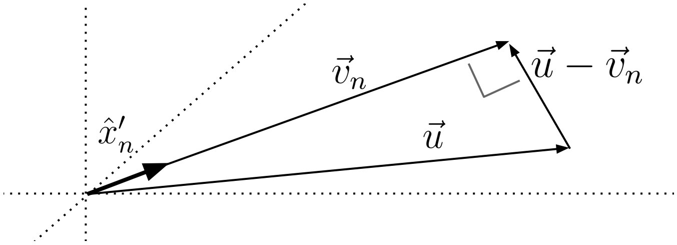


Figure 5.1: Vector representation of quantities relevant for velocity reconstruction. $\hat{x}'_n$: line of sight unit vector from radar n to voxel at $\vec{x}$, $\vec{u}$: true velocity in voxel at $\vec{x}$, $\vec{v}_n$: radial velocity measured by sensor n.

Since v_n is known and $v_n = \vec{v}_n \cdot \hat{x}'_n$, $\vec{v}_n$ can be substituted for $\vec{p}$ and $\hat{x}'_n$ for $\hat{n}$ in (5.9) to solve for the constant $c_{i,n}$:

$$\vec{v}_n \cdot \hat{x}'_n + c_{i,n} = 0 \tag{5.10}$$

$$c_{i,n} = -v_n. \tag{5.11}$$

A set of N planes can now be defined that, in an ideal case, would intersect in a single point $\vec{u}$:

$$\vec{p} \cdot \hat{x}'_n - v_n = 0, \tag{5.12}$$

where $\vec{p} \in \mathbb{R}^3$ and $n \in \mathcal{N}$.

In the non-ideal case of real targets, an optimisation problem can be set up to find the point with minimum distance to the set of N planes, i.e., to determine the best fitting intersection point $\tilde{u}$. First, a cost function is defined as the sum of squared distances of a point $\vec{p}$ to the set of planes as per (5.12). This is formulated as follows:

$$E(\vec{p}) = \sum_{n \in \mathcal{N}} \left(\vec{p} \cdot \hat{x}'_n - v_n\right)^2 \tag{5.13a}$$

$$= \sum_{n \in \mathcal{N}} \left(\left(\vec{p} \cdot \hat{x}'_n\right)^2 - 2\vec{p} \cdot v_n \hat{x}'_n + v_n^2\right) \tag{5.13b}$$

$$= \sum_{n \in \mathcal{N}} \left(\left(\vec{p}^T \hat{x}'_n \hat{x}'^T_n \vec{p}\right) - 2\vec{p}^T v_n \hat{x}'_n + v_n^2\right). \tag{5.13c}$$

Matrix notation is used from equation (5.13c) onward for the sake of clarity. The derivative of the cost function with respect to $\vec{p}$ is given by:

$$\frac{dE}{d\vec{p}} = 2 \sum_{n \in \mathcal{N}} \hat{x}'_n \hat{x}'^T_n \vec{p} - 2 \sum_{n \in \mathcal{N}} v_n \hat{x}'_n, \tag{5.14}$$

and setting this derivative to 0 yields the best fitting intersection point, namely the vector $\tilde{u}$, as:

$$\sum_{n \in \mathcal{N}} \hat{x}'_n \hat{x}'^T_n \tilde{u} - \sum_{n \in \mathcal{N}} \vec{v}_n = 0 \tag{5.15a}$$

$$\tilde{u} = \left(\sum_{n \in \mathcal{N}} \hat{x}'_n \hat{x}'^T_n\right)^{-1} \sum_{n \in \mathcal{N}} \vec{v}_n. \tag{5.15b}$$

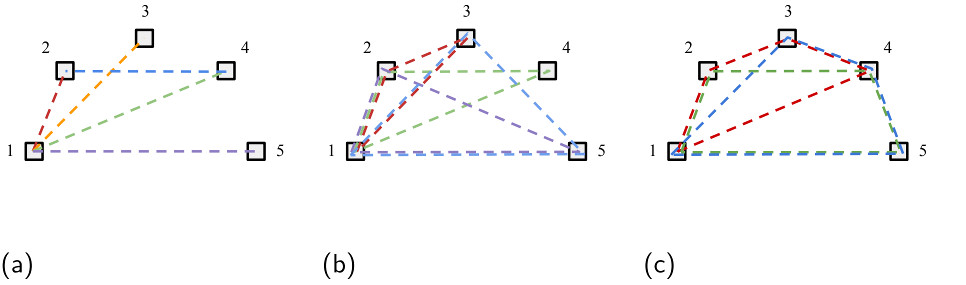


Figure 5.2: Schematic representation of all the evaluated sensor network geometries with the five available radar nodes in the network. Each considered geometry is color-coded and the constituent sensors are connected by dashed lines. (a) Two node geometries; (b) Three node geometries; (c) Four node geometries.

The resulting vector $\tilde{u}$ is the assumed dominant velocity in the voxel. This velocity reconstruction can be applied to all voxels in the considered volume of interest of the radar network, thus yielding a vector field $\tilde{u}(\vec{x})$. The velocity field can in principle be computed for all voxels, including those with only noise. Hence it can be decided to only compute, or visualise, the field at those voxels where the corresponding intensity map exceeds a user-defined threshold, i.e., $I(\vec{x}) > \epsilon$.

After the computation of the scalar reflection intensity map and velocity vector field at all voxels, the final representation generated by the proposed method is given by both $I(\vec{x})$ and $\tilde{u}(\vec{x})$. This representation constitutes a fusion of the data from the N sensors into three spatial dimensions, and can be subsequently used for tasks such as classification of activities.

5.3. Experimental Case Study in 2D

To demonstrate the efficacy of the proposed method, an experimental case study is conducted. The objective of the study is to gauge the accuracy of the reconstructed velocity distribution and reflection intensity maps for human activity classification tasks. Specifically, the intensity map and velocity field are computed according to the method described in Section 5.2 and used as input to a deep learning model with the aim of classifying different human activities. The dataset used for the 2D case study is described in Chapter 2, and features 120 s sequences of nine different human activities of unconstrained duration and direction. The data is captured with a network of five pulsed Ultra Wideband (UWB) radar sensors, operating as distributed monostatic nodes. For classification, a hybrid CNN-BiLSTM architecture is employed to process the fused representations of intensity maps and velocity fields.

5.3.1. Sensor Geometry

Since the proposed method is inherently based on the fusion of data from different radar nodes, various sensor network geometries are evaluated. As the dataset has been collected with the fixed five-node geometry from Figure 2.2, only subsets of these nodes can be evaluated. These subsets can be realised by leaving data from specific nodes out during the processing. Due to computational constraints, not all subsets that are possible with five nodes have been tested experimentally. Hence, subsets that are in the same rotational symmetry group, or that exhibit a bilateral symmetry along a horizontal axis, are represented in this study by a single member of the respective symmetry group. This is done under the assumption that the activities in the experimental scene were performed in fully random trajectories, so that there is minimal bias in activity location or orientation. As an example, the configuration $(1,2,4)$ is rotationally symmetric to the configuration $(2,3,5)$ and only the former is included in the experiment. As an exception to this, configurations $(1,3)$ and $(2,4)$ are both included with the aim of gauging any bias in the location and direction of activities in the dataset. All the geometries that are evaluated in this chapter are displayed in Figure 5.2.

5.3.2. Pre-processing

Pre-processing is performed on the data of each radar sensor prior to the implementation of the proposed method described in Section 5.2. First, the real-valued vector output by each of the sensors is Hilbert-transformed, and a fast-time/slow-time matrix with complex values is constructed, which is essentially considered as a range-time matrix $\mathscr{R}_{r,t}$, as in Section 2.2. The slow-time indices corresponding to one coherent processing interval T_{CPI} are then selected from this larger matrix for Doppler processing. Finally, an FFT along the slow-time dimension is performed to obtain a range-Doppler representation. This process is repeated along the sequence of original data with constant T_{CPI} to generate a sequence of range-Doppler matrices $\mathscr{RD}_{r,v}$. As no angular information can be extracted from the data of a single radar, the tensor defined in Section 5.2 is in this case only represented in two dimensions, as: $\mathscr{D}\left(r,\theta,\phi,v\right) = \mathscr{D}\left(r,v\right)$. Hence, the proposed method will operate in this case on bidimensional range-velocity matrices $\mathscr{D}\left(r,v\right)$, equivalent to range-Doppler maps $\mathscr{RD}_{r,v}$.

5.3.3. Method Parameters & Classification Approach

Since the radar sensors offer no angular information and are located in the same horizontal plane, it is decided to consider only the horizontal xy-plane for this study. The vertical z-direction is thus effectively projected onto the horizontal plane. It is hypothesised that in the recorded scenarios enough movement occurs in this horizontal plane in order to effectively perform activity classification, i.e. the activities can be distinguished by their horizontal velocity and reflection intensity distributions.

For the case study, the method parameters employed for obtaining the intensity map and velocity field are specified as follows. The coherent processing interval T_{CPI} is set to 0.26 s, which at the PRF of 122 Hz corresponds to 32 slow time samples. This value is based on proven effectiveness in previous research [114] and balances Doppler and time resolution for the highly dynamic nature of human motions. In order to process the full 120 s sequences, a series of intensity maps and velocity fields are computed. These are spaced at 32 slow time sample intervals, i.e. one per coherent processing interval without overlap. This interval will for the remainder of this chapter be referred to as a 'time step'. The measurement area is divided into a grid of 15 cm × 15 cm cells. As

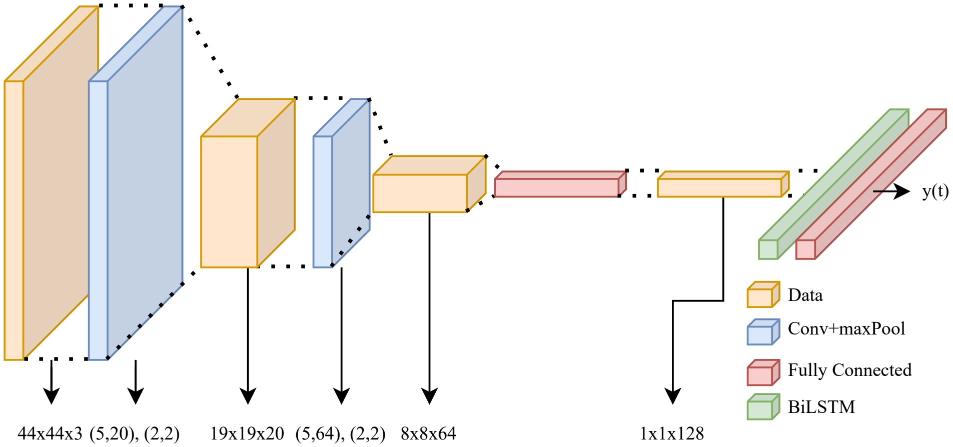


Figure 5.3: Schematic representation of the hybrid CNN-BiLSTM network utilised in the case study. Data are represented as 3D tensors with dimension $Width \times Height \times Channels/Depth$. The size of convolutional and maxpooling layers is shown as (CNN Kernel size, Channels), (maxpooling kernel size, maxpooling stride size). The output after the last fully connected layer is a time series of class predictions $y(t)$.

mentioned before, the range resolution of each sensor is about 7 cm and the grid size is selected as a compromise between reconstruction accuracy and computation time. Finally, the velocity field that is computed at each step is masked using the intensity map. This ensures that only velocities associated with the presence of a detected target are used as input to the classification pipeline. Specifically, the velocity vectors that are kept are those at the coordinates where the intensity values exceed the 95th percentile of the intensity values distribution for that time step. The size of the extended target after applying a threshold between the 90th and 99th percentile remains largely unchanged, due to a relatively steep intensity boundary.

