

Fady Ashraf William Aziz

»Investigating Radar-based Micro-motion Signatures for Human Detection and Identification in Short Range Indoor Environments«

Fady Ashraf William Aziz

Investigating Radar-based Micro-motion Signatures for Human Detection and
Identification in Short Range Indoor Environments

Herausgeber

Univ.-Prof. Dr.-Ing. Thomas Bauernhansl^{1,2}

Univ.-Prof. Dr.-Ing. Dipl.-Kfm. Alexander Sauer^{1,3}

Univ.-Prof. Dr.-Ing. Kai Peter Birke⁴

Univ.-Prof. Dr.-Ing. Marco Huber^{1,2}

¹ Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA, Stuttgart

² Institut für Industrielle Fertigung und Fabrikbetrieb (IFF) der Universität Stuttgart

³ Institut für Energieeffizienz in der Produktion (EEP) der Universität Stuttgart

⁴ Institut für Photovoltaik (ipv) der Universität Stuttgart

Kontaktadresse:

Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA
Nobelstr. 12
70569 Stuttgart
Telefon 0711 970-1100
info@ipa.fraunhofer.de
www.ipa.fraunhofer.de

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliografische Daten sind im Internet über <http://dnb.de> abrufbar.

Zugl.: Stuttgart, Univ., Diss., 2023

D 93

2023

Druck und Weiterverarbeitung:

Fraunhofer-Druckerei, Stuttgart
Für den Druck des Buches wurde chlor- und säurefreies Papier verwendet.



Dieses Werk steht, soweit nicht gesondert gekennzeichnet,
unter folgender Creative-Commons-Lizenz:
Namensnennung – Nicht kommerziell – Keine Bearbeitungen
International 4.0 (CC BY-NC-ND 4.0).

**Investigating Radar-based Micro-motion Signatures for
Human Detection and Identification in Short Range Indoor
Environments**

**Von der Fakultät für Konstruktions-, Produktions- und Fahrzeugtechnik
der Universität Stuttgart
zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Abhandlung**

Vorgelegt von

Fady Ashraf William Aziz

aus Kairo, Ägypten

Hauptberichter: Univ-Prof. Dr.-Ing. habil. Marco F. Huber

Mitberichter: Dr. Sevgi Zubeyde Gurbuz,
Assistant Professor,
The University of Alabama

Tag der mündlichen Prüfung: 07.03.2023

Institut für Industrielle Fertigung und Fabrikbetrieb IFF. der Universität
Stuttgart

2023

Acknowledgements

I am deeply honored to have reached this significant milestone in my academic journey as I document my research work with radar, which has been an unwavering companion over the past five years. I firmly believe in the power of collaboration and recognize that this work is not solely attributed to me; I am merely its presenter. First and foremost, I extend my heartfelt gratitude to my beloved mother for her unwavering support and constant encouragement. To my dear father, who has been my rock, I am immensely grateful for his guidance and unwavering belief in my abilities. I am also indebted to my brother, Andrew Aziz, whose exceptional example has served as a guiding light in my life. I would like to extend my heartfelt appreciation and gratitude to three dear friends who have played a significant role as key contributors to the success of this thesis: Youssef Ghally, Monika Markos, and Jerome Ghally.

I would like to express my deepest gratitude and appreciation to my dear friend Sherif Abdulaif for being an outstanding support and work partner throughout my academic journey. Together, we have accomplished great success and shared countless memorable moments along the way. I would like to express my sincere appreciation to all my friends who have stood by me throughout this arduous journey, especially during the challenging times. Their names hold immense significance, and I would like to acknowledge their unwavering support: Mina Abdallah, Fady Derias, Marina Albert, Mirna Osama, Monica Ramsis, Mina Michel, Mina Hany, Mina Raafat, George Bassem, Karim Medhat, Kiro Samy, Fady Matta, Laura Asharf, Freddy Matta, Lara Matta, Bassam Elmakhazangy, Ahmed Talaat, John Philip, Sarah Ghattas, Karim Abdallah and Veronica Mawhoub. I would like to express my heartfelt gratitude and appreciation to two dear friends who hold a special place in my heart: Michael Hany and Olivia Ryad. I would like to extend my deepest gratitude and appreciation to two of my most beloved friends, Youstina Nagy and Mario Morad, who hold a special place in my heart. To Andrew Nashaat, Fady Elgawly, Nadine Elgawly, Carla Karim, and Kiro Adel, my cherished friends, I am grateful for the profound meaning of friendship that you have brought into my life. Special thanks go to

Carla Karim and Sandra Soliman, who played instrumental roles in the preparations for my defense. I would like to extend my sincere appreciation to Marina Atef, an exceptional graphic designer and dear friend, for her invaluable assistance in creating the figures for my thesis. I would like to express my deepest gratitude and appreciation to my best and dearest friend, Philip Nader. I am grateful for the countless moments of laughter, shared experiences, and unwavering support that Philip has provided me with. His presence in my life has made a significant and positive impact, and I am forever grateful for his friendship. Thank you, Philip, for being there for me every step of the way.

I would like to express my heartfelt gratitude to the following individuals who have played pivotal roles in my academic journey: Dr. Urs Schneider, the head of my department, deserves special recognition for his exceptional support and care throughout the completion of this thesis. Dr. Schneider's guidance and unwavering dedication have made a significant impact on my success. I am truly indebted to him for sharing his extensive professional knowledge and for treating me like a member of his own family. I am grateful for the opportunity to have started my professional career under his supervision for the first five years.

I extend my sincere appreciation to Dr. Christophe Maufroy for his unwavering support and assistance throughout this journey. His guidance and mentorship have been invaluable. Additionally, I would like to express my gratitude to Dr. Maufroy for the incredibly thoughtful gift he presented me with on the day of my defense—the best graduation hat ever. It is a cherished reminder of this significant milestone.

I am immensely thankful to Professor Marco F. Huber for his continuous support throughout the entirety of my academic journey. Professor Huber's unwavering belief in my abilities and his genuine concern for my success have been truly inspiring. His guidance and mentorship have played a vital role in shaping my career.

Lastly, I want to express my deepest gratitude to Jesus Christ, my dearest and best friend throughout my life. His presence, guidance, and unwavering faith have provided me with strength, comfort, and inspiration every step of the way.

To all these individuals and everyone who has supported me along this challenging and rewarding journey, I extend my heartfelt appreciation for their invaluable contributions.

Abstract

The study of radar for human recognition based on deep learning is becoming increasingly popular. It has been demonstrated that the micro-Doppler (μ -D) spectrograms effect can reflect walking human gait by capturing the periodic micro-motions of the limbs. The research scope was extended to include human recognition for variable activities, and hence, a broad number of applications have been investigated, such as fall detection. In addition to this, there are two main factors that have introduced the radar as a powerful sensor for such applications. First is the radar detection capability that is not affected by any environmental limitations. Second, the multiple-input-multiple-output transmission protocol that enabled the radar detection and tracking for multiple humans.

Radar sensors are to be investigated for indoor human detection and identification in this thesis. A short-range analysis ($\leq 10m$) is the main focus of all of the proposed approaches. As a start of the research, the main characteristics of the captured walking μ -D signature are extracted and analyzed. Initial investigations are conducted utilizing a combined half-simulated signal model that we have developed. The model simulates a typical single-input-single-output radar for monitoring a single walking target. The range and velocity are the first two radar features that are investigated. Human-robot classification is used as a case study to test the viability of the research findings. The range feature is evaluated to reflect the micro motion behavior to ensure an enhanced detection capability. Due to the behavior of body parts within the μ -D signatures, the results of the initial feasibility studies formed the basis of the concept of human detection. In addition, it drew our attention to the significance of utilizing multiple radar features for enhanced detection. Walking is the primary activity in all of our investigations; it has been referred to as the most vital activity for studying the μ -D effect in radar.

Human identification is the second main research goal due to its critical requirement in most security systems. The walking activity is fixed, and hence, similar μ -D signatures are expected among humans. Accordingly, this task is regarded as the hardest and most generic, which can be generalized to other scenarios afterward. As a result, the designed

approach can be easily applied to other human recognition tasks. For a single decision, there is always a trade-off between the number of included subjects and the length of the captured spectrogram for the identification task. As a result, the research goal is to overcome these limitations by combining multiple radar features to reduce classification ambiguity while increasing the number of classes and achieving real-time classification. To accomplish this, the radar's estimated angle of arrival is investigated in order to generate a newly-measured micro angular velocity spectrogram that reflects the behavior of the body's micro motions on the angle of arrival feature.

Consideration is also given to the significant degradation in spectrogram quality when captured from different aspect angles with respect to the radar. The main concept is to look into a feasible deep learning approach to ensure an adaptive solution. The generalization and robustness of the final approach are assessed by applying it to an activity classification task that includes fall detection. For this aspect, the thesis work foresees two major aspects. First, reduce the acquired training dataset. Second, the real-time aspect is taken into account in such a way that all classification is done on a single occurrence of the activity, where a single walking step lasts for $\approx 0.5s$.

Contents

List of Acrynomes	xiv
List of Symbols	xv
List of Figures	xviii
List of Tables	xxi
1 Introduction	1
1.1 Motivation	2
1.2 Research Objectives	4
1.3 Thesis Overview	5
1.4 Thesis Organization	7
2 Background	9
2.1 Overview of Radar Systems	9
2.1.1 The Un-modulated Continuous Wave Radar	10
2.1.2 Frequency Modulation Continuous Wave Radar	13
2.1.3 Multiple-Input-Multiple-Output Transmission Protocol	16
2.2 Micro-Motion Signatures in Radar	21
2.2.1 Time-Frequency Representation	22
2.2.2 Micro-Motion Signature of a Moving Rigid Body	23
2.2.3 Micro-Motion Signature of a Moving Nonrigid Body	27
2.3 Human Walking Gait Analysis	28
2.3.1 Walking Human Kinematics	29
2.3.2 Human Walking Emperical Mathematical Model	30
2.4 Deep Learning Background	37
2.4.1 Deep Feedforward Network	38
2.4.2 Deep Convolutional Neural Network	42

2.5	Literature Review	45
2.6	Summary	57
3	Radar Signal Processing	59
3.1	The Main Radar Features	61
3.2	Spectrogram Computation	64
3.2.1	Micro-Doppler Spectrogram	65
3.2.2	Micro-Angular Spectrogram	67
3.3	Activity Adaptive Monitoring	69
3.3.1	Spectrogram Preprocessing	70
3.3.2	Keypoint Vector Extraction	72
3.3.3	Spectrogram Adaptive Segmentation	74
3.4	Summary	76
4	Human Detection While Walking	77
4.1	Combined Signal Model	77
4.1.1	Global Human Walk Model	78
4.1.2	Qualisys Walking Model	80
4.1.3	Human Body Segment Representation	81
4.2	Radar Features Interpretation	84
4.2.1	Comparison Between Simulations and Real Measurements	88
4.3	Decomposing Limbs Trajectories	89
4.3.1	Methodology	90
4.3.2	Results	92
4.3.3	Conclusion	93
4.4	Human-Robot Classification	93
4.4.1	Methodology	94
4.4.2	Results and Discussion	102
4.4.3	Summary	103
5	Feasibility Study for Human Identification While Walking	105
5.1	Human Identification Limitations	105

5.2	Methodology for Human Identification	106
5.2.1	Radar System Parametrization	107
5.2.2	Treadmill Experimental Setup	107
5.3	Autoencoder for Signature Investigation	109
5.4	Person Identification Network Architecture	111
5.5	Results and Discussion	112
5.6	Summary	114
6	Metric Learning for Free-motion Person Recognition	117
6.1	Metric Learning Approach	118
6.1.1	Fundamentals of Triplet Loss	118
6.1.2	Implementation of Triplet Loss	120
6.1.3	Triplet Mining	122
6.2	Neural Network Overview	125
6.2.1	Inception Architecture	125
6.2.2	Deep Convolutional Residual Inception Network	126
6.3	Human-ID Task	127
6.3.1	Problem Analysis	128
6.3.2	Radar Setup and Paramaterization	129
6.3.3	Experimental Setup	132
6.3.4	Results and Discussion	134
6.3.5	Embedding Space	136
6.4	Activity Recognition task	137
6.4.1	Experimental Setup	137
6.4.2	Slicing Techniques	139
6.4.3	Results and Discussion	140
6.4.4	Summary	145
7	Summary	147
7.1	Research Limitations and Future Work	149
	Bibliography	150

A	Appendix	i
A.1	Description of the Limb Decomposition Process	i
A.2	Datasheet of Ancortek Radar	i
A.3	Histogram-based Thresholding	xxv
A.4	Datasheet of Radarbook2 Radar	xxv

List of Acrynomes

AoA	Angle of Arrival
BMI	Body Mass Index
CAE	Convolutional Autoencoders
CDMA	Code Division Multiple Access
CFAR	Constant False Alarm Rate
CIP	Class Inclusion Probability
CM	Body Center of Mass
CoG	Center of Gravity
CW	Continuous Wave
DB-SCAN	Density-Based Spatial Clustering Of Applications with Noise
DDRN	Deep Discriminative Representation Network
FDMA	Frequency Division Multiple Access
FFT	Fast Fourier Transform
FMCW	Frequency Modulation Continuous Wave
FOV	Field of View
FSL	Few-shot Learning
GAN	Generative Adversarial Network
IF	Intermediate Frequency
ILSVRC	ImageNet Large-Scale Visual Recognition Challenge

K-NN	K-Nearest Neighbors
LIDAR	Light Detection and Ranging
LOS	Line of Sight
LPF	Low Pass Filter
MIMO	Multiple-Input-Multiple-Output
MLP	Multilayer Perceptron
mm-wave	Millimeter Wave
MRA	Multi-Scale Residual Attention Network
NB	Naïve Bayes
PRF	Pulse Repetition Frequency
RA	Range-Angle map
RCN	Reservoir Computing Networks
RCS	Radar Cross Section
RD	Range-Doppler map
RNN	Recurrent Neural Networks
SIMO	Single-Input-Multiple-Output
SISO	Single-Input-Single-Output
SNR	Signal to Noise Ratio
STFT	Short Time Fourier Transform
SVM	Supported Vector Machine
TDMA	Time Division Multiple Access

List of Symbols

Greek Symbols

Symbol	Unit	Description
α	-	Margin for fulfilling the distance of the triplet loss
β	Radians	The induced phase difference due to multiple receiving in MIMO
δ	m	The travelled distance by the signal to be received by consecutive Rx antenna
λ	m	Transmission wavelength
ω	Radians/s	Angular velocity of a rotating rigid body
$\psi_{IF}(t)$	Radians	The phase of the signal at the intermediate frequency
$\psi_{Rx}(t)$	Radians	Phase shift induced in the received signal due to a moving target
σ_R^2	-	Variance of range data in RD map.
σ_R	-	Standard deviation of range data in RD map.
σ_v^2	-	Variance of velocity data in RD map.
σ_v	-	Standard deviation of velocity data in RD map.
θ_{res}	Radians	The angular resolution in MIMO radar
ζ	Radians	Constant phase shift in the received signal of the FMCW radar
f_D	Hz	Doppler frequency
Δt	s	Time shift in the received radar signal
$\rho(x, y, z)$	V	The reflectivity function of the scattering point
μ -D	m/s	Micro-Doppler
μ -R	m	Micro-Range

Symbol	Unit	Description
$\mu\text{-}\omega$	Radians/s	Micro-Omega

Latin Symbols

Symbol	Unit	Description
A_i	V	Approximation method for RCS of ellipsoid at time instant i
A	V	Amplitude of the transmitted signal
B	Hz	Transmission bandwidth
G	-	The antenna gain
$I_{Rx}(t)$	V	In-phase component of the received radar signal after down-conversion
N_p	chirps/frame	Number of chirps per frame in the FMCW radar
N_s	samples/chirp	Number of samples per chirp in the FMCW radar
N_{FFT}	Samples	Number of samples for the FFT operation
N_{ab}	samples	The number of AoA bins that are suspected to include targets
N_{rb}	samples	The number of range bins that are suspected to include targets
P_{Tx}	W	The average transmitted power
P_{inc}	W	The power density of the incident signal
$Q_{Rx}(t)$	V	Quadrature component of the received radar signal after down-conversion
R_{max}	m	Maximum detectable range by the FMCW radar
R_{res}	m	Range resolution in the FMCW radar
R	Hz	Target range from the radar
T_p	s	Transmission sweep time
V_{ba}	-	Valid number of triplets for the batch all
V_{bh}	-	Valid number of triplets for the batch hard

Symbol	Unit	Description
V	m/s	Translational velocity of a rotating rigid body
$s(t)$	-	The reflected echo from a scatter point with respect to the space-fixed system
σ	-	The Object's RCS
\underline{R}_p	m	The range vector pointing from the two origins of the space-fixed system and the body-fixed system
\underline{v}	m/s	The radial velocity unit vector
a	-	Anchor sample of the triplet loss
c	m/s	The speed of light in free space
f_R	m/s	The Doppler shift induced due to rotational motion
f_T	m/s	The Doppler shift induced due to translational motion
f_r	Hz	Frequency shift induced due to the target's range
f_s	Hz	Sampling frequency of the received chirp by the FMCW radar
$f_t(t)$	Hz	Transmitted radar signal
n	-	Negative sample of the triplet loss
p	-	Positive sample of the triplet loss
$s_D(t)$	V	Complex received radar signal
$s_{Rx,even}$	V	received RD maps due to even receivers
$s_{Rx,odd}$	V	received RD maps due to odd receivers
$s_{Rx}(t)$	V	Received radar signal
$s_{Tx}(t)$	V	Transmitted radar signal
$s_{if}(t)$	V	Signal at intermediate frequency after the down-conversion
v_{max}	m/s	Maximum detectable velocity by the FMCW radar
v_{res}	m/s	Velocity resolution in the FMCW radar
v_{wr}	m/s	Relative walking velocity for a human
v	m/s	Radial velocity

List of Figures

2.1	In CW radar, the structure of transmitted and received signals.	10
2.2	Signal comparison between transmitted and received signals.	13
2.3	Angle of arrival (AoA)	17
2.4	Principle of a SIMO radar.	18
2.5	Spectral angle resolution according to different numbers of Fourier coefficients.	19
2.6	The principle of a MIMO radar.	21
2.7	Rotation of a rigid body with respect to a space-fixed system (X,Y,Z)	24
2.8	Simulated motion of free oscillating pendulum	25
2.9	Simulated motion of damping oscillating pendulum	26
2.10	μ -D signature induced due to the bird's flying activity	27
2.11	The two phases of the walking gait cycle for both body sides.	28
2.12	Translational trajectories.	33
2.13	Motion trajectories of the pelvis within single gait cycle	34
2.14	Stretching trajectories of the lower body joints	36
2.15	Stretching trajectories of the upper body joints.	37
3.1	Cubic representation of raw radar data.	60
3.2	Range Profiles of MIMO-FMCW Radar.	61
3.3	Range-Doppler map for walking motion within swinging phase.	62
3.4	Example for an Range-AoA map for a MIMO radar.	63
3.5	Reformation of radar raw data for CDMA analysis.	64
3.6	The μ -D spectrogram computation.	65
3.7	The non-linearity of a gamma correction for different values of γ	66
3.8	The μ - ω spectrogram computation.	67
3.9	Micro-motion signatures of a complete walking gait cycle in the LOS direction.	69
3.10	Examples of the thresholding techniques on a gray-scale spectrogram.	72
3.11	A visualization of the morphological processing chain.	73

3.12	Visualization of keypoint vectors on a grayscale μ -D spectrogram slice.	74
4.1	Body orientation with respect to the body-fixed system (x,y,z).	78
4.2	The human body segmentation based on the global human walk model.	79
4.3	Qualisys tools used for motion capture.	80
4.4	Velocity trajectories captured by Qualisys.	81
4.5	The body segment orientation with respect to the radar.	83
4.6	μ -D signatures of upper body parts.	85
4.7	μ -D signatures of lower body parts.	86
4.8	Micro motions interpretations on the main radar features	87
4.9	Comparison between μ -D signatures of simulations and real measurements. . .	88
4.10	Examples for the RD maps after power removal.	90
4.11	Representation of μ -D and μ -R features induced due to Qualisys model.	91
4.12	Comparison between RD maps induced due to walking human and moving robot.	96
4.13	Accuracy curves of conventional learning methods.	97
5.1	BMI-based distribution of both genders participating in the investigation. . .	108
5.2	Schematic for the treadmill experimental setup, and the corresponding μ -D signature.	108
5.3	The architecture of the proposed CAE.	109
5.4	Two clusters are observed on each subject from the right and left body sides.	111
5.5	Architecture of the utilized ResNet-50 network.	111
5.6	Examples of half gait μ -D signatures due to different BMIs.	113
5.7	A classification accuracy of 98% and 84% for the high and varying SNR, respectively.	113
6.1	The architecture of the triplet loss.	120
6.2	The three different types of negative samples: hard, semi-hard and easy. . . .	121
6.3	The inception network main building block.	126
6.4	The residual inception network architecture.	127
6.5	Comparison between vertical and horizontal radar polarization for walking in LOS.	129
6.6	Comparison between vertical and horizontal radar polarization for walking in 30°.	130

6.7	The effect of the rotation of the radar sensor frontend on the captured spectrograms.	132
6.8	Comparison of the μ -D captured due to walking activity.	133
6.9	Experimental setup for the non-constrained walking.	133
6.10	Error rate performance due to the dataset complexity and the observation window length.	135
6.11	t-SNE visualizations of the embedding space.	136
6.12	An example for the adaptive extraction of single instant of bending activity. .	138
6.13	The μ -D and μ - ω signatures for the main 8 activities.	140
6.14	Falling/Standing from falling signatures.	141
6.15	Confusion matrix for all the micro-motion signatures constellations.	142
6.16	t-SNE for the adaptive-sliced and combined μ -D and μ - ω signatures.	143
6.17	Confusion matrix after applying FSL including the 10 activities.	144
A.1	Algorithm of limbs decomposition presented in Section 4.3 and published in Abdulatif et al. 2017	i
A.2	Histogram-based thresholding techniques of Python library <i>skimage.filters</i> . . .	xxv

List of Tables

2.1	Body trajectories based on global human walk Model	32
4.1	Radar module parametrization.	89
4.2	Limbs decomposition confusion matrix.	92
5.1	The radar parametrization for μ -D acquisition.	107
6.1	The parameterization for the utilized MIMO radar.	131
6.2	Performance of the error rate due to training dataset sizes.	135
6.3	The classification performance due to different slicing techniques.	143

1 Introduction

There is currently a strong desire to incorporate a smart feature into any technology that involves human interaction. The human should be recognized with high freedom within the intended parameters of the application for such intelligent engagement. Due to the human body's remarkable flexibility and ability to change shape in a variety of circumstances that don't usually call for a human to even move, the recognition task is not seen as a simple one. An example of this would be to track a person's vital signs while being stationary to determine how emotionally and physically stressed they are, as described in Basjaruddin et al. 2021. As a result, a significant area of research is devoted to examining various perception systems with an excellent application of artificial intelligence for satisfying the desired smart features. For a deeper view of sensory systems used for different scopes of human recognition, e.g., tracking and activity recognition, some surveys are presented in Zhou & Hu 2008; Teixeira et al. 2010; Filippeschi et al. 2017, and Reich et al. 2020

The perceptual generalization is typically addressed as one key criterion, and this is done by looking into a stand-alone solution that works in many situations. Smoke, dust, darkness, and sunlight were among the environmental elements that were thought to have the most impact on the usefulness of the investigations as discussed in different studies e.g., Ho et al. 2005; Starr & Lattimer 2014; Zhang et al. 2019, and Zhao et al. 2020. A multi-sensory solution is frequently looked into due to the significant advancement in the many sensory solutions in terms of size, effectiveness, and price. Therefore, it is easier to overcome climatic circumstances or motion compensation for machines with strong movements by adding a sensor than improving the proposed algorithm as discussed in Martin et al. 2006, and Cho et al. 2014. In a smart perception system, this indicates that not all of the sensors will contribute equally, but the main operating sensor will be the one that can work without being affected by the environment.

Radar sensors have been investigated by many researchers as the main perception sensor as it is not yet an out-of-the-shelf sensor, and they can offer their data in multiple formats that can be utilized by an endless number of algorithms in the fields of deep learning and

signal processing as discussed in Pisa et al. 2016; Santra et al. 2018, and Li et al. 2019. Additionally, the radar capability of operating under any environmental circumstances e.g., darkness, dust and smoke never fails to introduce it as a powerful candidate compared to other perception systems as discussed in Dickmann et al. 2015, and Fischer et al. 2018. For example, a lighting system must be added to a camera to operate in a lower level of light; however, even the strong sunlight affects the performance of the camera systems in some scenarios. Other interesting features are the easiness of integration with other sensors, as still most of the sensors are in the form of development kits. Also, the harmless radiations, which introduced it powerfully in the medical sector as explained in Habash 2001. The radar sensors used for the bio-medical field usually operate at a carrier frequency $f_c \geq 10$ GHz, which are expected to have a limited penetration effect at the body tissues under the skin as explained in Laakso et al. 2012; Cho et al. 2007. Finally, the privacy aspect that is never broken by the radar compared to all other camera-based perception systems.

1.1 Motivation

Identifying a moving target and characterizing its motions is becoming increasingly important in a variety of industrial fields. Such systems, which are required in many locations for movement identification, are known as biometric systems. Some applications target at distinguishing humans from other moving things, such as a robot, while others require distinguishing between individuals. All of these applications fall under the umbrella issue of gait recognition. As seen in Kozlowski & James 1977, the primary concern of gait recognition is identifying the human from the walking style, as each human has a unique walking style.

Radar has been introduced as an effective technique for gait recognition applications in safety and surveillance due to its capability of capturing meaningful micro motion signatures, especially for walking activity, as explained in Chen 2019. Such signatures are known as micro Doppler (μ -D) signatures and have been proved to be feasible for many scenarios that require walking human detection, e.g., human and robot differentiation as discussed in Gurbuz et al. 2007, and Abdulatif et al. 2018. Others have used it for human identification, i.e., comparing μ -D signatures captured from different walking subjects,

such as the studies presented in Papanastasiou et al. 2021; Cao et al. 2018; Pegoraro et al. 2020, and Qiao et al. 2020.

The majority of radar-based activity identification applications are anticipated to be classification jobs, necessitating the deployment of deep learning approaches. The majority of these applications must be implemented in indoor conditions, where the backscattered radar signal typically contains a significant amount of noise due to multipath reflections. The capacity of the signal processing chain to eradicate these artifacts is a vital element that distinguishes a superior system from a just adequate one. In numerous instances, signal processing procedures, such as those provided in Du et al. 2013; Kronauge & Rohling 2013; Du et al. 2015, and Foued et al. 2017, have been examined for signal enhancements. However, relying solely on signal processing will restrict analysis to the investigated artifacts and hinder adaptation to new environments. Consequently, some researchers turned to deep learning to incorporate such adaptive sensing into their methods, such as the research published in Rock et al. 2019; Qu et al. 2019, and Abdulatif et al. 2019.

Another difficulty in the indoor environment is the limited space and the fact that numerous people are performing comparable activities simultaneously. To circumvent such limitations, improvements in radar hardware, such as those reported in Klauder 1960; Rohling & Moller 2008; Choi et al. 2011; Futatsumori et al. 2013, and Bae et al. 2020, are continually researched with the purpose of enhancing the resolutions of the three primary radar features, which are the range, velocity, and angle of arrival. However, increasing the resolution capabilities will always increase the data size and necessitate a more complex analysis of the acquired data, such as when dealing with the four-dimensional radar module proposed in Geibig et al. 2016, and Kueppers et al. 2020. To circumvent such requirements, certain analyses are performed, such as the inclusion of virtual antennas for improved angular resolution. However, such strategies will still be constrained by the current hardware capabilities; hence, deep learning has been introduced as a strong tool to overcome such limitations. For example, the studies published in Armanious et al. 2019; Geiss & Hardin 2020; Qian et al. 2021, and Ma et al. 2022 have demonstrated deep learning approaches for de-noising and super-resolution tasks. Hence, the focus of our research is on acquiring useful radar data and employing deep learning techniques to ensure the adaptability of our offered solutions.

1.2 Research Objectives

The research objective of the thesis is to analyze the feasibility of using radar sensor for indoor human detection and identification. Such a task is categorized as a human recognition task; hence, μ -D spectrograms are the most investigated radar feature. Walking is considered the main activity for investigation, as all body parts contribute to it in a periodic manner, and it is always a good indicator of the robustness of any system for human recognition. All of the proposed methods are intended for short-range analysis, which does not exceed 10 m. The research begins with the extraction of the primary characteristics of the captured walking μ -D signature. For such extraction, we have enhanced the commonly-used simulation signal model for representing the μ -D signatures in such a way as to give a more realistic μ -D representation. The main research goal was capturing the walking motion characteristics from the fundamental radar features, which are the Doppler and the range, as well as the mapped range-Doppler data. The feasibility study was applied by investigating the capability of machine learning for extracting the motion trajectories of the different body parts within the walking gait cycle. The feasibility is then evaluated by applying the research findings to a real-world use case scenario, namely human-robot classification based on a real radar.

The identification of humans is the second main research objective due to two significant factors. First, the application has recently become crucial for all security systems, such as mounting a sensor to detect strangers at a building's front door. Second, this task is regarded as the most difficult and generic, as the walking activity is fixed and similar μ -D signatures are anticipated among humans, making it difficult to classify. Therefore, the designed approach can be utilized easily afterward for other human detection applications. For the identification task, there is always a trade-off between the number of included subjects and the length of the captured spectrogram for a single decision. Therefore, the objective of the research is to overcome these limitations by combining multiple radar features to reduce classification ambiguity while increasing the number of classes and achieving real-time classification. The human identification task was divided into two phases. The first phase was designing a suitable experimental setup that is capable of capturing sufficient dataset in a feasible way; thus, a treadmill was utilized for such a task,

and the extracted μ -D signature was examined to determine if it reflects the walking style characteristics.

The feasibility study in the first phase was based on a signature of a single walking step, which is considered the minimum unit of representation for the walking motion. The number of shared subjects is maximized to overcome what is presented in the state-of-the-art studies. The second phase addresses the limitations of the free-walking scenario. For achieving this, the estimated angle of arrival (AoA) by the radar is investigated to derive a newly-measured micro angular (μ - ω) spectrogram reflecting the behavior of the body micro motions on the AoA feature.

Consideration is also given to the significant degradation in the quality of the captured spectrograms when captured from different aspect angles with respect to the radar. The main idea is to investigate a feasible deep learning approach in order to ensure an adaptive solution. The final approach's generalization and robustness are evaluated by applying them to an activity classification task that includes fall detection. Two main aspects are followed in this thesis work. First, decrease the acquired training dataset as much as possible. Second, the real-time aspect is considered in such a way that all classification is performed on a single occurrence of the activity, which lasts for ≈ 0.5 s for a single step of walking.

1.3 Thesis Overview

Following is an overview for the major points that are addressed in this thesis:

- Developing a signal model that is half measured and half simulated. The realistic walking motion is captured by an infrared-based motion capture system, and the radar signal is simulated by considering the body parts as a group of the connected ellipsoid as proposed by Chen 2019. The model is named the combined signal model as it combines both the realistic sense of the real motion and the simulation of the radar signals. The technique is different from the commonly-used model presented in Chen 2019 as it mainly relies on synthetic walking signatures based on Boulic et al. 1990. The idea of combining both realistic motion and synthetic radar data has been discussed in several publications such as the studies presented in Majumder

et al. 2006 with a general focus on radar signals, and Erol & Gurbuz 2015 with a main focus on μ -D signatures. Nevertheless, the presented technique is still different as the utilized motion capture system is of higher accuracy, and is widely used for other fields e.g., monitoring complex targets such as horses, as presented in Boye et al. 2014, Hardeman et al. 2019 and Rhodin et al. 2018. Accordingly, the proposed signal model can be extended to other fields.

- The radar signal model that is presented in Chen 2019 is adopted to generate Range-Doppler maps. Those maps were used to test the feasibility of the range feature for reflecting the micro motion behavior by using them as the main features for the task of real-time body limbs decomposition based on the complex trees machine learning approach. The body was divided into four main classes, which are arms, legs, feet, and base. The class definition was based on the swinging behavior as explained in Chen 2019. A brief discussion about the combined signal model and the results were published in Abdulatif et al. 2017.
- The outcomes of the first study were deployed for a real-life scenario involving a real radar for human-robot differentiation. The task is accomplished based on a single Range-Doppler map in which the feet displayed the most significant harmonics on both the range and Doppler features that were used for accurate human detection. Although the feet behavior does not appear clearly except within the swing phase, which lasts for $\approx 40\%$ of the walking gait cycle, the utilized neural network approach showed satisfactory performance in differentiating between a walking human and a moving mobile robot. The utilized neural network could achieve the classification task based on a single Range-Doppler map, compared to other machine learning techniques that required \approx ten maps to reach acceptable performance. A brief discussion about the experimental setup and the behavior of the different approaches is published in Abdulatif et al. 2018.
- Human identification based on walking human targets on a treadmill to capture clear μ -D signatures was the subject of a feasibility study. The research was conducted on 22 participants while applying the classification task only to half gait cycles based on the deep neural network architecture presented in He et al. 2016. The study aimed to test the effect of the walking style on the captured μ -D signatures; thus, the 22

subjects were of different body dimensions and both genders. The study showed a direct correlation between the body mass index and the captured μ -D signatures; thus, classification confusion is expected to occur for persons with comparable body dimensions. The study, including the utilized radar characteristics, the experimental setup, and the results, are discussed and published in Abdulatif et al. 2019. All the research activity till that step was based on a single-input-single-output radar.

- While studying the free-moving artifacts, a thorough feasibility study is conducted on human identification as the following step for the study presented in Abdulatif et al. 2019. The μ -D signatures are captured from various aspect angles while walking freely and not on a treadmill in a line of sight scenario. A multiple-input-multiple-output radar is used to ensure detection capability from various aspect angles, and a time-frequency computation is applied to the AoA dimension revealing a newly-measured micro angular velocity spectrogram. To the best of our knowledge, such spectrograms were not presented in any other state-of-the-art studies. For analyzing the characteristics of spectrograms, an adaptive slicing technique is used to mark the beginning and end of each walking step. As a solution to the issue of degraded μ -D spectrograms caused by walking at varying aspect angles with respect to the radar, a metric learning approach is investigated. The study, including the radar characteristics, the experimental setup, and the results, are discussed and presented in Weller et al. 2021.
- The robustness of this last architecture, including the radar type, the experimental setup, the adaptive slicing, and the metric learning approach, is applied to a human activity recognition task due to the high usefulness of the μ -D signatures for this research area. The study, including the radar characteristics, the experimental setup, the network architecture, and the comparison of the different slicing techniques, is presented in Aziz et al. 2022.

1.4 Thesis Organization

The rest of the thesis is organized as follows:

- Chapter 2: the research background for all the utilized tools in the practical thesis work is presented with a discussion about the previous related studies.
- Chapter 3: the radar signal processing for estimating and demonstrating the main radar features that are used in the practical thesis work is discussed in detail.
- Chapter 4: the main concept of radar-based human detection and a use case scenario about human-robot classification are discussed in detail.
- Chapter 5: a feasibility study about the concept of radar-based human identification while walking is presented.
- Chapter 6: a discussion about the last feasibility study that includes utilizing a metric learning approach for unconstrained human identification. The feasibility of the approach is tested on a use-case scenario of activity recognition.

2 Background

In this chapter, the theoretical background of the radar system and the utilized algorithms in this thesis work are discussed. An overview of the radar system and the different transmission protocols is presented in Section 2.1. The concept of investigating the main radar features to capture the micro motion behavior is discussed in Section 2.2. Since the walking is considered the main activity for investigation in this thesis work, the characteristics and the human body behavior through the walking gait cycle are discussed in Section 2.3. The theoretical background of the deep learning algorithms that are utilized in this thesis work is presented in Section 2.4. Finally, an overview of the state-of-the-art techniques that are related to our tackled research fields is presented in Section 2.5.

2.1 Overview of Radar Systems

The operation concept of radar systems consists of transmitting an electromagnetic signal at a specific frequency and receiving the scattered echoes reflected from various targets. The structure of the received signal is recognized to determine whether the target is in motion or stationary. As explained by Raemer 1996, the radar analysis allows for the observation of the target's range, azimuth and elevation angles, size, and speed. Radar systems should have different block diagrams and processing techniques based on the desired target characteristics to be extracted. The notation used in the thesis is underlined \underline{x} for vector, bold symbol \mathbf{X} for matrices, and ordinary x for scalars.

Pulsed radar is a common type of radar system that operates by transmitting a high-frequency, high-power pulsed signal. It has a pause between each pulse so that it can receive the echoes reflected back from the target. Pulsed radar can be classified based on its Pulse Repetition Frequency (PRF) as high, medium, or low. As described by Parker 2011, the low PRF radar can be used for target range identification, while the high PRF radar can be used for target velocity identification.

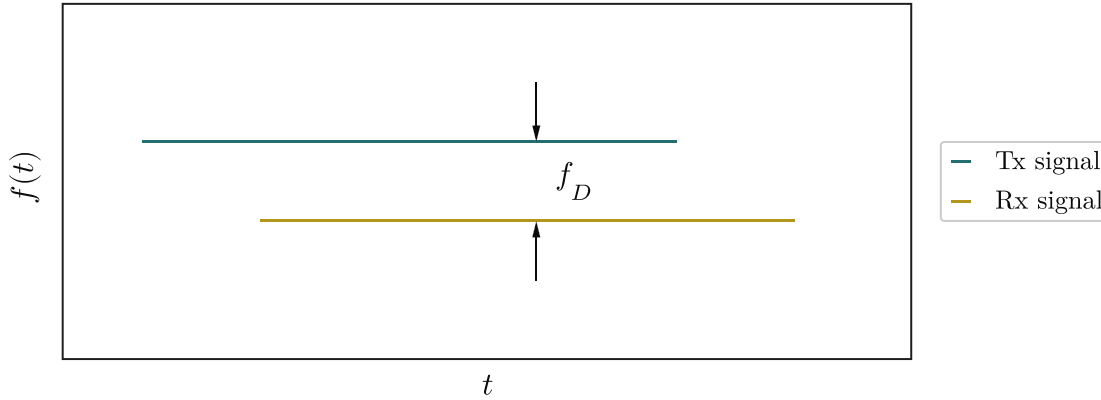


Figure 2.1: In CW radar, the structure of transmitted and received signals.

Continuous Wave (CW) radar is a frequently used type of radar system for gait recognition applications. The CW radar operates on the principle of continuous transmission and reception of reflected echoes with no silence intervals. The CW radar is divided into two primary categories, which are:

- The un-modulated CW radar that is used to determine target velocity.
- The frequency modulation CW (FMCW) radar is used for target joint range-velocity identification.

As a result, both the pulsed radar and the FMCW radar can determine the target range and velocity. The preferred radar system is determined by the application. The pulsed radar can identify a target's range at greater distances than the FMCW radar. The operation of frequency-modulated and un-modulated CW radars is discussed in the following section.

2.1.1 The Un-modulated Continuous Wave Radar

As demonstrated in Figure 2.1, the un-modulated CW radar operates by transmitting a signal with constant frequency and amplitude. According to Skolnik 2001, due to the continuous processing of the CW radar, the transmitter (Tx) and the receiver (Rx) should be perfectly isolated. The CW radar is based on the concept of detecting any frequency shift caused by the target's motion in the received signal. This phenomenon is known as the Doppler effect, which will be covered in the following section.

Doppler Effect

According to Da Andrade 1959, Christian Doppler discovered and characterized the Doppler Effect as the frequency change in the sound wave transmitted by a sound source due to the velocity of the source or the receiver towards the other. For instance, the Doppler effect can be detected when approaching or moving away from a functioning horn. According to Hughes & Cowley 2017, as human approaches the horn, the recognized sound pitch increases, and as the human moves away, the sound pitch decreases. As discussed in Eden 1992, the definition is generalized so that it applies to all propagating waves. Depending on the direction of motion of a moving light source, the spectrum of light perceived from such a source can range from red to blue. Therefore, the frequency shift induced due to the Doppler effect is dependent on the speed and direction of the target's motion. If the target is moving with a given radial velocity, it will have a velocity component in the direction of the radar Line of Sight (LOS), and the received echo will have a Doppler frequency shift (f_D) that is computed by the radar in order to identify the target's velocity.

In the case of stationary radar, a moving target with radial velocity v induces a Doppler shift ($-v/\lambda$) in the signal, where λ is the wavelength of the transmitted radar signal, as described in Chen 2019. As illustrated in Equation (2.1), f_D is calculated to be double the Doppler shift of each individual path. If the target is traveling away from the radar, a negative f_D will be produced since the radial velocity will be positive.

$$f_D = -2\frac{v}{\lambda}. \quad (2.1)$$

Doppler Frequency Shift Estimation in Radar

The time-frequency relationship illustrated in Figure 2.1 explains the signal transmitted by the CW radar. The operation frequency of the sent signal is constant concerning time. The radar recognizes the signal received from a moving target away from the target as having constant frequency $f_c + f_D$.

In the CW radar, a quadrature detector can be used to estimate the f_D according to Chen 2019. The received signal is down-converted using a form of the transmitted signal

and another form of the signal that is 90° out of phase. The received signal ($s_{Rx}(t)$) is a function of both the carrier frequency and the moving target's induced f_D , according to:

$$\begin{aligned} s_{Rx}(t) &= A \cos(2\pi(f_o + f_D)t) \\ &= A \cos(2\pi f_o t + \psi(t)) . \end{aligned} \quad (2.2)$$

where A is the amplitude of the transmitted signal and $\psi_{Rx}(t)$ is the phase shift caused by the target's movement. The down-conversion is performed on both the in-phase ($I_{Rx}(t)$) and quadrature ($Q_{Rx}(t)$) channels, and the output signal is taken out after a Low Pass Filter (LPF), according to:

$$\begin{aligned} I_{Rx}(t) &= \frac{A}{2} \cos(\psi(t)) \\ Q_{Rx}(t) &= -\frac{A}{2} \sin(\psi(t)) , \end{aligned} \quad (2.3)$$

The output of the two channels can be combined to form a complex signal ($s_D(t)$) for f_D extraction, according to:

$$s_D(t) = I_{Rx}(t) + jQ_{Rx}(t) = \frac{A}{2} e^{-j\psi(t)} = \frac{A}{2} e^{-j2\pi f_D t} . \quad (2.4)$$

For efficiency and simplicity of implementation, the f_D shift can be estimated using Fast Fourier Transform (FFT) analysis. As explained in Mahafza 2017, zero-padding can be applied to the signal to increase the resolution of the spectrum. As shown in Equation (2.5), the f_D can be defined as the first derivative of any detected phase difference ($\psi(t)$) between the received and transmitted signals.

$$f_D = \frac{1}{2\pi} \frac{d\psi(t)}{dt} . \quad (2.5)$$

The function proposed in Equation (2.5) cannot be used if the received signal contains more than one induced f_D . The time-frequency representation described in Section 2.2.1 can then be used to detect all of the induced frequency shifts.

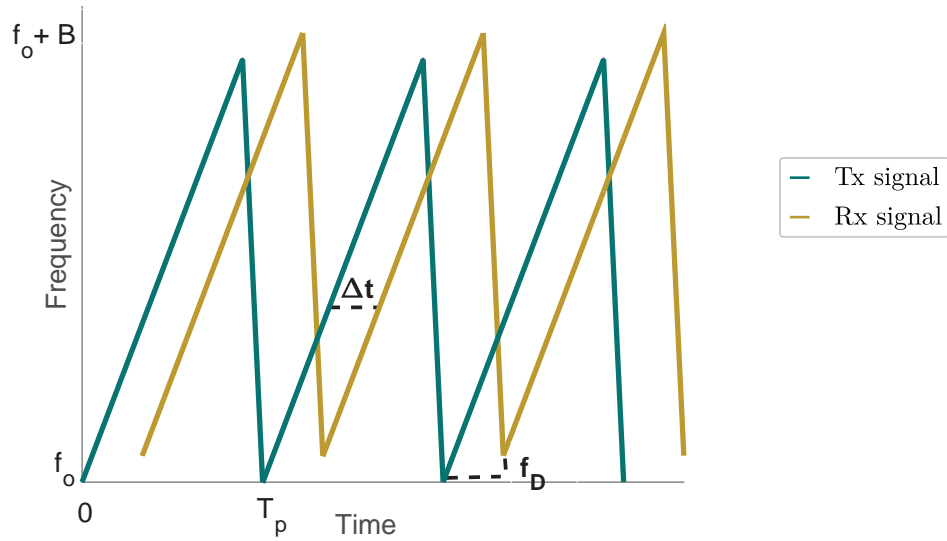


Figure 2.2: Signal comparison between transmitted and received signals.

2.1.2 Frequency Modulation Continuous Wave Radar

Some applications necessitate not only determining the velocity of a moving target, but also determining its range. Frequency modulation of the transmitted signal is required for the radar to detect the range of the target. As a result, this type of radar transmission architecture is known as the FMCW radar. To detect the range of the target, the radar must have a time system that can mark the transmitted and received signals, as described in Da Andrade 1959. The FMCW radar operates by sending periodic chirps while changing the transmission frequency linearly with respect to time, as illustrated in Figure 2.2. The frequency changes within each sweep according to the signal bandwidth, as explained in Lipa & Barrick 1990 and Da Andrade 1959. The frequency of the transmitted signal ($f_t(t)$) at any time instance can be expressed according to:

$$f_t = f_0 + \frac{B}{T_p}t, \quad (2.6)$$

where B represents the transmission bandwidth and T_p represents the sweep time. The n^{th} sweep's transmitted signal ($s_{Tx}(t)$) can be expressed according to:

$$s_{Tx}(t) = \cos \left(2\pi f_0 t + \frac{\pi B(t - nT)^2}{T_p} \right). \quad (2.7)$$

The reflected signal from a moving target is time delayed with Δt , which can be used for range identification. If the signal is also frequency shifted by f_D , it can be utilized for velocity calculation, as shown in Figure 2.2. The time delay Δt can be described as a function of the radial velocity of the target v and the target's range (R) according to:

$$\Delta t = \frac{2(R + vt)}{c}, \quad (2.8)$$

where c is the speed of light in free space. The received signal ($s_{Rx}(t)$) of the n^{th} sweep has the same form as the transmission signal but is Δt delayed. The signal received can be described according to:

$$s_{Rx}(t) = \frac{1}{R^2} \cos \left(2\pi f_o [t - \Delta t] + \frac{\pi B(t - \Delta t - nT_p)^2}{T_p} \right). \quad (2.9)$$

To create a down-converted signal, the received signal is mixed with the original form of the transmitted signal. The mixed signal is passed through the LPF to produce the intermediate frequency (IF) signal ($s_{if}(t)$), after which signal processing techniques can be used to estimate target range and velocity. The phase of the signal at the intermediate frequency ($\psi_{IF}(t)$) is used for range estimation, which is derived in Winkler 2007 and given according to:

$$\begin{aligned} \psi_{IF}(t) &= \psi_{Tx}(t) - \psi_{Rx}(t) \\ \frac{\psi_{IF}(t)}{2\pi} &= f_o t + \frac{Bt^2}{2T_p} - f_o(t - \Delta t) - \frac{B(t - \Delta t)^2}{2T_p} \\ &= f_o \Delta t + \frac{B}{T_p} t \Delta t - \frac{B}{2T_p} (\Delta t)^2. \end{aligned} \quad (2.10)$$

A range target $R = 10m$ with $v = 6m/s$ yields a very small Δt compared to T_p for a radar operating at $f_o = 25GHz$ with $B = 2GHz$ and $T_p = 0.5ms$. As a result, the final term in Equation (2.10) can be ignored as $\Delta t \ll T_p$, as explained in Winkler 2007. The relationship is extended to include the target range and velocity by substituting

Equation (2.8), which yields:

$$\begin{aligned}\frac{\psi_{IF}(t)}{2\pi} &= \frac{2f_o R}{c} + \left(\frac{2f_o v}{c} + \frac{2BR}{T_p c} \right) t + \frac{2Bv}{T_p c} t^2, \\ \frac{\psi_{IF}}{2\pi} &= \zeta + f_r t + f_D t + \frac{2Bv}{T_p c} t^2,\end{aligned}\tag{2.11}$$

where ζ is a constant phase offset, f_r is the frequency shift caused by the target range, and f_D is the frequency shift caused by the target velocity. The last term in Equation (2.11) can be ignored with the same operating bandwidth, sweep time, and carrier frequency proposed in previous equations, as assumed in Winkler 2007.

As a result, for target range determination, f_r will be estimated within one chirp and for target velocity determination, f_D will be estimated over multiple chirps. The chirp is sampled digitally with a sampling frequency of f_s and a number of samples per chirp of N_s . The number of chirps required for velocity determination is N_p , which must be settled in a single matrix (\mathbf{x}_{IF}). Each matrix element is described below:

$$x_{IF}(n_s, n_p) = A e^{j2\pi \left(\frac{2BR}{T_p c} T_s n_s + \frac{2f_o v}{c} T_p n_p \right)},\tag{2.12}$$

where $n_s = 0 \dots N_s - 1$, $n_p = 0 \dots N_p - 1$, and A is a constant referring to any amplitude effect in the received signal to estimate R and v , the approach presented here is based on applying a 2D-FFT to the time-domain matrix. The signal processing will be explained into details in Chapter 3, and the matrix is referred to as (\mathbf{X}_{IF}). Each matrix element in the frequency domain is given according to:

$$X_{IF}(q_r, q_v) = \sum_{n_p=0}^{N_p-1} \sum_{n_s=0}^{N_s-1} [x_{IF}(n_s, n_p) e^{-j2\pi q_r n_s / N_s} e^{-2\pi q_v n_p / N_p}],\tag{2.13}$$

where $q_r = 0 \dots N_s - 1$ and $q_v = 0 \dots N_p - 1$ are spectrum sampling bins. The signal power will be concentrated at f_r and f_D after this 2D processing. As explained in Lipa & Barrick 1990, zero-padding can be applied to the signal to increase the signal to noise ratio (SNR) of the spectrum against white noise, making estimation of f_r and f_D easier. As a result, the maximum estimated target range (R_{max}) and velocity (v_{max}) are derived as functions of the two bins $q_r = N_s - 1$ and $q_v = N_p - 1$ containing the maximum of the spectrum,

according to:

$$R_{max} = \frac{(N_s - 1)c}{2B}, \quad (2.14)$$

$$v_{max} = \frac{(N_p - 1)c}{2f_o N_p T_p}. \quad (2.15)$$

The radar parameters are adjusted and tuned based on the desired tracking target's characteristics. After applying 2D-FFT to the measurement matrix, a range-Doppler (RD) map is produced. In this matrix, each range-velocity cell has the indices q_r for range and q_v for velocity. The step sizes of Δq_r and Δq_v with reference to Equations 2.14 and 2.15 are used to obtain the resolutions for range (R_{res}) and velocity (v_{res}) according to:

$$R_{res} = \frac{c}{2B}, \quad (2.16)$$

$$v_{res} = \frac{c}{2f_o T_p N_p}. \quad (2.17)$$

Equations 2.14 and 2.15 can be reformulated to provide the following relations for B , T_p , N_s and N_p , which are utilized in radar design:

$$B = \frac{c}{2R_{res}}, \quad (2.18)$$

$$T_p = \frac{c}{4f_o \hat{V}_{max}}, \quad (2.19)$$

$$N_s = \frac{2B\hat{R}_{max}}{c}, \quad (2.20)$$

$$N_p = \frac{c}{2f_o T_p V_{res}}. \quad (2.21)$$

2.1.3 Multiple-Input-Multiple-Output Transmission Protocol

When an FMCW radar has a single Tx and a single Rx, referred to it as single-input-single-output (SISO) module, it can be utilized to estimate the target's range and velocity. Multiple targets may be included in the detection capability so long as they are separated by the resolution parameter of either the range or velocity parameter, as described in Section 2.1.2. As described in Robey et al. 2004, a Multiple-Input-Multiple-Output (MIMO) communication protocol is utilized to add the capability of tracking multiple targets. As

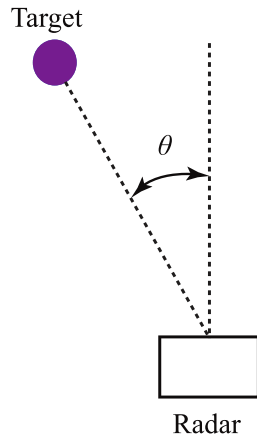


Figure 2.3: Angle of arrival (AoA)

depicted in Figure 2.3, the main concept involves equipping the radar with many receivers to enable the determination of the Angle of Arrival (AoA) of the target. Using several receivers is also intended to improve the mitigation of signal fading, interference suppression, and resolution enhancement. As stated in Li & Stoica 2008, an improvement in parameter estimation, target detection, target recognition, and target tracking is anticipated.

An MIMO radar can be modeled using a variety of transmission protocols, including time division multiple access (TDMA), frequency division multiple access (FDMA) and code division multiple access (CDMA). This thesis employs the CDMA protocol, due to the high capability of this protocol in SNR enhancement, which makes it as the most-commonly used protocol by the available MIMO radar modules. The CDMA protocol will be explored in length in the next section, while additional information on the other methodologies can be found in Sun et al. 2014. According to Bialer et al. 2021, the CDMA protocol is based on a modulation technique in which various Tx antennas transmit signals based on a series of orthogonal phase codes. According to Singhal 2012, the Hadmard code, commonly known as Walsh-Hadmard or Walsh code, is a good representation of the CDMA waveform.

Angle Estimation

The AoA can be estimated if the radar module is featured with a Single-Input-Multiple-Output (SIMO) protocol that includes a single Tx and multiple Rxs, as described in Richards et al. 2010. Figure 2.4 depicts an example for one Tx and three Rx, where θ is the AoA. The reflected signal must travel an additional distance (δ) to reach the

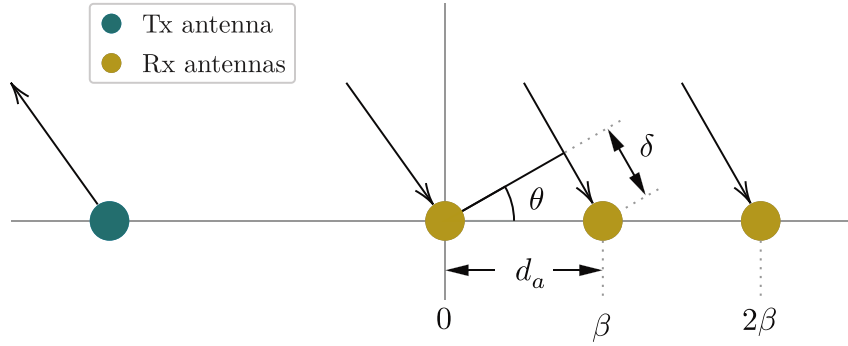


Figure 2.4: Principle of a SIMO radar.

second adjacent Rx antenna in order to be received by the first two adjacent Rx antennas. This distance induces a phase difference β between adjacent Rx antennas, which can be calculated by means of:

$$\delta = d_a \mathbf{sin}(\theta) , \quad (2.22)$$

$$\beta = \frac{2\pi}{\lambda} d_a \mathbf{sin}(\theta) . \quad (2.23)$$

Consequently, the AoA can be estimated based on the phase difference according to:

$$\theta = \mathbf{arcsin} \left(\frac{\beta\lambda}{2\pi d_a} \right) . \quad (2.24)$$

By substituting the limits of α for the interval $[-\pi, \pi]$, the unambiguous field of view (FOV) of the radar can be calculated according to:

$$\theta_{FOV} = \pm \mathbf{arcsin} \left(\frac{\lambda}{2d_a} \right) . \quad (2.25)$$

As explained by Li & Stoica 2008, and Instruments 2018, a design aspect is considered for the antenna spacing ($d_a = \lambda/2$) in order to maximize the radar's FOV to $\theta_{FOV,max} = \pm 90^\circ$. However, this perspective is hypothetical. As explained by Tahmouh & Silvius 2009, a FOV of $\theta_{FOV,max} = \pm 60^\circ$ is used in this thesis because it is feasible for activity recognition.

Generally, radar sensors of the present day have more than two Rx antennas. Figure 2.4 depicts a SIMO system with one Tx antenna and three Rx antennas. As depicted, a signal received by all three Rx antennas has an additional β phase shift at each successive

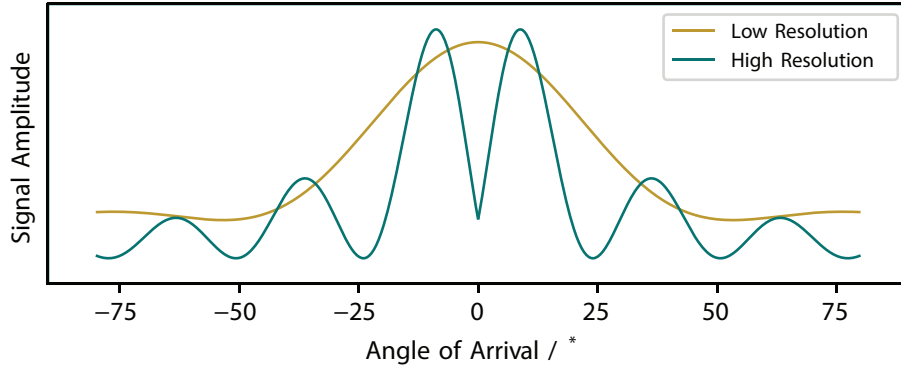


Figure 2.5: Spectral angle resolution according to different numbers of Fourier coefficients.

antenna relative to its preceding antenna. This results in a linear phase progression of the signal across all N_{Rx} antennas for equidistant Rx antennas. Referring to Figure 2.4, phase progression is a vector given by $\underline{\beta} = [0, \beta, 2\beta]$, where 0 denotes the first Rx antenna as the reference point. Consequently, the AoA can be reliably estimated by sampling the received signal across the channels, or antennas, and then analyzing the sampled data in the spectral domain. This can be achieved by applying an FFT to the channel information, which is one reason why a greater number of Rx antennas results in a higher angular resolution. The greater the number of channels that can be sampled, the more data points the FFT has to work with. This permits a greater number of Fourier coefficients, which ultimately leads to sharper peaks in the spectral domain, thereby increasing the resolution. Figure 2.5 demonstrates this exact point. While a radar sensor with a low angular resolution is unable to distinguish between the two objects at $\theta \approx \pm 9^\circ$, one with a higher angular resolution can.

The angle resolution is derived by considering two point-like targets (A and B): The first is at AoA of θ and the second is at an AoA of $\theta + \Delta\theta$. Signals reflected from these targets exhibit a phase shift between consecutive Rx antennas according to:

$$\beta_A = \frac{2\pi}{\lambda} d_a \mathbf{sin}(\theta) , \quad (2.26)$$

$$\beta_B = \frac{2\pi}{\lambda} d_a \mathbf{sin}(\theta + \Delta\theta) , \quad (2.27)$$

and have a difference that can be approximated by

$$\begin{aligned}\Delta\beta &= \beta_A - \beta_B , \\ &\approx \frac{2\pi}{\lambda} d_a \cos(\theta) \Delta\theta .\end{aligned}\tag{2.28}$$

The assumption that two objects separated by $\Delta\beta$ have distinct peaks in an FFT with N_{Rx} Fourier coefficients as long as the peaks are separated by more than $\frac{2\pi}{N_{Rx}}$ specifies the condition that must be satisfied to resolve the two individual targets in the spectral domain. The condition is stated according to:

$$\Delta\beta > \frac{2\pi}{N_{Rx}} .\tag{2.29}$$

and can be substituted into Equation (2.28) resulting in:

$$\Delta\theta > \frac{\lambda}{N_{Rx} d_a \cos(\theta)} .\tag{2.30}$$

The AoA is considered the third main feature that can be captured by the MIMO-FMCW radar after the range and velocity. Similarly, the resolution aspect is parameterized as the first two features as discussed in Sec. 2.1.2. Angular resolution (θ_{res}) can be defined as the minimum AoA between two targets to be separated as explained in Li & Stoica 2008. As explained in Milligan 2005, the angular resolution is dependent on the number of Rx antennas. It can be parameterized according to:

$$\theta_{res} = \frac{2}{N_{Rx}} .\tag{2.31}$$

Thus, the hardware design may limit the achievable angular resolution. Consequently, transmission protocols for formulating the waveform, e.g., TDMA or CDMA are utilized to enhance the angular resolution without the need of featuring the radar hardware with more physical Rx antennas. For formulating the radar waveform in such a way, the Tx antennas are included and a MIMO protocol is utilized.

In contrast to SIMO systems, which require twice as many Rx antennas, a MIMO radar can achieve the same result with only one additional Tx antenna. This provides a simple,

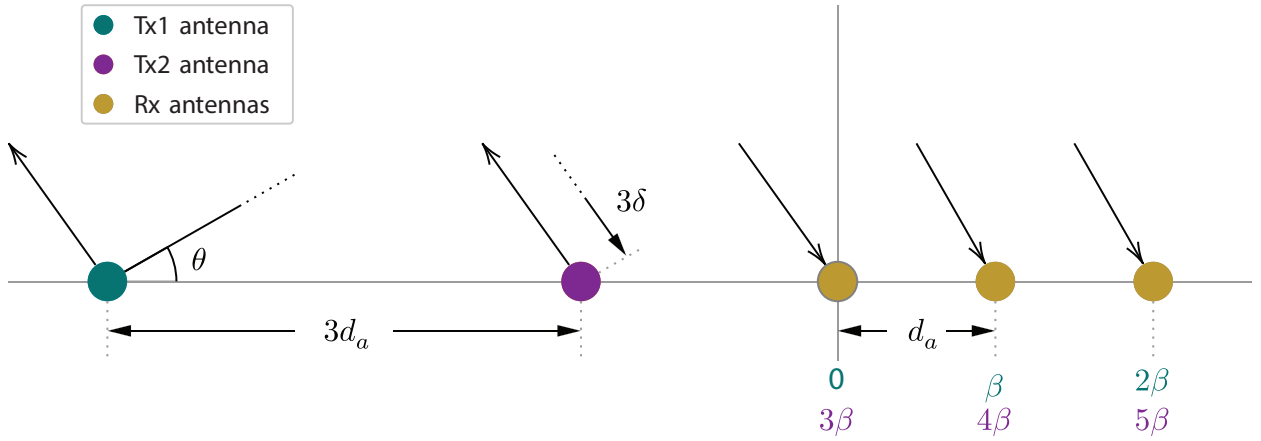


Figure 2.6: The principle of a MIMO radar.

and cost-effective method for improving angular resolutions. Figure 2.6 depicts a MIMO radar with two Tx and three Rx antennas. The Tx antenna is positioned at an arbitrary distance from the first Rx antenna, and the Tx antennas are separated by $3d_a$. Each Rx antenna is situated at a distance of d_a from its neighboring antenna.

A signal transmitted from Tx1 and reflected from a target will result in a phase progression vector across the Rx antennas of $\underline{\beta}_{Tx1} = [0, \beta, 2\beta]$. Similarly, antenna Tx2 emits a signal that is reflected by the same target back to the sensor. Since this antenna is located at $3d_a$ away from Tx1, its signal must travel $\delta = 3d_a \sin(\theta)$ farther than the signal from Tx1. Therefore, its phase progression vector is shifted by 3β , yielding $\underline{\beta}_{Tx2} = [3\beta, 4\beta, 5\beta]$. Combining these two vectors yields $\underline{\beta}_{Tx1, Tx2} = [0, \beta, 2\beta, 3\beta, 4\beta, 5\beta]$, which is identical to the phase progression vector of a SIMO radar with one Tx antenna and six Rx antennas. The MIMO radar depicted in Figure 2.6 generates an antenna array consisting of one Tx and six Rx antennas as described in Richards 2014.

2.2 Micro-Motion Signatures in Radar

According to Chen 2008, the Micro-Doppler (μ -D) effect is the Doppler effect caused by rotational, vibrational, or swinging motions of the particles comprising a moving bulk body. An excellent illustration of this is a person walking: the body may be viewed as a mass object with a definite translational velocity, while the arms and legs have a periodic swinging velocity that is contributed to the translational velocity component of the walking motion. As vibrations create certain micro motions in the Doppler effect, they also induce

some micro motions in the target range identification, defining the Micro-Range (μ -R) effect in radar, as described in Abdulatif et al. 2018. Consequently, a moving body that is accompanied by a number of small vibrating bodies will have an effect on both the velocity and range dimensions. According to Chen 2019, a moving body can be divided into two categories: rigid and nonrigid body movement. The μ -D effect can be evaluated physically in radar using the same techniques as the Doppler effect, as described in the previous section. In contrast, the μ -D signature is typically represented using the Short Time Fourier Transform (STFT) analysis described in the next section.

2.2.1 Time-Frequency Representation

The time-frequency representation is frequently employed to examine the μ -D signatures in order to analyze the movements of various targets. Implementing the time-frequency representation by applying STFT to the incoming signal. The time-frequency is represented by spectrograms that reflect the motion characteristics, which are commonly utilized for human activity recognition, as demonstrated in several research such as Javier & Kim 2014; Kim & Ling 2009; Kim & Ling 2008; Narayanan & Zenaldin 2015, and Tahmoush 2015. In addition, the captured μ -D signatures for walking activity have been investigated in various studies, such as classifying walking styles as discussed by Saho et al. 2017 and Seyfioglu et al. 2018 or human identification as discussed in multiple studies, such as Yang et al. 2019 and Abdulatif et al. 2019. Using the STFT permits the recognition of the multiple frequency components in the received radar signal, which can be estimated according to:

$$X[n, r] = \sum_{k=-\infty}^{\infty} x[k]w[k-n]e^{-j2\pi kr/N}, r = 0, 1, \dots, N-1 \quad (2.32)$$

where n represents the time index, r represents the frequency index, N represents the FFT size, and $\mathbf{w}[\mathbf{n}]$ represents the Gaussian windowing function. The windowing is implemented by multiplying $\mathbf{w}[\mathbf{n}]$ by $\mathbf{x}[\mathbf{n}]$ and then applying the FFT to the output. The time-frequency relationship can be identified by adjusting the windowing function to obtain the entire received signal $\mathbf{x}[\mathbf{n}]$. The spectrogram is the most frequent representation of the STFT, which has been utilized in numerous research, such as Kingsbury et al. 1998, and Greenberg

& Kingsbury 1997. The spectrogram is defined as the square magnitude of the STFT output signal and is mathematically expressed according to:

$$S[n, r] = |X[n, r]|^2 . \quad (2.33)$$

The window size should be chosen to match the needs of the application while balancing the time and frequency resolutions. By decreasing the length of the window function, greater temporal resolution can be achieved, and vice versa for frequency resolution. In order to achieve a smoother image of the spectrogram, sliding panes are overlapped. By utilizing zero-padding, resolution can be improved.

Window Function Types

The degree of window function smoothness is a compromise between spectral leakage and spectrogram resolution. The spectral leakage is described as the energy diffusion of some frequency components into adjacent frequency components, which distorts the spectrogram as explained in Hongwei 2009. Two window functions are offered to represent the degree of smoothness and spectral leakage of the window function:

- *Rectangular* window: offers the best resolution but with high spectral leakage effect.
- *Hamming* window: good compromise between the resolution and the spectral leakage.

The Hamming window is a regularly employed window function for STFT processing and will be utilized in this thesis for spectrogram representation.

2.2.2 Micro-Motion Signature of a Moving Rigid Body

As detailed in Chen 2019, a moving rigid body is a solid body that is not deformed and the distance between all of its pieces remains constant during motion. The mass of a rigid body is equal to the total of the masses of its spinning or vibrating components. The similar principle applies to the motion of a rigid body, which is typically characterized as a collection of micro translational and rotational motions. Micro rotational motions alter the phase of the reflected signal. As detailed in Chen & Ling 2002, the radar can employ μ -D analysis to determine the rotations of the pieces that comprise a moving rigid body.

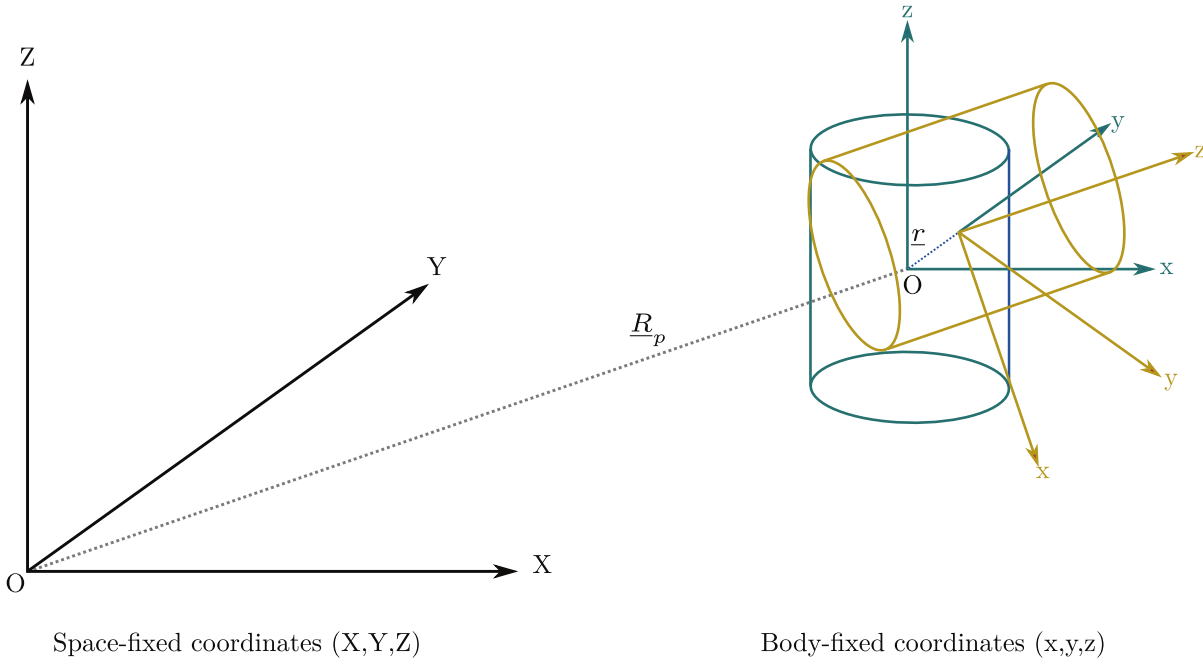


Figure 2.7: Rotation of a rigid body with respect to a space-fixed system (X,Y,Z)

As demonstrated in Figure 2.7, the rigid body motion may be mathematically represented using two coordinate systems: the space-fixed system (X,Y,Z) and the body-fixed system for the body itself (x,y,z). The absolute value of the range vector pointing from the two origins of the space-fixed system and the body-fixed system is represented by \underline{R}_p . According to Chen et al. 2006, the origin of the body-fixed system is the target's center of mass (CM). If \underline{r} represents the position vector of a random point with respect to the body-fixed system, then its position with respect to the space-fixed system can be described as $\underline{R}_p + \underline{r}$ and its velocity is expressed according to:

$$\underline{v}_s = \frac{d}{dt}(\underline{R}_p + \underline{r}) = \underline{v} + \underline{\omega} \times \underline{r} . \quad (2.34)$$

The CM of the rigid body in motion is moving with translational velocity (\underline{V}) and angular velocity ($\underline{\omega}$). According to Equation (2.34), the translational and rotational micro-motions are combined to form the motion of the moving rigid body's point. According to Chen 2019, a moving rigid body can be subdivided into a group of scattering points, and the analysis of the μ -D effect of the body's movement in radar can be performed by combining the μ -D effect of each scattering point. As an illustration of the mathematical model, place the radar at the origin of the space-fixed system. Then, the reflected echo ($\underline{s}(t)$)

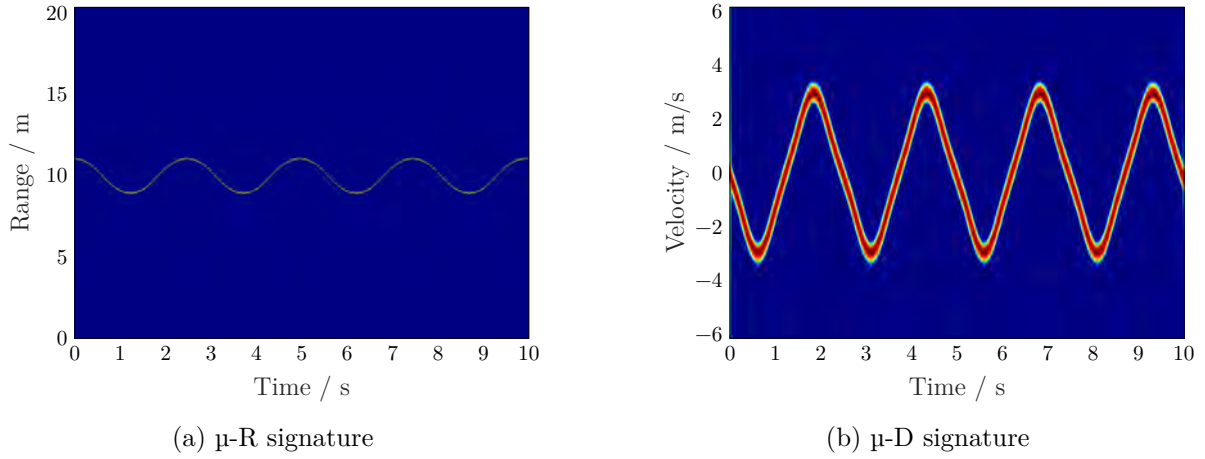


Figure 2.8: Simulated motion of free oscillating pendulum

from the scattering point is a function of the distance between the radar and the point $R = |\underline{R}_p + \underline{r}|$ and can be expressed according to:

$$\mathbf{s}(t) = \rho(x, y, z) e^{\frac{j4\pi f_o R}{c}} = \rho(x, y, z) e^{j\psi(t)}, \quad (2.35)$$

where $\rho(x, y, z)$ is the reflectivity function of the scattering point with respect to the coordinates of the body-fixed system. The f_D can be calculated as the phase's derivative with respect to time according to:

$$\begin{aligned} f_D &= \frac{1}{2\pi} \frac{d}{dt} \psi(t) \\ &= \frac{2f_o}{c} \frac{d}{dt} R \\ &= \frac{2f_o}{c} \frac{d}{dt} \left\{ (\underline{R}_p + \underline{r})^T \cdot \underline{v} \right\} \\ &= \frac{2f_o}{c} [\underline{V} + \underline{\phi} \times \underline{r}]^T \cdot \underline{v}, \end{aligned} \quad (2.36)$$

where \underline{v} is the radial velocity unit vector, which is used to determine the radar's LOS direction. The following additional derivation of 2.36 can be applied:

$$\begin{aligned} f_D &= \frac{2f_o}{c} [\underline{V} + \underline{\phi} \times \underline{r}] \\ &= \frac{2f_o}{c} \underline{V} + \frac{f_o}{c} \underline{\phi} \times \underline{r} \\ &= f_T + f_R, \end{aligned} \quad (2.37)$$

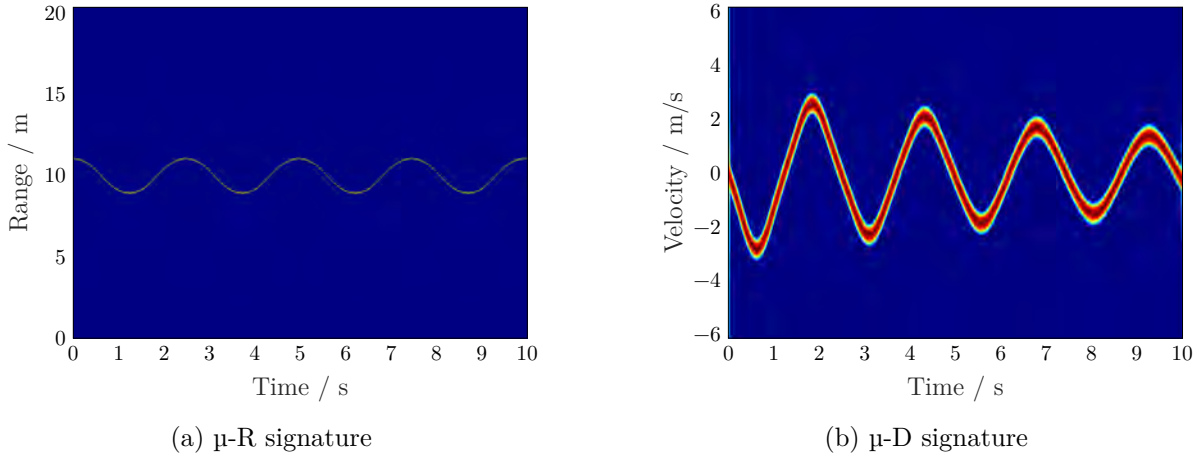


Figure 2.9: Simulated motion of damping oscillating pendulum

where f_T is the Doppler shift induced due to translational motion and f_R is the Doppler shift induced due to rotational motion. According to Equation (2.37), the total f_D induced by the scattering point is equal to the sum of the Doppler shifts induced by the translational and rotational motions.

A pendulum that oscillates is an example of a moving rigid body. When a pendulum structure consists of a heavy weight of a particular shape, such as a sphere or an ellipsoid, attached to a pivot via a rope. The radar received response from spherical and ellipsoidal geometries is analyzed using Radar Cross Section (RCS) approximation methods, which are briefly explained in the following section. The simulation model presented in Chen 2019 distinguishes between the free oscillation and the damping states of the oscillating pendulum. As depicted in Figure 2.8 and 2.9, the RCS approximation method yields accurate estimates of the captured μ -R and μ -D signatures.

The μ -R and μ -D signatures can be used to distinguish between various pendulum motions, such as damping and free oscillation. The simulation of μ -D and μ -R signatures of a rigid body, such as an oscillating pendulum consisting of an oscillating sphere or ellipsoid, can be used to simulate a more complex body. A swinging arm is simulated as a free oscillating pendulum, for example, when walking with constant relative velocity. When simulating the deceleration action of a human walker, the swinging arm is modeled as a dampened oscillating pendulum.