The classification approach aims to utilise a hybrid CNN-BiLSTM model. The CNN first extracts features from the 2D intensity map and velocity field for each time step, i.e., each CPI. Subsequently, the BiLSTM network performs the prediction step with access to the time series of feature vectors of the full sequence. The CNN-BiLSTM architecture benefits from the translation-equivariance of the CNN to effectively extract features from a spatial domain. In the time domain, the BiLSTM is able to process the extracted features as a time series spanning the full duration of the activity sequence. Such hybrid spatial-temporal architectures are used effectively in literature [61, 108, 115], and the model for this study is inspired by [33]. The hybrid architecture is shown schematically in Figure 5.3.

The velocity vector field generated by the proposed method at each time step is decomposed into two scalar 2D fields of magnitude and angle. Together with the corresponding intensity map, they are stacked, forming a $44 \times 44 \times 3$ input array. This input size is specifically the result of the chosen 15 cm × 15 cm grid cell size, and the 6.38 m baseline of the sensor network. The activities in the dataset are performed in all directions and at random locations. To further mitigate potential bias in the orientation of the movement, all inputs are randomly rotated in the horizontal plane by integer multiples of 90°. Within the network, two sequential convolutional blocks extract spatial features and are paired with maxpooling layers for dimensionality reduction. The hyperparameters for the two convolutional blocks are shown in Figure 5.3 and are based on the LeNet-5 architecture [116]. A fully connected layer subsequently yields a feature vector of 128 elements for this time step. Finally, a BiLSTM layer with 168 hidden units processes all feature vectors and yields a class prediction at each time step. The feature vector length and the number of hidden units are based on [33] and [38], respectively.

For training and validation, a leave-one-person-out (L1PO) scheme is adopted. Data from one participant is kept as a testing set, using the remaining participants' data for training the model. This process is repeated for all participants.

5.4. Experimental Results and Discussion

In this section, results of the case study will be discussed, with namely: an example of a typical intensity map and velocity field, the L1PO benchmark result, and the results for the study on sensor geometry. Despite the lack of angular capabilities of the individual radar sensors, favorable results are achieved thanks to the proposed method that leverages data from the radar network. The results are then compared to alternative methods in literature in Section 5.4.2. Finally, the results are further discussed in the form of a classification error analysis, and computational considerations will be outlined.

In Figure 5.4, a typical intensity map and velocity vector field output is shown for the proposed method. The five sensors are marked with red squares and the color scale is normalised to the maximum intensity value. The subject can be seen at the experimental area centre point, and is in this CPI moving in the negative y-direction. At half of the maximum intensity, it can be seen that the target takes up a space of approximately 60 cm × 30 cm. Based on the target orientation, these measurements correspond to expectation, i.e., approximately 60 cm shoulder width and 30 cm torso depth. The reconstructed velocities also correspond to the motion performed at this specific moment in time, with the vector field representing the bulk torso motion in the negative y-direction. A processing artifact can be seen in the form of a ring centered around middle sensor 3, with a section of this ring contained in the dashed ellipse. This artifact is the result of the relatively strong contribution of sensor node 3 in this particular situation, paired with the inability of differentiating returns from multiple azimuth directions.

The results for the L1PO testing approach are summarised in Figure 5.5. Test accuracy and Macro F1-score are shown for each participant in the dataset, averaged over all sequences. The average test accuracy and Macro F1-score over all participants are also indicated with horizontal lines. Two notable outliers with low F1-score can be identified in the figure, namely participants 11 and 13, with respect to the other participants. Inspection of the test results for these participants reveals that the excess errors primarily stem from transition events between e.g., walking and a subsequent fall. As the ground truth is recorded by the participants themselves by clicking on a remote controller, there is variability in when a transition is indicated. For the two outliers it is noted that this variation is relatively large compared to the remaining 12 participants, leading to increased ambiguity. In spite of the outliers, the average performance metrics show suitability of the method for classification. Section 5.4.2 compares the attained result with

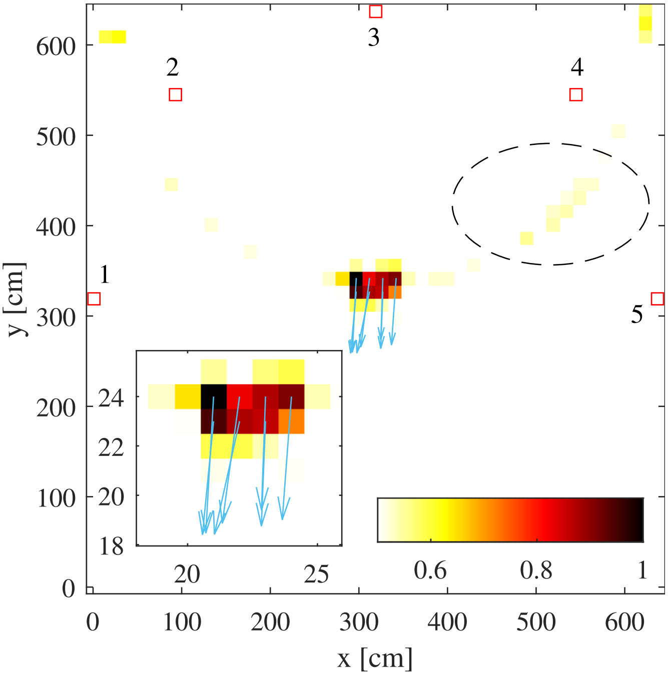


Figure 5.4: Typical intensity map and velocity vector field output with zoomed inset on the human target area. Sensors are numbered and marked with red squares. The human subject is located in the centre and is moving in the negative y-direction at this specific CPI. Velocity vectors are only shown where reflection intensity exceeds 80% of the maximum. A dashed ellipse is used to show processing artifacts.

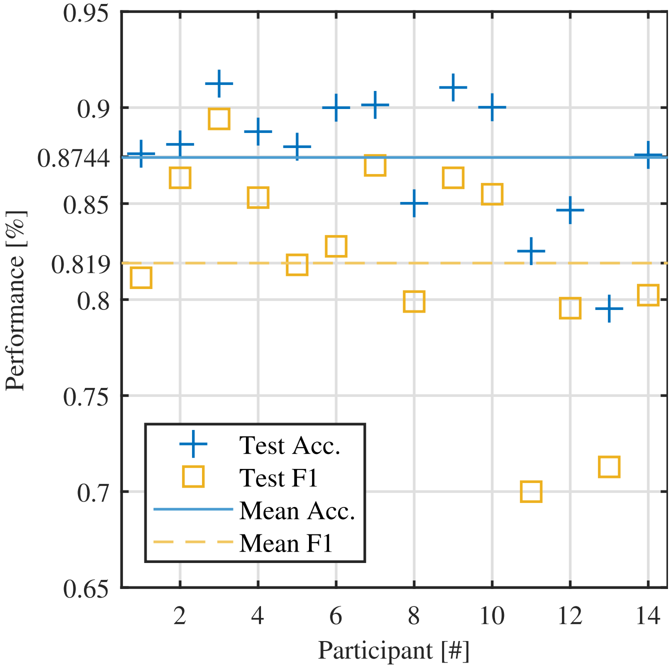


Figure 5.5: Results for the classification task, evaluated using a L1PO testing scheme. Test accuracy and Macro F1-score results are shown for each participant. Averages are indicated with horizontal lines.

alternative methods in literature on the same dataset, demonstrating improved classification performance of the proposed method.

5.4.1. Sensor Geometry Evaluation

A summary of the experiments pertaining to the sensor network geometry is shown in Figure 5.1, where the different utilised geometries are shown in Figure 5.2. For each amount of radar nodes used to reconstruct intensity and velocity fields, the average test accuracy and macro F1-score over all unique geometries is displayed, along with the standard deviation for both metrics. For every individual geometry experiment, the L1PO scheme is used. As such, $14 \cdot (5+4+3+1) = 182$ models are trained in total.

It is noted that performance increases with a larger amount of radar sensors used for the proposed data fusion method. This can be explained by two main factors:

1. Increased spatial diversity within the sensor network provides a greater range of observation angles, mitigating the effects of (partial) occlusion of the human target and individual body parts. This enhanced spatial diversity improves target illumination, resulting in a stronger intensity distribution following the application of the proposed method. In contrast with the surrounding empty space, this enhanced distribution enables the application of the threshold to more accurately identify the target volume. As a result, a more accurate selection of velocity vectors originating from the target is achieved, providing the classification model with

Table 5.1: Results of the study on sensor network geometry. The utilised 2, 3, and 4 node geometries can be seen in Figures 5.2a, 5.2b, and 5.2c respectively.

Nodes	Test Accuracy	Macro F1-score
(1,2)	0.758	0.631
(1,3)	0.768	0.647
(1,4)	0.791	0.677
(1,5)	0.797	0.691
(2,4)	0.773	0.643
Mean	**0.777±0.016**	**0.658±0.025**
(1,2,3)	0.804	0.731
(1,2,4)	0.812	0.709
(1,2,5)	0.821	0.723
(1,3,5)	0.834	0.756
Mean	**0.822±0.011**	**0.729±0.024**
(1,2,3,4)	0.834	0.748
(1,2,4,5)	0.848	0.775
(1,3,4,5)	0.843	0.767
Mean	**0.842±0.007**	**0.763±0.014**
(1,2,3,4,5)	0.866	0.804

the most reliable information regarding both the spatial extent of the target and its velocity distribution.