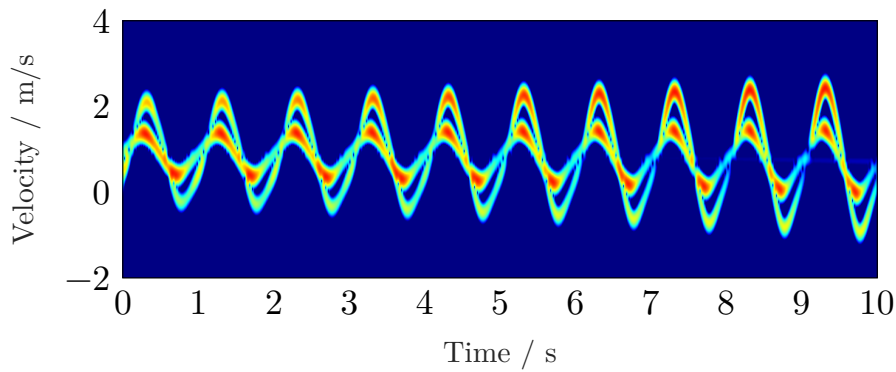


Figure 2.10: μ -D signature induced due to the bird's flying activity

2.2.3 Micro-Motion Signature of a Moving Nonrigid Body

A deformable moving body can be defined as a nonrigid body. Correspondingly, the distance between all of its components may change over time as it moves. Radar can analyze the μ -D effect of a moving nonrigid body as the μ -D effect of a group of moving rigid bodies connected to each other. Human movement is an important example of a nonrigid body in motion that must be analyzed by the radar.

A bird's flying activity is analyzed as periodic motion. A bird typically flies with the same translational velocity as a human when walking. The flapping wings induce oscillating Doppler effects around the f_D induced by the flying bird's translational velocity. Based on the model described in Denavit & Hartenberg 1955 and Chen 2019, a simulation describing the kinematic motion of the bird is presented. A bird's wing is described as being composed of two ellipsoids. This approximation provides a precise indication of the μ -D signatures detected by radar, as illustrated in Figure 2.10.

The μ -D signature received as a result of a flying bird are a scaled-down version of the μ -D signatures received as a result of a walking human. The human limbs swing similarly to bird wings, with a few exceptions. The human limbs must undergo the same translational motion as the human torso when walking. Upon closer inspection, there are brief moments between each step in which both legs are firmly planted on the ground. In these instances, the limbs have the same translational velocity as the entire human body. This makes the relative walking velocity the radar's strongest received component. For this thesis, walking is considered to be the primary investigational activity. Thus, the

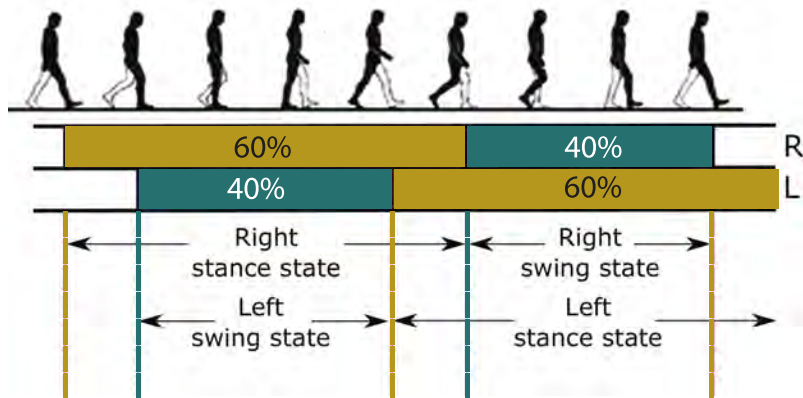


Figure 2.11: The two phases of the walking gait cycle for both body sides.

walking motion is explained briefly in the following section, and the simulation model used in our published feasibility study in Abdulatif et al. 2017 is described in Chapter 4.

2.3 Human Walking Gait Analysis

As explained in Chen 2019, walking is considered the primary activity for investigating the μ -D signatures captured by the radar because all body parts contribute on a periodic basis. Consequently, it is selected as the primary activity for investigation in this study, as it is the primary activity performed by a human in an indoor environment. This is the same case as other studies that investigated the μ -D signatures for activity recognition, such as Gurbuz & Amin 2019 and Gurbuz et al. 2017. During walking, the human body does not maintain a static shape, as the distances between body parts typically change as the body moves. Therefore, a walking human body can be viewed as a deformable, nonrigid body, and the analysis of human motion by any device is a complex procedure. The human gait can often be identified using visual image techniques. However, these visual techniques have a number of limitations, including distance and light variations Nixon & Carter 2006. These limitations make the radar a useful instrument for analyzing human motions in numerous experiments, as it operates normally through walls and is unaffected by light changes during the day.

Walking is described as a succession of periodic cycles. According to Chen 2019, the human walking gait cycle can be fully described by the stance and swing states, respectively. As shown in Figure 2.11, and explained in Mummolo et al. 2013, the stance state accounts

for sixty percent of the cycle, while the swing state accounts for forty percent. The stance state is when one foot is completely on the ground and the other foot is swinging or only touching the ground with its heel or toe. There are three phases used to describe the stance state:

- *Double support*: both feet are touching the ground. The start of the stance state in which one heel is striking the ground.
- *Single limb stance*: one foot is touching the ground and the other is above the ground which happens between the *double support* and the *second double support*.
- *Second double support*: last action when both feet are again touching the ground.

The swing state is the acceleration or deceleration phases of the human. The swing state is including only one foot touching ground and the other foot is swinging. Therefore, there is no *double support* phases within the swing state. While taking a step, the human arms and legs are swinging periodically. However, not all humans have the same walking motion and this give the idea for differentiating between people from their walking style as explained in Nixon & Carter 2006 and Cutting & Kozlowski 1977. Thus, identifying humans while walking is considered one of the main objectives in this thesis work. Human gait analysis can also be used to recognize the emotional state of a human as explained in Venture 2010. The walking motion of the human as a nonrigid body can be manipulated as a group of rigid bodies connected to each others. For this analysis, robotics forward kinematics can be used, that will be explained in the next section.

2.3.1 Walking Human Kinematics

The mechanical structure of a serial robot arm accurately represents human limbs. As explained in Michniewicz & Reinhart 2016, the structure of a robot arm manipulator consists of a series of rigid bodies, commonly referred to as links, that are connected to one another via robot joints. Therefore, the primary components of a robot manipulator are an arm that performs the robot's mobility action, a wrist that provides motion smoothness, and an end-effector that is responsible for performing the robot's required action. When analyzing the mechanical structure of a human arm, the upper and lower arm segments

can be viewed as robot links, the elbow and wrist as robot joints, and the hand as a robot end-effector.

Position and orientation are sufficient to completely describe a rigid body in space, as explained in Siliciano et al. 2010. According to Chen 2019, forward and inverse kinematics can be used to manipulate human walking motion. The fundamental kinematic parameters for human motion can be classified as either linear or angular. Position, velocity, and acceleration are required for a comprehensive description of the translational walking motion of a human in a three-dimensional space with respect to a reference space. Since the position of every point on the human body varies with respect to time, these three linear kinematic parameters can be used to describe the translational motion in its entirety.

The angular orientations of the body's segments (limbs) should also be taken into account during movement. For the description of these orientations or segment angles, it is necessary to calculate the angular velocity and acceleration. The angular velocity can be described as the rate of change of angular orientations over time and angular acceleration is described as the rate of change of the angular velocity over time. By detecting and calculating these angular kinematic parameters, it is possible to completely describe the rotational motion of any link or segment of the human body.

The linear velocity and acceleration can be calculated by analyzing the angular parameters of a component of a rigid body that is rotating with an angular velocity and acceleration. These two angular kinematic parameters adequately describe the angular rotation of a segment of the human body. Calculating the normal and tangential velocity components at this point allows one to determine the linear velocity of any point or particle of this rigid body component. The center of rotation is determined relative to the reference frame. At any point on a segment of a rotating rigid body, the tangential velocity and acceleration can be calculated using the point's position relative to the rotation center, the angular velocity, and the angular acceleration.

2.3.2 Human Walking Empirical Mathematical Model

According to Chen 2019, the global human walking model created by Boulic et al. 1990 can be combined with a signal model to represent realistic μ -D signatures. This thesis considers a similar procedure for defining the primary body characteristics that influence

radar-based human detection, which is briefly discussed in Chapter 4. In this section, the key features of the empirical model of global human walking are discussed briefly. The Denavit-Hartenberg (D-H) convention described in Hartenberg & Denavit 1964 is a kinematic convention that can be used to explain the positions of the rigid components that make up the flexible human body. The global model of human walking can be used to determine the kinematic parameters discussed in the previous section. The model divides the human body into seventeen rigid segments and defines the motion trajectories for each segment. The model is parameterized with the values proposed by Chen 2019 for better interpretation, and the motion trajectories are plotted to reflect the reaction of the body parts during the entire gait cycle.

The global model of human walking is based on the average parameters of a vast quantity of experimental data. This average perspective presents the model as an efficient and universal representation of the human body, regardless of body dimensions. The walking human can be viewed as a collection of interconnected rigid bodies whose positions change over time. In the presented model, the translational motion and orientations of certain human body segments must be considered. The model is not based on solving time-dependent motion equations. The primary objective of the model is to provide the positions of any segment of a rigid body at any given time. Therefore, the model's functions used to determine the positions and orientations of any rigid body part with respect to time. According to the model, as shown in Table 2.1, 12 trajectories are necessary to describe the human walking motion. All the presented equations and figures in this section are based on the study presented in Boulic et al. 1990, which were derived mainly through capturing the motion of different test subjects more than applying a mathematical proof.

When used, these functions can describe the positions and orientations of body parts at any point during a human's walking cycle. Therefore, the determined values are based on the human walking speed. According to the global walking model presented by Boulic et al. 1990, the twelve trajectories have two mathematical representations. Six trajectories are described as sinusoidal functions, and the remaining six as cubic spline functions.

Before beginning the motion, it is necessary to determine several key parameters for human walking, including the human walking relative velocity (v_{wr}), which is normalized by the human leg height and given in meters per second. Consequently, it can be used to

Trajectory	Translation	Body Rotation	left Rotation	Right Rotation
Vertical translation	$Tr_{vert}(t)$			
Lateral translation	$Tr_{lat}(t)$			
Translation forward/backward	$Tr_{f/b}(t)$			
Rotation forward/backward		$\theta_{f/b}(t)$		
Rotation left/right		$\theta_{l/r}(t)$		
Torsion rotation		$\theta_{torsion}(t)$		
Stretching of the hip			$\theta_{hip}(t)$	$\theta_{hip}(t + 0.5)$
Stretching of the knee			$\theta_{knee}(t)$	$\theta_{knee}(t + 0.5)$
Stretching of the ankle			$\theta_{ankle}(t)$	$\theta_{ankle}(t + 0.5)$
Motion of the thorax		$\theta_{thorax}(t)$		
Stretching of the shoulder			$\theta_{shoulder}(t)$	$\theta_{shoulder}(t + 0.5)$
Stretching of the elbow			$\theta_{elbow}(t)$	$\theta_{elbow}(t + 0.5)$

Table 2.1: Body trajectories based on global human walk Model

determine the walking cycle length, which is expressed as $R_{cyc} = 1.346\sqrt{v_{wr}}$. Therefore, the cycle duration can be calculated as $T_{cyc} = R_{cyc}/v_{wr}$. Support and double support can be expressed as $T_{supp} = 0.752T_{cyc} - 0.143$ and $T_{Dsupp} = 0.252T_{cyc} - 0.143$, respectively. During walking, the body must have a local body-fixed coordinate system, as described in Section 2.2.2, with the human spine base as the origin. The body-fixed origin should be positioned at 58% of the average human height. As previously explained, trajectories are divided into translational and rotational trajectories, whose mathematical representations will be briefly explained in the following sections.

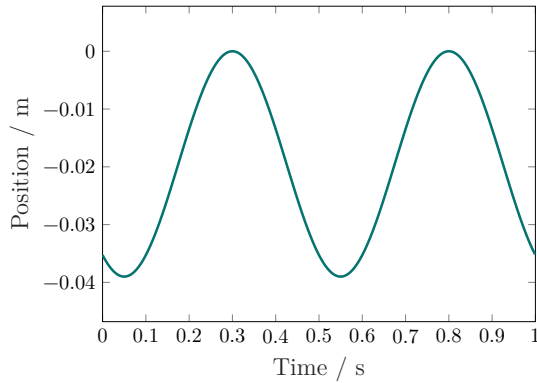
Translational Trajectories

Table 2.1 defines the first translational motion as the vertical displacement between the base of the spine and the head. It is mathematically expressed according to:

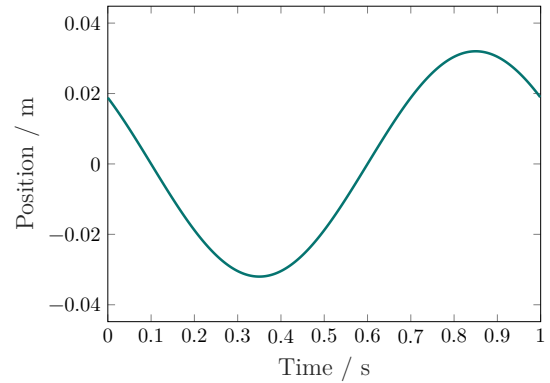
$$Tr_{vert}(t) = -b_v + b_v \sin[2\pi(2t_{rel} - 0.35)] , \quad (2.38)$$

in which $b_v = 0.015v_{wr}$. The relationship described in Equation (2.38) is graphically represented in Figure 2.12a. The second translational trajectory reflects the lateral vibration of the spine's origin and is depicted according to:

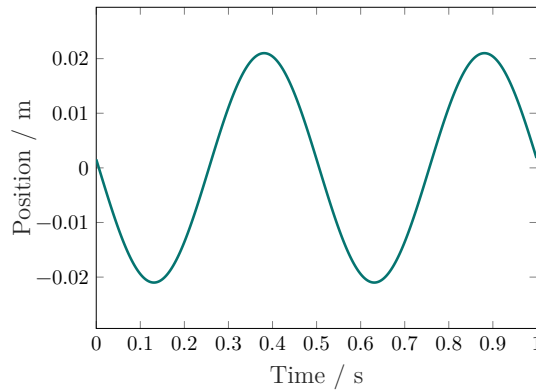
$$Tr_{lat}(t) = b_1 \sin[2\pi(2t_{rel} - 0.1)] , \quad (2.39)$$



(a) Vertical translation of the spine base.



(b) Lateral vibration of the spine origin.



(c) Translation forward/backward acceleration.

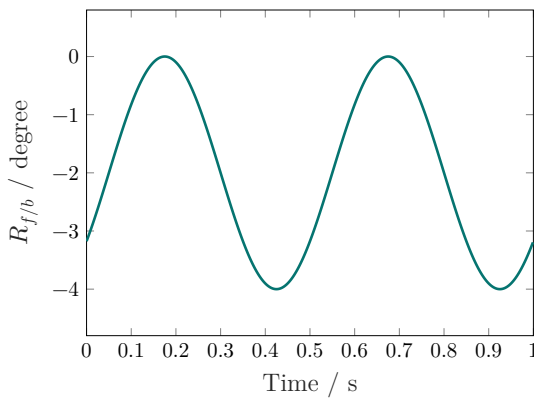
Figure 2.12: Translational trajectories.

$$b_1 = \begin{cases} -0.128V_{wr}^2 + 0.128v_{wr}, & v_{wr} < 0.5 \\ -0.032 & v_{wr} > 0.5 \end{cases}. \quad (2.40)$$

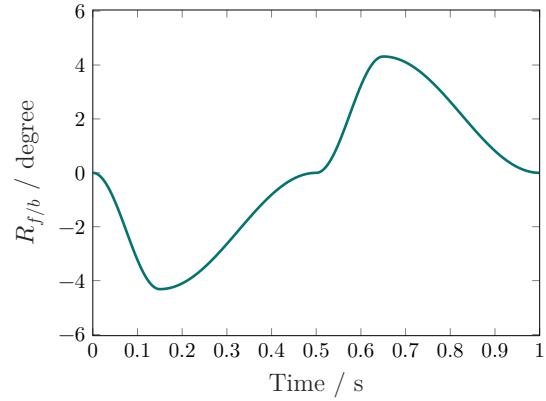
The trajectory used to describe the human acceleration and deceleration during the swing phase of the walking gait cycle is forward/backward translation. It is mathematically expressed according to:

$$Tr_{f/b}(t) = b_{f/b} \mathbf{sin}[2\pi(2t_{rel} + \phi_{f/b})], \quad (2.41)$$

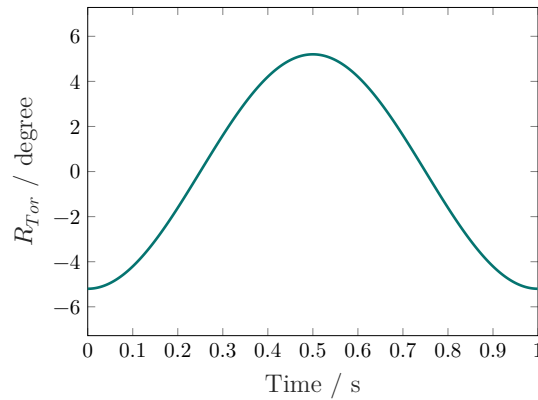
$$b_{f/b} = \begin{cases} -0.084v_{wr}^2 + 0.084v_{wr}, & v_{wr} < 0.5 \\ -0.021 & v_{wr} > 0.5 \end{cases}. \quad \phi_{f/b} = 0.625 - T_{supp} \quad (2.42)$$



(a) The back behavior with respect to the pelvis



(b) The pelvis behavior while taking a step



(c) Pelvis orientation with respect to the spine

Figure 2.13: Motion trajectories of the pelvis within single gait cycle

The three translational trajectories are represented as sinusoids, as described in Equations (2.38), (2.39), and (2.41). Figure 2.12 depicts a graphical interpretation of all the functions, taking into account the average human dimensions for clarity.

Rotational Trajectories

The previous translational trajectories primarily described the major translational movements of the entire body while walking. The back and torso are the primary mass-bearing parts of a walking human that produce a discernible vibration. The hip, which is a component of the pelvis, connects the upper body to the lower body limbs, as explained in Anatomy 2016. Due to the unique vibration of the hip during walking, the first two discussed rotational trajectories describe the movement of the human body relative to the pelvis. The first rotational trajectory is the Rotation forward/backward, which describes the bending behavior of the back relative to the pelvis at the start of a step and is defined

according to:

$$R_{f/b} = -br_{f/b} + br_{f/b} \mathbf{sin}[2\pi(2t_{rel} - 0.1)] , \quad (2.43)$$

$$br_{f/b} = \begin{cases} -8v_{wr}^2 + 8v_{wr}, & v_{wr} < 0.5 \\ 2 & v_{wr} > 0.5 \end{cases} . \quad (2.44)$$

The pelvis of a walking human typically falls on the side that is being used to take a step. This bending is characterized by the Rotation left/right trajectory and the following function describes its behavior:

$$R_{l/r} = \begin{cases} -br_{l/r} + br_{L/R} \mathbf{cos}[2\pi(10t_{rel}/3)] , & 0 \leq t_{rel} < 0.15 \\ -br_{l/r} - br_{L/R} \mathbf{cos}[2\pi(10(t_{rel} - 0.15)/7)] , & 0.15 \leq t_{rel} < 0.5 \\ -br_{l/r} - br_{L/R} \mathbf{cos}[2\pi(10(t_{rel} - 0.5)/3)] , & 0.5 \leq t_{rel} < 0.65 \\ -br_{l/r} - br_{L/R} \mathbf{cos}[2\pi(10(t_{rel} - 0.65)/7)] , & 0.65 \leq t_{rel} < 1 \end{cases} , \quad (2.45)$$

where $br_{l/r} = 1.66v_{wr}$ holds. In Figure 2.13, the two functions describing the body's rotation behavior with respect to the pelvis are plotted.

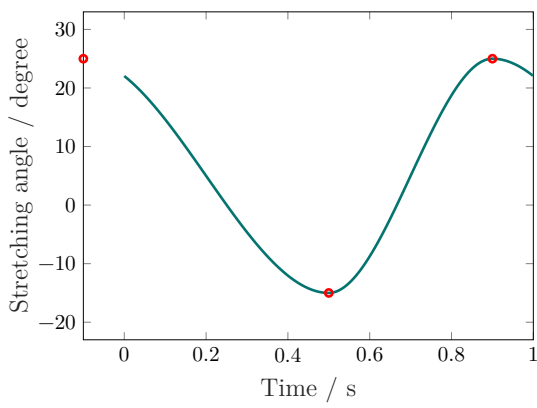
The following rotational trajectory is the torsion rotation, which describes the orientation of the pelvis relative to the spine during a step. The following defines the rotational behavior describing function:

$$R_{Tor} = -br_{Tor} \mathbf{cos}[2\pi(t_{rel})] , \quad (2.46)$$

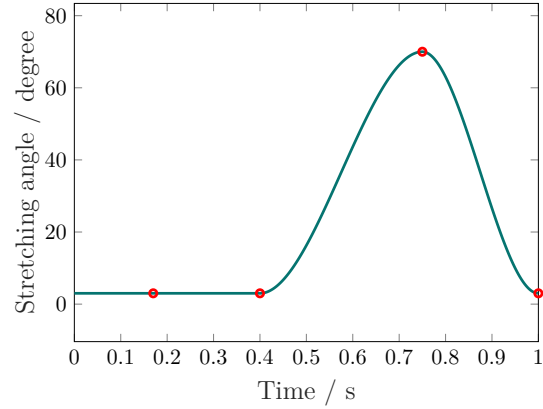
where $br_{Tor} = 4v_{wr}$. The function presented in Equation (2.46) is plotted and shown in Figure 2.13c.

The functions used to define trajectories can be described using sinusoids or cubic splines, which are described as third-order polynomials with a number of control points that the polynomial must pass through, as explained in Spline 2016. As shown in Figure 2.14, the next three trajectories are used to describe the bending action of the lower body joints for the hip, knee, and ankle using cubic splines.

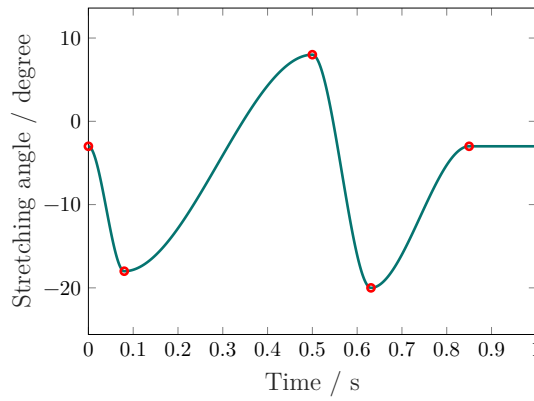
The cubic spline functions are also used to describe the behavior of the last three trajectories, which define the bending or stretching actions at the thorax, shoulder, and



(a) Stretching of the hip.



(b) Stretching of the knee.



(c) Stretching of the ankle.

Figure 2.14: Stretching trajectories of the lower body joints

elbow of the upper body. The three trajectories consist of the stretching of the thorax, which has four control points, and the stretching of the shoulder, which has rotational motion in both the left and right directions.

$$R_{Shoulder} = 3 - br_{Shoulder} \cos[2\pi(t_{rel})], \quad (2.47)$$

where $br_{Shoulder} = 9.88v_{wr}$ is true. The final trajectory is the elbow extension. As shown in Figure 2.15c, the elbow stretches similarly to the shoulder, but only with positive values and five control points.

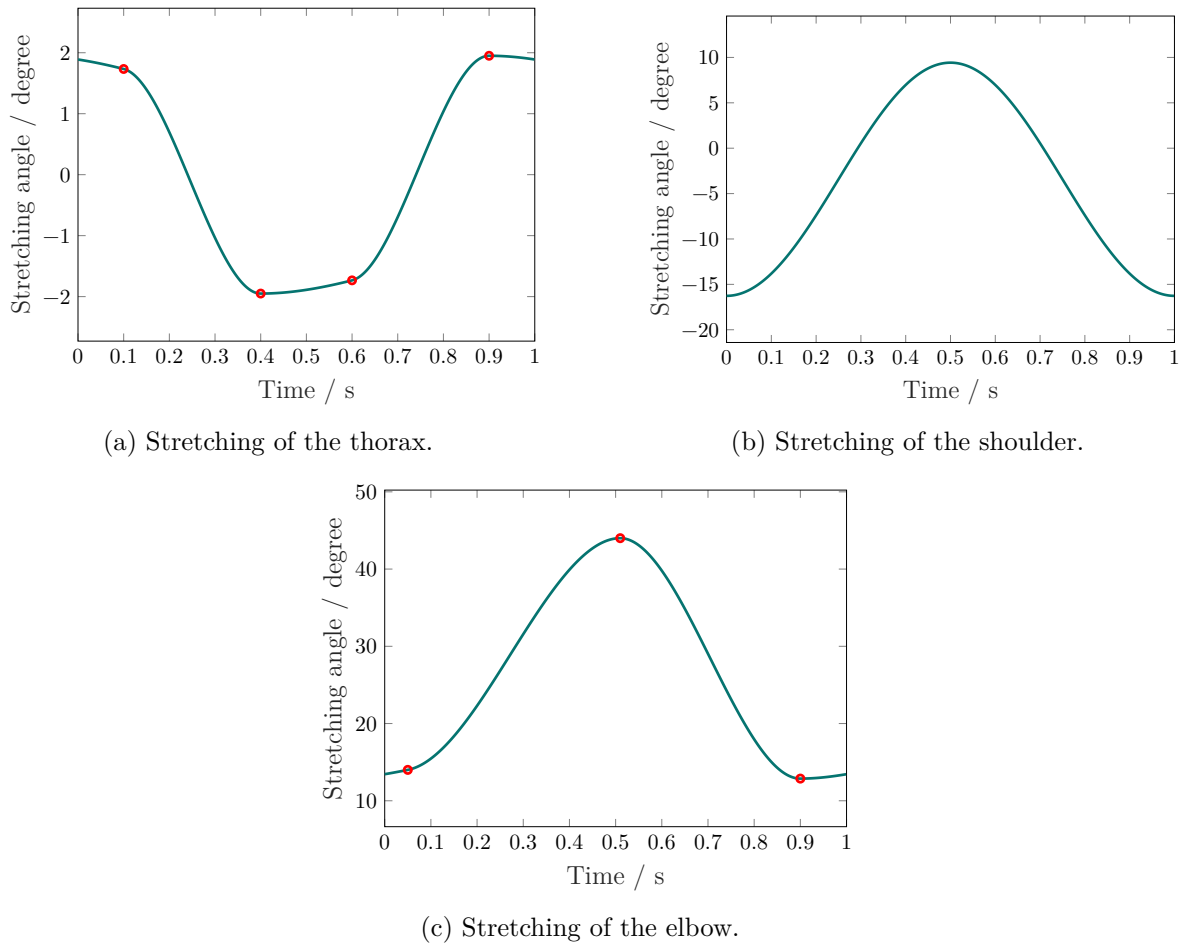


Figure 2.15: Stretching trajectories of the upper body joints.

2.4 Deep Learning Background

In this section, the theoretical background of the deep neural network aspects is discussed since it is the main approach used in all the applications that are addressed in this thesis. The section is divided into two main parts as follows:

- The main aspects of the deep feedforward network are presented in Section 2.4.1 as it is considered the most fundamental network architecture for the deep learning approaches.
- The main aspects of the deep convolutional neural network are discussed in Section 2.4.2 as it is the most commonly used in radar applications and in our thesis as well.

2.4.1 Deep Feedforward Network

The most fundamental component of deep learning is the deep feedforward network, which is sometimes referred to as a feedforward neural network or a multilayer perceptron (MLP). The purpose of a feedforward network is to get as close as possible to the desired function. For instance, the formula $y = f(x)$ can be used to transfer a set of inputs to a category. A feedforward network is one that defines a mapping, such as $y = f(x; \Theta)$, and then learns the values of the parameters (Θ) that produce the most accurate approximation of the function as described in Goodfellow et al. 2016.

These kind of models are referred to as feedforward models because information travels from the function that is being evaluated from x , through the intermediate calculations that are used to define f , and then finally to the output y . There are no feedback connections at all, thus the model's outputs are not being fed back into themselves in any way. While, Recurrent Neural Networks (RNN) refers to the networks that result when feedforward neural networks are expanded to incorporate feedback connections as explained in Goodfellow et al. 2016.

Practitioners of machine learning place a significant emphasis on feedforward networks. They serve as the foundation for a wide variety of significant commercial applications. As an illustration, the convolutional networks that are utilized for object detection from images are a specific form of feedforward network. On the conceptual road to recurrent networks, which are the driving force behind many natural language applications, feedforward networks serve as a conceptual stepping stone.

The reason why feedforward neural networks are referred as networks is because they often express themselves by piecing together a wide variety of operators. The model is connected to a directed graph, which explains how the functions are put together as a whole. These operators are organized in layers, in which the first layer in the chain is called the input layer, the last one is called the output layer, and the layers between them are called the hidden layers. The depth of the model can be determined by the total length of the chain. This phrase eventually gave rise to the deep learning as mentioned in Goodfellow et al. 2016.

The intermediate layers are referred to as hidden layers since the training data does not demonstrate the expected output for each of these levels. During the training of the

neural network, the main task is match $f(x; \Theta)$ to $f(x)$. The training data supplies us with erratic and approximative examples of $f(x)$ evaluated at a variety of inputs. The training examples detail what the output layer should accomplish at each point x . The learning algorithm is responsible for deciding how to make use of the hidden layers in order to execute an approximation as effectively as possible.

Finally, these networks are referred to be neural because neuroscientific principles serve as a general source of inspiration for them. In most cases, the vector values at each hidden layer of the network are consisting of several units that operate in parallel, with each unit representing a vector-to-scalar function. This is an alternative way of thinking about the layer than thinking of it as representing a single vector-to-vector function. Each component is analogous to a neuron in the sense that it takes in information from a large number of other components and then calculates its own activation value. In order to extend linear models such that they can represent nonlinear functions of x , we can apply the linear model not to x itself but rather to a transformed input instead $\Phi(x)$. Thus, the network tries to learn $\Phi(x)$ for a single layer according to:

$$y = f(x, \Theta, \Omega) = \Phi(x; \Theta)^T \Omega + b , \quad (2.48)$$

where Ω is the weight vector for each layer, $\Phi(x; \Theta)$ is the activation function, and b are the biases. The main task of the activation function is adding a non-linear capability to the neural network. The selection of the activation function depends on the desired task. In this section, some commonly used activation functions that may be used in this thesis work are described.

Sigmoid Activation Function

The sigmoid function was widely used in many use-cases. The sigmoid function converts the inputs that are being received to a range that varies from zero to one. the sigmoid function is defined as:

$$\textit{sigmoid}(x) = \frac{e^x}{1 + e^x} . \quad (2.49)$$

The sigmoid function has two main drawbacks. First, it has the problem of vanishing gradients, which relates to the fact that the activation of the neurons saturates either at zero or one. A deeper overview for the gradient definition and the vanishing gradient problem are presented in Hecht-Nielsen 1992. This can be illustrated as the derivative of the sigmoid function approaches an extremely low value for large negative and positive input values. Because of the near-zero derivative, the gradient of the loss function would be very minimal in this scenario, which would prevent the update to the weights and therefore the entire learning process. Second, the outputs of the function are not zero-centered, which is another trait of the sigmoid activation that should be avoided because it is an undesirable attribute. In most cases, this results in the training of the neural network being more challenging and unstable.

Tanh Activation Function

The tanh function is another activation function that is frequently utilized in deep learning. The following equation describes how a real-valued number can be mapped to the interval $[-1, 1]$ based on the following function:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (2.50)$$

Similar to the sigmoid function, the neurons reach their maximum capacity for both big negative and positive values, at which point the function's derivative reaches zero. However, in contrast to the sigmoid, its outputs are centered around zero. As a result of this, the tanh non-linearity is usually preferred over the sigmoid non-linearity in practical applications.

Rectified Linear Unit

Recently, the Rectified Linear Unit (ReLU), has been frequently used. The activation is limited to zero by setting the threshold according to:

$$R(x) \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (2.51)$$

One benefit of deploying the ReLU function is that when compared to alternative activation functions, the use of ReLU in practice speeds up the convergence of gradient descent towards the local minimum of the loss function. This is owing to the fact that it is linear and does not saturate. Other benefit is that it doesn't include very computationally expensive operations such as exponentials, when compared to other activation functions such as the tanh and sigmoid.

However, there is also a drawback in the ReLU activation function. As a result of the fact that the outputs of this function are zero for input values that are less than zero, the neurons that make up the network can become extremely fragile during the training process. During the process of updating the weights, it is possible that the weights will be adjusted in such a way that the inputs for certain neurons will always be less than zero. Therefore, those neurons are guaranteed to be zero at all times and do not contribute in any way to the process of training. This indicates that the gradient that is being carried through these ReLU neurons will also be zero.

Leaky Rectified Linear Unit

Leaky ReLU is only an enhanced variant of the ReLU activation function. In order to overcome the issue of deactivating some neurons, leaky ReLU includes a minor change in the function according to:

$$R(x) = \begin{cases} x & \text{if } x \geq 0 \\ Ax & \text{otherwise} \end{cases} . \quad (2.52)$$

The main idea behind the leaky ReLU is replacing the horizontal line for values below zero a non-horizontal linear line. This linear line's slope can be modified based on the parameter A as described in Equation (2.52). Accordingly, when including such leaky linear line, the zero gradients are avoided.

Softmax Activation Function

Softmax activation function is only used in the last layer of a neural network if the prediction of probability scores is required to be achieved by the neural network. Accordingly, the

softmax activation function induces the output neurons values in the range of $[0,1]$, so they represent probabilities. When classifying input features into distinct classes, the fact that these classes are mutually exclusive must be taken into account. For such analysis, the probability scores of all output neurons must accumulate in total to one, which is achieved by the softmax activation function.

2.4.2 Deep Convolutional Neural Network

Deep convolutional neural network (DCNN) are a type of neural network with a grid-like structure used to process data. Time-series data, which can be viewed as a 1-D grid sampling at regular intervals, and image data, which can be viewed as a 2-D grid of pixels, are two examples. In multiple real life scenarios, convolutional networks have been proved to be effective. The DCNN suggests that the network performs a mathematical procedure known as convolution, which is a type of linear operation. Accordingly, DCNN are neural networks that substitute convolution for ordinary matrix multiplication in at least one layer.

DCNNs have the benefit over standard feedforward neural networks of capturing spatial and temporal dependencies in an image and lowering runtime and space complexity. The convolutional filter, often known as the kernel K , is the fundamental principle underlying the calculations of a DCNN. It computes linear convolutions on the data points that covers a receptive field of the whole input data matrix and computing linear convolutions on the data points. The kernel traverses the image with a specified stride until the entire image has been explored. This convolution procedure inherits easily the exploitation of two invariant properties of the data, which are spatial locality and translational invariance. Spatial localization describes the topological proximity of relevant information for a given feature. Translational invariance indicates that visual characteristics can be detected regardless of their position. While the former is used as a result of the filter covering an enclosed part of the input data, the latter is achieved by sliding K across the entire image. Thus, a single K is reused across the entire image, and a single parameter set is shared across all of the image's data points. This is known as parameter sharing, and it drastically minimizes the quantity of parameters that must be stored. Traditional NNs, on the other hand, utilize each parameter only once as explained in Ankile et al. 2020. The

most generic form of convolution can be described as an operation that is applied between two functions of real-valued parameters, and is commonly symbolized by an asterisk. The generic mathematical representation can be expressed according to:

$$\begin{aligned} s(t) &= (x * K)(t) \\ &= \int x(a)K(t-a) \mathbf{d}a , \end{aligned} \quad (2.53)$$

where $s(t)$ represents the convoluted signal, x represents the function mapping to an input value, and a represents the age of the signal's value. The convolutional process gives greater weight to values that are closer to the present time t than others. In fact, signals and their processing chain are time-discrete rather than time-continuous. If x and w are defined solely on integer multiples of t , the discrete convolution can be expressed as:

$$\begin{aligned} s(t) &= (x * K)(t) \\ &= \sum_{a=-\infty}^{\infty} x(a)K(t-a) . \end{aligned} \quad (2.54)$$

The input to the convolution and its kernels in DL applications are typically multidimensional arrays, which are known as tensors as explained in Goodfellow et al. 2016. To address this, the convolution will be performed using a multidimensional filter on several dimensions. Within the scope of this thesis, the extracted spectrograms or RD maps are in the form of two-dimensional grayscale images, in which multiple features may be required to be concatenated together for some analysis. An example is the metric learning approach that is investigated for both the human identification and activity recognition tasks that are presented in Chapter 6. For such architecture, concatenation of both the captured spectrograms of the Doppler and AoA features is applied. Thus, a three-dimensional convolution operation must be computed as explained in Goodfellow et al. 2016. For such specific example, the mathematical relation can be expressed according to:

$$\begin{aligned} S(p, q, r) &= (I * K)(p, q, r) \\ &= \sum_l \sum_m \sum_n I(l, m, n)K(p-l, q-m, r-n) , \end{aligned} \quad (2.55)$$

where I is the image tensor. The typical configuration of a layer in a convolutional network includes three steps. At the beginning of the process, a layer will carry out a number of convolutions in parallel in order to generate a set of linear activations. In the second phase, each linear activation is sent through a nonlinear activation function, such as the ReLU activation function. This completes the linear activation process. This stage is also referred to as the detector stage as explained in Goodfellow et al. 2016. The output of the layer is further modified by the pooling function that is applied in the third step of the process. Pooling is used to further minimize the spatial dimension of the convoluted features and reduce the necessary processing power. Additionally, it may extract prominent features that are rotationally and spatially invariant. There are two distinct types of pooling: average and maximum. The average pooling function delivers the average of all values covered by K , whereas the maximum pooling function returns only its maximum value. During the training process, the DCNN seeks to extract just the most dominating features from the input data. If average pooling is currently being utilized, these characteristics are mixed with non-dominant characteristics, resulting in a counterproductive impact. However, maximum pooling eliminates all non-dominant features, thereby confirming the concept behind DCNN training. Due to the fact that it disregards any values other than the greatest within the current receptive field, it also functions as a noise suppressor as explained in Ankile et al. 2020.

The final stage of a DCNN is the classifier. It is composed of dense layers, where each neuron in one layer is coupled to each neuron in the next layer. This classifier is used to discover the non-linear combinations of high-level characteristics that are output by the convolutional portion of the network via a potentially non-linear function. The learning procedure is completed by feeding a tensor through the network and applying error backpropagation at each iteration step. Traditionally, the responsibility of lower layers is to capture low-level characteristics such as edges, colors, and gradient orientations. For the scope of the thesis work, applying deeper layers can be illustrated as the extraction of higher level characteristics such as the swings induced in the μ -D spectrogram due to the hands and feet motion as explained in Saha 2018.