2. More sensors results in more instances of radial velocity measurement. This overdetermined system, paired with the Least Square Error (LSE) based approach used for the velocity reconstruction, results in mitigation of, e.g., measurement error in individual Doppler components. This in turn increases the accuracy of the velocity reconstruction in a grid cell.

In the case of two node geometries, it is hypothesised that a geometry that provides orthogonal measurements would give the highest performance. As the sensors provide no azimuth information, an orthogonal pair of range measurements will yield the most localised intensity profile. Inspection of the results in Table 5.1 however reveals the opposite, with geometries $(1,2)$ and $(1,3)$ yielding a performance that is almost two standard deviations lower than geometries $(1,4)$ and $(1,5)$. A possible explanation for this is the higher importance of larger differences in observation angles, allowing the target to be more completely illuminated.

To gauge potential bias in the data towards specific orientations or areas, geometries $(1,3)$ and $(2,4)$ are compared. The attained accuracy is 76.8 % and 77.3 % respectively, within the two-sensor standard deviation of 1.6 %.

For the three and four node geometries, the increase in performance is accompanied by a decrease in the standard deviation of the results. This lower effect of sensor positioning may in this case be indicative of partial redundancy. Of interest is that for three

Table 5.2: Test accuracy and macro F1-score results for the proposed classification method versus two reference methods. All results are based on the same dataset [66] and the same L1PO validation scheme. The results in this table are for the full nine-class classification problem. *:[37],†:[33].

Classifier	Test Accuracy	Macro F1-score
Proposed	**0.874**	**0.819**
Point Transformer*	0.869	0.787
CNN-BiGRU†	0.851	-

and four node geometries, the best performance is again achieved for those geometries that include the sensor pair (1,5). It is thus concluded that a large azimuthal opening angle of the sensor geometry is beneficial for the application of the proposed method for activity classification.

Summarising, the study on sensor network geometry reveals a roughly linear trend in performance increase per added sensor. At the maximum of five sensors, no saturation of performance is observed. This seems to indicate that further expansion of the sensor network can increase the classification performance. Based on the error analysis in Section 5.4.3, promising locations for additional sensors would be those that offer boresight axes that are not in the horizontal plane. These additional sensors would be able to provide information pertaining to the vertical velocity distribution. A secondary note on the geometry study is the relatively low effect of sensor geometry in comparison to the amount of sensors. The former yields performance variations on the order of 2 %, whereas the latter brings about changes in excess of 5 %. This may be due to the range of studied geometries all being relatively restricted to a half-circle, as opposed to e.g. on a sphere or more complex topology. Finally, not all possible network configurations are evaluated, as mentioned in Section 5.3. Under the assumption that the dataset is unbiased in activity location and orientation, this should have a limited impact. However, individual nodes observe different clutter in the laboratory space which potentially affects their contribution to the fusion process.

5.4.2. Comparison to Alternative Approaches

A comparison is made with alternative fusion-classification approaches. Results are obtained on the same dataset, and employing the same L1PO testing scheme. It should be noted that some of the reference methods performed their evaluation on a constrained 5-class problem, where activities like "Falling from walking" and "Falling from standing" are merged into "Falling". Where this is the case, the results for the proposed method are also evaluated for the 5-class problem. Table 5.2 shows the 9-class comparison, and Table 5.3 the 5-class.

For the full 9-class classification task, the reference fusion-classification methods comprise one based on the Point Transformer architecture [37], and one on a hybrid CNN-BiGRU (Gated Recurrent Unit) architecture [33]. For the former, radar data is processed into a point cloud representation with dimensions range, radial velocity, time, and reflection intensity. Point clouds from various sensors are fused by simply adding the points from all individual sensors into a single point cloud object. No data asso-

Table 5.3: Test accuracy and macro F1-score results for the proposed classification method versus two reference methods. All results are based on the same dataset [66] and the same L1PO validation scheme. The results in this table are for a five-class classification problem. *:[37],†: Signal Fusion [38], ‡: Feature Fusion [38].

Classifier	Test Accuracy	Macro F1-score
Proposed	**0.930**	**0.898**
Point Transformer*	-	0.862
GRU†	0.909	0.778
LSTM†	0.910	0.769
bi-GRU†	0.933	0.844
BiLSTM†	0.931	0.836
BiLSTM‡	0.924	0.840

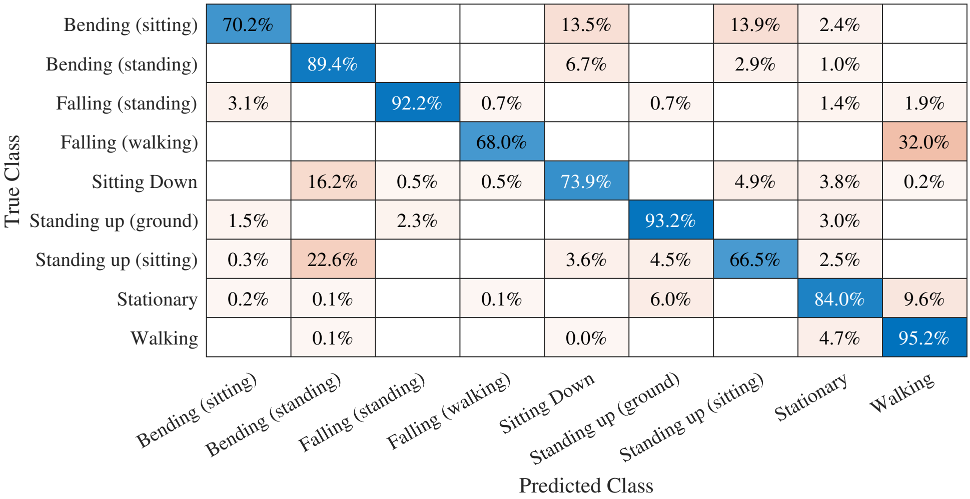
ciation is performed, or an attempt to estimate a more comprehensive representation of reflection intensity and/or velocity distribution, as in the proposed method. In [33], fusion of data from N sensors is achieved through a maxpooling operation over N feature maps. The feature maps are the result of the CNN module processing spectrogram representations of the sensor data. Both reference methods feature a fused representation that is at a higher abstraction level than the proposed method with respect to the kinematics of the observed experimental scene. Because of this, the versatility of the proposed fused representation is assumed to be greater in terms of applicability to other tasks, such as for instance tracking or specific fall detection methods, such as the one proposed in [56]. In addition, the classification performance of the proposed method is improved with respect to the reference methods.

In the case of the constrained 5-class classification problem, favorable results are also achieved with the proposed method. The reference methods are again based on the Point Transformer architecture [37], and a selection of RNN's [38]. Two types of fusion are employed in [38]. First, signal fusion entails element-wise addition of N complex-valued range-time representations from N sensors. A single spectrogram is then computed based on the fused range-time matrix. The second type is feature fusion and comprises the concatenation of N spectrogram representations along the Doppler dimension. In terms of macro F1-score, improvements of almost 4 %pt (percentage point) are achieved when utilizing the proposed method. The test accuracy is within 0.4 %pt of the best performing model.

Finally, the benefits of the velocity reconstruction are illustrated by training a model using only the intensity map data, and without the accompanying velocity vector fields. Sequences are randomly distributed following a 80%/20% ratio into a training and a testing set. This approach yields an accuracy of 50.04 % and a Macro F1-score of 48.45 %, clearly indicating the superiority of the proposed method.

5.4.3. Error Analysis and Computational Considerations

Figure 5.6 shows a representative confusion matrix for a single L1PO result, i.e., with test data from a single participant. The five largest errors in this matrix will be discussed. Most prevalent among the classification errors is the confusion between "Falling (Walk-



	Bending (sitting)	Bending (standing)	Falling (standing)	Falling (walking)	Sitting Down	Standing up (ground)	Standing up (sitting)	Stationary	Walking
Bending (sitting)	70.2%				13.5%		13.9%	2.4%	
Bending (standing)		89.4%			6.7%		2.9%	1.0%	
Falling (standing)	3.1%		92.2%	0.7%		0.7%		1.4%	1.9%
Falling (walking)				68.0%					32.0%
Sitting Down		16.2%	0.5%	0.5%	73.9%		4.9%	3.8%	0.2%
Standing up (ground)	1.5%		2.3%			93.2%		3.0%	
Standing up (sitting)	0.3%	22.6%			3.6%	4.5%	66.5%	2.5%	
Stationary	0.2%	0.1%		0.1%		6.0%		84.0%	9.6%
Walking		0.1%			0.0%			4.7%	95.2%

Figure 5.6: A confusion matrix for the test result for a single participant.

ing)" and "Walking". A likely explanation for this error is the ambiguous transition point between these activities. Since predictions are made at every time step (0.26 s), the ambiguous time between walking and falling constitutes a relatively large amount of equally ambiguous predictions. This ambiguity is compounded by variability in the ground truth due to each participant indicating transition points themselves by clicking on a remote controller. The next four most frequent errors are between activities that all share an initial forward rotation of the torso, including standing up from sitting. Since the case study is constrained to the 2D horizontal plane, these errors can be attributed to a lack of information on both the vertical posture of the subject, as well as a lack of a detailed vertical velocity profile. It is noted that standing up from the ground is kinematically distinct to the other activities with a strong vertical component and as such exhibits fewer errors during classification.

Typical computational requirements, based on the classification case study and available hardware, are reported here. Fusion processing of a full 120 s sequence on a single core of a 3.40GHz i5 CPU takes on average 51 s. However, since time steps are independent, multi-core processing yields significant improvements. Typical CNN-BiLSTM models are on the order of 10MB, and inference time for a full sequence is less than 0.1 s on a NVIDIA Tesla V100S.