2.5 Literature Review

In recent years, there has been a strong demand to add intelligent capabilities to applications that require more flexible human-object interaction. Such a feature necessitates a perceptual system that can recognize humans performing numerous activities, as explained in Dubois & Charpillet 2013. In indoor environments, the area coverage is typically limited; nonetheless, there are several factors that affect the detecting capabilities, such as walls and varying light and temperature conditions as explained in many studies e.g., Bernard et al. 2013, Falanga et al. 2020 and Arrue et al. 2000. For these reasons, radar has been highly investigated in the scope of human recognition due to its capturing capability that is not affected by environmental conditions or strong indoor obstacles such as walls, e.g., the studies presented in Chen et al. 2014; Li et al. 2019; Cheng et al. 2020, and Chetty et al. 2017. Moreover, radar has a great advantage of capturing meaningful signatures reflecting the micro motion behavior of the human body that have been investigated on different scales. One aspect was investigating the captured μ -D signatures for defining its characteristics, e.g., the studies presented in Seifert et al. 2018a; Seifert et al. 2018b; Seifert et al. 2019a; Seifert et al. 2019b and Seifert et al. 2020. Another research field has grasped the attention of many researchers for radar-based activity recognition, e.g., the studies presented in Erol & Amin 2019; Singh et al. 2019; Zhu et al. 2018; Wang et al. 2021, and Shah & Fioranelli 2019. However, differentiating the human from another moving object, e.g., robots as explained in Gurbuz et al. 2007 or the activity recognition are not considered the hardest tasks as the induced signatures are different. The harder task that has aroused the attention of multiple researchers is human identification. Such a task is harder due to the similarity between the captured μ -D signatures as they are all induced from the same activity, which is normally walking, e.g., the studies presented in Cao et al. 2018; Dong et al. 2020; Qiao et al. 2020, and Ni & Huang 2021.

This thesis work concludes that radar is the primary perception sensor that should be investigated because it is not affected by any environmental conditions. In addition, processing the radar data provides a simpler and more feasible method for capturing the micro-motion behavior of human body parts. The thesis work focuses primarily on investigating the possibility of capturing the micro motion signatures on various

radar features that are commonly used for human detection, identification, and activity recognition. Consequently, this section discusses the state-of-the-art radar-based techniques associated with such application fields.

Human Detection

Some camera systems are used for human detection, and they do so by recognizing objects in the scene based on the feet and heads of those objects as described in Suhr & Jung 2017. In order to improve the effectiveness of night vision human detection in surveillance applications, the study presented in Takeda et al. 2013 makes use of infrared cameras. Nevertheless, every single one of the vision-based sensors has restrictions in a variety of lighting and climatic conditions. LIDAR, which stands for light detection and ranging, is the standard sensor for safe human detection in industrial safety applications. The reflections from human legs at knee level are used to determine whether or not there is a person present in the area that is being scanned, as explained in Taipalus & Ahtiainen 2011. However, LIDAR suffers from a number of drawbacks, including difficulties in detecting reflections from dark surfaces and issues with operating in harsh outdoor environments.

The system described in Zhao et al. 2006 is a 3D-based technique that employs multiple cameras to create a 3D human model in order to bypass the obstacle limitation. For motion tracking, a local optimization algorithm is used, and recognition is accomplished through linear time normalization. Consequently, relying on a limited number of sensors to do the desired task is regarded as an additional key factor by numerous studies. Due to its ability to see through obstacles, radar is proposed by numerous studies as a possible replacement for a camera e.g., Zetik et al. 2006, Ram & Ling 2008 and Amin 2017.

When creating a system for gait recognition, the system's cost and computing efficiency are two crucial design factors to consider. The application described in Ahmadi et al. 2015 is based on wearable sensors and individualized kinematic modeling. By monitoring the motion of the lower limbs and describing the 3D trajectories of individual leg segments, the reported work satisfies these two objectives. A sensor attached to the foot is then utilized to distinguish between the stance and swinging walking states. The alignment of the leg joints is then established. Lastly, the detected motion is modified by employing a

specialized kinematic model. The described technology is economical and computationally efficient. All processing, however, is performed using advanced methods.

Another use of wearable ultrasonic sensors for the analysis of different walking phases is provided in Qi et al. 2015. The walking activity is typically separated into two phases, which are the stance and the swinging, as explained in Chen 2019. During each phase, the human movement consists of a double support phase in which both feet are on the ground and a single support phase in which one foot is on the ground while the other is swinging. By detecting the positions of the ultrasonic sensors, the different states can be identified. These sensors may be detected using a fixed-anchor network, and their identification is further improved using the recursive *Newton-Gauss* technique and *Kalman Filter*. Using a commercial optical motion tracking system, the performance of the system is examined. The device is suitable for use in medical applications to distinguish between damaged and uninjured individuals. It can also be utilized in the sectors of sports and rehabilitation engineering. However, wearable sensors are not appropriate for many applications and are seen as a significant constraint.

A motion capture system may also be utilized for human detection, such as Qualisys Qualisys 2015, which demonstrates high accuracy by recognizing the fixed infrared markers on the walking human using high-tech cameras and algorithms. As a result of the mandatory availability of cameras and infrared markers, the capturing system is viewed as more laboratory-like. Additionally, Qualisys is expensive compared to competing systems, and it still has the same restriction as wearable perception systems. It can be inferred that the cost, computing efficiency, and computational complexity of a competent gait tracking system limit its use to certain locations and fields of application. In order to circumvent such limitations, radar has been investigated as a feasible sensor for extracting the body limb trajectories based on the μ -D signatures as described in Chen 2019. The concept has been also further evaluated by us in our study published in Abdulatif et al. 2017.

Through-the-Wall radar is a form of radar used in surveillance applications. This method necessitates the detection of living beings from the exterior of a structure behind a wall. Radar can detect unique μ -D signals that are produced by human respiration. Consequently, the μ -D signatures of a human's breathing can be utilized in conjunction with such radar to detect a living human, as detailed in Radzicki et al. 2016. The laboratory

experiment conducted in the study consisted of a constructed concrete wall on both sides, of which sat the radar and a seated human. The outcomes were examined using the STFT method.

As described in Wang & Fathy 2011, one application of the discovered μ -D is the recognition of a moving limb, such as the arm. A hardware radar design capable of detecting the μ -D signatures of swinging arms is given. Using radar, a comparison was made between two walking techniques, one arm or two arms swinging. Combining the range profiles with the μ -D signatures improved the results. Heartbeat recognition demonstrates that the radar system is also capable of differentiating between stationary and walking humans.

Through μ -D signature recognition, a variety of human movements may be evaluated. However, walking recognition is one of the most significant topics of μ -D analysis involved in a wide range of security and surveillance applications. A difficult aspect of walking μ -D analysis is the dissection of various bodily components, as discussed in Chen 2019. The study presented in Guldogan et al. 2011 provides an example of the decomposition of body segments. The study decomposed body parts using the STFT approach in conjunction with particle filtration. Utilizing the Monte-Carlo particle filtering technique, the torso and the legs were successfully distinguished.

Raj and Chen report a study on μ -D signatures decomposition in Raj et al. 2009 and Raj et al. 2010. The paper provided two approaches based on g-Snake modeling of the time-frequency representation. The first technique uses a Gaussian g-Snake model to derive μ -D signatures from the smoothed manifold curves of moving body parts; this is a parametric method for disentangling body part signatures. The steepest descent method is used to minimize mean-square error and estimate the values of these curves, which are represented as functions of various parameters of interest. The first technique relied on prior knowledge; the second, based on a nonparametric approach, proposes using a peak tracking algorithm instead. These suggested algorithms require a time frame of at least half a human gait cycle ($\approx 0.5s$). Therefore, this approach cannot be utilized in a real-time safety-critical application.

In radar, the application for recognizing human gait is commonly referred to as μ -D signature. The μ -D signature received from a walking human is the superposition of the

effects caused by various body parts. Each bodily part, particularly the human limbs, has been seen to induce a unique μ -D signature distinct from the others. These characteristics are investigated using a time-frequency representation, as described in Hornsteiner & Detlefsen 2008. The proposed work demonstrates that the mobility of human body parts induces very specific measured velocities, such that even the velocity of the torso, the largest body part, cannot be seen as a constant velocity. The work is also elucidating the effectiveness of the spectrogram representation in depicting the velocity change of small cross-section body parts, such as the legs.

All of the provided state-of-the-art techniques only account for the unique μ -D signatures created by each body part, as may be deduced. While walking, however, human body components generate a unique μ -R signature that must be taken into account for the decomposition of various body parts. Moreover, a unique micro motion signature is expected on the AoA dimension if a MIMO protocol is used and the AoA can be estimated. Thus, the main focus of the thesis work is investigating the main characteristics of the captured micro motion signatures on the main radar features, which are range, Doppler, and AoA. The human detection is considered the first stage of the research in this dissertation, where the basic case of single target detection is examined on a SISO radar. The research outcome is applied to a human-robot classification use case.

Human Identification

Utilizing the properties of millimeter-wave (mm-wave) backscattered radar has become a concept of keen interest for human identification. The successful adaptation of DL to this application domain has resulted in significant classification accuracy gains. In Cao et al. 2018, the authors compare supervised learning techniques such as supported vector machines (SVM) and naïve Bayes (NB) to one of the earlier NN architecture, AlexNet presented in Krizhevsky et al. 2012. As the learning dataset for the various techniques, μ -D signatures of targets running at LOS have been collected for comparison purposes. In contrast, NNs that perform end-to-end classification techniques such as SVM and NB heavily rely on manually crafted features. To extract the most descriptive features from a given dataset, the latter requires a substantial amount of time and specialized domain knowledge. Consequently, results will vary based on the extracted features. This

severely restricts the applicability of machine learning techniques such as SVM and NB to specialists in the relevant fields. Due to the fact that NNs perform all feature extraction themselves, the obstacles of SVM and NB are circumvented, making NNs frequently the superior option.

The majority of research conducted in the field of human identification using mm-wave radar employs μ -D spectrograms and a DCNN architecture. In Papanastasiou et al. 2021, an X-band radar and a VGG-16 DCNN are utilized to classify the μ -D spectrograms of targets walking at LOS to the radar sensor. The work in Qiao et al. 2020 utilizes a LOS dataset of μ -D spectrograms and employs AlexNet for the classification task. In this instance, the radar sensor's polarization is aligned along the vertical plane to account for the human's vertically elongated body structure. Important to this study is the μ -D algorithm for separating limb movements from torso movements. Thus, the authors were able to demonstrate that the limbs' motions contain more distinct characteristics than the body's bulk motion. The research in Vandersmissen et al. 2018 removed the restriction of only being able to move along LOS path to the radar sensor, allowing for unrestricted movement along any path. Similar to the previous two studies, they employ a custom-built DCNN with a basic architecture. Even though human gait is by far the most commonly used biometric for classification with electromagnetic mm-wave, experiments have been conducted based on other biometric characteristics, such as the heartbeat as described in Cao et al. 2019.

Common DCNN network architectures that achieved remarkable results in the ImageNet large-scale visual recognition challenge (ILSVRC) described in Russakovsky et al. 2015 or custom DCNN network architectures employing elementary structures have been utilized in most of the previous studies. However, research has been conducted to evaluate the advantages of more advanced DCNN architectures for the mm-wave human identification task based on μ -D spectrograms. Consequently, the authors of Yang et al. 2019 use an adaptive weight learning strategy that is motivated by the attention mechanism as explained in Hu et al. 2018, whereas Jalalvand et al. 2019 introduces reservoir computing networks (RCN) to the topic. These are desirable for time-series data processing, as μ -D signatures are, due to their high efficiency in capturing temporal information and their good noise-handling properties. On their dataset, the bi-directional RCN developed in

Jalalvand et al. 2019 outperforms conventional DCNNs. It was captured by a SISO-FMCW radar and consisted of five people walking in an unrestricted and spontaneous manner. The dataset has been made publically accessible and named IDRad dataset. Another method based on this dataset is described in Addabbo et al. 2020, with the objective of discovering the causal relationship underlying the evolution of gait signal. Deep temporal CNNs are employed for this purpose, which are distinguished primarily by their sequence length and the casualness of their convolutional architecture.

The backscattered radar signal typically contains a fair amount of noise, especially in indoor environments, due to furniture and multipath reflections. The ability of the signal processing chain to eliminate these interferences is a crucial element that distinguishes a superior system from a merely adequate one. Using a maximum energy ridge extraction method, the research in Ni & Huang 2020 proposes a solution in which the range domain is used to extract the precise motion trajectory of the target using the range domain described in Iatsenko et al. 2016. For the subsequent Doppler FFT, only the range bins surrounding the trajectory path where the actual target exists are utilized, thereby rejecting noise artifacts and ghost targets. In addition, the paper demonstrates the efficacy of transfer learning in the context of mm-wave-based person identification. Data-wise, μ -D signatures and the images comprising the ImageNet dataset are different from each others. Feeding the spectrograms as RGB images to the network, which was initialized with the ImageNet weights, alleviated concerns that this would cause issues with the transfer learning of a previously trained ResNet-50 as explained in He et al. 2016 architecture. As reported, fine-tuning this pre-trained network significantly outperforms training from scratch.

The study presented in Yang et al. 2020 provides an approach that makes direct use of the range feature in which only the lower limb movement of the target is considered. The step length, step duration, instantaneous lower limb velocity, and inter-lower limb distance are intrinsic to the lower limb movement and can be used as another biometric. The authors argue that the lower limb motions of humans are more consistent and stable even over the course of days, as they rarely interfere with actions such as typing on a smartphone, looking at a watch, carrying various types of bags, or putting hands in pockets. The ability to identify multiple individuals simultaneously is a benefit that has not been observed in previous research. Consequently, the same paper implements a spatial-temporal silhouette

analysis for step segmentation based on AoA data and unsupervised learning techniques in order to determine the optimal segmentation strategy.

The fact that human motion induces time-variant frequency modulations in the backscattered radar signal enables not only human identification but also the classification of different types of motion e.g., the studies presented in Zhang et al. 2016; Amin & Erol 2018, and Liu et al. 2020. As both of these tasks have been successfully accomplished individually, research works such as Lang et al. 2019 and Li et al. 2019 demonstrate the advantages of combining person identification and motion classification. Therefore, Lang et al. 2019 proposes a DCNN architecture consisting of three fundamental components: The first is a backbone subnetwork inspired by the DenseNet architecture described in Huang et al. 2017. The second component is fundamental to the network’s performance attention modules, and the third is a hierarchical feature fusion structure. By sharing weights and jointly learning to classify tasks and identify individuals, the network’s overall performance can be enhanced. This is supported by the research presented in Li et al. 2019, which describes a multi-scale residual attention network (MRA) with a multi-scale learning component and a residual attention mechanism. The MRAs are ideally suited for multi-task learning because they can extract features of varying granularity and thus learn more accurate representations of the various tasks.

In real-life applications such as access control and surveillance systems, the likelihood of unidentified individuals entering the frame is high. While closed-set machine learning systems perform well on the same set they were trained on, they frequently exhibit erroneous behavior when presented with untrained instances. The authors of Ni & Huang 2021 propose a solution in which an unknown subject is classified as unknown rather than being mapped to the closest known subject. A probabilistic discriminant model is utilized for this purpose. It consists of a deep discriminative representation network (DDRN) with cosine margin loss to map images into an embedding space and a class inclusion probability (CIP) model to bind known class regions within this embedding space. The CIP then returns the probability of class membership, which is then thresholded to determine whether the query sample is a member of a class or has an unknown identity. For this strategy, it is essential to learn a discriminative embedding space in which samples

of the same class are mapped to close proximity and those of different classes to distant locations.

As described in Section 6.1, it is not always possible to collect large amounts of data for neural networks to learn from. However, all previously presented works relied on large datasets of μ -D signatures that necessitated extensive data collection time. As a solution, the research in Niazi et al. 2021 proposes a Gaussian prototypical network architecture to solve the task of human identification and classification using few-shot learning and different types of luggage. This is considered an uncommon technique compared to the other investigated approaches in the state-of-the-art.

The μ -D spectrograms that have been utilized thus far display an equivalent width of range 1-4 s of spectrogram data, with the majority falling within the three-second range. While a spectrogram of that size provides the NNs with a wealth of information from which to learn, it inherits three flaws, the first of which is its size. A three-second μ -D spectrogram sample contains roughly three complete gait cycles. If only a small amount of raw data is available and there is no overlap, the dataset will consist of very few samples. If, however, the continuous spectrogram data were sliced without overlap every half second, the resulting dataset would be six times larger. The periodic nature of the spectrogram, which is approximately $2Hz$, could be utilized as a side effect. A second issue that arises with non-periodically sliced spectrograms is that gait cycles that lie near one sample's outer boundaries may be cut at random within the half gait cycle interval. The result is a noisier dataset, which can lead to inferior training outcomes. The third issue that must be addressed is the amount of memory space the NN requires. The larger the input samples, the greater the required memory capacity. Consequently, hardware constraints frequently necessitate a small NN architecture, which may not be adequate for satisfactory performance. Our feasibility study presented in Abdulatif et al. 2019 identifies humans using half-second μ -D samples. Its primary objective is to comprehend the effect of human body characteristics on walking style. Targets walk on a treadmill while a radar sensor is positioned at predetermined distances. ResNet-50 is used for classification, and it has been discovered that a human's μ -D signature correlates with its body mass index (BMI). The study is discussed and presented in details in Chapter 5.

The next step is to track the in-frame targets in addition to the successful identification of humans using neural networks primarily trained on large datasets. This is precisely what the study in Janakaraj et al. 2019 does use two subsystems: one for tracking and one for the identification task. The tracking is based on RD maps obtained by computing a two-dimensional FFT on the raw radar data and a sliding window that predicts the position of the target in the next RD map instance. In addition to tracking, it measures the distance traveled, which is proportional to the size of the window. A residual DCNN is then employed for identification purposes. The study presented in Wagner et al. 2017 demonstrates additional work for joint target tracking and human identification based on conventional spectrum estimation. A clustering stage combines subsequent detections and generates individual target representations. Features can be extracted from the representations and then fed to a NN classifier. The tracking is accomplished with a Kalman filter. The authors of Pegoraro et al. 2020 present an approach comparable to Wagner et al. 2017 but with more sophisticated algorithms. The tracking is still performed using a Kalman filter, but the Hungarian algorithm is added to improve track association capabilities. Utilizing a custom DCNN architecture based on the inception residual network family, the identification task is completed. By connecting the tracking and identification modules, significant algorithmic improvements are obtained. Thus, correct labels can be reassigned even if targets are temporarily unable to be tracked due to factors such as occlusion or intersecting motion trajectories. Studies on the significance of angle data for simultaneous tracking and identification have concluded that the range-Doppler-azimuth space produces better results than the relying only on the range-Doppler space.

The use of spectrogram representations has been the primary focus of research on mm-wave-based human identification. One reason is the well-known effectiveness of neural networks in image classification tasks. However, their computation requires intensive computations through applying multiple FFT operations, resulting in high computational demands. Therefore, satisfying real-time requirements is frequently challenging. Using sparse radar point clouds, the method in Zhao et al. 2019 alleviates the computational burden of the system by proposing a solution to human identification during unrestricted and spontaneous walking. These are clustered by the density-based spatial clustering of applications with noise (DBSCAN) algorithm, which employs a modified Euclidean

distance that accounts for the human’s greater elongation along the vertical axis compared to their horizontal elongation for improved clustering outcomes. Then, a Kalman filter is used to predict where the target will move in the subsequent frame, and the Hungarian algorithm is employed for inter-frame object association. A sequence of occupancy grids is ultimately utilized by a deep recurrent neural network (DRNN) for human identification. They are based on the clustering of radar point clouds and their association with an existing human track. The authors of Pegoraro et al. 2020 refined their system and are now basing their new research presented in Pegoraro & Rossi 2021 on sparse radar point cloud sequences. For target tracking, a converted-measurements Kalman filter is now utilized, which estimates not only the target’s position but also its horizontal extent. In addition, joint probabilistic data association with the nearest neighbors is a component of the tracking block, and the Hungarian algorithm is still employed for track association. The tracking and identification modules are interconnected, with the identification module consisting of a novel DL classifier that was designed to operate effectively on sparse radar point clouds. Extensive studies demonstrate the efficacy of the proposed method, which works well within real-time constraints.

The presented work aims to fill a gap in the literature by establishing a radar mm-wave-based human identification system that requires only sparse data for training and performs satisfactorily with only a small amount of gait information for each sample. Consequently, two types of spectrograms, μ -D and μ - ω , are computed from the raw radar data. These are subsequently sliced into half gait cycle samples and fed to a DCNN network structure based on the one proposed in Pegoraro et al. 2020. To satisfy the requirements of few-shot learning, the network is trained using the triplet loss presented in Schroff et al. 2015, an unheard-of loss function in the thesis’s research field. Lastly, a CIP model similar to that proposed in Ni & Huang 2021 is applied to determine the class to which a query sample belongs.

Activity Recognition

Previous studies relied on deploying conventional machine learning techniques with radar signals for classifying different human activities e.g., the SVM utilized in Zenaldin & Narayanan 2016, and random forest for gesture recognition in automotive applications in

Smith et al. 2018. However, traditional machine learning had a number of drawbacks and limitations that limited the effectiveness of these techniques. Traditional machine learning techniques require manual feature extraction, which heavily relies on the application domain expertise of the researchers. Recent advances in computational power and advancements in the field of deep learning have enabled the achievement of remarkable results in a variety of fields and the resolution of many of the limitations and issues associated with traditional machine learning. Recently, a variety of deep learning techniques have been investigated for radar-based systems.

It has been shown that DL techniques outperform more traditional learning methods, as indicated in Gurbuz & Amin 2019. It was mentioned that the training dataset's size was one of the key challenges. Such an issue necessitated the adoption of transfer learning, where simulated data were employed as the base for training. In Erol et al. 2019, an alternative solution was described in which a generative adversarial network (GAN) was utilized for data augmentation. One of the primary reasons for expanding the training dataset was to reduce classification confusion between similar behaviors, such as (bending and sitting) or (standing and walking) or (crawling and creeping). Using additional radar elements, such as range profiles, was another method for resolving the classification ambiguity, as explained in Erol & Amin 2019. In the aforementioned investigations, the classification task was based on a single occurrence of each activity. Thus, operating the radar to take single snapshots individually, as described in Campbell & Ahmad 2020, and Gurbuz & Amin 2019. Another approach for extracting single occurrence of each activity is based on slicing a long-period captured signature as discussed in Li et al. 2019. In realistic circumstances, the actions are likely to occur sequentially, necessitating adaptive slicing, such as the recurrent neural network (RNN) proposed in Gurbuz & Amin 2019. Similarly, in Li et al. 2019, streams of several activities were sliced using a sliding window with a 50-70% overlap. There are more factors to consider, such as data generalization and the detection angle, according to Gurbuz & Amin 2019.

Since the research scope of the thesis is directed in extracting meaningful radar features that can reflect the micro motion behavior; it has a direct correlation with the activity recognition task. Since both the μ -R and the μ -D signatures have been investigated in previous studies and were proved to show classification ambiguity for similar activities, the

feasibility of including our newly measured μ - ω signatures is investigated. Additionally, such tasks suffer from the obligation of capturing a sufficient dataset size for training, and hence, our scope is directed towards investigating a metric learning approach for the classification task. Such an approach is based on the triplet loss architecture, which has a two-fold benefit. First, it is used by us to test the concept of few shot learning (FSL) for including hard activities to be collected e.g., falling. Second, the concept of combining the captured signatures from multiple AoA is tested to extend the classification to more generic and realistic use-cases and not limited to the LOS scenarios.

2.6 Summary

In this chapter, the theoretical background of both the radar system, as well as the algorithms that were applied in the work for this is discussed. A general introduction to the radar system as well as an explanation of the various transmission protocols can be found in Section 2.1. In the Section 2.2, the idea of studying the primary radar features in order to capture the micro motion behavior is described in greater detail. Walking is considered to be the primary activity that is investigated in this thesis work; therefore, the characteristics of the human body as well as the behavior of the human body during the walking gait cycle are presented in Section 2.3. The theoretical foundation of the deep learning algorithms that were used in the work that was done for this thesis can be found in Section 2.4. A review of the most state-of-the-art approaches in the relevant study domains that we have focused on can be found in Section 2.5.

3 Radar Signal Processing

In Section 2.1, the estimation of the radar’s primary characteristics has been discussed from theoretical and physical perspectives. In this section, the signal processing procedure for estimating the primary radar characteristics will be discussed briefly. The MIMO-FMCW transmission protocol is discussed in this chapter. The radar receives the reflected chirps from the radar’s surroundings and processes them as a raw data cube, as depicted in Figure 3.1. According to Li & Stoica 2008, each dimension of the raw data cube represents a main radar feature. The first dimension reflects the received chirp with its samples, as described in Section 2.1.2, which is responsible for the estimation of the range. As demonstrated in Figure 3.1, the processing within a single chirp is commonly referred to as the fast time dimension, as described by Richards 2014. The second dimension is responsible for analyzing consecutive chirps through time, referred to as slow time in Figure 3.1, and is used to estimate velocity. The third dimension of the radar data cube represents the simultaneous reception of chirps from multiple Rx antennas. Consequently, this dimension is responsible for AoA estimation, which in Figure 3.1 is referred to as spatial sampling. And therefore, processing the radar cube enables the mapping between any pair of radar features or all three if the application requires it.

Multiple features increase the algorithm’s applicability, as demonstrated in Lin & Ling 2006; Cui & Dahnoun 2021; Ram et al. 2007 and Lin & Ling 2006, in which the Doppler feature is utilized as an indicator for activity detection and the AoA feature is used for real-time tracking. Such goals indicate the viability of the MIMO radar for applications requiring high real-time performance and present the radar as a powerful stand-alone solution for multiple-human recognition. To preserve computational resources, however, the radar cube must be processed with great wisdom. We conducted the deployment of MIMO radar for prosthetic leg integration with traversing objects in a study that was published in Aziz et al. 2021. Such research reveals that the MIMO radar is ideally suited for the powerful swinging motion of the human leg during walking. However, evaluating the entire radar cube necessitates enormous computational resources that are typically

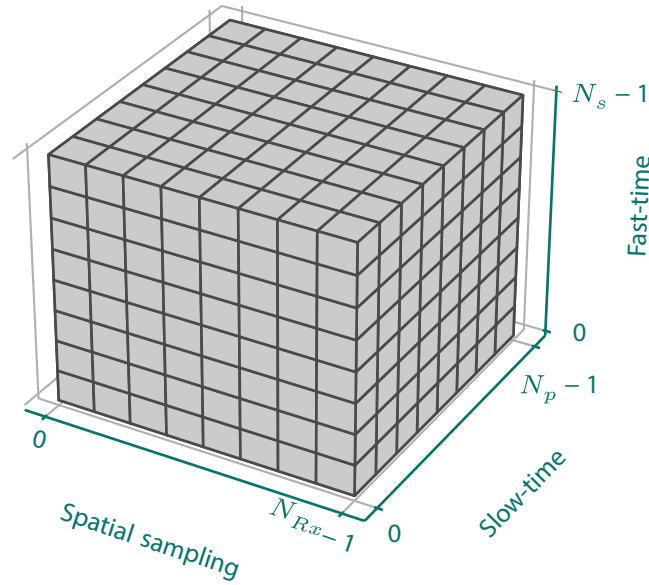


Figure 3.1: Cubic representation of raw radar data.

unavailable in such circumstances. Consequently, the Constant False Alarm Rate (CFAR) was utilized for optimized computation efforts.

In the majority of cases, human recognition applications require the use of deep learning, and MIMO radar is regarded a powerful tool since it can provide data in several formats, such as 2D maps, 3D matrices, and signals as has been stated in Aziz et al. 2022 or Weller et al. 2021 or Aziz et al. 2021. The radar is a viable sensor for integration with camera systems, such as infrared in Ulrich et al. 2018 and RGB cameras in Li et al. 2017, due to its flexibility in data formats and real-time performance. In this section, therefore, the signal processing of the data received by the MIMO radar is examined, as it is the most generalized radar type that may be employed as either a SISO or a MIMO. The application of AoA for human tracking is outside the scope of this study. To provide the most effective real-time performance, this section discusses the importance of assessing the acquired micro motion signatures on all the radar's primary features for a single instance of human activity e.g., a single walking step. Consequently, the outcome of the thesis can be extended to include integration with a tracker or integration of the radar with other perception systems.

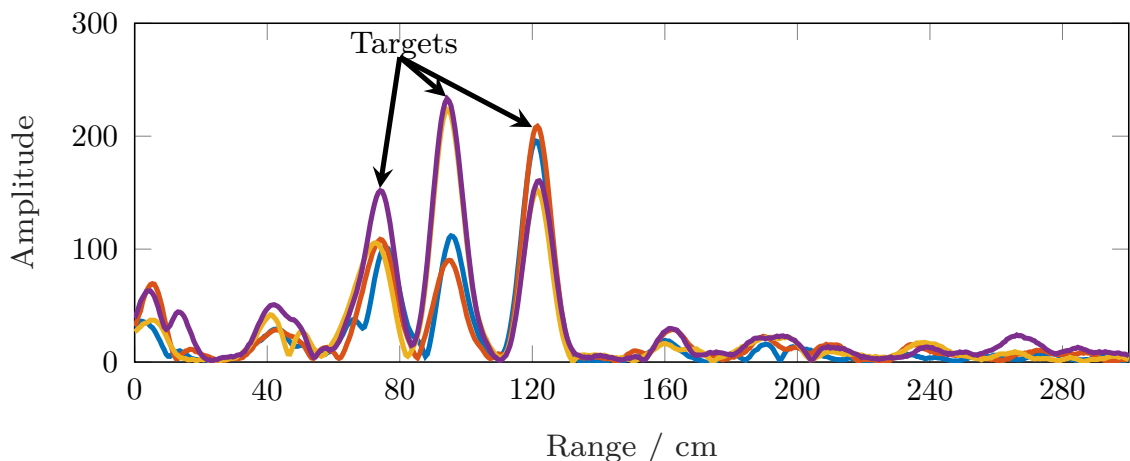


Figure 3.2: Range Profiles of MIMO-FMCW Radar.

3.1 The Main Radar Features

Analyzing the frequency variation in the IF signal, Section 2.1.2 describes the concept of range estimation. Consequently, the range can be estimated by applying a direct FFT operation to the received signal and evaluating the range profile, as depicted in Figure 3.2. Due to the MIMO transmission protocol, Figure 3.2 displays multiple range profiles, which are subsequently processed for AoA estimation. Before the FFT is applied, each bin on the fast time dimension is defined by R_{res} that is estimated based on Equation (2.16). Following the application of the FFT, each frequency bin is defined based on the parameterized number of samples (N_{FFT}), which is solely responsible for the smoothing quality of the range profiles, as described in Richards 2014. The estimated range profile is constrained by the parameterized R_{max} derived from the Equation (2.14).

The second FFT is applied to the slow time dimension in order to revile the phase change between consecutive range bins, which reflects the Doppler effect described in Section 2.1.2. To facilitate comprehension, the third dimension of the radar data cube is ignored for the time being, and an example of a processed RD map is displayed in Figure 3.3. Similar to the range dimension, the v_{res} and v_{max} share the same characteristics. According to Richards 2014, the velocity characteristic is typically used to distinguish between two targets located at the same distance from the radar. As discussed in Section 4.3, it can also be used as an indicator of the micro-motions of the body during walking for the purposes of the thesis. As will be discussed in Section 4.4, this characteristic has been shown to

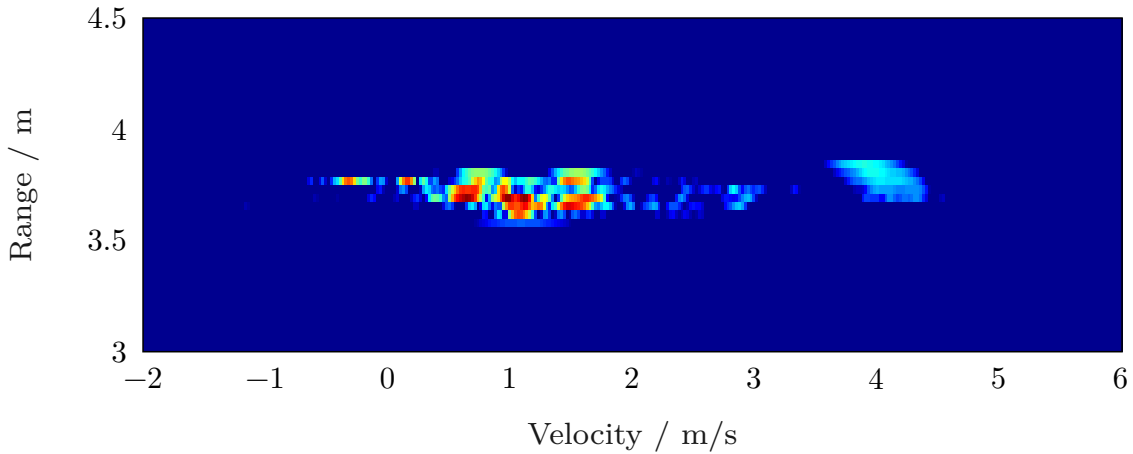


Figure 3.3: Range-Doppler map for walking motion within swinging phase.

be very useful for distinguishing humans from other moving targets, such as robots, as will be shown in Section 4.4. Figure 3.3 illustrates an example of the Range-Doppler captured for a walking human during the swing phase, where the micro motions of the limbs, particularly the feet, can be seen as multiple detected points on both the velocity and range dimensions.

Similar to how FFT is applied to the first two dimensions to obtain the range and velocity features, a third FFT is applied to the spatial sampling dimension to estimate the AoA of approach, as described in Robey et al. 2004. In the majority of instances, the AoA estimation is used as a scanning feature in the radar, in which a Range-Angle (RA) map is created. For better understanding the velocity dimension is neglected for now, and an example for the Range-Azimuth map is shown in Figure 3.4. As described in Section 2.1.3, the CDMA transmission protocol is utilized to ensure a higher angular resolution. Moreover, as explained in Chiao et al. 1997, utilizing the CDMA protocol has a direct effect on improving the SNR. In this thesis, the Hadamard code described in Section 2.1.3 is utilized. Estimating the initial phase for the available Tx antennas is performed using the Hadamard code. The MIMO radar utilized for this thesis work is featured with two Tx and 16 Rx, thus, such structure is used for illustrations in this section. For the available two Tx, an example of a 2×2 Hadamard code is given as:

$$H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (3.1)$$

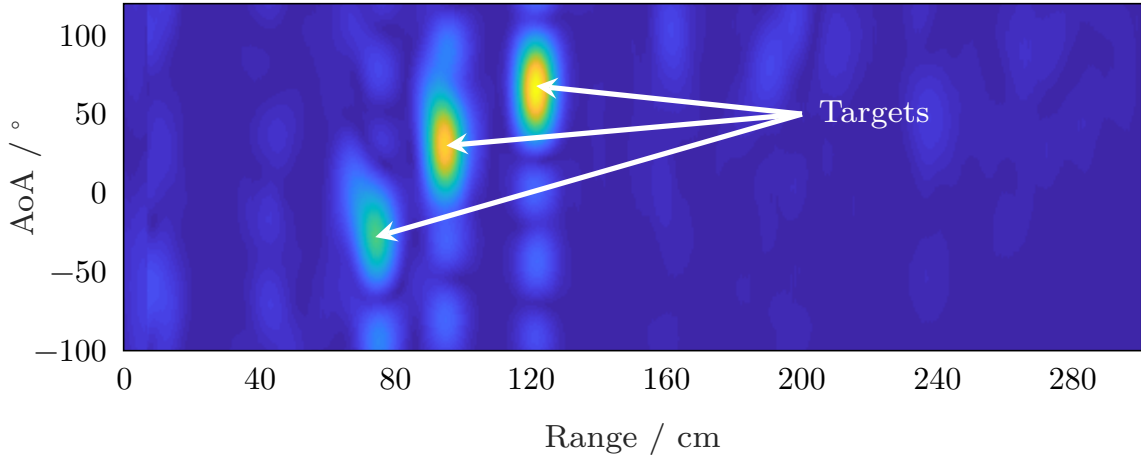


Figure 3.4: Example for an Range-AoA map for a MIMO radar.

For each radar frame, the RA maps are received consecutively based on the parameterized N_p . The process of realizing the virtual antennas starts by separating the received maps due to the odd ($s_{Rx,odd}$) and even ($s_{Rx,even}$) antennas according to Equations (3.2), and (3.3), respectively:

$$s_{Rx,odd} = s_{Tx1} + s_{Tx2} , \quad (3.2)$$

$$s_{Rx,even} = s_{Tx1} - s_{Tx2} . \quad (3.3)$$

Using these two received signals at their respective Rx antennas, the individual Tx antenna components ($s_{Tx}(t)$), where n refers to a Tx antenna, can be distinguished by means of:

$$s_{Tx,1} = \frac{1}{2}(s_{Rx,odd} + s_{Rx,even}) , \quad (3.4)$$

$$s_{Tx,2} = \frac{1}{2}(s_{Rx,odd} - s_{Rx,even}) . \quad (3.5)$$

As an illustration, this separation process is depicted in Figure 3.5 for four RA maps. Transposing the matrices improves their visibility. All 16 Rx antennas sample the two phase-modulated Tx signals at the same time and chirps are concatenated along the fast-time dimension. This data stream contains the superimposed signal of Tx1 and Tx2. These equations (3.4), and (3.5) are used to transform the 16 physical Rx channels into 32 virtual ones. The first 16 channels now represent Tx1's signal, while the second 16

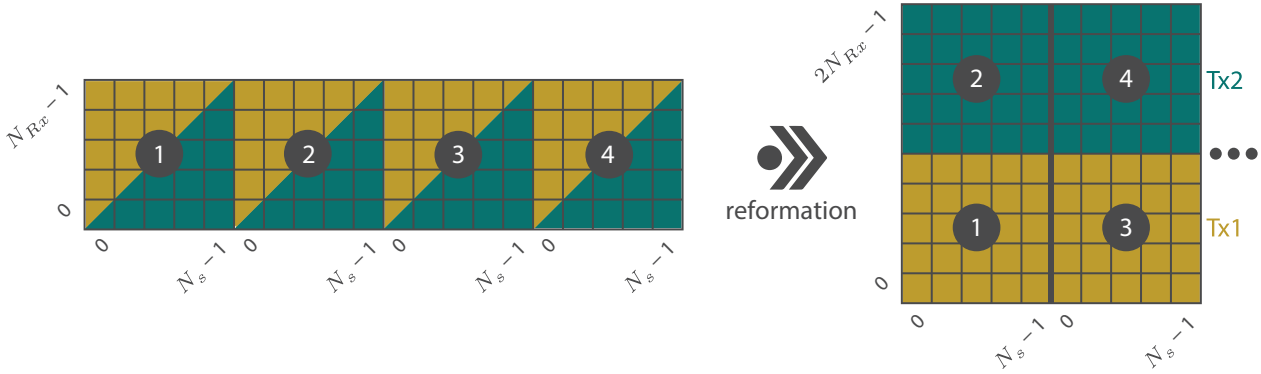
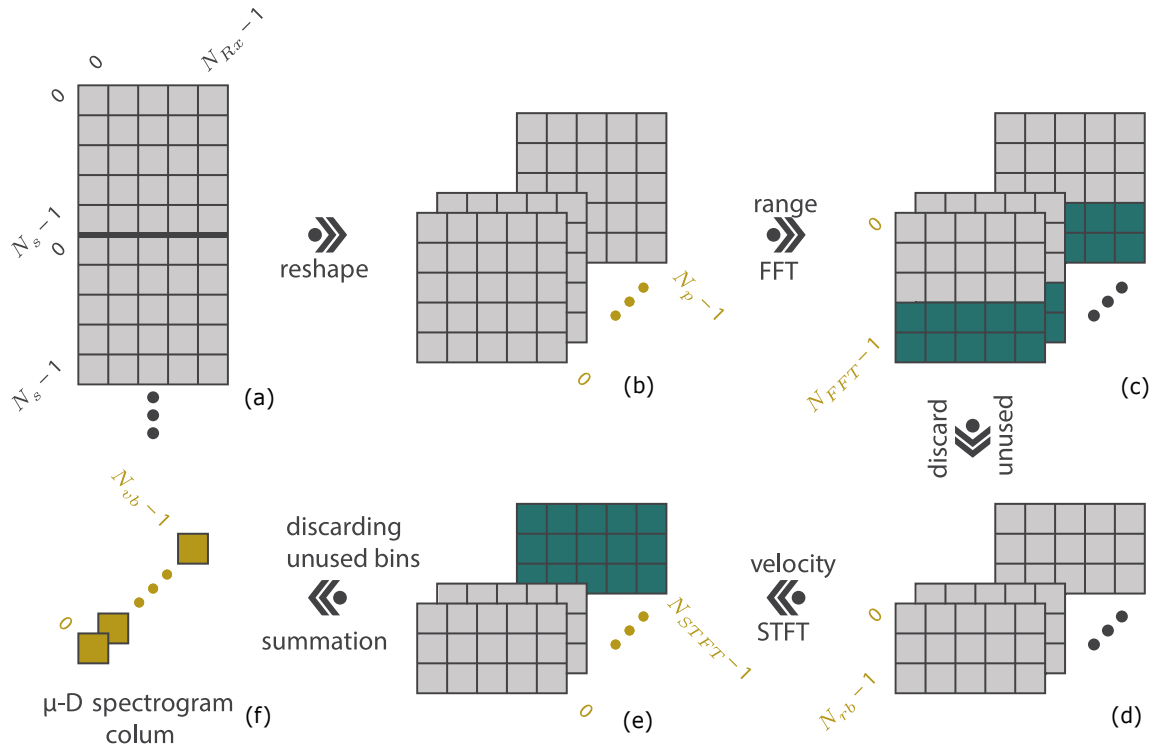


Figure 3.5: Reformation of radar raw data for CDMA analysis.

channels represent Tx2's signal. The two channels that overlap and are delineated by the two white lines carry the same information. Due to the fact that the raw data matrix is now twice as wide but only half as long as the original data stream, the angular resolution has been doubled while the maximum unambiguous velocity has been halved.