5.5. Experimental Feasibility Study in 3D

To demonstrate the ability of the proposed method to yield three-dimensional representations of extended targets and motions, a feasibility study is performed. For this study, data is captured with a four-sensor network that extends not only in the horizontal plane, but in the vertical direction as well. The measurement setup is shown in Figure 5.7, and the sensors are the same as those used in the 2D case study. Two scenes are captured: a metal sphere of diameter 30 cm suspended 120 cm from the ground, and a human sub-

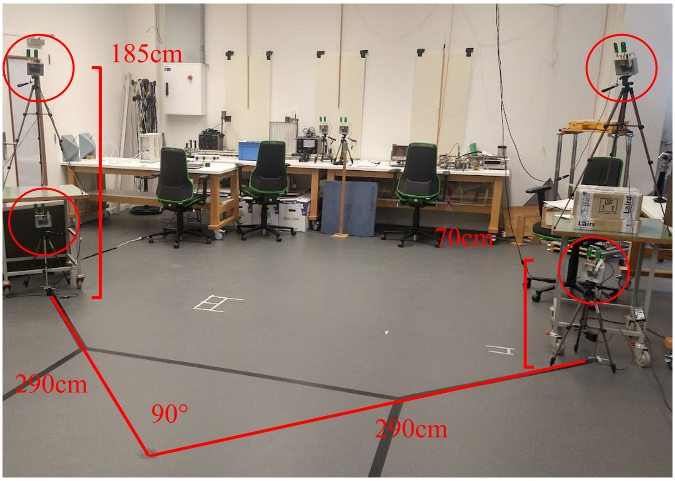


Figure 5.7: The experimental setup used in the 3D feasibility study. Sensors are placed at the base of an isosceles right triangle with sides of 290 cm. The lower sensors are located at a height of 70 cm, the upper sensors at 185 cm. Targets are positioned at the apex of the triangle.

ject performing a series of squats, hence exhibiting movements towards the upward and downward direction periodically.

The data are processed following the proposed method, and the results are shown in Figure 5.8. Subfigures a-d represent 3D intensity and velocity fields at key points during the squatting activity, with the maximum intensity of the thresholded 3D matrix projected on 3 orthogonal axes to facilitate visualisation. The four states of standing upright (a), moving down into squatting position (b), sitting in squat (c), and moving back up to the standing position (d), feature intensity and velocity distributions that correspond to the physical expectation of the human movement. Noise levels in this experiment are however higher, indicating that an alternative sensor geometry or additional preprocessing may be required to fully utilise the proposed method in three dimensions for subsequent classification purposes. Figures 5.8e and 5.8f contain the metal sphere and corresponding 3D intensity distribution respectively. The distribution is visualised by means of two 2D slices of the 3D intensity matrix. The sphere is seen at the correct height and surface area.

5.6. DISCUSSION

In this section, the known and expected method drawbacks will be discussed, based on the obtained results and the theoretical foundations of the approach. Potential solutions are subsequently proposed.

1. For the reconstruction of the velocity in a voxel, a single Doppler value is selected from the full spectrum. In this work, the assumption is made that a single dominant scatterer with a largely localised velocity spread is present in a voxel, and is simultaneously observed by the majority of the radar sensors. When this assumption does not hold, the velocity reconstruction in the respective voxel is performed

with projections that do not yield a consistent true velocity.

2. When multiple targets are present, ghost targets may appear when the number of sensors with line of sight to the different targets is equal to or lower than the amount of spatial dimensions considered.

3. Currently, no calibration of the various sensors' contributions to the intensity field is performed. As such, close proximity of a target to one sensor yields a very strong contribution to the intensity field of the respective sensor. This causes the reconstructed target shape to be distorted, and causes artefacts in the resulting image, as seen in e.g., Figure 5.4.

Regarding possible solutions to the above, first, the current simplistic approach of selecting the Doppler component with the highest signal amplitude can be improved by implementing a peak detection algorithm such as, e.g., a CFAR-based one. Secondly, if two values for each sensor are selected instead of the current single value, the LSE-based approach allows for a straightforward reconstruction of two velocities in each voxel. To this end, the cost function (5.13c) is evaluated for all $\binom{2N}{N}$ combinations. The combination yielding the lowest cost function is then taken to compute the two velocities. Due to the closed form of (5.15b), this approach is computationally viable for reasonable N. For the problem of potential ghost targets, sensor network topology is a key consideration. If a target is obscured for one of the sensors in the network, the intensity map will feature lower values at the target location, making the real target less distinguishable from ghost targets that may be present in the intensity map. The ghost target problem is caused by ambiguities in the angular coordinates. As such, the utilisation of MIMO sensors with the ability to determine Direction of Arrival strongly alleviates potential issues relating to ghost targets. Notably, in the case of Human activity classification and when the amount of people present is not of primary concern, the presence of a ghost target can result in a false alarm, but never in a missed detection of a critical event.

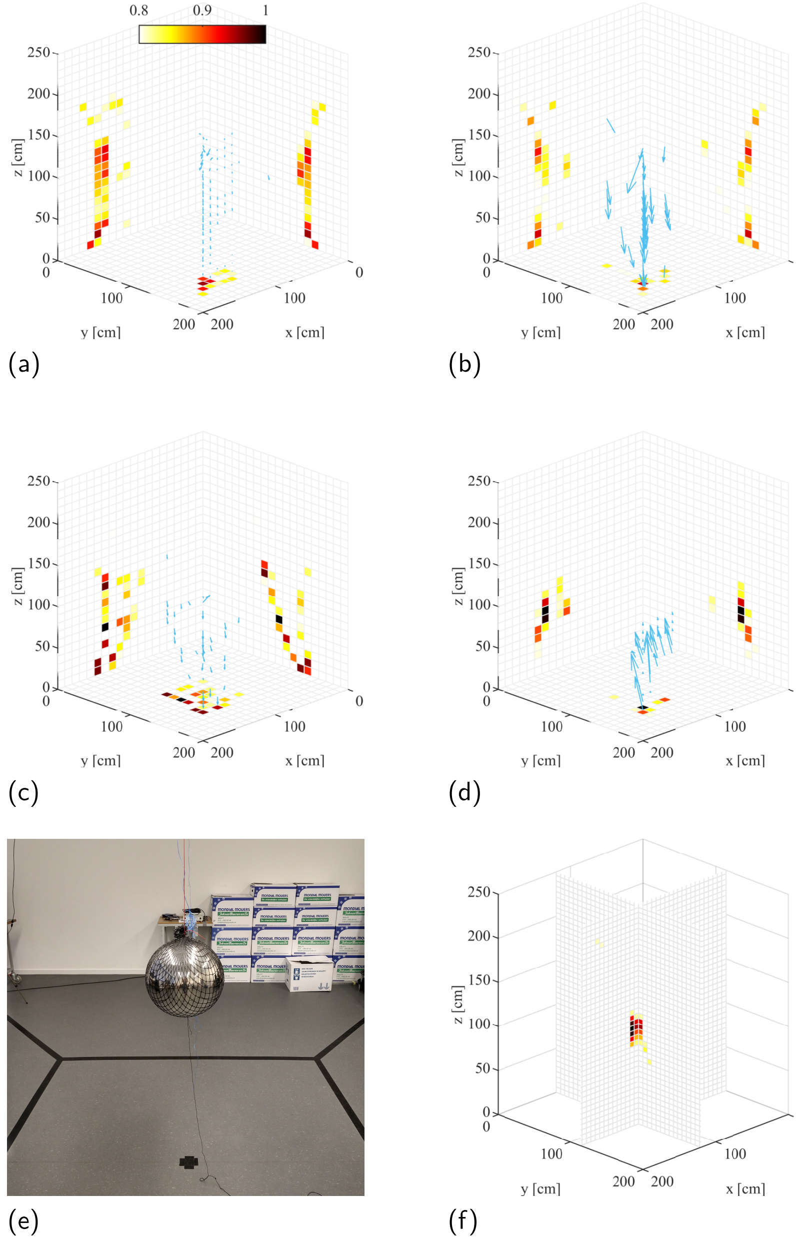
With regards to sensor calibration, an initial improvement can be the inclusion of range compensation according to the radar range equation, or the addition of compensation based on the antenna patterns of the sensors used for the network. The latter can even be employed to anticipate occluded areas in the measurement space.

5.7. Conclusion

This chapter proposes a novel sensor fusion method that processes data from a network of radar sensors and yields three-dimensional representations of both reflection intensity and velocity distribution. The fused representation can be obtained regardless of the capacity of the individual radars to determine angle-of-arrival. The fused data representation are easily linked to the kinematic of the observed target, and allows for versatile application to various tasks, from tracking to human activity classification.

The proposed method is evaluated in a case study, where it is applied to a human activity classification problem. For this task, sequences of human activities from a publicly available dataset are processed using the method. Subsequently, classification is performed by means of a hybrid CNN-BiLSTM model. For a leave-one-person-out testing scheme, a test accuracy and macro F1-score of 87.4 % and 81.9 % are achieved, out-performing alternative fusion methods on the same dataset. An extensive study on the number and position of radar nodes in the network is performed to evaluate their effect on classification performance. It is found that for the five-node geometry utilised in the case study, additional nodes can likely further increase prediction accuracy. Most notably, error analysis indicates that additional sensors that are not in the horizontal plane are expected to improve classification most substantially. An experimental feasibility study is also conducted, successfully demonstrating the proposed method's capability to produce 3D representations of extended target shapes and velocity distributions.

In future work, a network of multiple-input-multiple-output (MIMO) sensors will be utilised to fully exploit the capabilities of the method. Specifically, inclusion of velocity components in the vertical direction is expected to mitigate classification errors between activities that feature distinctive vertical motions.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 5.8: Results for the 3D feasibility study. For subfigures a-d, 3D intensity and velocity fields are represented with the maximum intensity of the thresholded 3D matrix projected on 3 orthogonal axes. (a) Human subject standing upright. (b) Human subject moving down into squatting position. (c) Human subject sitting in squat. (d) Human subject moving back up to the standing position. (e) Metal sphere used for feasibility study. (f) Intensity distribution of metal sphere, represented with two 2D slices of full 3D intensity matrix.

6

CONCLUSIONS

6.1. Major Results and Novel Contributions

The contributions of this PhD research are summarised in the following paragraphs.

- *A novel radar data representation using a point cloud structure in an atypical feature space (Chapter 3)*

  Radar data in uncompressed formats are rarely used for the purpose of activity classification due to their inherent sparsity and the computational overhead involved in neural networks that can process them directly [33, 38]. Point cloud representations are beneficial in scenarios where feature spaces are sparse, as they scale favourably when compared to full matrix representations. Whilst point clouds have been prevalent in radar applications in automotive, their utilisation has primarily been limited to conventional 3D Cartesian spaces. **In contrast, this work proposes a novel approach that leverages point cloud representations, but in an atypical feature space with coordinates representing radar-specific features such as range, Doppler, and time.** Notably, this novel representation allows for the utilisation of a singular Single Input Single Output (SISO) radar sensor, as no localisation of the target in 3D space is required. Importantly, noise and clutter cancellation techniques are applied that directly reduce the size of the input data. This enables the preservation of input data at full radar resolution, rather than suppressing noisy data points but having to allocate memory to them regardless. With the introduction of new variables for classification, the data in a point cloud scales linearly, rather than exponentially. With the aforementioned proposed method, a Leave-one-Person-Out (L1PO) test accuracy and macro F1-score of 86.9 % and 78.7 % are demonstrated on a classification task based on a challenging, publicly available dataset.

- *Demonstration of classification approaches using fixed or dynamically changing windows as viable methods for continuous activity classification. (Chapters 3 & 4)*

  A large part of existing literature on radar-based activity classification focuses on single activity classification, which is inadequate for continuous classification tasks, where transitions between activities are unknown. In recent years, Recurrent Neural Networks (RNNs) have largely been employed to address this challenge, as they are capable of handling continuous sequences of activities, making predictions at time scales in the millisecond range. **This research demonstrates that this high temporal resolution for making predictions is not strictly necessary, and that powerful single activity classifiers can be utilised effectively.** Specifically, windowing approaches with fixed and adaptive window sizes are studied in conjunction with a powerful Point Transformer network. On a five-class activity classification task, fixed-window and adaptive window approaches yield macro-F1 scores of 86.2 % and 88.0 % respectively, compared to 84.4 % for the best RNN-type classifier in the literature.

- *A suitable segmentation algorithm paired with a strong classifier achieves a balance between performance and computational efficiency (Chapter 4)*

  The common approach in the literature of predictions at time intervals in the millisecond range is a strategy that can be computationally prohibitive. The neural networks required for such tasks often necessitate the downsampling of input data, which can degrade classification performance. **This research proposes an alternative approach that incorporates an activity sequence segmentation algorithm in conjunction with a classifier designed to handle the obtained segments.** Sequence segmentation is based on monitoring fluctuations in a quantity derived from the micro-Doppler spectrogram, the Rényi entropy. Classification is achieved by means of a Point Transformer network. The proposed approach is computationally lightweight compared to conventional methods, allowing for increased classification performance while maintaining a reasonable network size. On a five-class human activity classification task, the proposed method is able to attain a macro F1-score of 88.0 % with a model size of ~17 MB, compared to 84.0 % for a 25 MB Bidirectional Long Short Term Memory network.

- *Reconstruction of extended target location and velocity distribution in 3D space (Chapter 5)*

  For the classification of human activities, accurate interpretation of human kinematics is essential. Differentiating the velocities of various limbs can aid classification performance, but only radial velocity in point cloud representations has been utilised in previous literature. **This research proposes a fusion method that reconstructs the location, shape, and velocity profile of extended targets, such as human subjects, in full 3D.** This fused data representation can be obtained regardless of the capacity of the sensors to determine angle-of-arrival, i.e., with SISO sensors. The fused data representation can furthermore be linked to the kinematics of the observed target, and allows for versatile application to various tasks, from tracking to human activity classification. The processing is applied to a publicly available dataset of human activities to obtain the reconstructed intensity and velocity maps, and these representations are utilised for a nine-class classification task. The L1PO test accuracy and macro F1-score for this task reach 87.4 % and 81.9 % respectively.

6.2. Recommendations for Future Research

In the following paragraphs, recommendations for future research are summarised. They include improvements of methods proposed in this PhD research, as well as novel research directions.

- *Improvement of 3D intensity and velocity reconstruction.* Several avenues for enhancing the method presented in Chapter 5 are identified to improve the accuracy of intensity and velocity reconstruction. First, the contribution of each sensor in terms of signal amplitude are currently not range-compensated, leading to less accurate reconstruction of extended target shape, particularly when close to one of the sensors. The primary reason for this omission is the strong influence of background noise at longer ranges. Therefore, noise suppression is required for effective range compensation.

  Secondly, the selection of the Doppler component used for velocity reconstruction can be improved. Currently, only the largest Doppler component in terms of signal amplitude is selected for velocity vector reconstruction. This is however not always the correct choice due to, e.g., occlusion effects. A potential improvement may be found in the least squares approach used to reconstruct the velocity vectors: a low value of the cost-function would imply that the constituent Doppler components likely originate from the same velocity vector.

Thirdly, in the current approach, the selection of ‘physical’ velocity vectors from the reconstructed vector field is performed by setting a threshold on the intensity. A high intensity is assumed to imply that accompanying velocity vectors likely originate from a real target. A better implementation could involve assigning a weight to each vector in the field derived from the reconstructed intensity at that location, eliminating the need for a user-defined threshold while still reflecting the likelihood of the vectors having physical significance.

Finally, the incorporation of radar antenna patterns in the reconstruction could allow for more accurate reconstruction of the reflection intensity in the 3D space. Similarly to the range compensation, compensation for signal strength of a given sensor in a certain point in space will allow for a better-weighted contribution to reconstructed reflection intensity of that specific sensor.

- *Multi-label classification.* This research demonstrates that segmentation of activity sequences yields favourable results in terms of classification performance and computational efficiency. However, multiple activities in a segment are unavoidable, in part due to the temporal uncertainty in the location of a transition point between activities. In these ambiguous cases, allowing for multiple predictions of a classifier can improve the accuracy of the classification task, as demonstrated for example in [50, 54]. As an initial approach, a multi-label classification method may simply be based on classifier confidence. Implementing a detection threshold on classifier output expands the prediction from being solely based on the single most likely class, to including potentially multiple classes in the case of an ambiguous interval.

- *Unsupervised data labelling.* With current prevalence of machine learning approaches, there is an increasing demand for large, annotated radar data sets. Manual labelling of data is a costly endeavour but is still the primary means of annotation [64, 71]. Unsupervised labelling of radar data may for instance be achieved with an accompanying camera. Video activity classification is able to discern comparatively many activities [117], and can therefore be utilised to generate labels to train networks for radar-based approaches, which can then be employed when camera usage is not feasible. Since radar data for different environments can vary substantially based on e.g., presence of clutter and multipath characteristics of the area, a challenge is to have large diversity in video-data.

- *Hierarchical classification.* For long term monitoring of subject health, fine-grained predictions of a fundamental nature such as *Walking* and *Standing Still* may not be optimal for healthcare professionals. Overarching activity types such as *Preparing Meals* or *Wandering* can be considered to be more informative. A hierarchical classification approach could function in a similar vain to Multi-Label classification, where predictions are made on the level of individual motions, but simultaneously on a higher level of activity archetypes. For the latter, continuous classification of a stream of activities is a necessity, further highlighting the importance of this research.

- *Open set classification and anomaly detection.* The majority of current research on activity classification focusses on single-label predictions on a closed set of activities. The accompanying assumption is that all motions can be assigned to a single representative of such a set, which is not realistic. Expanding the cardinality of the closed set allows for the assumption to be increasingly satisfied, but comes at the cost of greatly increased complexity of the discriminating algorithm as well as the labelling of the data for training. Open set classification and anomaly detection are avenues of research that shift the classification problem away from a restrictive set of fundamental activities. Open set classification allows for much greater flexibility in the case of atypical motion types, and anomaly detection constrains the problem to a binary scale that by design encompasses all human motions.

  The idea of fundamental motion types does not necessarily have to be abandoned in order to utilise, e.g., anomaly detection. An abstract activity space can be defined, with dimensions representing for example prediction confidences on a closed set of activities. Within this space, anomalous regions can be defined that represent linear combinations of the overarching set of activities.

- *Application to real scenarios.* All methods in this thesis work have been benchmarked on a carefully crafted experimental dataset of human activities. Despite its advantages compared to other datasets in comparable literature, some aspects are unavoidably artificial in nature. Data capture in a real facility rather than a university environment poses several challenges. First, long-term monitoring is required to collect an amount of rare (critical) events that is large enough to be useful for ML training purposes. For example, in [118], data capture lasted for 11 months. Such long term monitoring will require an automated data labelling approach, but the utilisation of cameras for this purpose comes with privacy and other ethical concerns. Additionally, the amount of data output by radar sensors is large enough to warrant live processing, potentially by constantly retraining a classification model with new data as it becomes available. This next step in radar-based HAR poses significant challenges, both computational to manage the amount of data, and methodological to formulate suitable retraining approaches.

Bibliography

[1] S. Z. Gürbüz and M. G. Amin, "Radar-Based Human-Motion Recognition With Deep Learning: Promising Applications for Indoor Monitoring," *IEEE Signal Processing Magazine*, vol. 36, pp. 16–28, jul 2019.

[2] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante, "Radar and RGB-Depth Sensors for Fall Detection: A Review," *IEEE Sensors Journal*, vol. 17, pp. 3585–3604, jun 2017.

[3] S. Ahmed, S. Abdullah, and S. H. Cho, "Advancements in Radar Point Cloud Generation and Usage in Context of Healthcare and Assisted Living Domain: A Review," *IEEE Sensors Journal*, pp. 1–1, 2024.

[4] S. Dong, L. Wen, Y. Ye, Z. Zhang, Y. Wang, Z. Liu, Q. Cao, Y. Xu, C. Li, and C. Gu, "A Review on Recent Advancements of Biomedical Radar for Clinical Applications," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 5, pp. 707–724, 2024.

[5] C. Li, V. M. Lubecke, O. Boric-Lubecke, and J. Lin, "Sensing of Life Activities at the Human-Microwave Frontier," *IEEE Journal of Microwaves*, vol. 1, pp. 66–78, jan 2021.

[6] 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.

[7] United Nations, Department of Economic and Social Affairs, Population Division, "World population prospects 2019," tech. rep., United Nations, 2019.

[8] G. Bergen, M. R. Stevens, and E. R. Burns, "Falls and fall injuries among adults aged ≥ 65 years - united states, 2014," *Morbidity and Mortality Weekly Report*, vol. 65, no. 37, pp. 993–998, 2016.

[9] B. V. Watts, B. Shiner, Y. Young-Xu, and P. D. Mills, "Sustained effectiveness of the mental health environment of care checklist to decrease inpatient suicide," *Psychiatric Services*, vol. 68, no. 4, pp. 405–407, 2017.