3.2 Spectrogram Computation

As discussed in Section 2.5, μ -D signatures are the most frequently employed characteristic for radar-based applications that include human activity recognition, particularly walking. In stead of applying the second FFT for analyzing the μ -D signatures, the STFT is applied to the radar data cube to revile μ -D spectrograms. As explained in Chen & Ling 2002; Chen et al. 2006; Chen 2008 and Chen 2019, the central concept of STFT is to demonstrate the frequency responses that occur in a short amount of time, which will reflect the micro-motions of the human body. It is a similar concept of applying a Zoom-FFT that is explained in Moyer & Stork 1977, in which the extracted frequency responses are reflecting the micro rate of change occurring on each range bin. A similar process can be applied to captured AoA bins in which the μ - ω signature is anticipated, as shown in Figure 2.7. The μ - ω signature is investigated for the first time in this dissertation, and it has been demonstrated to improve human detection and identification, as will be discussed in Chapter 6. However, this feature is only advantageous if it does not violate the radar detection capability of not moving laterally relative to the radar. This must be distinguished from the technique developed by Nanzer 2010, which estimates the lateral angular velocity independently of the AoA estimation. In the following sections, the signal

Figure 3.6: The μ -D spectrogram computation.

processing chains for estimating the μ -D and μ - ω signatures, respectively, are presented and briefly discussed.

3.2.1 Micro-Doppler Spectrogram

The MIMO transmission protocol is still considered for the creation of the μ -D signatures, in which the signal processing chain is described in Figure 3.6. However, analyzing the chirps received from only one Tx antenna is considered, and a SIMO protocol is considered. The MIMO feature has been abandoned for this computation as the μ -D signature to decrease the computation effort, as it is only dependent on the resolution aspects that were presented in Section 2.2.1. Additionally, the inclusion of multiple Tx antennas in the process may cause some signal distortion as the initial phase shift is not perfectly of $\Delta\varphi = \pi$ as explained in Robey et al. 2004. As a first step, the data is reshaped to a raw data cube similar to the one shown in Figure 3.5. Before processing the first FFT on the fast time axis, a Hamming window and zero padding are applied to avoid the spectral leakage and offer a smoothed range profile as proposed by de Jesus Romero-Troncoso 2016 and Mitra & Kuo 2006.

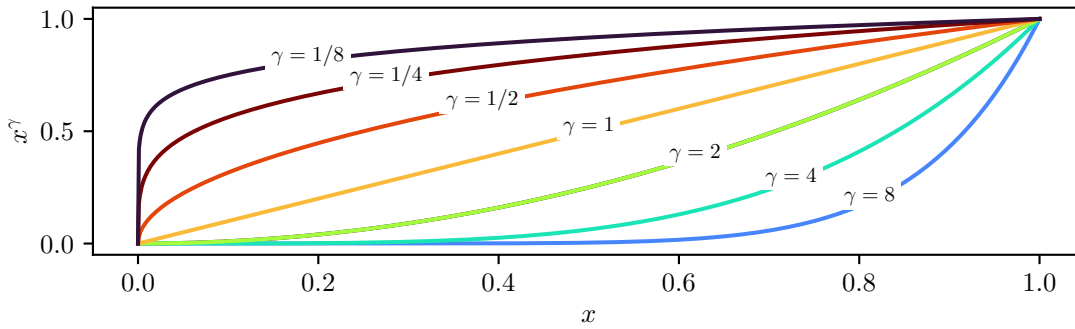
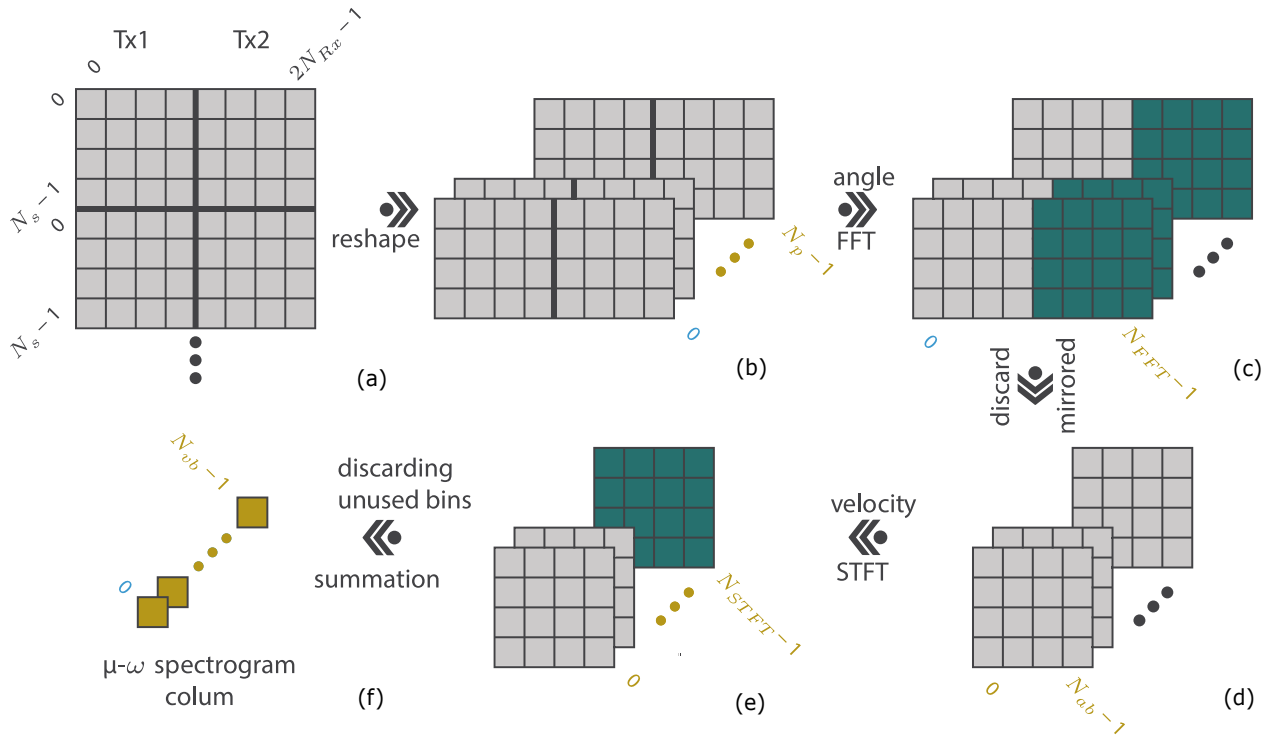


Figure 3.7: The non-linearity of a gamma correction for different values of γ .

As shown in Figure 3.6b, the data cube dimensions has changed to $D \in \mathbf{D}^{N_{FFT} \times N_{Rx} \times N_p}$, where the first axis represent the range profile. In order to accommodate for the introduced DC offset due to the Hamming window, the mean amplitude of the zero target is divided by the sum of all the range bins amplitudes. Then, the range bins that are of no interest are discarded, which are marked with green in Figure 3.6c representing in this specific example the far range bins from the radar. Other commonly-used technique, is applying a CFAR to the range profile for identifying the range bins that include targets as proposed by many studies e.g., Aziz et al. 2021, Wang et al. 2019 and Kueppers et al. 2020. This processing has a two-fold benefit of decreasing the computation efforts and diminishing the received multi-path scattered signals and the range profile dimension changes to N_{rb} as noted in Figure 3.6. Afterwards, the STFT is applied to the slow time dimension on each range bin as described in Section 2.2.1. The STFT is applied on each range bin included in the range profile along the slow time as explained in Section 2.2.1, and the data cube is now of size $D \in \mathbf{D}^{N_{rb} \times N_{Rx} \times N_{STFT}}$ as shown in Figure 3.6e. Finally, a dimension reduction is applied for both N_{RX} and N_{rb} dimensions giving a single column in the final μ -D spectrogram as shown in Figure 3.6f. The gamma correction algorithm is applied for enhancing the SNR, which has a direct effect of suppressing the low magnitude values that are expected to be noise and enhancing the high magnitude values that are expected to be the micro-motion harmonics as explained by Guo et al. 2004. The gamma correction function can be represented by different values of gamma (γ) as shown in Figure 3.7.

Figure 3.8: The μ - ω spectrogram computation.

3.2.2 Micro-Angular Spectrogram

The μ -D spectrograms have been utilized in multiple radar-based human detection applications, especially when dealing with different DL approaches as was discussed in Section 2.5. However, the researchers have faced different limitations, when relied only on the μ -D feature. For example, the study presented by Erol & Amin 2019 have recorded a magnificent confusion between similar activities e.g., falling and sitting. Another example for the human identification, in the study presented by Cao et al. 2018, have faced a trade-off between either increasing the number of classes or increasing the duration for the μ -D signature for each classification instance. Thus, other studies e.g., Erol et al. 2017 have proposed relying on both the range and Doppler features for enhancing the activity recognition applications. However, such technique still has limitations for boosting the possible classes with acceptable performance. For these reasons, in this thesis work, the AoA in the azimuth dimension is investigated in a similar style to the range dimension. The main concept is investigating the possibility of the body micro motions to induce harmonics in the time-frequency representation. To the best of the author's knowledge, such signatures were not investigated in any other study.

The signal processing chain for estimating the μ - ω signatures is presented in Figure 3.8. Unlike the processing of the μ -D signature, all the available Tx antennas should be included to ensure the best angular resolution as explained in Section 2.1.3. The utilized MIMO radar in this thesis work is featured with two Tx antennas as shown in Figure 3.8a. The raw data is reshaped to a data cube of dimensions $D \in \mathbf{D}^{N_s \times (2N_{Rx}) \times N_p}$ as shown in Figure 3.8b. Afterwards, a similar procedure to realizing the range profile is held on the AoA dimension as shown in Figure 3.8c. The procedure is based on applying a Hamming window, zero-padding and FFT for realizing an AoA profile to change the data cube dimensions to $D \in \mathbf{D}^{N_s \times N_{FFT} \times N_p}$. Since the FFT is applied to real values, a mirroring occurs in the spectrum domain, and half of the spectrum is discarded as shown in Figure 3.8c, and explained in Schmid 2012, where the discarded bins are marked with green. In a similar way to the processing of the μ -D signature, a DC compensation is done to the angle profile as the range profile. A similar procedure to the one explained in Section 3.2.1, the only sensible angle and range bins are kept in the data cube to change the dimensions to $D \in \mathbf{D}^{N_{rb} \times N_{ab} \times N_p}$ as shown in Figure 3.8d, where both N_{rb} and N_{ab} are standing for the number of sensible bins on both the range and AoA dimensions, respectively.

The time-frequency analysis represented by the STFT function is applied to each angle bin to investigate any induced harmonics due to the body micro motions through the fast time dimension as shown in Figure 3.8e. Finally, a dimension reduction is done to output a single column for the μ - ω spectrogram as shown in Figure 3.8f. An example for the extracted micro motion signatures is shown in Figure 3.9, where the full walking gait cycle in LOS is presented. As can be observed that the μ - ω signature is showing a comparable characteristics to the μ -D signature. As described in Figure 2.7 and Equation (2.37), the velocity detected by any swinging human limb is expected to have translational and rotational velocity components. The latter can be explored by realizing the μ -D signature as explained by Chen et al. 2006, Chen 2008 and Chen 2019. The newly-realized μ - ω signature is different from the spectrograms derived by Nanzer 2010, in which extra added feature is added for the radar to be capable of capturing the motions in lateral direction. As can be observed in those studies that the realized micro angular spectrograms in a lateral scenario are showing the exactly the same performance as the μ -D signatures in LOS scenario. This indicates the meaningfulness and the sensibility of the newly

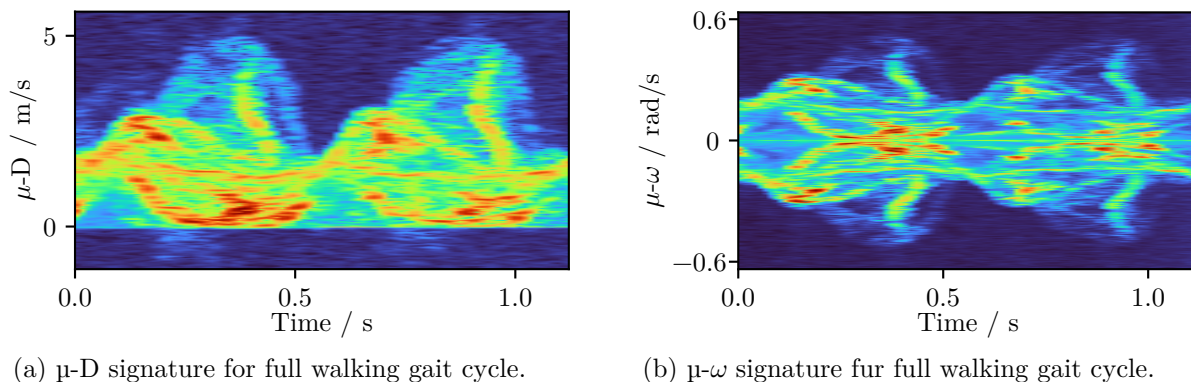


Figure 3.9: Micro-motion signatures of a complete walking gait cycle in the LOS direction.

derived $\mu\text{-}\omega$, since it is expected to show a similar behavior to the $\mu\text{-D}$ signature but not identical as can be observed in Figure 3.9. Thus, a similar behavior is expected in the radar-based applications that rely on the $\mu\text{-D}$ signature as the main feature and combining both is expected to enhance the performance e.g., decreasing the classification ambiguity for activity recognition. The feasibility and the reliability of the newly-captured $\mu\text{-}\omega$ signatures are investigated in Chapter 6 in two applications, which are the walking human identification and activity recognition. For the $\mu\text{-}\omega$ signature, a non-identical mirroring effect can be observed around the zero-velocity axis. To the best of the current knowledge, this is due to the variability of the detected points on the azimuth dimension based on the available angular resolution. Such mirroring didn't have any negative effect on the usage of the $\mu\text{-}\omega$ signatures, nevertheless, it showed a performance enhancement in the use-cases presented in Chapter 6.

3.3 Activity Adaptive Monitoring

As has been discussed in Gurbuz & Amin 2019, applying the classification on single occurrence of the activity shows great enhancements in the accuracy. Such segmentation aspect adds to the algorithm the feature of performing on realistic scenarios as the activities are expected to happen in series e.g., sitting then standing then walking. This fact is evaluated by us in this thesis work and applied in differed studies that are published in Abdulatif et al. 2019 and Weller et al. 2021 for human identification on single walking step. Moreover, it is evaluated by us in Aziz et al. 2022 for activity recognition including fall

detection. The concept of the μ -D signature is explained mainly on the walking activity Chen 2019, in which the realized spectrogram includes the main walking velocity that is induced by all the body parts and velocity harmonics that are reflecting the swinging behavior of the different limbs. Such spectrogram structure for walking can be generalized on different activities as has been discussed in different studies such as Seyfioglu et al. 2018; Erol & Amin 2019; Aziz et al. 2022, where each activity will have induced harmonics due to the micro motions that can be tracked for single occurrence extraction. Consequently, the signal processing aspect is discussed briefly in this section as it will be used in Chapter 6.

For applying the adaptive monitoring to the captured μ -D signatures, the spectrograms must be processed for envelope detection that reflects mainly the feet behavior through the walking gait cycle as has been explained in Sections 2.3, and 3.2. The envelope detection plays a main role in identifying the start and the end of single walking step; thus, an adaptive slicing algorithm can be applied. The signal processing algorithm used in this thesis for adaptive slicing is based on the study presented in Seifert et al. 2018a. In Section 3.3.1, processing the spectrogram is discussed, the keypoint vector extraction is explained in Section 3.3.2, and the adaptive segmentation of the spectrogram is introduced in Section 3.3.3.

3.3.1 Spectrogram Preprocessing

The μ -D signatures are usually utilized in DL applications, where the classification is based on capturing a long period of μ -D signature and then slicing based on the desired sample length. The slicing technique is considered a main key factor for the classification accuracy as it defines how many occurrences of each activity is included in the slice as explained in Gurbuz & Amin 2019. The adaptive slicing is proved to grant the best performance as explained in many studies e.g., Seyfioglu et al. 2018; Erol & Amin 2019; Aziz et al. 2022. However, analyzing the μ -D signatures for slicing is not foreseen as an easy task in most of the studies, and to bypass such process, some researchers relied on capturing single occurrences as the study presented by Seyfioglu et al. 2018 and Campbell & Ahmad 2020. Such occurrence-based snapshots is not feasible when dealing with only the walking activity, and the most of the researchers who dealt with walking-based human identification relied on a fixed-duration slicing technique as described in Zhang et al. 2016;

Addabbo et al. 2020; Ni & Huang 2020. Such technique is considered feasible for the walking human identification, but the adaptive slicing is still highly recommended to ensure the best performance. Since both the μ -D and μ - ω signatures are timely-synchronized, the extraction of characteristics for adaptive slicing is performed on the μ -D signatures.

The spectrogram is first converted to grayscale for subsequent binarization by normalizing its values to $[0, 1]$. Thresholding is used to get closer to a binary spectrogram. This is one of the simplest and most computationally efficient methods for distinguishing objects from the background. Thresholding is classified into two types: histogram-based and local. The histogram of pixel intensity is used for the former. It is used to make assumptions about its properties and to define the threshold value. The second category computes the threshold value locally and makes decisions based solely on neighboring pixels. Because the latter approach is computationally more expensive than the former, histogram-based thresholding is used to solve the problem at hand. The Python library *skimage.filters* contains implementations of various approaches. All of them are compared in the appendix in Figure A.2. During the computation, the black represents the value zero and white represents the value one.

The mean and minimum methods are histogram-based thresholding techniques that could separate the actual μ -D harmonics from the background as shown in Figure 3.10, and are thus investigated further. Figure 3.10a shows the original grayscale spectrogram, while Figures 3.10b and 3.10c show the results of mean and minimum thresholding techniques, respectively. Figure 3.10d displays the corresponding histogram with the threshold values for the mean and minimum algorithms. While both methods produce nearly identical results, the minimum method includes slightly less noise without discarding any relevant μ -D data points. Its operation is based on iterative smoothing until the histogram has only two peaks. The threshold value is the minimum value between them. As shown in Figure 3.10d, the text μ -D data has very evenly distributed intensities when the background noise peak at the spectrogram's upper end is ignored. This behavior is unfavorable for the minimum thresholding approach because it increases the likelihood of only one peak. After that, the algorithm will fail to complete. On the contrary, the spectrogram's mean value is determined solely by the spectrogram matrix and can thus be

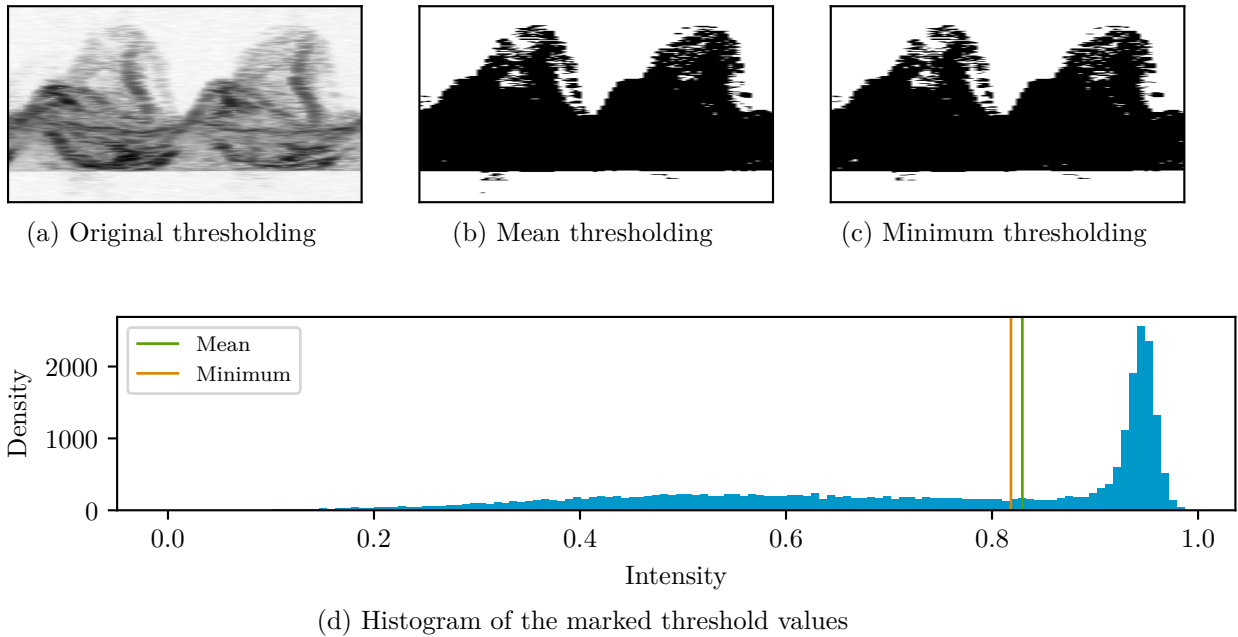


Figure 3.10: Examples of the thresholding techniques on a gray-scale spectrogram.

computed at any time. It is also faster to compute than the minimum approach, and it is eventually chosen as the thresholding technique.

For analyzing the geometrical characteristics of images, morphological image processing is commonly utilized as explained in Seifert et al. 2018a. In the case of the μ -D spectrograms, both the dilation and erosion approaches are combined together to define the region of interest that contains the main induced harmonics due to body micro motions as explained in Sundararajan 2017, and an example for the output of the processing is shown in Figure 3.11. The morphological closing that is applied in this thesis work has a main functionality for offering a clearer output for the spectrogram to enhance the signature envelop detection.

3.3.2 Keypoint Vector Extraction

For efficient adaptive slicing, it is necessary to extract the primary motion style characteristics that reflect body characteristics. As discussed in Section 2.3, the human body can be divided into two main parts, the bulky body part (torso, shoulders, and hips) and the swinging limbs (legs and arms), when walking is used as the basic activity for studying the μ -D signatures. The bulky body part is represented in the μ -D signature as the highest

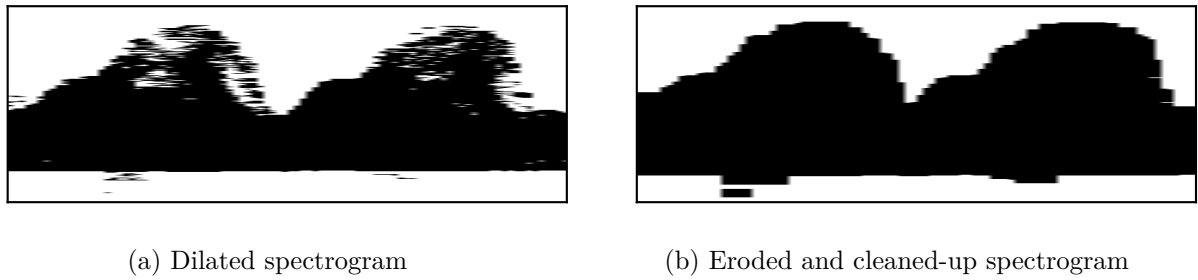


Figure 3.11: A visualization of the morphological processing chain.

intensity velocity harmonic of the walking relative velocity, whereas the swinging limbs are represented as swinging frequency harmonics that are primarily represented by the envelope of the spectrogram. Consequently, when extracting a key point vector, it will incorporate the center of gravity (CoG), primary and secondary envelopes to reflect the behavior of the bulky and swinging body parts during the walking gait cycle. The technique can be applied to other activities, with the CoG reflecting any translational behavior and the primary and secondary envelopes reflecting any rotational motion. Figure 3.12 is an example of the extracted key point vector for the walking activity.

Center of Gravity

The gray-scaled spectrogram is multiplied by the morphologically processed spectrogram to extract the CoG behavior. The primary function of such multiplication is noise reduction. The CoG behavior corresponds to the relative walking velocity and walking direction. The expected output of this processing is the values with the highest intensities, which represent the CoG, while some atypical scenarios are considered for precise extraction. The algorithm may result in a rough output of the CoG vector, which is refined by an algorithm for smoothing. This smoothing has been found to be extremely advantageous as it directly reduces the influence of the swinging limbs on the CoG behavior. This primarily occurs during walking because, during the stance phase, all swinging limbs have very minor swings that manifest as small oscillations around the CoG vector.

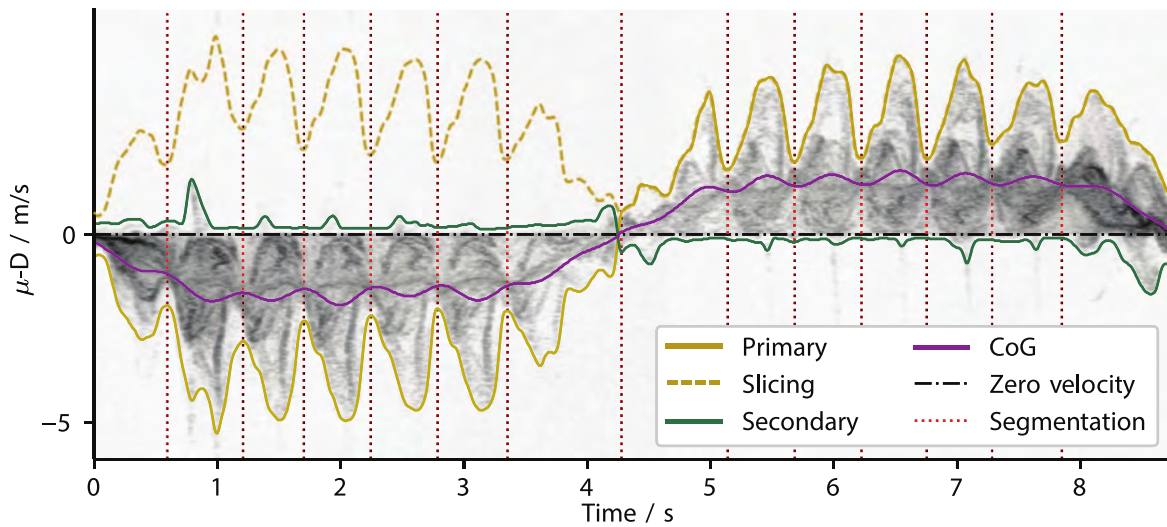


Figure 3.12: Visualization of keypoint vectors on a grayscale μ -D spectrogram slice.

Primary and Secondary Envelope

Principal factors affecting the estimation of the primary and secondary envelopes are the morphologically processed spectrogram and the direction of the target's motion. As described in Section 2.1.1, when a human walks towards the radar, the radar induces a positive velocity value and vice versa when moving away from the radar; this value can be estimated based on the CoG behavior. The algorithm is based on analyzing the binary spectrogram that has been morphologically processed. Searching for the points where a change in state occurs in order to draw the main vector for both the primary and secondary envelopes is the key concept. The depicted spectrogram is searched in a double-iterative manner, with the first iteration applied to the rows and the second to the columns. Figure 3.12 depicts an example of the extraction of both key point vectors, in which a person walks normally towards the radar during the first period, and then switches the motion direction during the second period.

3.3.3 Spectrogram Adaptive Segmentation

The spectrogram adaptive segmentation is based on peak detection through interpreting the extracted key point vectors. For the walking signatures, the frequency peaks are induced due to the feet swinging as has been discussed in Section 3.3. However, if the

human is walking away from the radar, the velocity components will be induced by the radar as peaks in the negative domain that can be bypassed through mirroring the primary key point vector around the zero axis if detected as shown in Figure 3.12. Thus, the half gait cycles represented by single occurrence of a peak can be easily monitored and sliced. A refining for the slicing technique is achieved through taking more arguments into consideration e.g., height of the peaks, the distance between consecutive peaks and the peak's threshold with respect to the neighboring point. This allows an adaptive slicing algorithm, which is more feasible as a normal human is not expected to show a fixed period performance for each single occurrence as shown in Figure 3.12.

To further refine the slicing positions for spectrogram segmentation, the peak detection algorithm accepts arguments such as the height and prominence of the peaks, their threshold to neighboring data points, and the distance between consecutive peaks, among others. For the current application, prominence and distance are utilized. The former indicates the height of a local maxima from the highest point of its surrounding local minima. While the optimal value for prominence is determined empirically, the distance between two peaks can be calculated mathematically. The frequency of successive half gait cycles is approximately 2 Hz. Therefore, one half gait cycle requires $\approx 0.5s$ to complete. This period of time can be converted to an integer number of consecutive μ -D spectrogram columns, thereby defining the minimum distance between two consecutive slicing positions. As shown in Figure 3.12, the combination of the CoG, the truthful and mirrored primary envelopes, and the peak detection algorithm with its two parameters prominence and distance permits a robust and accurate estimation of half gait cycles.

Only the μ -D spectrogram has been processed thus far. In favor of redoing the entire spectrogram processing for the μ - ω signature, the slicing positions from the μ -D signature are transferred without modification to the μ - ω signature. This not only ensures time synchronization but also saves a substantial amount of computation time, which is advantageous for a processing chain capable of real-time processing. Due to the absence of envelopes, it is impossible to isolate the μ - ω signature from its surrounding noise. As DL has demonstrated an aptitude for dealing with noise such as the studies presented in Yang et al. 2017, Abdulatif et al. 2019, and Gurbuz et al. 2020, this trade is not considered a drop.

3.4 Summary

In this chapter, the radar signal processing algorithms that are used in this thesis in different studies and phases are explained. In Section 3.1, estimating the main radar features is explained. In Section 3.2, formulating both the μ -D and the μ - ω spectrograms is discussed. In Section 3.3, the spectrogram adaptive monitoring for single occurrence extraction is discussed in detail.

4 Human Detection While Walking

As discussed in Chapter 3, the μ -D signature of a human body motion can be studied as the superposition of various body segment trajectories. Human motion can be detected, categorized, and tracked, in particular for safety applications, through the extraction of real-time limb μ -D trajectories. In this chapter, we combine two approaches to model a human walker's μ -D signatures. A decomposition feasibility study for μ -D limbs is then presented, using both range and Doppler radar features. The limb decomposition task is achieved through analyzing the micro motions behavior on the RD maps. Because of this, micro-Range (μ -R) behavior is induced by human limbs, which is analogous to the well-known μ -D signatures. The feasibility study uses a decision tree classifier to categorize human limbs into four groups (base, arms, legs, feet). The classifier has been validated, and it can separate the μ -D signatures of individual limbs from a signature of a human in motion. The results of the study is validated through relying on the Range-Doppler maps for enhanced human-robot classification. The detailed studies about extracting the limbs trajectories and the human-robot classification are published in Abdulatif et al. 2017 and Abdulatif et al. 2018.

4.1 Combined Signal Model

As proposed in Chen 2019, a simulation model can be used to investigate the body's behavior during the walking motion in detail. Based on the global human model published in Boulic et al. 1990, which is explained in Section 2.3.2, this model defines the displacement of the body segments during the walking gait cycle. This simulation provides a good indication of the structure of the μ -D signature, but it is not realistic and cannot be easily extended to additional motions because a new model must be defined for each motion. On the other hand, obtaining a true radar capable of recording relevant μ -D signatures is not always possible and involves complex processing on occasion. In response, we have developed a combined measured and simulated signal model in which the realistic motion

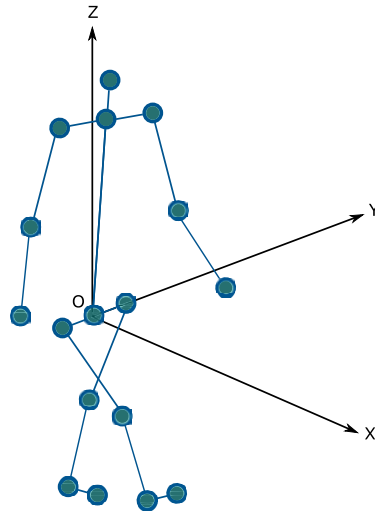


Figure 4.1: Body orientation with respect to the body-fixed system (x,y,z) .

of the human body is observed and a signal model is used to simulate the back-scattering effect for each body part. Our developed Qualisys walking model is based on the global walking model's subdivision of the human body into 16 body segments. Consequently, the segmentation of the body is defined in the subsequent section.

4.1.1 Global Human Walk Model

The global human walk model discussed in Section 2.3.2 is an empirical mathematical model used to describe the kinematics of a walking human. It is commonly utilized in radar applications for gait recognition analysis. According to Gurbuz et al. 2007, the model is utilized to classify the detected target as a person or another moving item. Using the motion trajectories described in Section 2.3.2, an animated simulation model is constructed from the global model. Therefore, the human body can be split into 17 major markings. In a three-dimensional Cartesian coordinate system, the motion trajectories can be used to determine the positions of the body's major points at any given moment. As illustrated in Figure 4.1, the local body-fixed system (x,y,z) of the walking human body has its beginning at the base of the spine, and the person is traveling in the positive x direction.

The 17 primary body markings are divided into three categories:

- The body *end-points*: five marks representing the head, two hands, and two feet.
- The body *joints*: 10 markings, which include the shoulders, elbows, hips, knees, and ankles.

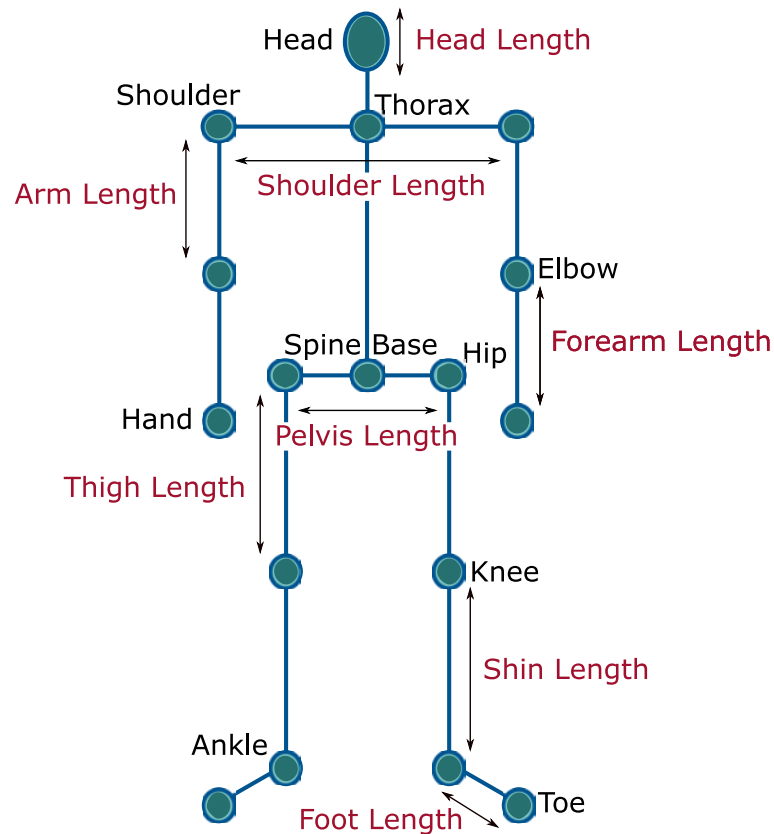


Figure 4.2: The human body segmentation based on the global human walk model.

- The body *centered-points*: two marks, the neck and the base of the spine.

The connection between these primary points might form the segments of the human body. As seen in Winter 2009, the length of each body segment is expressed as a value standardized to human height. The simulation of the human body's structure is depicted in Figure 4.2.

The human body can be viewed as a nonrigid body in motion. As described in Section 2.2.3, the motion should therefore be expressed as a set of rigid bodies moving in connection with one another for motion analysis. Each body segment must be represented by a rigid body in order to create an animated walking human model. Consequently, the human walking simulation is based on the rigid body motion described in Section 2.2 in conjunction with the trajectory functions of Boulic et al. 1990 described in Section 2.2. In the following section, the various depictions of body segments will be discussed.

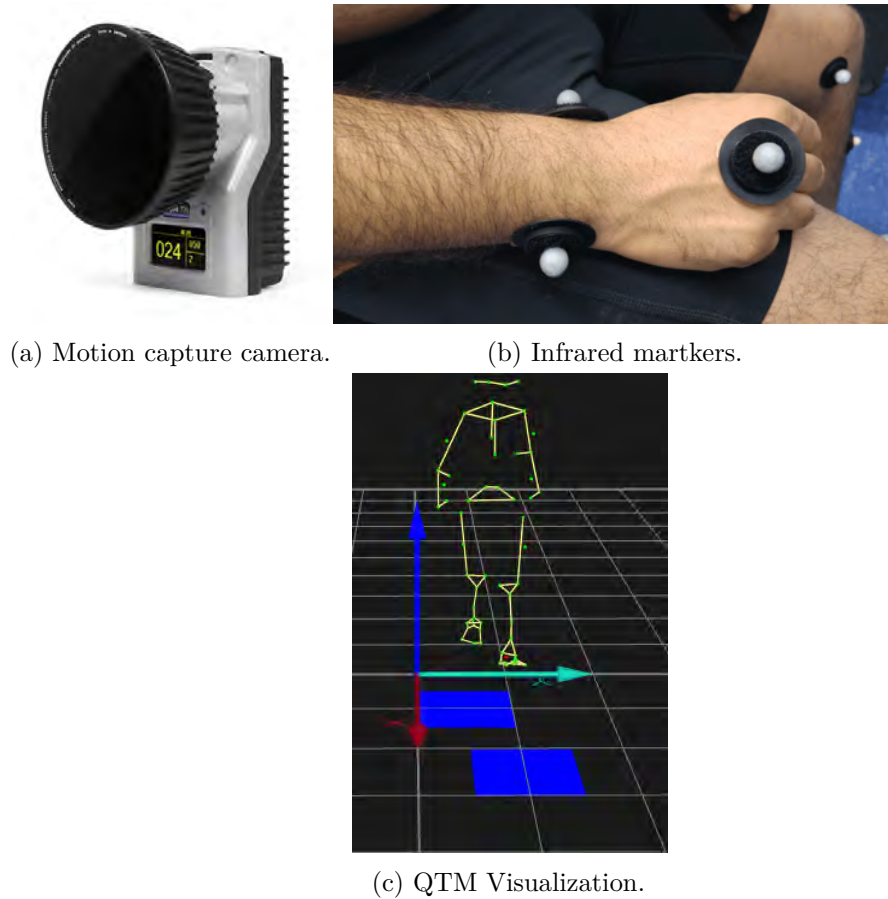


Figure 4.3: Qualisys tools used for motion capture.

4.1.2 Qualisys Walking Model

Qualisys is an optical motion capture technology that is used to precisely track the positions and velocities of moving objects. The system consists of three primary components: motion capture cameras, software required for motion analysis, and infrared markers put on the objects to be tracked, as depicted in Figure 4.3 and guided by Qualisys 2015. Qualisys is utilized in numerous application sectors that need the capture of moving objects with realism and precision. In the science of bio-mechanics, for instance, capturing human motion is used to detect and prevent accidents. Another illustration is Qualisys sports performance, which provides a 360° analysis of the player, and more information about it is presented in Qualisys 2015. The precise capture of an athlete's motion is utilized to analyze his performance by providing extensive information about his movements.

As illustrated in Figure 4.3, the Qualisys human walk model was created so that infrared markers are placed on the 17 major body segments recommended by Boulic and Thalmann.

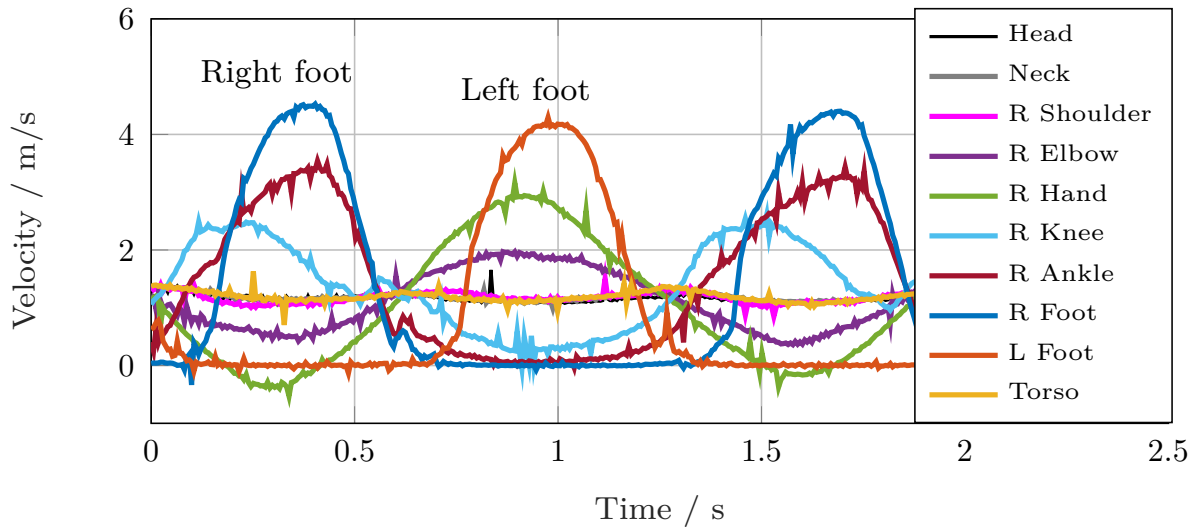


Figure 4.4: Velocity trajectories captured by Qualisys.

The developed chamber for capturing human movements consists of 12 cameras. The cameras are evenly spaced along the chamber's 15 m length. Using the software from Qualisys 2015, the infrared marker becomes a legitimate location point when it is caught by at least two cameras.

The software used for reading the captured positions of the infrared markers is Qualisys Track Manager (QTM) that is demonstrated by Qualisys 2015. The captured positions are measured with respect to one reference origin placed in the middle of the measurement chamber. According to the Qualisys model, the 12 motion trajectories that were explained in Section 2.3 are replaced in this simulation with a real motion behavior of each body segment. Therefore, the analyzed μ -D and μ -R signatures is realistic and dependent on the captured walking style. The normalized lengths of the body parts with respect to the human height is also replaced with the real lengths of the different body segments. The velocity trajectory of each body part is determined based on the captured positions of the infrared markers as shown in Figure 4.4.

4.1.3 Human Body Segment Representation

The segments of the human body are depicted as rigid bodies. The most common representational geometrical shapes are spheres, cylinders, and ellipsoids as shown in Figure 4.5. The depiction of body parts as recognizable geometric shapes facilitates radar

measurements. These representations, along with the motion trajectories defined in Boulic et al. 1990, facilitate the construction of a simulation model for radar analysis, as described in van Dorp & Groen 2003. For radar analysis, the RCS described in the following section should be utilized.

Radar Cross Section (RCS) Approximation Methods

The scattered waves can be separated into two distinct polarization types. The first portion of the waves will have the same polarization as the radar antenna, while the second portion will be orthogonal to it as illustrated in Mahafza 2005. The object's RCS is estimated based on the intensity of the signal portion that has the same polarization as the radar antenna. If the power density of the incident signal is P_{inc} , then the reflected power from a target at a distance R from the radar can be defined according to:

$$P_{ref} = \sigma P_{inc} , \quad (4.1)$$

where σ represents the RCS of the object. The radar equation presented in Skolnik 2001 can be used to express the relationship between the radar's received power and the object's RCS according to:

$$P_{object} = \frac{P_{Tx} G^2 \lambda^2 \sigma}{(4\pi)^3 R^4 L} , \quad (4.2)$$

where P_{Tx} is the average of the transmitted power, G is the antenna gain, and L refers to the losses. Using prediction methods, the RCS of known geometries can be precisely estimated. Due to the scattering effect, however, these prediction techniques must be solved using complex mathematical techniques and boundary conditions. Therefore, approximated RCS prediction methods can be used in computer simulation analysis. The method of approximation prediction depends on the detected target's shape. If the detected target shape is an ellipsoid with axes a , b , and c in the x , y , and z directions, respectively. The expression for the ellipsoid RCS approximation method is according to:

$$\sigma_{ellipsoid} = \frac{\pi a^2 b^2 c^2}{(a^2 \sin^2(\theta) \cos^2(\phi) + b^2 \sin^2(\theta) \sin^2(\phi) + c^2 \cos^2(\theta))^2} . \quad (4.3)$$

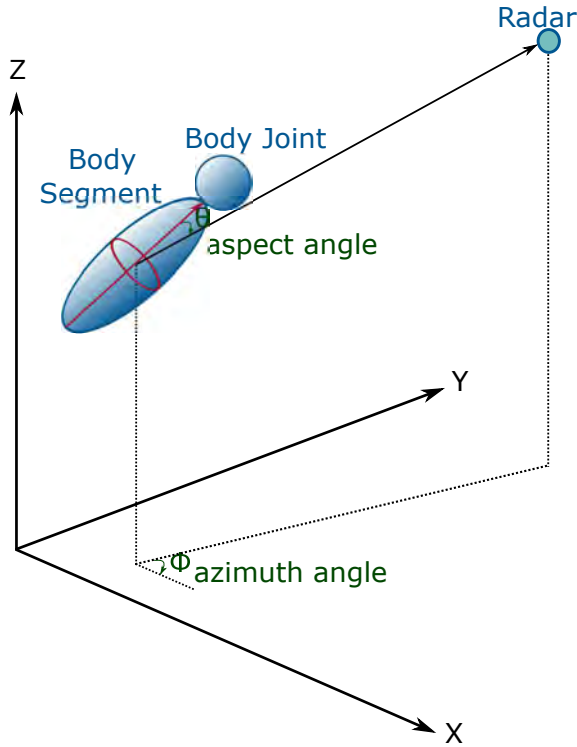


Figure 4.5: The body segment orientation with respect to the radar.

The sphere RCS approximation method corresponds to Equation (4.3) with $a = b = c$ and can be expressed according to:

$$\sigma_{sphere} = \pi c^2 . \quad (4.4)$$

A circular cylinder with a radius of r is the third shape that can be used to represent a body segment. The RCS method of approximation can be stated according to:

$$\sigma_{cylinder} = \frac{\pi b'^4 c'^2}{(a'^2 \sin^2(\theta) + c'^2 \cos^2(\theta))^2} . \quad (4.5)$$

The segments of the human body can be represented by either an ellipsoid or a cylinder, while the connecting points between them are represented as spheres. The ellipsoidal representation of body segments in the simulation model is more realistic than the cylindrical representation as shown in Figure 4.5. Due to the normal orientation with respect to the radar signal, a body segment represented as a cylinder produces stronger reflections. According to van Dorp & Groen 2003, the smoother reflections that result from the ellipsoidal representation of the body segment presents the simulation in a realistic

way, when compared to the real measurements. As shown in Figure 4.5, the two angles presented in Equation (4.3) describe the ellipsoid orientation with respect to the radar.

The simulation's RCS approximation method depends solely on the ellipsoidal body segment. By referencing the trajectory functions presented in Section 2.2, the body joint is viewed as a point that is used to obtain the various positions and rotations actions with respect to time. The aspect angle (θ) and azimuth angle (ϕ) are calculated using both the radar location and the positions of each body segment according to:

$$\theta = \mathbf{arctan}\left(\frac{\sqrt{x^2 + y^2}}{z}\right), \quad (4.6)$$

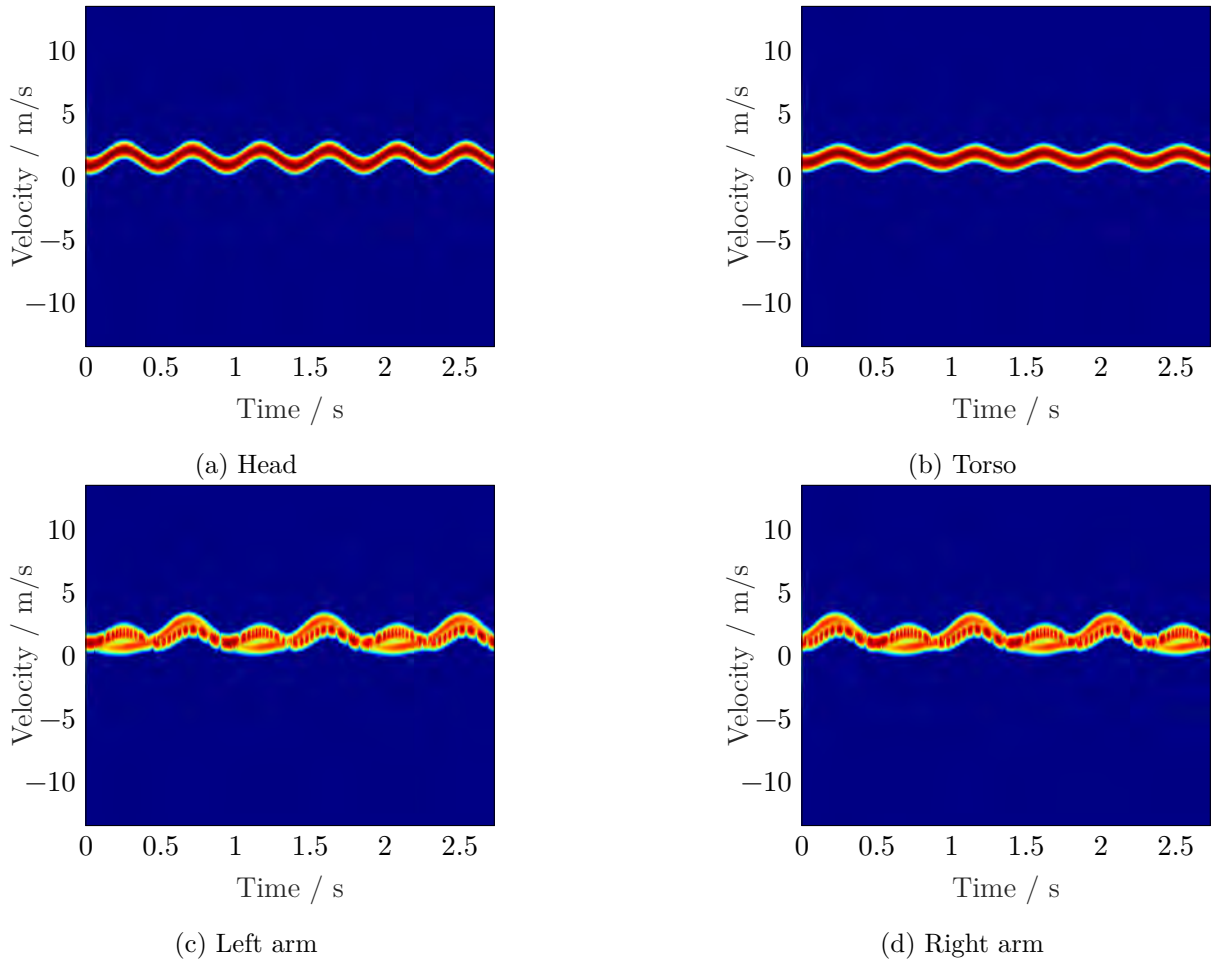
$$\phi = \mathbf{arctan}\left(\frac{x}{y}\right), \quad (4.7)$$

where $\sqrt{x^2 + y^2}$ represents the projection of the range between the rigid body and the radar in the (X-Y) plane, whereas z represents the height of the body segment from the (X-Y) plane. On the basis of the human body structure depicted in Figure 4.2, the length of each body segment can be estimated. Therefore, all of the parameters necessary for RCS approximation as shown in Equation (4.3) are available. The subsequent section explains the μ -D and μ -R signatures analyzed as a result of the simulation model described.

4.2 Radar Features Interpretation

The presented signal model is capable of estimating μ -D signatures based on both the global and Qualisys human walking models. This section's interpretations describe a person walking toward or away from the radar along the LOS. The μ -D analysis can make use of the RCS approximations described in Equation (4.3), as the amplitude and phase of the scattering effect are estimated by the square root of the RCS, as explained in van Dorp & Groen 2003. At each instance of time, the signal model describing the RCS approximation method is expressed as A_i for each body part represented as an ellipsoid.

$$A_i = \sqrt{\sigma_i} e^{\frac{-j4\pi R_i}{\lambda}} = \frac{\sqrt{\pi abc}}{a^2 \sin^2(\theta_i) \cos^2(\phi_i) + b^2 \sin^2(\theta_i) \sin^2(\phi_i) + c'^2 \cos^2(\theta_i)} e^{\frac{-j4\pi R_i}{\lambda}}. \quad (4.8)$$

Figure 4.6: μ -D signatures of upper body parts.

As shown in Figures 4.6, and 4.7 the μ -D signature is periodic, due to the periodic behavior of each body part through the walking activity. The μ -D signatures of a walking human with relative walking velocity v_{wr} can be separated into those caused by less swinging body parts and those caused by swinging limbs. The μ -D signatures of the less mobile body parts, such as the head and torso, can be utilized to detect the v_{wr} as shown in Figures 4.6a, and 4.6b. The μ -D signatures of the swinging arms and legs can indicate that some parts of the human body have velocities greater than v_{wr} , which can be used to adjust the radar parameters based on the maximum velocity detected as shown in Figures 4.6c, 4.6d, and 4.7. However, none of the μ -D signatures presented in Figure 4.6 can fully describe the human walking locomotion alone. Although the μ -D signatures of swinging arms in Figures 4.6c and 4.6d, respectively, demonstrate the swinging effect of the

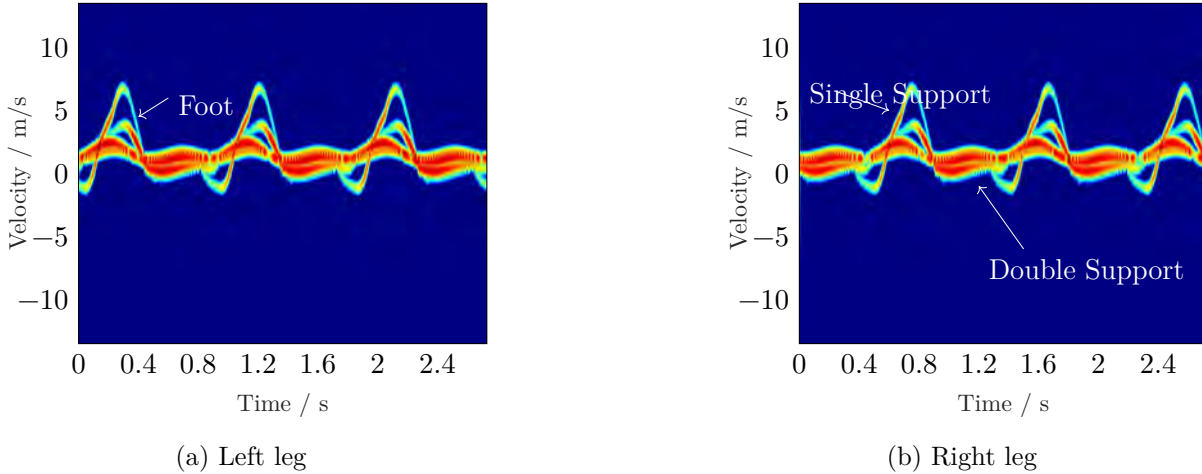


Figure 4.7: μ -D signatures of lower body parts.

arms, this is insufficient because the feet are expected to demonstrate a more pronounced swinging effect as shown in Figure 4.7.

The μ -D signature of the leg from the hip to the foot is shown in Figure 4.7. The states of the walking gait cycle are evident in the μ -D signatures of the legs in Figure 4.7. During walking, the human body alternates between the single support state and the double support state. It is possible to analyze the transition between the single support state, in which only one foot is swinging and the other is touching the ground, and the double support state, in which both feet are touching the ground. The maximum velocity component results from the unique motion of the feet, as depicted in Figure 4.7.

Even though the arms and legs are swinging, their μ -D signatures reveal that they induce the relative walking velocity v_{wr} the majority of the time. The detected μ -D signatures of a human walker consist of the superposition of all body parts involved in walking. Due to the fact that all body parts induce the same μ -D signatures, the decomposition of body segments is exceedingly complex. The primary velocity component detected with the most power is the human walking velocity. The arm and leg swinging effects are analyzed as periodic velocity components oscillating around the human walking velocity. As shown in Figure 4.8, the different power values of the velocity components can be clearly analyzed, as each body part induces velocity components that differ from the other body parts. Accordingly, relying only on the μ -D signatures for extracting the limbs trajectories is not enough and the range feature is interpreted in the next section.

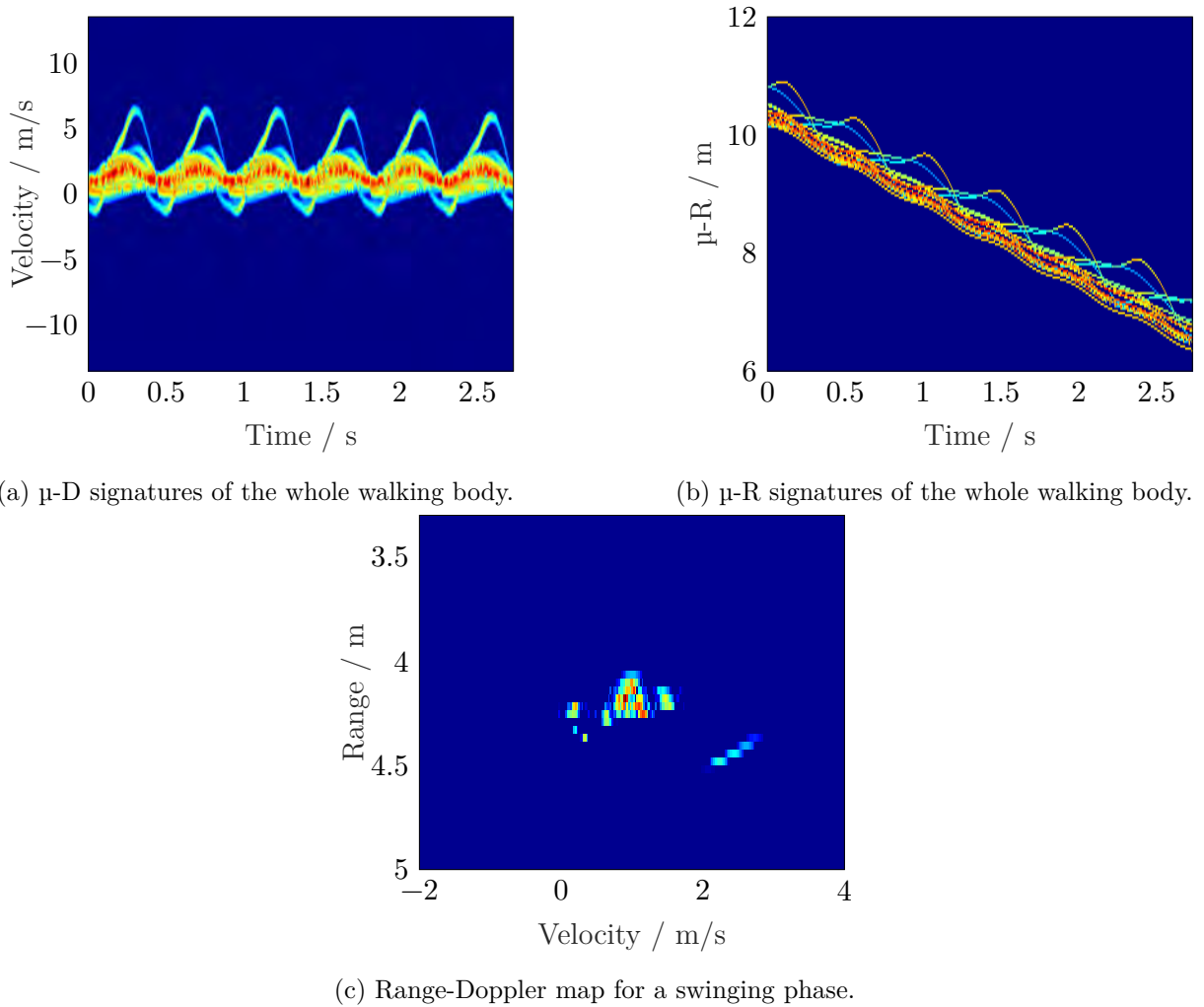


Figure 4.8: Micro motions interpretations on the main radar features

The signal model presented in Equation (4.8), is used for interpreting the μ -R behavior through walking. The μ -R signature induced due to each body part can be seen in Figure 4.8b, in which a periodic micro motion behavior can be seen clearly. Since the range is usually interpreted in the radar, while being mapped to the velocity in a form of RD map, the signal model was adopted to get an RD map that is shown in Figure 4.8c. The RD map is derived through the signal processing operation that is described in Section 3.1, through applying a second FFT through the derived range profiles through time after parameterizing the model with a suitable N_p . Such analysis can be applied to both the global and the Qualisys walking models. As a conclusion, for an enhanced human detection, both the μ -D and μ -R signatures will be used for the limbs decomposition application that is presented in Section 4.3, and the human-robot differentiation that is presented in

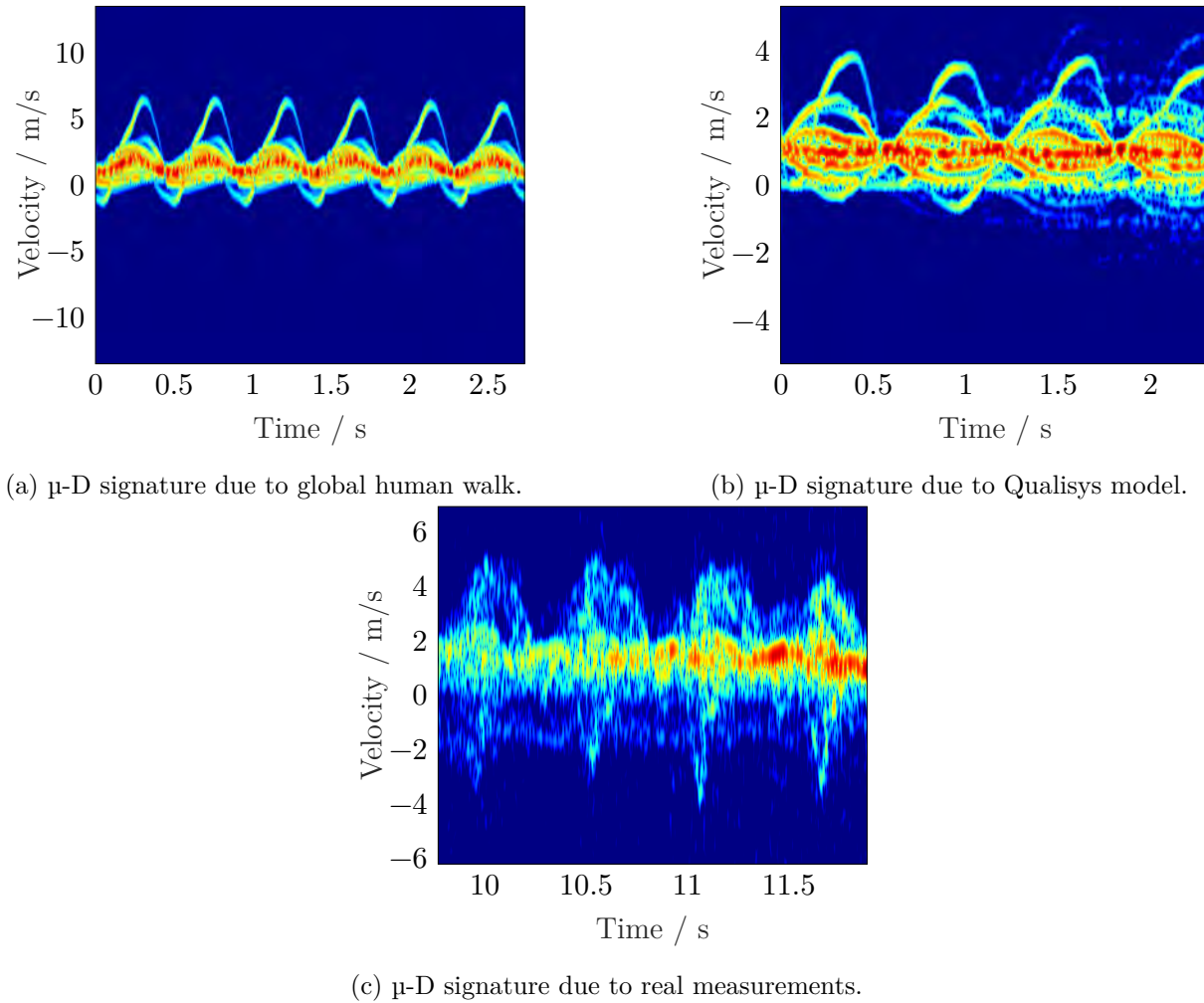


Figure 4.9: Comparison between μ -D signatures of simulations and real measurements.

Section 4.4. A comparison between the models is presented in the following section to test how our novel combined signal model is realistic, when compared to real measurements.

4.2.1 Comparison Between Simulations and Real Measurements

A comparison is shown in Figure 4.9 between the simulated μ -D signatures of the two models and a measured μ -D signature from a real radar to determine which simulation model is more suitable for classification. In the previous section, the μ -R was presented as a crucial parameter to be included in the human detection procedure. Thus, an FMCW radar is used to compare the μ -D signature in order to be capable of capturing the μ -r signatures if needed. The tested human for motion capture in the Qualisys model is the same human walking towards the radar, so that the same walking style can be analyzed via simulation

and real measurement. Figure 4.4 depicts a human walking at $v_{wr} \approx 1.5m/s$ with a maximum foot velocity of $\approx 4.5m/s$. The radar is parameterized based on Equations (2.14), (2.15), and (2.16). The resulted parameterization is presented in Table 4.1:

As a conclusion from Figure 4.9, the global human walk model is feasible to analyze the concept of the μ -D signatures. However, it cannot be considered as a generalized model as it only relies the human height and the walking relative velocity. On the other hand, the Qualisys walking model is showing a more realistic signature when compared to the real-radar measurement. The Qualisys model also is not lacking the sense of the walking style as it is capturing the real motion of the human. Additionally, the Qualisys model is not limited to only walking motion, it can be extended to other activities. For these reasons, the combined model based on combining the Qualisys walking model and the simulated signal model is utilized in our feasibility study that is published in Abdulatif et al. 2017, and discussed briefly in the following section.

4.3 Decomposing Limbs Trajectories

As has been discussed in previous sections, the μ -D signature is captured by the radar as superimposed harmonics induced due to all body limbs. However, the real-time extraction of those motion trajectories is considered vital for many fields e.g., hand detection for hazard industrial areas as has been investigated by different studies such as Yu et al. 2020; Wang et al. 2020, and Ho & Gader 2002. Other example is presented by Seifert et al. 2018a, where the walking irregularity is detected through analyzing the feet reaction. However, this is still not considered a pure extraction as the feet trajectory can be monitored through envelope detection of the captured μ -D signature. Moreover, the perception systems that are used for capturing the limbs motion e.g., Qualisys are foreseen more as

Table 4.1: Radar module parametrization.

Radar Parametrization		Attributes	
Carrier frequency f_o	25 GHz	Maximum range (R_{max})	19 m
Bandwidth (B)	0.2 GHz	Range resolution (R_{res})	7.75 cm
Chirp duration (T_p)	0.5 ms		
Samples per chirp (N_s)	256	v_{res}	0.1 m/s
Chirps per frame (N_p)	128		

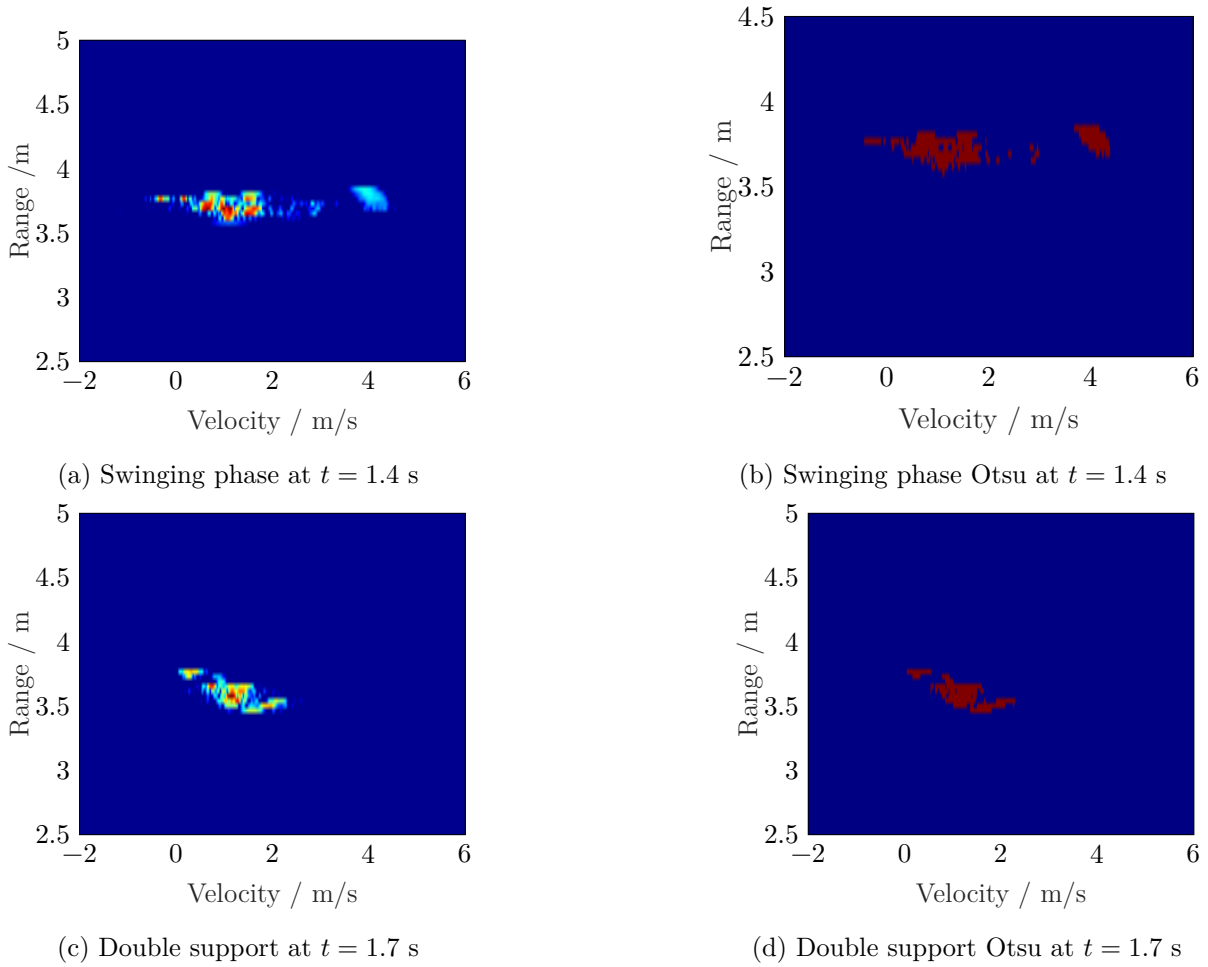


Figure 4.10: Examples for the RD maps after power removal.

laboratory devices and they lack the freedom of motion as normally, attaching reflectors or wearable sensors to the body is required. Thus, a feasibility study about real-time limbs decomposition is presented in our publication Abdulatif et al. 2017 based on the RD maps extracted from the combined signal model that is created and briefly described in this section.

4.3.1 Methodology

The radar combined model is parameterized based on Table 4.1, and real humans of height $\approx 1.7 - 1.9m$. The model is utilized afterwards, where RD maps for the walking motion are derived as shown in Figure 4.10. The color information in the μ -D spectrogram shows the radar's received power as a result of different parts of the body moving. However, limb decomposition is not something that should be viewed in terms of power. For one

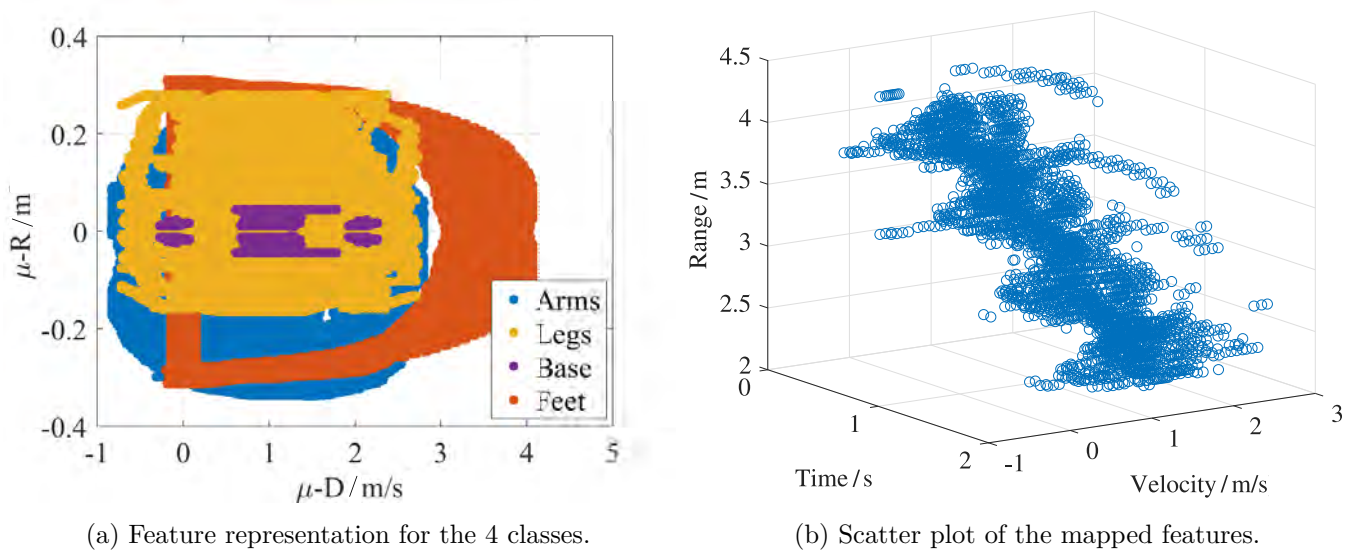


Figure 4.11: Representation of μ -D and μ -R features induced due to Qualisys model.

thing, the μ -D value at any given time is the sum of reflected signals from different parts of the body. Moreover, the changing human position relative to the radar will result in a power fading effect in actual radar measurements. Therefore, power will not factor into the classification procedure.

The power is removed in two stages. First, the Gamma-transformation technique described in Gonzalez 2009 is applied to the RD map to reduce the noise effect. Afterwards, the essential μ -D harmonics are then extracted based on the Otsu's thresholding technique presented in Otsu 1979. Figure 4.10 demonstrates that both μ -R and μ -D functionality are maintained upon power loss. The shift in range that results from the human body's constant motion is readily apparent. Both the range and the velocity of the swing can be affected by the movement of the different limbs.

The body is divided into four classes, and the scatter plots depicted for both the μ -D and the μ -R is shown in Figure 4.11. The classes are divided based on the following:

- Base: minimum velocity and swings induced due to torso, head, and neck.
- Arms: higher swinging effect, which are elbows, upper and lower arms.
- Legs: this class incorporates the knee, lower, and upper legs.
- Feet: the highest swinging inducing the highest velocity component.