[10] W. G. on Suicidal Behaviors, "Practice guideline for the assessment and treatment of patients with suicidal behaviors," *American Journal of Psychiatry*, vol. 160, no. Suppl. 11, pp. 1–60, 2003.

[11] Q. Lin, D. Zhang, L. Chen, H. Ni, and X. Zhou, "Managing elders' wandering behavior using sensors-based solutions: A survey," *International Journal of Gerontology*, vol. 8, no. 2, pp. 49–55, 2014.

[12] S. Z. Gürbüz, M. M. Rahman, E. Kurtoglu, and D. Martelli, "Continuous Human Activity Recognition and Step-Time Variability Analysis with FMCW Radar," in *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 01–04, IEEE, sep 2022.

[13] S. A. Shah, A. Tahir, J. Le Kernec, A. Zoha, and F. Fioranelli, "Data portability for activities of daily living and fall detection in different environments using radar micro-doppler," *Neural Computing and Applications*, vol. 34, pp. 7933–7953, may 2022.

[14] Y. Yao, C. Liu, H. Zhang, B. Yan, P. Jian, P. Wang, L. Du, X. Chen, B. Han, and Z. Fang, "Fall Detection System Using Millimeter Wave Radar Based on Neural Network and Information Fusion," *IEEE Internet of Things Journal*, pp. 1–1, 2022.

[15] M. M. Rahman, D. Martelli, and S. Z. Gürbüz, "Gait Variability Analysis with Multi-Channel FMCW Radar for Fall Risk Assessment," in *2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 345–349, IEEE, jun 2022.

[16] H. Ji, C. Hou, Y. Yang, F. Fioranelli, and Y. Lang, "A One-Class Classification Method for Human Gait Authentication Using Micro-Doppler Signatures," *IEEE Signal Processing Letters*, vol. 28, pp. 2182–2186, 2021.

[17] A.-K. Seifert, M. G. Amin, and A. M. Zoubir, "Toward Unobtrusive In-Home Gait Analysis Based on Radar Micro-Doppler Signatures," *IEEE Transactions on Biomedical Engineering*, vol. 66, pp. 2629–2640, sep 2019.

[18] Z. Ni and B. Huang, "Open-Set Human Identification Based on Gait Radar Micro-Doppler Signatures," *IEEE Sensors Journal*, vol. 21, pp. 8226–8233, mar 2021.

[19] H. Li, A. Mehul, J. Le Kernec, S. Z. Gürbüz, F. Fioranelli, S. Z. Gurbuz, F. Fioranelli, S. Z. Gürbüz, and F. Fioranelli, "Sequential Human Gait Classification with Distributed Radar Sensor Fusion," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7590–7603, 2021.

[20] T. Sakamoto, P. J. Aubry, S. Okumura, H. Taki, T. Sato, and A. G. Yarovoy, "Noncontact Measurement of the Instantaneous Heart Rate in a Multi-Person Scenario Using X-Band Array Radar and Adaptive Array Processing," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, pp. 280–293, jun 2018.

[21] X. Dang, Z. Chen, and Z. Hao, "Emotion recognition method using millimetre wave radar based on deep learning," *IET Radar, Sonar & Navigation*, vol. 16, pp. 1796–1808, nov 2022.

[22] K.-C. Peng, M.-C. Sung, F.-K. Wang, and T.-S. Horng, "Noncontact Vital Sign Sensing Under Nonperiodic Body Movement Using a Novel Frequency-Locked-Loop Radar," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, pp. 4762–4773, nov 2021.

[23] 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.

[24] X. Li, Z. Li, F. Fioranelli, S. Yang, O. Romain, and J. Le Kernec, “Hierarchical Radar Data Analysis for Activity and Personnel Recognition,” *Remote Sensing*, vol. 12, p. 2237, jul 2020.

[25] B. Erol, S. Z. Gürbüz, and M. G. Amin, “Motion Classification Using Kinematically Sifted ACGAN-Synthesized Radar Micro-Doppler Signatures,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, pp. 3197–3213, aug 2020.

[26] M. M. Rahman, S. Z. Gürbüz, and M. G. Amin, “Physics-Aware Design of Multi-Branch GAN for Human RF Micro-Doppler Signature Synthesis,” in *2021 IEEE Radar Conference (RadarConf21)*, pp. 1–6, IEEE, may 2021.

[27] K. Ahuja, Y. Jiang, M. Goel, and C. Harrison, “Vid2Doppler: Synthesizing Doppler Radar Data from Videos for Training Privacy-Preserving Activity Recognition,” in *2021 CHI Conference on Human Factors in Computing Systems*, (New York, NY, USA), pp. 1–10, ACM, may 2021.

[28] Z. Zeng, M. G. Amin, and T. Shan, “Automatic Arm Motion Recognition Based on Radar Micro-Doppler Signature Envelopes,” *IEEE Sensors Journal*, vol. 20, pp. 13523–13532, nov 2020.

[29] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, “Continuous Human Activity Classification from FMCW Radar with Bi-LSTM Networks,” *IEEE Sensors Journal*, vol. 20, pp. 13607–13619, nov 2020.

[30] E. Kurtoglu, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. S. Crawford, and S. Z. Gürbüz, “ASL Trigger Recognition in Mixed Activity/Signing Sequences for RF Sensor-Based User Interfaces,” *IEEE Transactions on Human-Machine Systems*, pp. 1–14, nov 2021.

[31] N. Nguyen, V.-S. Doan, M. Pham, and V. Le, “SRCNN: Stacked-Residual Convolutional Neural Network for Improving Human Activity Classification Based on Micro-Doppler Signatures of FMCW Radar,” *Journal of Electromagnetic Engineering and Science*, vol. 24, pp. 358–369, jul 2024.

[32] 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.

[33] 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.

[34] Z. Guo, R. G. Guendel, A. Yarovoy, and F. Fioranelli, “Point Transformer-Based Human Activity Recognition Using High-Dimensional Radar Point Clouds,” in *Proceedings of the IEEE Radar Conference*, vol. 2023-May, Delft University of Technology, 2023.

[35] G. Tiwari and S. Gupta, “2-D Point Cloud Transformation Technique to Achieve Range–Angle Invariant Physical Activity Classification Using a Single mmWave Radar,” *IEEE Sensors Letters*, vol. 8, pp. 1–4, jan 2024.

[36] Y. Kim, I. Alnujaim, and D. Oh, “Human Activity Classification Based on Point Clouds Measured by Millimeter Wave MIMO Radar With Deep Recurrent Neural Networks,” *IEEE Sensors Journal*, vol. 21, pp. 13522–13529, jun 2021.

[37] N. C. Kruse, F. Fioranelli, and A. Yarovoy, “Radar Point Cloud Processing Methods for Human Activity Classification With Point Transformer Networks,” *IEEE Transactions on Radar Systems*, vol. 2, pp. 1–12, 2023.

[38] R. G. Guendel, F. Fioranelli, and A. Yarovoy, “Distributed radar fusion and recurrent networks for classification of continuous human activities,” *IET Radar, Sonar & Navigation*, apr 2022.

[39] L. Cao, S. Liang, Z. Zhao, D. Wang, C. Fu, and K. Du, “Human Activity Recognition Method Based on FMCW Radar Sensor with Multi-Domain Feature Attention Fusion Network,” *Sensors*, vol. 23, p. 5100, may 2023.

[40] Z. Liu, S. Guo, C. Ding, L. Tang, and G. Cui, “Human Activity Recognition Based on Multidomain Fusion Network for LFMCW Radar,” in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 10426–10430, IEEE, jul 2024.

[41] X. Qiao, G. Li, T. Shan, and R. Tao, “Human Activity Classification Based on Moving Orientation Determining Using Multistatic Micro-Doppler Radar Signals,” *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2021.

[42] A. Khasnobish, A. Ray, A. Chowdhury, S. Rani, T. Chakravarty, and A. Pal, “Novel Composite Motion Extraction from Velocity Signature of FMCW Radar for Activity Recognition,” in *2021 21st International Radar Symposium (IRS)*, pp. 1–8, IEEE, jun 2021.

[43] S. Z. Gürbüz, B. Erol, B. Çağlıyan, and B. Tekeli, “Operational assessment and adaptive selection of micro-Doppler features,” *IET Radar, Sonar & Navigation*, vol. 9, pp. 1196–1204, dec 2015.

[44] J. Zabalza, C. Clemente, G. Di Caterina, Jinchang Ren, J. J. Soraghan, and S. Marshall, “Robust PCA micro-doppler classification using SVM on embedded systems,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, pp. 2304–2310, jul 2014.

[45] Z. Li, J. Le 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, p. 3473, mar 2023.

[46] S. Björklund, H. Petersson, and G. Hendeby, "Features for micro-Doppler based activity classification," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1181–1187, 2015.

[47] M. B. Özcan, S. Z. Gürbüz, A. R. Persico, C. Clemente, and J. J. Soraghan, "Performance analysis of co-located and distributed MIMO radar for micro-Doppler classification," in *2016 13th European Radar Conference (EuRAD)*, (London), pp. 85–88, IEEE, oct 2016.

[48] Youngwook Kim and Hao Ling, "Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, pp. 1328–1337, may 2009.

[49] S. Z. Gürbüz, C. Clemente, A. Balleri, and J. J. Soraghan, "Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems," *IET Radar, Sonar & Navigation*, vol. 11, pp. 107–115, jan 2017.

[50] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Classification Strategies for Radar-Based Continuous Human Activity Recognition with Multiple Inputs and Multi-Label Output," *IEEE Sensors Journal*, 2024.

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

[52] X. Wang, P. Chen, H. Xie, and G. Cui, "Through-Wall Human Activity Classification Using Complex-Valued Convolutional Neural Network," in *2021 IEEE Radar Conference (RadarConf21)*, pp. 1–4, IEEE, may 2021.

[53] R. Yu, Y. Du, J. Li, A. Napolitano, and J. Le Kernec, "Radar-based human activity recognition using denoising techniques to enhance classification accuracy," *IET Radar, Sonar & Navigation*, vol. 18, pp. 277–293, feb 2024.

[54] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Radar-Based Continuous Human Activity Recognition with Multi-Label Classification," in *Proceedings of IEEE Sensors*, pp. 1–4, IEEE, 2023.

[55] L. Werthen-brabants, G. Bhavanasi, I. Couckuyt, T. Dhaene, and D. Deschrijver, "Split BiRNN for Real-Time Activity Recognition using Radar and Deep Learning," *Scientific Reports*, vol. 12, pp. 1–12, dec 2022.

[56] N. C. Kruse, R. Guendel, F. Fioranelli, and A. Yarovoy, "Distributed Radar Fusion for Extended Target Location and Velocity Reconstruction," in *2024 IEEE Radar Conference (RadarConf24)*, pp. 1–6, IEEE, may 2024.

[57] Y.-S. Chen, K.-H. Cheng, and Y.-A. Xu, “Transformer-Sequential-Based Learning for Continuous HMR with High Similarity using mmWave FMCW Radar,” in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2023.

[58] M. J. Bocus, K. Chetty, and R. J. Piechocki, “UWB and WiFi Systems as Passive Opportunistic Activity Sensing Radars,” in *2021 IEEE Radar Conference (RadarConf21)*, (Atlanta, GA, USA), IEEE, 2021.

[59] A. Gorji, H.-U.-R. Khalid, A. Bourdoux, and H. Sahli, “On the Generalization and Reliability of Single Radar-Based Human Activity Recognition,” *IEEE Access*, vol. 9, pp. 85334–85349, 2021.

[60] S. Z. Gürbüz, E. Kurtoglu, M. M. Rahman, and D. Martelli, “Gait variability analysis using continuous RF data streams of human activity,” *Smart Health*, p. 100334, oct 2022.

[61] Z. Yang, H. Wang, P. Ni, P. Wang, Q. Cao, and L. Fang, “Real-time Human Activity Classification From Radar With CNN-LSTM Network,” in *2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA)*, (Chengdu, China), pp. 50–55, IEEE, aug 2021.

[62] S. Hor, N. Poole, and A. Arbabian, “Single-Snapshot Pedestrian Gait Recognition at the Edge : A Deep Learning Approach to High-Resolution mmWave Sensing,” in *2022 IEEE Radar Conference (RadarConf22)*, (New York, NY, USA), pp. 1–6, IEEE, 2022.

[63] T. Stadelmayer, M. Stadelmayer, A. Santra, R. Weigel, and F. Lurz, “Human Activity Classification Using mm-Wave FMCW Radar by Improved Representation Learning,” in *Proceedings of the 4th ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, (New York, NY, USA), pp. 1–6, ACM, sep 2020.

[64] D. Gusland, J. M. Christiansen, B. Torvik, F. Fioranelli, S. Z. Gürbüz, and M. Ritchie, “Open Radar Initiative: Large Scale Dataset for Benchmarking of micro-Doppler Recognition Algorithms,” in *2021 IEEE Radar Conference (RadarConf21)*, pp. 1–6, IEEE, may 2021.

[65] N. Nguyen, T. Nguyen, M. Pham, and Q. Tran, “Improving Human Activity Classification Based on Micro-Doppler Signatures Separation of FMCW Radar,” in *2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp. 454–459, IEEE, nov 2023.

[66] R. G. Guendel, M. Unterhorst, N. Kruse, F. Fioranelli, and A. Yarovoy, “Dataset of continuous human activities performed in arbitrary directions collected with a distributed radar network of five nodes,” Nov 2021.

[67] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, “Point Transformer,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16239–16248, IEEE, oct 2021.

[68] A. Rényi, "On measures of information and entropy," *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability*, p. 547–561, 1960.

[69] Time Domain, *Datasheet PulsON410*, 11 2013.

[70] Y. He, *Human Target Tracking in Multistatic Ultra-Wideband Radar*. PhD thesis, 2014.

[71] M. J. Bocus, W. Li, S. Vishwakarma, R. Kou, C. Tang, K. Woodbridge, I. Craddock, R. McConville, R. Santos-Rodriguez, K. Chetty, and R. J. Piechocki, "OPERAnet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors," *Scientific Data*, vol. 9, p. 474, dec 2022.

[72] Y. Zhao, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Distributed Radar-based Human Activity Recognition using Vision Transformer and CNNs," in *2021 18th European Radar Conference*, no. April, (London), pp. 301–304, 2022.

[73] P. Svenningsson, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Calibration of Cognitive Classification Systems for Radar Networks for Increased Reliability," in *2022 IEEE Radar Conference (RadarConf22)*, pp. 1–6, IEEE, mar 2022.

[74] X. Wang, Y. Wang, S. Guo, L. Kong, and G. Cui, "Capsule Network With Multiscale Feature Fusion for Hidden Human Activity Classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.

[75] Z. Li, F. Fioranelli, S. Yang, L. Zhang, O. Romain, Q. He, G. Cui, and J. Le Kernec, "Multi-domains based human activity classification in radar," in *IET International Radar Conference (IET IRC 2020)*, pp. 1744–1749, Institution of Engineering and Technology, 2021.

[76] 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, pp. 8–12, jan 2016.

[77] Y. Lin, J. Le Kernec, S. Yang, F. Fioranelli, O. Romain, and Z. Zhao, "Human Activity Classification with Radar: Optimization and Noise Robustness with Iterative Convolutional Neural Networks Followed with Random Forests," *IEEE Sensors Journal*, vol. 18, no. 23, pp. 9669–9681, 2018.

[78] C. Loukas, F. Fioranelli, J. Le Kernec, and S. Yang, "Activity Classification Using Raw Range and I & Q Radar Data with Long Short Term Memory Layers," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, pp. 441–445, IEEE, aug 2018.

[79] M. Wang, Y. D. Zhang, and G. Cui, "Human motion recognition exploiting radar with stacked recurrent neural network," *Digital Signal Processing: A Review Journal*, vol. 87, pp. 125–131, apr 2019.

[80] M. G. Amin and R. G. Guendel, "Radar classifications of consecutive and contiguous human gross-motor activities," *IET Radar, Sonar & Navigation*, vol. 14, pp. 1417–1429, sep 2020.

[81] M. G. Amin, "Micro-Doppler classification of activities of daily living incorporating human ethogram," in *Radar Sensor Technology XXIV* (A. M. Raynal and K. I. Ranney, eds.), p. 14, SPIE, apr 2020.

[82] 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, pp. 20335–20347, sep 2021.

[83] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[84] K. Kehelella, G. Leelarathne, D. Marasinghe, N. Kariyawasam, V. Ariyarathna, A. Madanayake, R. Rodrigo, and C. U. S. Edussooriya, "Vision Transformer with Convolutional Encoder–Decoder for Hand Gesture Recognition using 24-GHz Doppler Radar," *IEEE Sensors Letters*, vol. 6, no. 10, pp. 1–4, 2022.

[85] J. Zheng, Z. Xu, and B. Li, "An Action Recognition Method Based on the Voxelization of Point Cloud From FMCW Radar," in *2022 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC)*, pp. 1–3, dec 2022.

[86] S. Chen, W. He, J. Ren, and X. Jiang, "Attention-Based Dual-Stream Vision Transformer for Radar Gait Recognition," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3668–3672, 2022.

[87] A. Dey, S. Rajan, G. Xiao, and J. Lu, "Fall Event Detection using Vision Transformer," in *2022 IEEE Sensors*, pp. 1–4, 2022.

[88] Y. Gao, N. Ziems, S. Wu, H. Wang, and M. Daneshmand, "Human Health Activity Intelligence Based on mmWave Sensing and Attention Learning," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pp. 1391–1396, dec 2022.

[89] Z. Guo, "Point Transformer-Based Human Activity Recognition using high-dimensional Radar Point Clouds," Master's thesis, Delft University of Technology, 2022.

[90] N. Kruse, F. Fioranelli, and A. Yarovoy, "Continuous Human Activity Classification with Radar Point Clouds and Point Transformer Networks," in *Accepted for 2023 20th European Radar Conference (EuRAD)*, European Microwave Association (EuMA), 2023.

[91] V. Tran-Quang, T. Ngo-Quynh, and M. Jo, "A lateration-localizing algorithm for energy-efficient target tracking in wireless sensor networks," *Ad Hoc & Sensor Wireless Networks*, vol. 34, pp. 191–220, 01 2016.

[92] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[93] J. Hoffmann, S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford, D. de Las Casas, L. A. Hendricks, J. Welbl, A. Clark, T. Hennigan, E. Noland, K. Millican, G. van den Driessche, B. Damoc, A. Guy, S. Osindero, K. Simonyan, E. Elsen, J. W. Rae, O. Vinyals, and L. Sifre, "Training compute-optimal large language models," 2022.

[94] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[95] S. W. Kang, M. H. Jang, and S. Lee, "Identification of human motion using radar sensor in an indoor environment," *IEEE Sensors Journal*, vol. 21, p. 2305, mar 2021.

[96] N. C. Kruse, R. G. Guendel, F. Fioranelli, and A. G. Yarovoy, "Segmentation of Micro-Doppler Signatures of Human Sequential Activities using Rényi Entropy," in *International Conference on Radar Systems (RADAR 2022)*, (Edinburgh), pp. 435–440, Institution of Engineering and Technology, 2022.

[97] A. Lentzas, E. Dalagdi, and D. Vrakas, "Multilabel classification methods for human activity recognition: A comparison of algorithms," *Sensors*, vol. 22, p. 2353, Mar. 2022.

[98] F. Fioranelli, J. Le Kernec, and S. A. Shah, "Radar for Health Care: Recognizing Human Activities and Monitoring Vital Signs," *IEEE Potentials*, vol. 38, pp. 16–23, jul 2019.

[99] H. Nematallah and S. Rajan, "Adaptive Hierarchical Classification for Human Activity Recognition Using Inertial Measurement Unit (IMU) Time-Series Data," *IEEE Access*, vol. 12, pp. 52127–52149, 2024.

[100] J. J. Steckenrider, B. Crawford, and P. Zheng, "GPS and IMU Fusion for Human Gait Estimation," in *24th International Conference on Information Fusion (FUSION 2021)*, (Sun City, South Africa), 2021.

[101] Q. Jian, S. Guo, P. Chen, P. Wu, and G. Cui, "A Robust Real-time Human Activity Recognition method Based on Attention-Augmented GRU," in *2021 IEEE Radar Conference (RadarConf21)*, pp. 1–5, IEEE, may 2021.

[102] M. Liuni, A. Röbel, M. Romito, and X. Rodet, "Rényi information measures for spectral change detection," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 3824–3827, 2011.

[103] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[104] N. Engel, V. Belagiannis, and K. Dietmayer, "Point Transformer," *IEEE Access*, vol. 9, pp. 134826–134840, 2021.