As a preprocessing step, the mean is removed from the μ -R data at each time step in order to extract only the swing effect without the linear regression behavior depicted in

Figure 4.8b. Our proposed method is independent of time, so features from the entire duration of the experiment are used to train the classifier. Due to body symmetry, the right and left body parts with corresponding μ -D and μ -R signatures will be very similar. This will result in undetermined left and right body part differentiation based solely on μ -D and μ -R. As shown in Figure 4.7, only the velocity phase shifts distinguish left from right body parts, as evidenced by the half-cycle difference in the feet’s motion during the gait cycle. In practice, however, the starting body side of the motion is unknown. Therefore, it is difficult to identify each side independently without prior knowledge. In addition, in previous decomposition techniques, the first moving side was treated as prior knowledge, and a side flip was performed every half cycle as explained in Raj et al. 2009 and Raj et al. 2010. Consequently, the left and right body parts will be grouped together.

4.3.2 Results

Each subject was asked to walk for three minutes so that features could be collected and labeled before classification for training and validation. The dataset was split into 75% for training and 25% for validation. The confusion matrix in Table 4.2 is obtained by training a decision tree classifier. The legs and arms classes are confused at times due to an overlap in the μ -D and μ -R planes, as depicted in Figure 4.11. However, the feet and legs and arms classes are confused with equal frequency. The base has the lowest performance due to the presence of low μ -D and μ -R components, which are also present in the legs and arms classes.

Table 4.2: Limbs decomposition confusion matrix.

True/Predicted	Arms	Feet	Legs	Base
Arms	65%	1%	32%	2%
Feet	14%	73%	13%	0%
Legs	19%	1%	79%	1%
Base	37%	0%	37%	26%

4.3.3 Conclusion

In this section, a feasibility study is presented about using the captured μ -D for limbs trajectories decomposition. For such application, our combined signal model presented in Section 4.1 is utilized. The human body is divided into four main groups based on the swinging intensity through the walking gait cycle, in which two main points are concluded. First, the combined signal model is showing a realistic behavior when compared to real measurements. The combined signal model has been created by us at that time to the unavailability of radar modules that are capable of capturing meaningful μ -D signatures at that time. It opens big room for researchers who suffer from the same problem or who want to simulate some scenarios in a realistic way. It is not used afterwards in the thesis work as both problems weren't applied for us. There were our trials for simulating the radar μ -D signatures that were mentioned, and discussed in Section 4.1. However, using other techniques for capturing the human motion such as Kinect-based model, presented by Erol & Gurbuz 2015, will not offer the same accuracy compared to the utilized Qualisys system. Additionally, it is limited only to human motion and cannot be extended to other moving objects such as horses. Second, for an enhanced human detection, the mapped μ -D and μ -R signature have to be used. This has another advantage of adding a real-time performance as capturing single RD map usually needs less than half a minute as will be discussed in the next section. Accordingly, the RD maps are investigated as the main features for human-robot classification application that is discussed in the next section. Nevertheless, our use-case scenario required a real radar; thus, the next study is based on a real FMCW radar in order to be capable of capturing both the Doppler and range radar features.

4.4 Human-Robot Classification

Target classification based on their unique μ -D and μ -R signatures will be possible for a wide variety of moving targets. Gait recognition, security, and monitoring are only a few examples of the many potential applications of this kind of categorization. In this section, our study published in Abdulatif et al. 2018 is discussed briefly in which the robustness and feasibility of the utilized RD maps for human limbs decomposition is tested for a

real-life use-case scenario of human-robot classification. The use-case scenario required integrating a radar in a production line in a factory, in which the area beside the machine is always required to be human-free, while if a mobile robot approaches the machine for any reason, the work flow shouldn't stop. For such task, a 25 GHz SISO radar module is used and parameterized as shown in Table 4.1. Relying on the RD maps is fulfilling the real-time constraint of our desired use-case scenario. A detailed analysis is provided that contrasts conventional learning techniques with those that rely on manually extracted features, ensemble classifiers, and deep learning. Without performing any additional feature extraction, ensemble classifiers like gradient boosting and random forest are fed reconstructed range and velocity profiles. Finally, a DCNN is employed, and RD maps are fed directly into the network. DCNN identifies humans and robots on a single RD map with an accuracy of $\approx 99\%$.

As discussed in Section 2.5, classification in real time was not addressed by earlier methods. The proposed classifier training dataset is also quite sizable, numbering in the hundreds of examples. Research shows that when training a DCNN, more training data leads to better results as explained in Goodfellow et al. 2016. Therefore, this amount is insufficient for designing robust classifiers, particularly in DCNN. Our presented study addresses the human-robot classification problem using FMCW radar RD maps that can be computed in much shorter time intervals ($\approx 0.1s$) as been utilized in Jokanovic & Amin 2017. Consequently, the real-time constraint can be satisfied, and the number of datasets used to train deep models has increased from hundreds to thousands. Finally, a comparison of various learning strategies is presented. The user manual including the datasheet for the Ancortek radar module used for this classification task is described in appendix in Section A.2.

4.4.1 Methodology

Dataset Formulation

The whole procedure of building up the dataset is explained briefly in Abdulatif et al. 2018. To collect the data, measurements were taken on 10 test subjects of varying heights and genders while they walked. There are innumerable types of machines referred to as

robots. Due to the proposed task’s focus on industry, two industrial robots are included for data collection and testing. The first robot is a mobile robot assistant (Care-O-bot) created by the robotics division at Fraunhofer IPA that can be referred to in Koay et al. 2014. The second robot is a six-axis robotic arm created by Stäubli that can be referred to at Robots 2017. During each experiment, the subjects moved at random aspect angles within the radar detectable region. To avoid the extreme radial velocity fading effects described in Bartsch et al. 2012, however, the case of exact lateral motion was omitted.

Each data sample is labeled with the current target class during data collection, which is either a human or robot. This labeling is required for supervised learning in order for models to understand the standard solution and consequently learn the correct model parameters. The collected data and associated labels are separated into two subsets. The first subset will be utilized for model construction. The training set is used to learn the model parameters, while the validation set is used to tune model hyper-parameters, such as the number of hidden layers and neurons in each layer of a neural network. The second subset (test set) is used to evaluate the generalization performance of the final model by comparing model predictions to the actual class labels.

The test set is constructed using the hold-out technique. For a hold-out split, the entire dataset D is divided into two non-overlapping subsets. One is utilized as the training set S and the other as the test set T . It can be expressed mathematically as $D = S \cup T, S \cap T = \emptyset$. The models are trained on S and evaluated on T to determine their generalization performance. Due to the relatively high sampling frequency, contiguous samples from a single measurement can be very similar. Therefore, a test on randomly selected samples from all measurements cannot accurately reflect the generalization performance of the model. Consequently, separate experiments of each object type are chosen to construct the test set, with each experiment comprising up to twenty percent of the total data for this object type. Thus, the entire test set comprises approximately 20% of all collected data.

RD Map Investigation

The RD map interpretation is discussed briefly in Abdulatif et al. 2018. As has been explained in Section 2.3, the human walking motion is described as a succession of periodic cycles, with two phases constituting a complete gait cycle. The initial phase (swinging) in

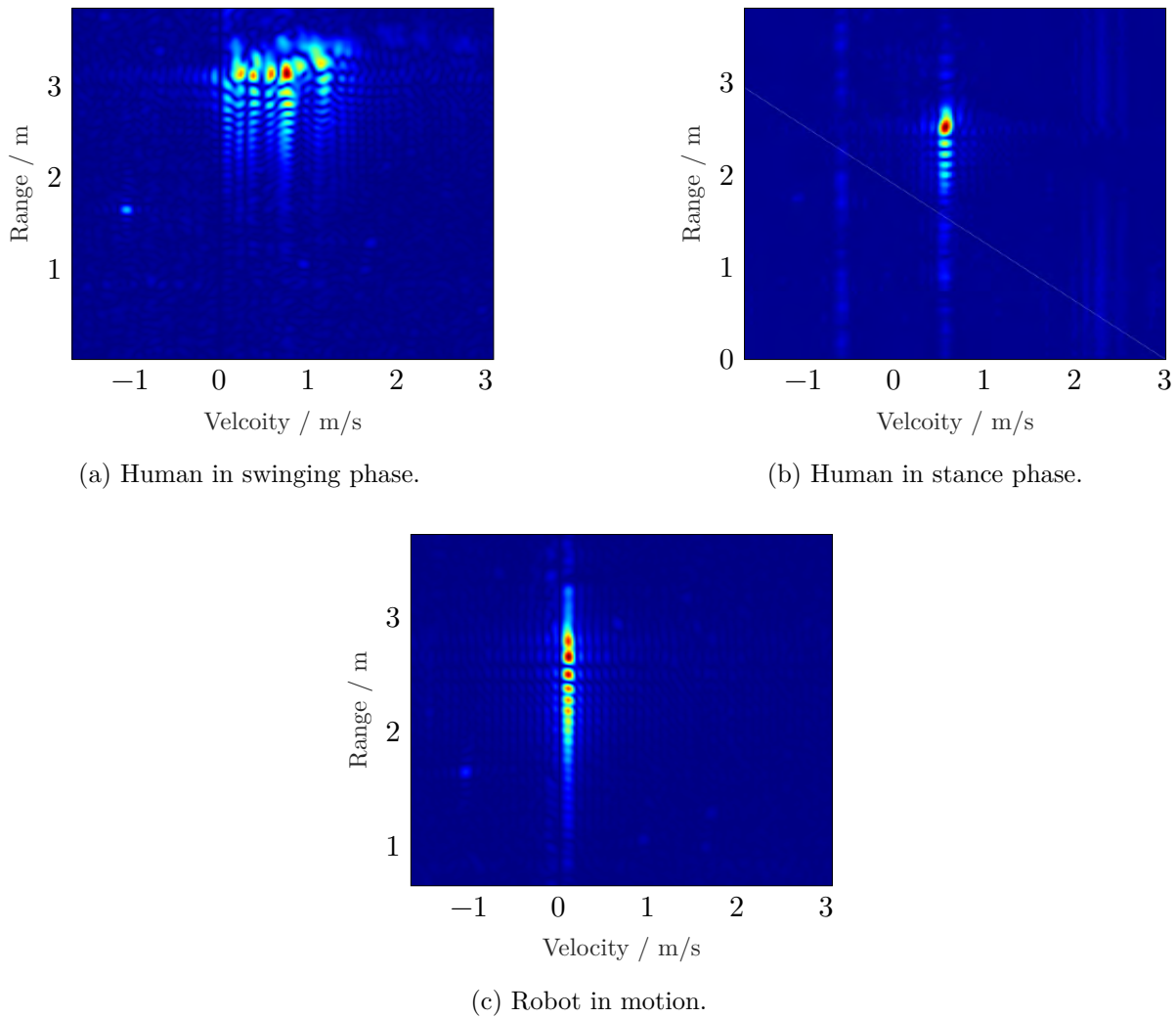


Figure 4.12: Comparison between RD maps induced due to walking human and moving robot.

which only one foot is in motion and the other is on the ground. As shown in Figure 4.12a, the human appears on the RD map as a broad distribution along both RD axes during this phase. This distribution represents a range of velocities due to the bulk moving body parts (torso and head) as well as the swinging effect of various body limbs (arms, legs and feet). In the second phase (stance), no swinging limbs are observed and only mass motion is observed. This phase accounts for 60% of the gait cycle and is the predominant phase. Comparing the robot motion and human stance phase depicted in Figures 4.12b and 4.12c, respectively. It is evident that the stance phase poses the greatest difficulty in our human-robot differentiation task, as both RD maps have a narrow horizontal distribution that is very similar.

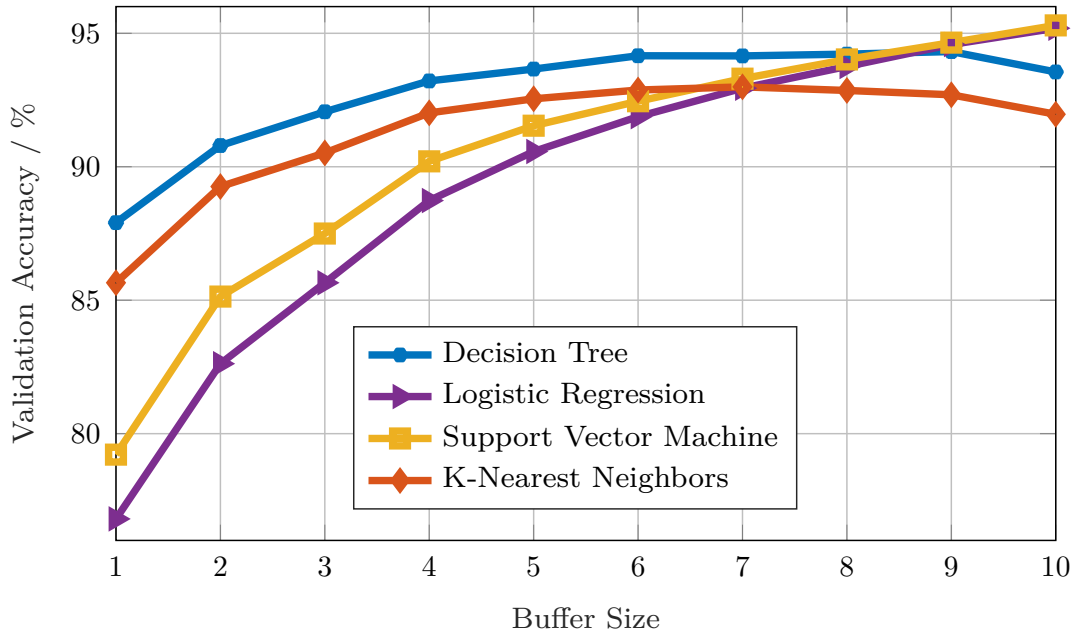


Figure 4.13: Accuracy curves of conventional learning methods.

Conventional Learning on Handcrafted Features

After acquiring the dataset as described in the preceding section, we evaluated several conventional machine learning techniques using features manually extracted from RD maps. Prior to feature extraction, the Otsu method is used as an unsupervised image thresholding on RD maps to separate the RD data corresponding to the target from the background that is described in Liao et al. 2001. Utilizing this technique, the original continuous RGB values in each RD map are quantized into 10 discrete levels. The five lowest levels out of ten are disregarded because they can be considered noise. The performance of the Otsu algorithm is discussed in Section 4.3, and can be observed in Figure 4.10.

Afterwards, features are extracted from each RD map to represent the desired classification between human and robot. To prevent overfitting, the features are also chosen so that they do not reflect an exact velocity or range. Thus, the extracted features only take into account the distribution across RD maps. The distribution can be represented as the difference between the maximum and minimum detected values in both RD dimensions, which are representing the two features for both range and velocity profiles. In addition, features such as the variance in velocities σ_v^2 and in ranges σ_R^2 can be viewed as the polynomial features of the standard deviation in velocities σ_v and ranges σ_R with a degree of 2 as described in Zhou 2021. In addition, the covariance between range and Doppler

values is taken into account as a feature. Finally, we have seven features that have been utilized with various conventional machine learning techniques.

Several classical machine learning methods were implemented, trained, and evaluated with the feature data following feature extraction. Decision Tree, Logistic Regression, SVM, and K-Nearest Neighbors (K-NN) are employed. However, depending on the characteristics of a single RD map, the performance of each method is subpar. This resulted in the use of a sequence of feature vectors compiled from several successive RD maps (a so-called sample buffer). Each sequence of feature vectors is concatenated with feature vectors extracted from each sample in a buffer. Therefore, the number of utilized features will increase from 7 (with a buffer size of 1) to 70. (in the case of a buffer of size 10). As demonstrated in Figure 4.13, the classification accuracies of all methods increase as the buffer size grows. The highest test accuracy using SVM is 95.3% at a buffer size of 10 bytes. However, such a large buffer size results in a latency of more than one second, causing issues in safety-critical real-time applications. It is also can be noticed in Figure 4.13, a drop in the accuracy performance for both the SVM and the K-NN approaches within the ninth and the tenth sample; however, it is only a redundant drop, which is less than 2%.

Ensemble Trees with Restructured RD Data

Using conventional machine learning methods with hand-crafted features has two major drawbacks. On the one hand, manually constructing features from raw data is time consuming and requires domain expertise. On the other hand, as discussed in the preceding section, adequate performance can only be achieved at the expense of a large buffer size, which is directly proportional to the inference latency.

In general, the predictive power obtained by combining multiple models is superior to using a single model alone. Ensemble learning, a family of machine learning methods that perform the learning task by constructing a group of individual learners and combining their outputs, is an illustration of this concept.

Always present in machine learning is the bias-variability trade-off. Bias and variance are two forms of predictive model error. A simple model is susceptible to underfitting the training data, resulting in a model with high bias and low variance. In contrast, complex models have a tendency to overfit the training set, resulting in low bias but high variance.

From this conundrum, two contradictory strategies for reducing prediction error can be derived: reducing the variance of complex models and reducing the bias of simple models.

Krogh and Vedelsby demonstrated in Krogh & Vedelsby 1994 that the error of an ensemble can be calculated according to:

$$\widehat{E} = \overline{E} - D, \quad (4.9)$$

where \overline{E} represents the average error of all individual learners and D measures the diversity of individual learners. Diverse learners are required to reduce the predictive error of an ensemble model. Common ensemble learning techniques include bagging and boosting. To accommodate the diversity of individual learners, bagging and boosting rely on diverse training sets for individual learner instruction. The distinction lies in how the different training sets are acquired.

Bagging comes from the idea of bootstrap aggregating which reduces prediction error by decreasing the variance of complex individual learners. Bagging entails constructing distinct training sets by randomly sampling the entire dataset with replacement (bootstrap sample). Following the construction of K predefined training sets, K learners will be trained on these K training sets. Therefore, the individual learners can be generated in parallel, and there is no strong interdependence between them. The hypothesis of the entire ensemble can be obtained by averaging the hypotheses of all K individual learners without weighting. Consequently, the estimated bias does not change, whereas the estimated variance decreases by a factor of $K - 1$ as explained in Wang et al. 2016. The utilized technique in this thesis work is the random forest approach.

Boosting enhances the performance of predictions by reducing the bias of weak individual learners. It generates diverse training sets by assigning weights iteratively to data samples. Each data sample's weight is proportional to how well this data sample can be predicted by the current ensemble. By doing so, the distribution of training data is modified, resulting in a greater focus on the data portion that has not been accurately predicted thus far.

Weak learners are simple models that can acquire training data with low accuracy (< 50%) with a high bias and a low variance. The study presented in Schapire 1990 that a group of weak learners can be combined to form a strong ensemble capable of arbitrarily high training accuracy. Boosting employs this concept and sequentially constructs a

group of weak learners. The current weighted training set obtained in the manner described previously is used to induce each individual weak learner. After generating the predetermined number of individual K learners, the ensemble hypothesis is obtained through a weighted vote of all predictions of weak learners. The bias and variance of an ensemble will gradually decrease as a result of the combination of incompetent learners who concentrate more on incorrectly predicted samples. The utilized technique in this thesis work is the gradient boosting approach.

Before feeding the RD map to the classifiers, it must be restructured. Before feeding the RD map into the ensemble classifiers, it must be flattened to a one-dimensional feature vector because it has a two-dimensional structure of 512×512 . To generate feature vectors from RD maps, the elements of each RD map are averaged along both dimensions. This yields two 512-dimensional vectors, one representing the Doppler profile (row vector) and the other, the range profiles (column vector). The hand crafted features are not applicable to be included in this case compared to the classical learning techniques as our main aspect was to reduce the computation effort and time in order always to be offering the best real-time performance.

Second, to ensure a classification based on the target's motion dynamics, as opposed to absolute values such as velocity or range. The information pertaining to absolute measurement values should be eliminated from the data, leaving only the RD distributions. We propose a method to eliminate such information containing concrete target motion parameters as follows: in both Doppler and range profile vectors, the elements corresponding to high power areas are shifted to the vectors centers. Since the positions of these elements in both the Doppler profile and range profile correspond to the target's absolute velocity and range, respectively. By shifting the large-valued components of both vectors to the center, information regarding the velocity and range of the target is concealed. This algorithm obtains a weights vector for both velocity and range profiles by normalizing their power values to the sum of all their elements, respectively. Then, a weighted average is applied to the velocity and range indices based on their respective weights in order to obtain an average value close to the high power region of each profile. Consequently, the profiles are shifted from the calculated indices to the centers of vectors. In order to reduce complexity, 128 elements are removed from both ends of the shifted vectors, resulting in

256 elements per profile. Both restructured Doppler and range profiles are concatenated into a single 512-dimensional feature vector following dimension reduction.

CNN Network Architecture and Training

CNNs have proven to be highly effective at image classification, and since RD maps are essentially images as well, this also applies to RD classification. Therefore, the use of CNNs in this scenario is reasonable. As previously depicted in Figure 4.12, a typical RD map of a human has a broad horizontal distribution that represents a variety of body part velocities. A comparison reveals that RD map induced due to robot motion will have a limited horizontal distribution due to its rigid body motion. The distinction between the two patterns is already discernible to human eyes. Consequently, a CNN is also capable of differentiating them.

The proposed CNN model utilizes an input image size of 200×200 . This provides a trade-off between processing time and performance. The grayscale color mode is used for the following reason: in RD maps, the color represents the back-scattered power, which correlates to the target's position relative to the radar and can be affected by metallic components on targets (e.g., wearable metal articles such as watches or rings on human targets). According to our strategy, none of this information should be considered. The RGB color mode is more sensitive to noise caused by unwanted objects and clutter than the grayscale mode. Using one-channel grayscale images as input reduces the computational complexity of both training and prediction.

The utilized network architecture is inspired by Lecun's "LeNet-5" presented in LeCun et al. 1998 . It has a stack of six convolutional layers with the ReLU activation function. A max-pooling layer follows each convolutional layer, which consists of 16 convolutional kernels each measuring 3×3 pixels in size. After the convolutional layer stack, there is a fully connected layer with 16 neurons. One sigmoid-activated neuron in the output layer at the end is fully coupled to the 16 neurons in the layer above.

The selection of the optimizer has an immediate impact on the training's outcome and required time. The modern adaptive optimizer Adam proposed in Kingma & Ba 2014 is used to train our proposed CNN. The Adam optimization algorithm is one of the most popular training algorithms for DNN. It improves the traditional stochastic

gradient descent algorithm by allowing the computation of individual adaptive learning rates for various parameters. Thus, the Adam optimizer achieves excellent optimization outcomes while maintaining rapid convergence rates. Dropout is a simple yet effective regularization technique proposed in Srivastava et al. 2014 to prevent a deep learning model from complex co-adaptations on training data, also known as overfitting. During training, it ignores neurons according to a predetermined ratio (the dropout rate). During the forward propagation of each training step, ignored neurons are unable to temporally contribute to the activation of their connected neurons in the final layer. During back propagation, the weights of the ignored neurons are not updated. In this instance, dropout is applied to the final fully-connected layer in all implementations, and a dropout rate of 0.5 is selected. This hyperparameter is determined by trial and error.

Both validation accuracy and loss begin to swing around the same level after the best validation accuracy, attained in the 30th training period. This shows that the model has already provided the best match to the data.

4.4.2 Results and Discussion

Random forest and gradient boosting are put to the test to see if ensemble learning on reconstructed RD maps is practical. They both stand out as excellent examples of bagging and boosting, respectively. The random forest performs with a classification accuracy of 93.4%, while the gradient boosting achieved a classification accuracy of 97.85%. There are two explanations for this:

- According to Caruana et al. 2008, boosted trees outperform random forests for low-dimensional problems with up to 4000 data dimensions.
- According to Fernández-Delgado et al. 2014, gradient boosting can outperform random forest in a binary classification scenario.

Gradient boosting is therefore thought to be a superior option for future research and comparisons. The suggested CNN outperforms the boosting-based technique and obtains the best performance. Accuracy levels for training and testing were 99.65% and 98.34%, respectively. As a result, the miss-classification rate for both classes is roughly 0.5%, which is acceptable for the task at hand.

4.4.3 Summary

In this section, a use-case scenario of differentiating a walking human from a moving robot is presented. The main research objective is validating the aspects for enhancing the human detection that are discussed in Section 4.3, and the study is published by us in Abdulatif et al. 2018. The study is based on achieving the classification task based on captured RD maps through comparing classical machine learning techniques with ensemble techniques, and a CNN approach. The CNN has achieved the best performance of 99% for single RD map that needs to be captured ≈ 0.1 s. In Section 4.4.1, the methodology is presented including the experimental setup and the different utilized approaches. In Section 4.4.2, a summarized discussion about the results is presented.

5 Feasibility Study for Human Identification While Walking

5.1 Human Identification Limitations

Due to its significance within the security domain, human identification was chosen as the next investigational step. Moreover, the completion of the task based on a radar-based perception system is anticipated to be of great interest due to the radar's ability to operate under any environmental conditions without compromising privacy. In this section, a feasibility study investigating the effect of human body characteristics on μ -D signature-based walking human identification is presented. As stated in Chapter 2, walking has been selected as the primary activity for investigation in this thesis because it is regarded as the fundamental motion for studying the μ -D behavior. The primary objective of the feasibility study is to analyze the influence of various factors on the behavior of the μ -D signature. As described in Chapter 2, the study is conducted on the smallest unit of the walking gait cycle, a half gait cycle. The following is a summary of the main aspects of the feasibility study:

- The real-time aspect since the duration of a half gait cycle is $\approx 0.5s$ for a normal walking velocity.
- The effect of the body dimensions on the walking style and the captured μ -D signature. Accordingly, 22 subjects have been included in the study.
- The main characteristics of the μ -D signature, in which a treadmill was used in the experimental setup for removing the average walking velocity and capturing only signatures reflecting the body micro motions.
- The classification ambiguity for multiple persons.

A treadmill is used in our proposed experimental setup to capture μ -D signatures from 22 subjects of both genders with different physiques. The latent space representation is

then extracted from the μ -D signatures using convolutional autoencoders (CAE). Using t-distributed stochastic neighbor embedding, it is then interpreted in two dimensions. Our study states a direct effect of the body mass index (BMI) on the μ -D signature of a walking subject. On the basis of the μ -D signature, a 50-layer deep residual neural network is then trained to identify the walking subject. We achieve 98% accuracy on the test set with a high SNR and 84% accuracy with varying SNR levels. Our study is published in Abdulatif et al. 2019, including figures, an explanation of the experimental setup, and the results. The feasibility study wasn't of a main concentration of comparing the effect of using multiple lengths of captured μ -D signatures as this was tested in the preliminary literature that were discussed in Section 2.5. nevertheless, multiple comparisons such as the effect of the length of the μ -D signatures or the included classes were included in the latter study that is presented in Weller et al. 2021, and discussed in Section 6.3. Examples for the comparisons are shown in Fig. 6.10.

5.2 Methodology for Human Identification

Due to the deformability of a moving human body, analyzing human gait is considered a difficult task. The human walking gait can be studied by measuring the associated temporal and spatial parameters. The primary characteristics of the human gait motion, such as stride length, walking velocity, and swing and stance phases, are represented by these spatio-temporal parameters. As a result, the motion of each body part can be represented by a translational or rotational trajectory. Then, as explained by the global human walking model, presented by Boulic et al. 1990, they can be used to determine the position and orientation of body parts over time. However, this model will only provide an indication of the body's behavior, not an accurate simulation. In addition, it is limited to the walking motion only. On the other hand, the μ -D signature can be used for any activity that is realistically identical for each individual.

In order to obtain an accurate study of human gait, body weight is also an important factor. According to Hwang et al. 2017 and Błaszczuk et al. 2011, the spatio-temporal parameters of the walking gait are influenced by body weight. As a result, if evaluated correctly, walking style can be a useful metric for human identification, as described in

Cutting & Kozlowski 1977. In order to obtain a representative sample of μ -D signatures, 22 contributors with varying BMIs based on their weight and height were chosen.

5.2.1 Radar System Parametrization

The proposed experiment employs a CW radar with a carrier frequency of $f_o = 25GHz$. The datasheet for the Ancortek radar module used in the experimental setup is presented in the appendix, in Section A.2. This study focuses primarily on analyzing the velocity components induced due to various body parts through the walking activity. Therefore, no range information is required to be captured, and the μ -D signature can be extracted without frequency modulation. Table 5.1 demonstrates the radar parameterization considered for this experiment. The velocity characteristics are computed using the equations presented in Section 2.1.2.

According to van Dorp & Groen 2003, when studying the μ -D signature of a non-rigid body motion, such as a walking human, the signature of each individual is anticipated to be unique due of changes in walking style and recognized RCS. The μ -D signature of a human gait is the superposition of the signatures coming from the motions of distinct body parts, such as the arms, legs, and feet. Consequently, the back-scattered radar signal would contain a multitude of superimposed frequency components with substantial temporal variations. A time-frequency analysis employing the STFT processing described in Chapter 2 is utilized to visualize the variations in velocity.

5.2.2 Treadmill Experimental Setup

The study requires a data set containing diverse human body features. Therefore, 22 individuals of varying weights and heights were considered. The sample consists of 5

Table 5.1: The radar parametrization for μ -D acquisition.

Radar Paramaters		Velocity Attributes	
Center frequency	25 GHz		
Sampling frequency	128 kHz	Maximum velocity	6 m/s
Pulse frequency	2 kHz	Velocity resolution	2 cm/s
Chirp duration	1 ms		

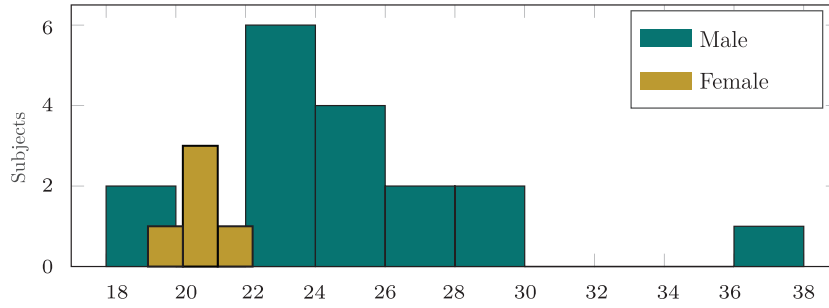


Figure 5.1: BMI-based distribution of both genders participating in the investigation.

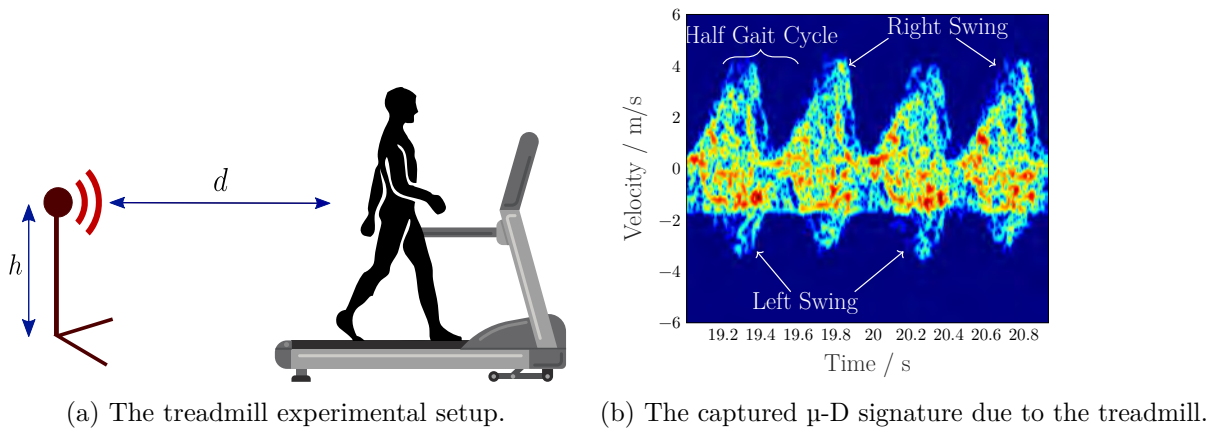


Figure 5.2: Schematic for the treadmill experimental setup, and the corresponding μ -D signature.

females and 17 males with weights ranging from 54 kg to 115 kg and heights between 1.62 m and 1.95 m. To monitor the relationship between height and weight, the BMI is calculated by dividing weight by height squared for each participant. As explained in Nuttall 2015, the BMI can be deceiving when measuring the subject's volume because it does not account for bone density, body fat, and muscle mass. To mitigate this effect, participants are selected based on a direct correlation between their BMIs and their body volumes. Figure 5.1 illustrates the BMI distribution of selected subjects.

As shown in Figure 5.2a, the radar is placed at a certain height and distance from the back of the treadmill. The subjects are instructed to walk away from the radar at an average speed of 1.6 m/s on the treadmill. All participants are tested, while the radar is at the same height and distance, which are 1 m and 3 m, respectively. As a result, the subject's micro motion behavior is the only variable in the experiment. As a result, the μ -D signature should be determined solely by the measured subject's walking style and

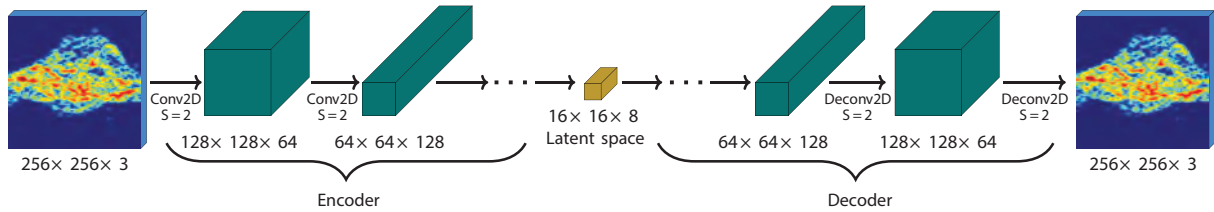


Figure 5.3: The architecture of the proposed CAE.

RCS. It is expected that a subject's RCS will correlate with its cross section and, as a result, BMI.

A full gait cycle's μ -D signature is interpreted as two main half cycles. Each half gait represents the fluctuating velocity components of various body parts. In comparison to the left leg and the right arm, the left foot generates the highest velocity component during one-half gait. As shown in Figure 5.2b, these effects are reversed in the subsequent half gait, while the periodicity of the full gait cycle is maintained. Due to the periodicity of each half gait cycle, the μ -D signature is segmented on a half-gait basis to ensure that the data accurately reflect the walking motion. Each subject's experiment acquisition time is 180 s. Depending on a subject's height, the number of steps taken per second can vary. However, an average of 0.5 s can be considered for a half gait cycle. Consequently, an average of 360 half gait cycles are recorded per subject. Each half gait cycle is then saved as an unsigned 16-bit integer RGB image with a $256 \times 256 \times 3$ dimension for use as an input for the proposed networks.

5.3 Autoencoder for Signature Investigation

A total of 7920 half gait μ -D images were gathered utilizing the designed experimental setup. Significant amounts of useful information regarding the motion characteristics can be extracted from each image. In order to examine the impact of human body characteristics on the μ -D signature, an unsupervised autoencoder is used to reduce the dimensionality of the input data to a specific latent space. Since the input μ -D signature is represented in the time-frequency space, a CAE is used to extract latent image features.

A CAE's encoder-decoder structure is described and illustrated in Le 2013. Encoder uses a sequence of convolutional layers to transform a high dimensional input into a latent space representation or bottleneck (maximum compression point). The Decoder upsamples

back from the latent space representation to reconstruct the high dimensional input. The primary objective of CAE is to minimize the cost function, which is represented as the mean square error between input and output. As shown in Figure 5.3, the proposed CAE uses a $256 \times 256 \times 3$ input image and a 4-layer convolutional network with a stride of 2 to reduce the input to a $16 \times 16 \times 8 = 2048$ latent space. To reconstruct the input at the decoder side, a mirrored deconvolutional network is utilized. Consequently, the 2048 dimensional feature vector extracted from the downsampled latent space represents the essential information in the half gait μ -D signature. This precise feature space size is obtained by optimizing the hyper-parameters of our proposed CAE. Consequently, further dimension reduction utilizing additional layers may result in an imperfect reconstruction of the input image and, consequently, the loss of vital information in the bottleneck latent space.

To visualize the distribution of the learned latent space representation, the t-SNE representation introduced in van der Maaten & Hinton 2008 is used to reduce dimensions further. The t-SNE can be defined as a nonlinear dimension reduction technique that can be used to visualize the distribution of the 2048 dimensional encoded latent space in a two-dimensional space. The primary objective of t-SNE is to generate two joint probability distributions for both the high-dimensional space and the low-dimensional space that has been mapped. As explained in Joyce 2011, the Kullback–Leibler divergence between the two distributions is minimized as a metric for measuring the dissimilitude between the two distributions. Therefore, it is highly probable that high-dimensional samples with similar global or local structures will be visualized as neighboring points in 2D space.

Based on this method, the dimensions of the encoded half gait signatures can be reduced to a 2D space that is easily interpretable. In Figure 5.4a, the reduced t-SNE is plotted for 14 out of 22 subjects in order to facilitate visualization. The t-SNE representation is capable of classifying each subject as two close, but distinct clusters for the left and right swing cycles. This demonstrates the efficacy of t-SNE as it identifies each half gait as distinct clusters within the same 2D space occupied by the subject. Moreover, neighboring clusters of two distinct subjects may be of different genders and heights or weights. In most instances, however, the BMI of neighboring clusters is comparable. As depicted in Figure 5.4b, the 22 subjects can be divided into five BMI groups. Clusters that belong to

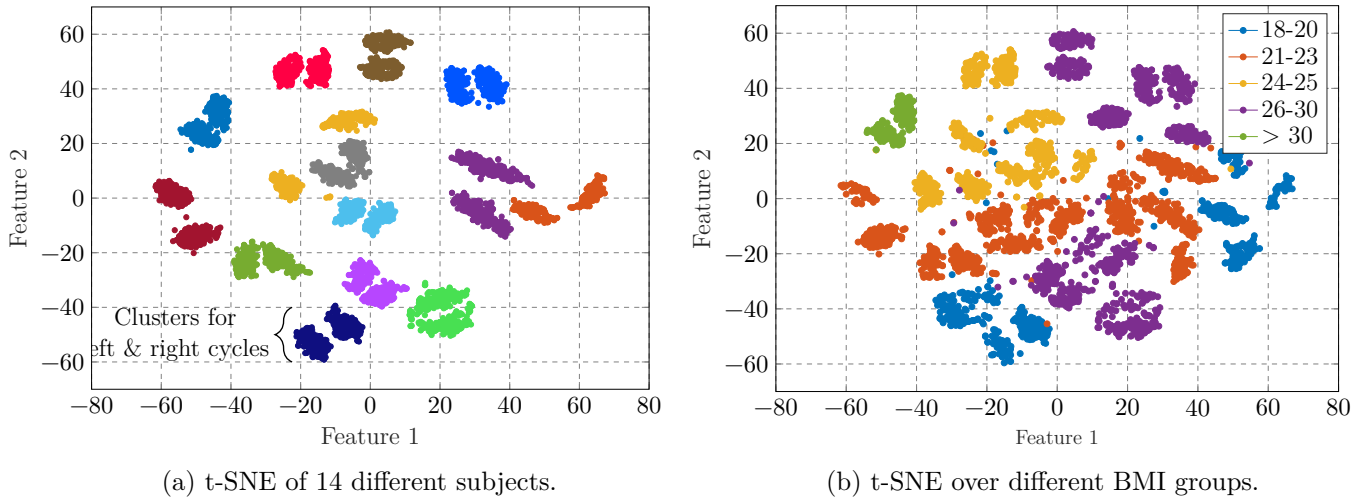
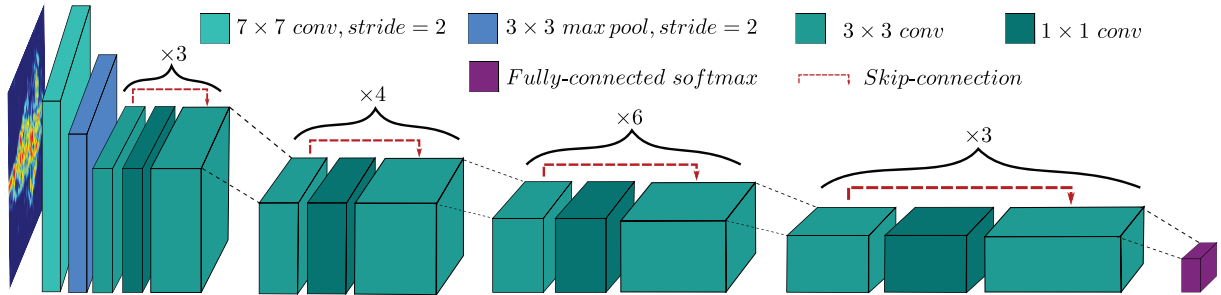


Figure 5.4: Two clusters are observed on each subject from the right and left body sides.



the same BMI group are typically distributed over the same area in 2D space. Due to the fact that walking style can have a substantial effect on the μ -D signature outline and local structure, some BMI groups tend to congregate in two major areas of the restricted space. The latent space distribution of the solitary participant belonging to the obese category (BMI over 30) can also be observed as a distinct cluster in 2D space.