[105] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, “PCT: Point cloud transformer,” *Computational Visual Media*, vol. 7, pp. 187–199, jun 2021.

[106] S. K. Koul and R. Bharadwaj, “UWB and 60 GHz Radar Technology for Vital Sign Monitoring, Activity Classification and Detection,” in *Wearable Antennas and Body Centric Communication*, ch. 7, pp. 191–215, Singapore: Springer, 1 ed., 2021.

[107] R. Zavorka, R. Marsalek, J. Vychodil, E. Zochmann, G. Ghiaasi, and J. Blumenstein, “Human activity classification via Doppler shift estimation from 60 GHz OFDM transmission,” in *2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–5, IEEE, apr 2022.

[108] C. Wang, X. Zhao, and Z. Li, “DCS-CTN: Subtle Gesture Recognition based on TD-CNN-Transformer via Millimeter Wave Radar,” *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[109] L. Wang, Z. Cui, Y. Pi, C. Cao, and Z. Cao, “Adaptive framework towards radar-based diversity gesture recognition with range-Doppler signatures,” *IET Radar, Sonar & Navigation*, vol. 16, pp. 1538–1553, sep 2022.

[110] S. Z. Gürbüz, M. M. Rahman, E. Kurtoglu, E. A. Malaia, A. C. Gurbuz, D. J. Griffin, and C. S. Crawford, “Multi-Frequency RF Sensor Fusion for Word-Level Fluent ASL Recognition,” *IEEE Sensors Journal*, pp. 1–1, 2021.

[111] Z. Li, Y. Liu, B. Liu, J. Le Kernec, and S. Yang, “A holistic human activity recognition optimisation using AI techniques,” *IET Radar, Sonar & Navigation*, sep 2023.

[112] Z. Yu, A. Zahid, A. Taha, W. Taylor, J. L. Kernec, H. Heidari, M. A. Imran, and Q. H. Abbasi, “An Intelligent Implementation of Multi-Sensing Data Fusion With Neuromorphic Computing for Human Activity Recognition,” *IEEE Internet of Things Journal*, vol. 10, pp. 1124–1133, jan 2023.

[113] 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)*, vol. 2021-May, pp. 1–6, IEEE, may 2021.

[114] N. C. Kruse, F. Fioranelli, and A. Yarovoy, “Continuous Human Activity Classification with Radar Point Clouds and Point Transformer Networks,” in *2023 20th European Radar Conference (EuRAD)*, no. September, (Berlin), pp. 302–305, IEEE, 2023.

[115] Z. Zhang, Z. Tian, and M. Zhou, “Latern: Dynamic Continuous Hand Gesture Recognition Using FMCW Radar Sensor,” *IEEE Sensors Journal*, vol. 18, pp. 3278–3289, apr 2018.

[116] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[117] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, M. Suleyman, and A. Zisserman, “The kinetics human action video dataset,” *CoRR*, vol. abs/1705.06950, 2017.

[118] E. Stone, M. Skubic, M. Rantz, C. Abbott, and S. Miller, “Average in-home gait speed: investigation of a new metric for mobility and fall risk assessment of elders,” *Gait Posture*, vol. 41, pp. 57–62, Jan. 2015.

ACKNOWLEDGEMENTS

With these closing words I want to thank those individuals who were vital for the realisation of this thesis, either through their direct contributions or their kind support over the years. I want to thank my promotor and supervisor Olexander Yarovyi, for his guidance and advice. You have always given me the assurance that I had the full support of yourself and the group. I never felt like I was on my own. Francesco, this acknowledgement section could easily span several pages just including all that you've done to support my PhD research. Your dedication to your flock of PhD students is commendable, as are your broader efforts in support of the group, the research field as a whole, and probably other commitments that I'm not even aware of. You show all the qualities that an exceptional supervisor should have. You advise, invent, motivate, review, and correct, all whilst attending the majority of the (in)formal social events. Many times I have walked out of our weekly meeting feeling more confident about the project.

My gratitude goes out to my doctoral committee, for their time and efforts in reading my thesis, and providing their expertise. Special thanks to Matt Ritchie who gave acute feedback to my thesis, presumably whilst sleep-deprived and with a newborn on one arm. My thanks as well to Julien le Kernec, for taking on a role of external supervisor in my first year and for our continued discussions of research topics in the years that followed.

I want to thank my colleagues in the MS3 group who accepted a stray physicist with open arms. My four years in the group have been an absolute pleasure, mostly thanks to you. I'm happy that the end of my PhD journey does not mark the end of our planned activities. I look forward to brewing an elixir of amnesia for the fermentation festival, and to our attempt at swimming 250 m in a straight line in May.

Ronny, you were my HAR companion for the majority of my four years and I want to thank you for the insightful discussions, for teaching me about close quarters combat socks, for always offering help wherever you could, and for your great sense of humour. Ignacio, thank you for the rabbit holes, for the spoiler-free book advice, and for the biweekly game of squash. I look back fondly to our travels together, and witnessing Scooter Guy together was a bonding experience in my regard. I sincerely hope we can keep up our weekly game of Civilization for many more years.

Thanks Alec, for your aid in making the GPU-machine the proper fire hazard that it was always meant to be. Your diligence has not gone unnoticed. Many thanks go to our invaluable support staff, Pascal, Peter, Fred, Antoon, Minke, and of course Esther.

To my dear friends from my time in Leiden, Delft, and Groningen, I say thank you. Thank you for the game nights and the parties, for the W&R Tuesdays (sorry I didn't attend more often) and the outings, for the shared travels and for Raki on a bridge. Above all, thank you for your support over the years.

Hinke, we started our academic journey together, in groep Paars. On our PhD adventure we again joined as travel companions. Thank you for giving me a standard of hard work and excellence to strive towards, and thank you for your boundless love and understanding.

I want to give thanks to my loving family. Thank you Tycho and Maaike, for your support and genuine interest over the years, and for giving the world Junis & Miro. I finally wish to thank my parents, Gerard and Julia, who were not only instrumental in bringing about my presence on earth, but who have always nurtured an environment of learning and curiosity. I owe everything to you.

About the Author

Nicolas Kruse received the B.Sc. degree in Applied Physics in 2017, at the Delft University of Technology, the Netherlands, and received the M.Sc. degree in Physics in 2020 at the University of Groningen, the Netherlands. He joined the Microwave Sensing, Systems, and Signals group at the faculty of Electrical Engineering of the Delft University of Technology in March of 2021, where he is currently researching classification algorithms for continuous human activity sequences through micro-doppler signatures.

List of Publications

Journal Papers

7. **Kruse, N. C.**, Daalman, A., Fioranelli, F., & Yarovoy, A. G. (2025). *Radar Point Cloud-based Continuous Human Activity Classification Using Rényi Entropy Segmentation Methods.* [Under Review] IEEE Transactions on Radar Systems.

6. **Kruse, N. C.**, Guendel, R. G., Fioranelli, F., & Yarovoy, A. (2025). *Reconstruction of Extended Target Intensity Maps and Velocity Distribution for Human Activity Classification.* In IEEE Transactions on Radar Systems, 3, 14-25.

5. Ullmann, I., Guendel, R. G., **Kruse, N. C.**, Fioranelli, F., & Yarovoy, A. (2024). *Classification Strategies for Radar-Based Continuous Human Activity Recognition with Multiple Inputs and Multi-Label Output.* In IEEE Sensors Journal, 24, 40251-40261.

4. Guendel, R. G., **Kruse, N. C.**, Fioranelli, F., & Yarovoy, A. (2024). *Multipath Exploitation for Human Activity Recognition using a Radar Network.* In IEEE Transactions on Geoscience and Remote Sensing, 62, 1-13.

3. Ullmann, I., Guendel, R. G., **Kruse, N. C.**, Fioranelli, F., & Yarovoy, A. (2023). *A survey on radar-based continuous human activity recognition.* In IEEE Journal of Microwaves, 3, 938–950.

2. **Kruse, N. C.**, Fioranelli, F., & Yarovoy, A. (2023). *Radar Point Cloud Processing Methods for Human Activity Classification with Point Transformer Networks.* IEEE Transactions on Radar Systems, 2, 1–12.

1. Fioranelli, F., Guendel, R. G., **Kruse, N. C.**, & Yarovoy, A. (2023). *Radar Sensing in Healthcare: Challenges and Achievements in Human Activity Classification & Vital Signs Monitoring.* In Bioinformatics and Biomedical Engineering (pp. 492–504). Springer.

Conference Proceedings

6. **Kruse, N. C.**, Guendel, R., Fioranelli, F., & Yarovoy, A. (2024). *Distributed Radar Fusion for Extended Target Location and Velocity Reconstruction.* 2024 IEEE Radar Conference (RadarConf24), 1–6.

5. **Kruse, N. C.**, Fioranelli, F., & Yarovoy, A. (2023). *Continuous Human Activity Classification with Radar Point Clouds and Point Transformer Networks.* 2023 20th European Radar Conference (EuRAD), September, 302–305.

4. Ullmann, I., Guendel, R. G., **Kruse, N. C.**, Fioranelli, F., & Yarovoy, A. (2023). *Radar-Based Continuous Human Activity Recognition with Multi-Label Classification.* Proceedings of IEEE Sensors, 1–4.

3. **Kruse, N. C.**, Guendel, R. G., Fioranelli, F., & Yarovoy, A. G. (2022). *Segmentation of Micro-Doppler Signatures of Human Sequential Activities using Rényi Entropy.* International Conference on Radar Systems (RADAR 2022), 435–440.

2. Guendel, R., **Kruse, N. C.**, Fioranelli, F., & Yarovoy, A. (2022). *Exploiting Radar Data Domains for Classification with Spatially Distributed Nodes.* NATO SET-312 Research Specialists' Meeting on Distributed Multi-Spectral/Statics Sensing.

1. Svenningsson, P., **Kruse, N. C.**, Fioranelli, F., & Yarovoy, A. (2022). *Calibration of Cognitive Classification Systems for Radar Networks for Increased Reliability.* 2022 IEEE Radar Conference (RadarConf22), 1–6.

DATASETS

1. Guendel, Ronny G.; Unterhorst, Matteo; **Kruse, Nicolas**; Yarovoy, Alexander; Francesco Fioranelli (2024): *Dataset of continuous human activities performed in arbitrary directions collected with a distributed radar network of five nodes.* 4TU.ResearchData. dataset. Available: https://doi.org/10.4121/eda41444-9ae8-4655-aac5-4d2541b2bcd8