5.4 Person Identification Network Architecture

On the basis of the presented study, it can be concluded that subjects with comparable BMI and walking styles are likely to have comparable μ -D signatures. In order to classify them as distinct subjects, a more complex architecture that can utilize local spectrum details is required. This was observed when the authors of Cao et al. 2018 employed a shallow 6-layer architecture to classify 20 subjects with a BMI range of 18 Kg/m²

and achieved an acceptable performance with only 4 subjects. Increasing the number of subjects causes BMIs and walking styles to become more similar, resulting in a decline in classification performance to 68.9% for 20 participants.

In this feasibility study, a more sophisticated architecture is used to classify the 22 participants based on their collected half gait μ -D signatures. Due to the importance of the real-time aspect, an architecture with fewer computations and a short inference time is required. Therefore, the ResNet-50 architecture presented in He et al. 2016 is utilized, as its number of parameters and operations is significantly lower than that of the VGG-19 network presented in Simonyan & Zisserman 2014. As depicted in Figure 5.5, the ResNet-50 architecture is based on residual blocks with a skip connection at every second layer defined as:

$$a[l + 2] = \text{RELU}(a[l] + z[l + 2]) , \quad (5.1)$$

where l is the index of the layer, $a[.]$ is the activation of the layer, and $z[.]$ is the output of the layer prior to ReLU activation. These connections enable the training of extremely deep networks without performance degradation. This can be explained as the network learning the identity mapping of the activation $a[l]$ in the case of low output weights $z[l + 2]$. In other instances, the network is capable of learning weights and biases that enhance performance. Therefore, the skip connections will either stabilize or enhance the network's performance.

Dimensional agreement is required for skip connections to be valid, which is not always the case. To address this issue, a zero padding is utilized to enable residual operations over these layers. Moreover, it is known that extremely deep networks can experience problems with vanishing gradients, which can be mitigated by batch normalization Ioffe & Szegedy 2015. The output of the final convolutional layer is then passed to a fully connected layer after being flattened. The likelihood of 22 classes corresponding to participants is then generated by a softmax output layer.

5.5 Results and Discussion

An Adam optimizer is used to train the ResNet-50 architecture on 7920 labeled half gait μ -D signatures. On a single NVIDIA Titan X GPU, the training of the model took

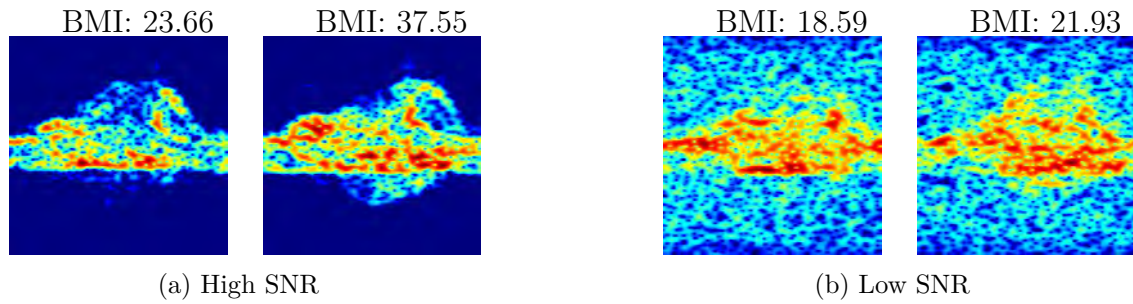
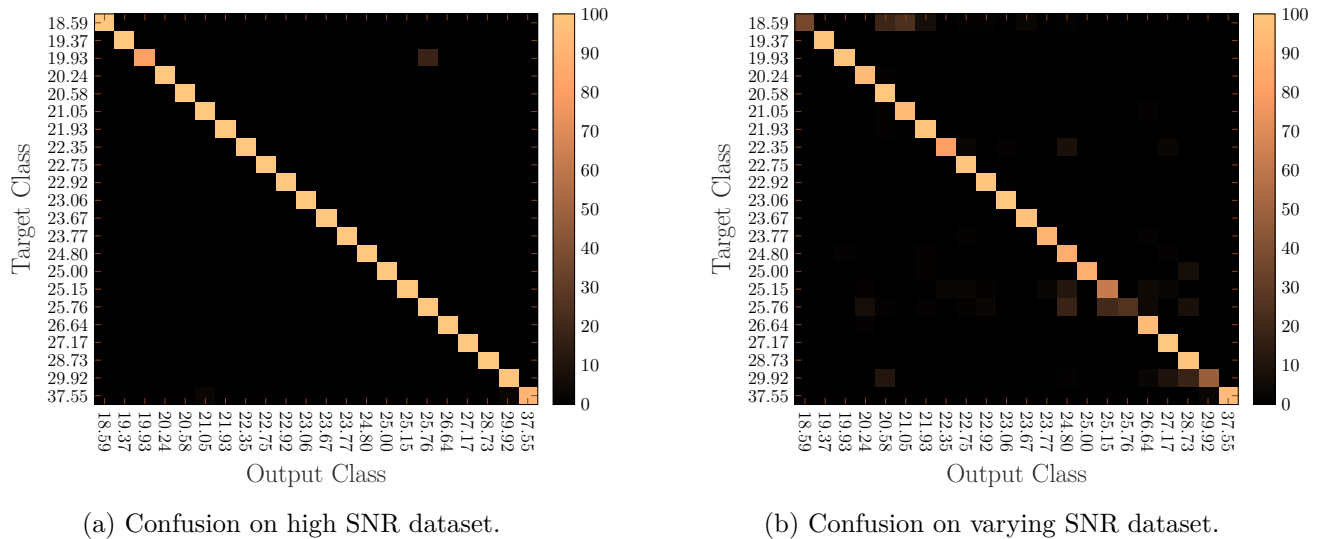
Figure 5.6: Examples of half gait μ -D signatures due to different BMIs.

Figure 5.7: A classification accuracy of 98% and 84% for the high and varying SNR, respectively.

approximately 30 minutes. In Figure 5.6a, examples of collected signatures from various BMI groups are displayed. Due to a relatively lower RCS, subjects with a low BMI tend to have less power within the signature local structure. In contrast, the global structure of the signature is primarily related to the velocity components of the swinging limbs and thus to the walking style.

To evaluate the performance of ResNet-50, a second experiment is conducted with the same 22 subjects, treadmill speed, and experiment duration in order to collect an additional 7920 signatures for testing. The radar is placed at the same distance from the treadmill, which is 3m. The network's overall accuracy on the unseen test set was 98%. The model inference time, or the time required to classify one half gait signature, is only 200 ms.

In Figure 5.7a, a normalized confusion matrix for the collected test set is displayed. It can be concluded that adding more layers to the proposed ResNet-50 architecture improved the classification performance for multiple subjects (> 20), despite the fact that

the majority of the participants belonged to the same BMI group ($21\text{Kg}/\text{m}^2$) and had comparable walking styles.

To validate the hypothesis demonstrated in the previous section that a subject's BMI can influence its signature, a more realistic experiment on person identification is conducted. In this experiment, the radar is placed 10 m further away from the treadmill. According to the equations presented in Richards 2014, increasing the target range from the radar causes a decrease in SNR. This is primarily observed in individuals with a lower BMI due to the lower RCS detected. In addition, the μ -D signatures of individuals with comparable BMI have partially lost their distinctive global and local structures. As shown in Figure 5.6b, the increased similarity between the half gait signatures is anticipated to make the classification task more difficult.

On the basis of this setup, an additional dataset is collected for all 22 subjects in order to evaluate the effect of the lower SNR. All other experimental conditions, including walking speed and data collection duration, are identical. The low-SNR and high-SNR datasets are combined and then divided into train and test sets. Consequently, the combined dataset contains more images with varying SNR values, making person identification more difficult. The variation in SNR forces the network to acquire more complex local μ -D structures to differentiate between distinct subjects. The mixed-SNR training set was used to train the same network structure described in Section 5.4. On the unknown test set, it achieves an accuracy of 84%. As depicted in Figure 5.7b, the confusion is greatest between subjects with comparable BMI values.

5.6 Summary

In this chapter, an experimental study illustrating the effect of human body characteristics on the μ -D signatures of walking individuals is presented. Using a CW radar, the μ -D signatures of 22 subjects of various genders and body types walking on a treadmill are measured. Consequently, the RCS and walking style of the moving subject are the primary factors that influence the measured μ -D signatures. The BMI is used as a metric for the body volume of selected participants, ranging from lean bodies starting from $[18-20]\text{ kg}/\text{m}^2$ to obese levels at $37\text{ kg}/\text{m}^2$. A CAE is utilized to encode the μ -D signatures of the half

gait into a latent space representation. To visualize the latent space distribution in 2D, a nonlinear dimension reduction based on t-SNE is applied. On the basis of visual analysis, the BMI of the subjects and their walking styles are found to have a direct effect on their μ -D signatures. A deep ResNet-50 architecture is trained on μ -D signatures for person identification. The architecture correctly identifies μ -D signatures with a 98% accuracy at low SNR values and an 84% accuracy at different SNR values, with confusions occurring predominantly in subjects with comparable BMIs.

On the basis of this setup, an additional dataset is collected for all 22 subjects in order to evaluate the effect of the lower SNR. All other experimental conditions, including walking speed and data collection duration, are identical. The low-SNR and high-SNR datasets are combined and then divided into train and test sets. Consequently, the combined dataset contains more images with varying SNR values, making person identification more difficult. The variation in SNR forces the network to acquire more complex local μ -D structures to differentiate between distinct subjects. The mixed-SNR training set was used to train the same network structure as in Section 5.4. On the unknown test set, it achieves an accuracy of 84%. As depicted in Figure 5.7b, the confusion is greatest between subjects with comparable BMI values. Nevertheless, the BMI is not considered of the only effect on the human identification task, but it was investigated as it has a direct correlation with the RCS. The feasibility study helped in putting the guidelines for the experimental setup for designing an experiment for human identification through unconstrained walking.

6 Metric Learning for Free-motion Person Recognition

MIMO radar modules have been employed for people identification, in which the interferometric analysis between separate receiving antennas is utilized to estimate the target's AoA as explained in Milligan 2005, and a tracking function is introduced to μ -D-based human-ID applications as described in Pegoraro et al. 2020, and Zhao et al. 2019. However, the rate of change of AoA profiles can be tracked over time to reveal the micro-motions of angular velocity. As far as we are aware, this was not previously examined as an additional element for improving the human-ID task. This method differs from the one described in Nanzer 2010, in which the tangential angular velocity was calculated independently of the angle of attack due to hardware restrictions. The primary objective was to increase the field of view (FOV) to incorporate tangential circumstances. Nonetheless, classification uncertainty between classes is still anticipated, since the newly-measured micro angular spectrogram in the tangential dimension demonstrates identical performance to the commonly-used μ -D signatures in the LOS, as detailed in Nanzer 2010; Nanzer & Zilevu 2014 and Nanzer 2016. In addition, the provided technique is based on estimating the micro angular velocity while remaining independent of the AoA.

This chapter evaluates the viability of depending on the μ - ω signature to improve human identification. In addition, the concept of employing a metric learning approach for overcoming restrictions such as detection from numerous aspect angles other than LOS, increasing the number of classes, and real-time classification is investigated. The real-time classification is based on the adaptive μ -D slicing described in Section 3.3.3, in which classification is performed only on a single walking step or half gait cycle, which lasts $\cong 0.5s$. The algorithm is initially applied to the task of human identification, for which the whole study and results are published in Weller et al. 2021. Then, the same method, comprising the metric learning approach, the μ - ω utilization, and the adaptive technique,

was utilized in a subsequent investigation for activity recognition, the results of which were reported in Aziz et al. 2022. Accordingly, the algorithm and context of metric learning are explained first. The problem analysis, experimental design, and outcomes for both use-case situations are then provided. The datasheet for the Radarbook2 module used for all the experiments held in this chapter is presented in the appendix, in Section A.4.

6.1 Metric Learning Approach

Relying on metric learning approaches has a main function of obtaining satisfactory classification compared to conventional DL techniques, but with a less amount of labeled data. The basis of metric learning is a similarity score between two or more provided examples as explained in Li et al. 2012 and Kaya & Bilge 2019. This similarity score is derived based on the embedding space representation of input data, which is trained to contain only the most significant characteristics. In contrast to standard learning techniques, these features do not mirror the structure of the input data, but instead focus on distinguishing characteristics across samples of various classes. Thus, an embedding space is generated in which intrinsic data points are mapped to near proximity while extrinsic data points are pushed further away.

6.1.1 Fundamentals of Triplet Loss

Typically, the number of classes in supervised learning is fixed. Common loss functions, such as the categorical cross-entropy loss, are only applicable to a limited number of classes, and NNs must be constructed accordingly. In certain circumstances, such as surveillance or access control systems, the ability to process a variable number of classes is essential. One approach to accomplish this is by comparing one sample to another and calculating their similarity score. Triplet loss is a loss function that can accomplish the aforementioned with a triplet of samples as described in Hermans et al. 2017.

The triplet loss was introduced in Schroff et al. 2015, where it was utilized to achieve face verification, recognition, and clustering. The datasets were comprised of facial images captured from widely varying angles, and the state-of-the-art performance was achieved with only 128 bytes per face, whereas previous systems required thousands. Triplet loss is

therefore a loss function that can learn an embedding space with high representational efficiency. Accordingly, it was selected for the (μ -D)-based human identification due to the availability of common constraints e.g., the limited amount of data and the urge of detecting from multiple AoA.

Triplet loss is built on the concept of metric learning and provides a loss function that is ideally suited to our thesis work. It is necessary to learn an embedding space in which embeddings of the same class are mapped to close proximity and those of different classes are kept well apart. The objective is not to merge embeddings of the same class into a single point. The only criteria is that, given two samples of the same class and one of a different class, the sample of the different class must be at least as distant as the sample of the same class. Samples of the same class will be referred to as positives (p) and samples of different classes will be referred to as negatives (n). The sample on which p and n are based is known as anchor (a). For fulfilling the aforementioned condition in the embedding space, a distance between the three samples should satisfy:

$$d(a, p) - d(a, n) \leq 0 . \quad (6.1)$$

However, Equation (6.1) can be easily satisfied in two ways that do not require the network to acquire any useful information. The first procedure may involve mapping each sample to the same point in the embedding space. Subsequently, the embedding vector of each sample will be identical. The second possibility is that each sample's embedding vector equals zero. In order to prevent this, a margin (α) is included such that:

$$d(a, p) - d(a, n) + \alpha \leq 0 \quad (6.2)$$

In order to effectively cluster the samples in the embedding space, the network must now learn a nontrivial solution. From this, the loss over a triplet is calculated:

$$\ell(a, p, n) := \mathbf{max} [d(a, p) - d(a, n) + \alpha, 0] . \quad (6.3)$$

Figure 6.1 provides a visual representation of the complete DCNN network architecture. The DCNN network is fed a concatenation of both the μ -D and μ - ω spectrograms for slices

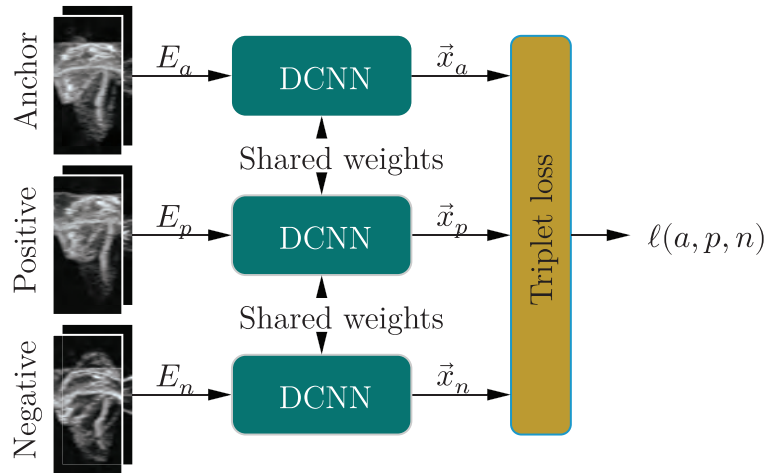


Figure 6.1: The architecture of the triplet loss.

of single steps, as depicted in Figure 6.1. As the triplet loss requires the embeddings of a , p , and n for each iteration, the DCNN processes all three concatenations. This can be shown as three DCNNs with identical weights, one for each concatenation. Each network is now identical to the other two. The embeddings are sent to the triplet loss, and the loss scalar is backpropagated through the network.

6.1.2 Implementation of Triplet Loss

According to the definition of the triplet loss, there are three types of triplets that may occur:

- Easy triplets with zero loss since $d(a, p) + \alpha < d(a, n)$.
- Hard triplets in which the positive is farther from the anchor than the negative $d(a, p) > d(a, n)$.
- Semi-hard triplets in which the positive is closer to the anchor than the negative, but not by more than the margin $d(a, p) < d(a, n) < d(a, p) + \alpha$.

The definitions of the three categories depend on the relationship between $d(a, n)$ and $d(a, p)$. Thus, they can be converted to the negatives, which are classified as easy, hard, and semi-hard. The regions of the embedding space that correspond to the three negative types can be seen in Figure 6.2.

The sampling of triplets is a different process from the training of neural networks. Consequently, it adds a computational expense that should be kept to a minimum. Choosing

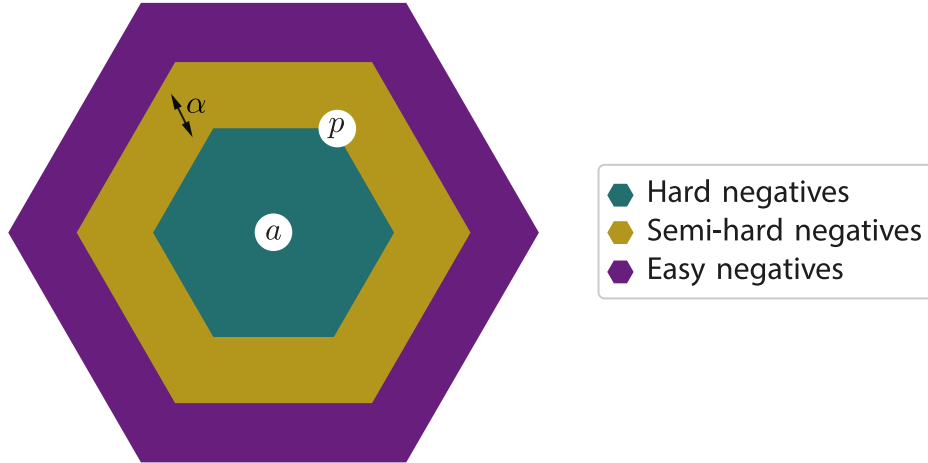


Figure 6.2: The three different types of negative samples: hard, semi-hard and easy.

B triplets and stacking the associated $3B$ samples into the batch is the most basic method for sampling a batch. To determine whether a negative is easy, semi-hard, or hard, it is necessary to compute the embeddings of all $3B$ samples. These are then utilized to construct the B triplets that contribute to the loss. Considering that $3B$ samples can be joined in $6B^2 - 4B$ different ways to generate valid triplets as explained in Hermans et al. 2017. Thus, it is not feasible to employ only B contributors to the loss. For this reason, the authors in Hermans et al. 2017 recommended organizing the triplet loss for more effective performance. The main concept for their approach is based on applying a random selection for K classes and N samples for each class. This will result in a mini-batch of size $B = KN$. The most basic way to train the NN is describe in Hermans et al. 2017, and called the batch all, where all the valid triplets (V_{ba}) are used and can be evaluated as ($V_{ba} = KN(N - 1)(KN - N)$).

However, the batch all technique is not considered the most feasible as it will include all the easy triplets that don't contribute to the learning process. As a solution another technique is applied known as the batch hard. This technique is based on selecting the most challenging positives and negatives for each anchor within the mini-batch. This results in producing only moderately challenging triplets, as they are the most challenging inside their mini-batch. Accordingly, the available triplets (V_{bh}) can now be evaluated as ($V_{bh} = KN$), and hence, it contributes to the loss as three times greater than the traditional technique. In the context of the human identification, the hardest triplets are

expected to occur between the μ -D signatures induced due to comparable BMI as has been presented by in our study in Abdulatif et al. 2019, and discussed in Chapter 5.

The implementation of batch all and batch hard triplet loss is based on the blog article presented in Schroff et al. 2015 and the corresponding Github repository in Moindrot 2018. To keep the overhead of triplet sampling to a minimum, computations are performed as efficiently as possible. Even though it requires slightly more computational effort, the non-squared version of the triplet loss is utilized in this thesis work. It is demonstrated in Wu et al. 2017 that the squared version has a negative effect on the relationship between the attractive and repulsive gradients of hard positive and hard negative pairs, respectively.

It is straightforward to implement the batch all triplet loss. For each batch of size B , N samples of K classes are selected and organized into a three-dimensional tensor. Nonetheless, not all triplets are valid. The result is a mask that is true if and only if the triplet is valid. The entries of all invalid triplets in the tensor are set to zero, and the sum of all tensor elements is calculated. However, not all triplets are valid, so a filtering for valid tensors is carried out. The entries of all invalid triplets in the tensor are set to zero, and the sum of all tensor elements is calculated. The implementation of batch hard triplet loss is more complicated because hard positives and negatives must be determined beforehand. To accomplish this, the Euclidean distance is first computed for each embedding of the B samples.

6.1.3 Triplet Mining

Good quality embeddings are the foundation of a well-performing system in metric learning. A considerable amount of research has been conducted solely on loss functions to obtain these. Their job is to extract all of the model's positive and negative characteristics and, in doing so, significantly affect the embedding space. Nevertheless, there is more to designing the optimal system for a given task. The authors of Wu et al. 2017 demonstrate that the selection of appropriate training samples is at least as important as the selection of an appropriate loss function. Experiments they conduct reveal that different loss functions behave similarly when sampling strategies are appropriate. It is demonstrated that the same loss function produces radically different solutions when combined with different sampling strategies.

Sampling in DL describes the logic used to determine which samples to select from which batches are constructed. These batches are subsequently used for training, validation, and testing. Sampling weights data points effectively and defines a heuristic objective function implicitly. This makes it difficult to apply and reproduce the insights to diverse architectures, optimization frameworks, and data sets. Different sampling strategies and loss functions are therefore implemented and evaluated empirically to determine the optimal combination.

Offline Mining

Offline mining is the most primitive method for locating triplets and identifies them offline. For example, mining can be performed prior to each epoch. During the mining process, embeddings are computed for every sample in the entire training dataset, and invalid triplets are discarded. In the subsequent step, semi-hard or difficult triplets can be appended. When completed, found triplets are stacked into batches and then fed into the NN. This technique is not very efficient because it requires computing $3B$ embeddings to obtain B triplets that contribute to the loss scalar. In addition, offline triplet mining must regularly update the offline mined triplets. Otherwise, the network won't generalize, and the mined triplets will quickly become simple ones.

Online Mining

The online mining of triplets is more complex and was first introduced in Schroff et al. 2015. The central concept is to compute triplets in real-time in parallel with the training process. To accomplish this, B embeddings are computed for each batch of size B , and a maximum of B^3 triplets are discovered. Evidently, the majority of them will be invalid and therefore will be discarded. To obtain the desired triplets, one of the two mining strategies batch all or batch hard can now be implemented. Compared to offline mining, online triplet mining requires the computation of a great deal fewer embeddings, making it significantly more efficient and requiring less memory and computational resources. Another advantage is that the embeddings are always computed using the most recently parameterized DCNN, even within epochs. Therefore, the definitions of easy, semi-difficult,

and difficult negatives change with each backpropagation cycle. It guarantees true easy, semi-difficult, and difficult triplets at every stage of the training process.

The online mining of triplets is contingent on the fulfillment of certain preconditions. Therefore, a meaningful representation of anchor-positive distances is essential for mining success. Specifying a minimum number of samples N of each class K to be present in each mini-batch B can ensure that this meaningful representation is maintained. Consequently, the batch size cannot be chosen arbitrarily and is instead defined by $B = KN$. In the practical work for this thesis, the online mining strategy for batch-hard triplets is implemented.

Distance-Weighted Mining

The study presented in Wu et al. 2017 proposes another suitable mining strategy for metric learning approaches called distance-weighted mining. It is based on two factors, the first of which is that gradients expose high variances and low SNRs when negatives are too intense. The second consideration is that completely random samples are frequently too far apart, making it difficult to provide the optimization algorithm with useful information. Therefore, distance-weighted mining draws samples uniformly based on their distances. This results in samples that are dispersed across the entire range of distances rather than concentrated in a small region. Typically, both ends of the distance range contain noisy samples. As they have a negative impact on the training process, these regions will be clipped.

The authors of Schroff et al. 2015 report that their semi-hard triplet loss scalar decreases at a slower rate after a period of training. This may be the result of their mining strategy no longer discovering informative triplets. Distance-weighted sampling, on the other hand, consistently produces informative samples while controlling the variance. Aside from this, distance-weighted mining has two major benefits. Firstly, it corrects the bias introduced by the geometry of the embedding space, and secondly, it ensures that every data point has a chance of being sampled. This allows for a reduction in gradient variance, resulting in more stable training. Regardless of the loss function, distance-weighted mining is capable of producing embeddings that are qualitatively superior.

6.2 Neural Network Overview

In this section, an overview about the utilized neural network is presented. This network architecture is used for both studies, which are the free walking human identification that is discussed in Weller et al. 2021, and activity classification from multiple aspect angles that is discussed in Aziz et al. 2022. In Section 6.2.1, the structure of the inception network is presented and the main concept of formulating the network in a wider architecture instead of a deeper one is discussed. Afterwards in Section 6.2.2, the idea of combining both the residual and inception architectures is presented and discussed.

6.2.1 Inception Architecture

Similar to residual architectures, the multi-path methodology can also be attributed to inception networks. They are founded on the principle of going wider rather than deeper, and is full described in Szegedy et al. 2015. To accomplish this, they first divide the input into parallel branches and then apply convolutions of varying scales. The data is then independently convoluted for each branch before being merged back together. The size of the convolutional kernels within each branch varies. This enables the extraction of spatial information at multiple scales. In addition, the convolution process is separated into two operations, allowing cross-channel and spatial correlations to be evaluated independently. Typically, a 1×1 convolution comes before convolutions with intermediate and larger kernel sizes. It is utilized to reduce computational complexity without compromising generalization abilities. Since convolutional layers are the fundamental building block of inception networks, sparse connectivity is utilized and redundant information is effectively omitted. Therefore, inception blocks enable a network to acquire more accurate representations of input data with fewer parameters and are less susceptible to over-fitting as described in Szegedy et al. 2015.

Figure 6.3 depicts the inception block utilized in the practical work of this thesis. It is based on the original research on inception networks described in Szegedy et al. 2015 and consists of four parallel branches. The first only uses one 1×1 convolution to extract spatial information with fine-grained resolution. The second and third branches reduce computational complexity by employing a 1×1 convolution. The subsequent 3×3 and

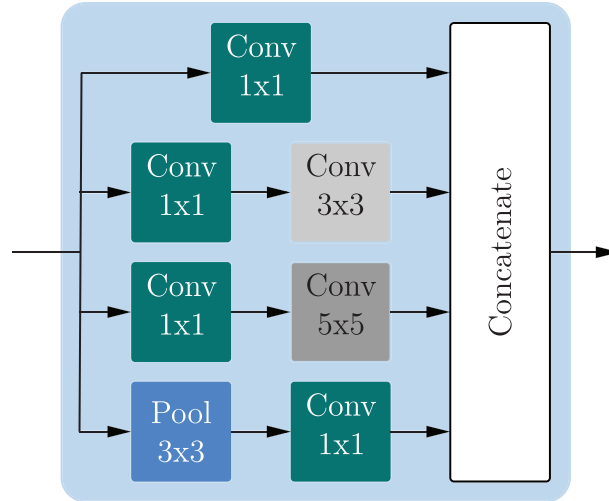


Figure 6.3: The inception network main building block.

5×5 convolutions extract spatial information at medium and large scales, respectively. A 3×3 max-pooling operation is followed by a 1×1 convolution to form the last branch. The max-pooling layer extracts dominant features from the input mapping, and the 1×1 convolution ensures that the size of this branch’s data is comparable to the other three. The output of the inception block is the concatenation of the mappings of the individual branches along the dimension of the filter.

6.2.2 Deep Convolutional Residual Inception Network

The concept of combining residual and inception structures was proposed by Szegedy et al. 2015. Experiments demonstrated that residual connections are not inherently required to train very deep NNs successfully. Nonetheless, they significantly reduce the time until training converges, which is a compelling argument for their use. The training results consistently outperformed those of simple residual and inception architectures.

The residual inception network deployed for this thesis work is inspired by the architecture described in Pegoraro et al. 2020 and is depicted in Figure 6.4. The concatenated half gait cycle μ -D and μ - ω spectrograms are initially fed into the initial stage of the network. The inception block, which is depicted in Figure 6.3, now processes the concatenations. A residual skip connection adds to its output the mappings preceding the inception block. In contrast to the skip connection depicted in Figure 6.4, the current connection computes a 1×1 convolution rather than feeding an exact identity mapping further down the

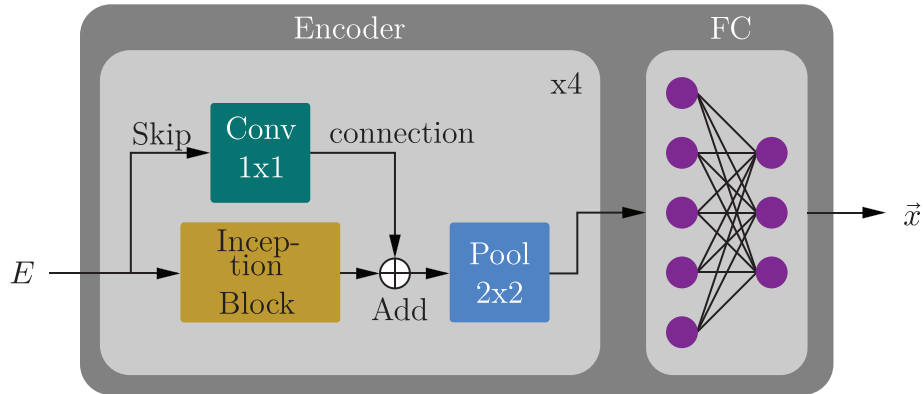


Figure 6.4: The residual inception network architecture.

network. This is required to match the dimensions of the skip connection mapping to those of the inception block output. After the element-by-element addition, a max-pooling operation with a 1×1 kernel is executed. It suppresses noise and results in a diminution of spatial dimension. This concludes the first stage, which is repeated three times more. The obtained fully convolutional network architecture can be viewed as an encoder at this point. The input data is mapped to an N-dimensional embedding space containing only its most distinctive characteristics. These mappings are flattened and fed through a two-layer deep fully connected (FC) network before they are presented to the loss function. Combining the encoder and FC network results in the DCNN architecture.

6.3 Human-ID Task

In this section, an overview of the human identification task based on free walking is presented. The study is discussed and published in details in Weller et al. 2021. This study is considered the following step to our feasibility study published in Abdulatif et al. 2019, and discussed in Chapter 5. Our main objective in the study is accomplishing the human identification task based on a free walking scenario in stead of the treadmill-based experimental setup used in Abdulatif et al. 2019 and discussed in Section 5.2.2. The limitations discussed in Section 5.1 were taken into consideration, when the experimental setup was designed and the utilized neural network was selected. Those limitations can be summarized as the real-time performance, effect of the body characteristics, capturing meaningful μ -D signatures and the classification ambiguity between different classes. The problem analysis is stated and discussed in Section 6.3.1. The radar paramterization and

the experimental setup are presented in Sections 6.3.2, and 6.3.3 , respectively. Afterwards, the results are presented and discussed in Section 6.3.4.

6.3.1 Problem Analysis

Deep learning has been successfully adapted to the topic of radar-based human identification, which has resulted in leaps in classification accuracy, as demonstrated by the study presented in Cao et al. 2018. The majority of current research utilizes μ -D spectrograms and DCNN architectures, such as Papanastasiou et al. 2021; Qiao et al. 2020 and Vandermissen et al. 2018. The advantage of more advanced DCNNs has been studied in Yang et al. 2019; Jalalvand et al. 2019 and Addabbo et al. 2020, while an explicit application of range-Doppler maps is presented in Yang et al. 2020 and Ni & Huang 2020. The latter demonstrates the efficacy of transfer learning in relation to radar-based human-ID. Thus, transfer learning is presented as a solution to the need for massive amounts of data for training DL approaches that can also be used for activity recognition. Few-shot learning (FSL) is utilized for human-ID in Niazi et al. 2021 for the same reason. FSL has the benefit of requiring fewer labeled samples to successfully train a DCNN and performing better with unseen data. Using μ -D signatures, however, reduces the observation window and increases the number of classes.

All the mentioned papers solved the radar-based human-ID task using at least 1 s of μ -D spectrogram data (one full gait cycle), with the majority falling within the 3 s range. Longer observation windows allow for the inclusion of additional classes and the identification of a less constrained walking style. Alternatively, our feasibility study, which is described in Chapter 5 and published in Abdulatif et al. 2019, completed the classification task on only half of the gait cycle, which lasts $\approx 0.5s$. However, the study was conducted on treadmills with restricted motion in LOS scenario. The research proposes residual neural networks as a viable alternative to relying on deep neural networks to solve such complex classification tasks. The study also revealed a correlation between the BMI and the μ -D signature, indicating that similar BMI can lead to classification ambiguity due to the similarity of the walking style. In conclusion, there is a trade-off between the real-time analysis and radar features, as well as the number of classes and unrestricted walking. Furthermore, it is recommended to rely on less training data as the human-ID

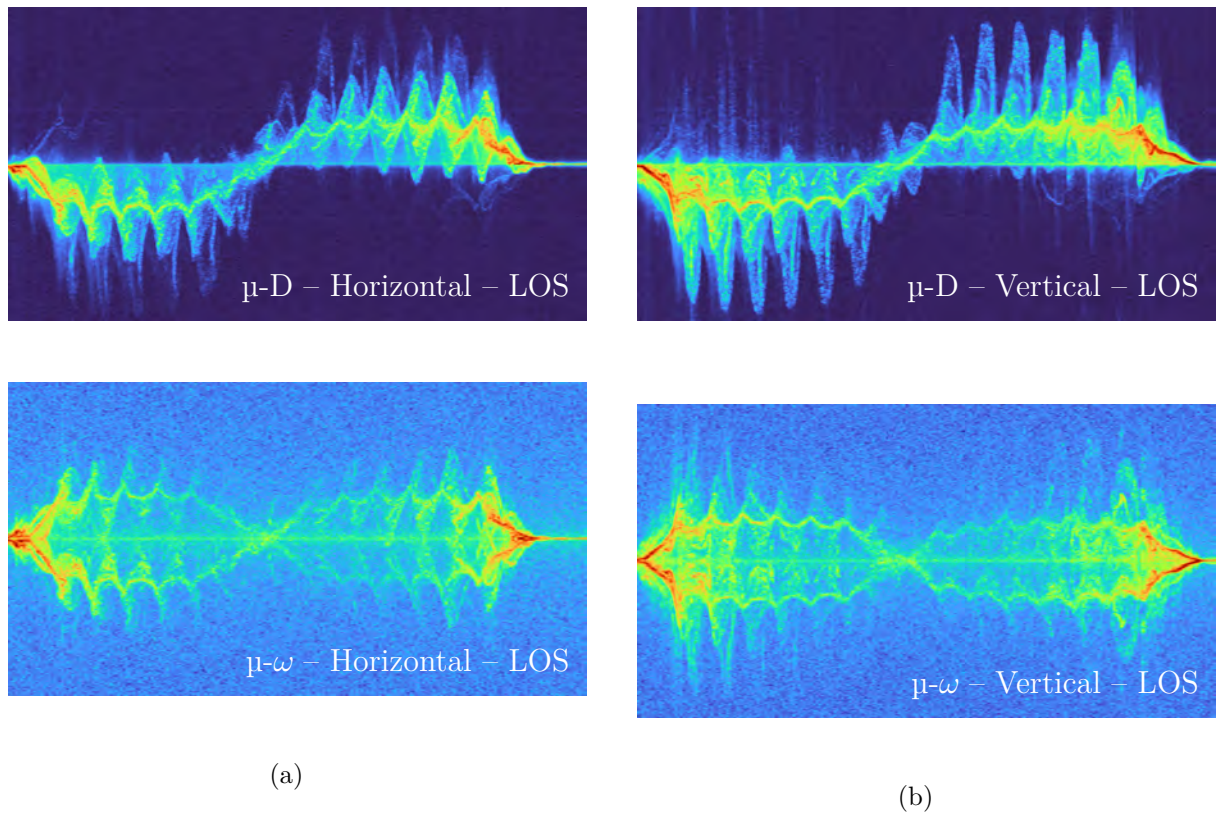


Figure 6.5: Comparison between vertical and horizontal radar polarization for walking in LOS.

application always requires the addition of new classes. Therefore, walking for extended periods of time and retraining the entire network are not applicable in the real world.

In this section, the methodology for the human-ID task is described, in which 22 subjects walk unconstrained according to their natural gait. The framework is designed based on a MIMO radar capable of capturing both $\mu\text{-}\omega$ and $\mu\text{-D}$ spectrograms. The overall captured signatures are divided into half walking gait cycles utilizing an adaptive technique. The experimental setup consists of walking at various aspect angles relative to the radar. The proposed method is based on a similarity score methodology that calculates a correlation between the walking signatures from various angles.

6.3.2 Radar Setup and Paramaterization

The utilized radar is a MIMO-FMCW module that is featured with a linear uniform antenna (LUA) in which the direction of orientation has to be chosen carefully to ensure capturing the micro motion signatures with th best SNR. Thus, a comparison is held

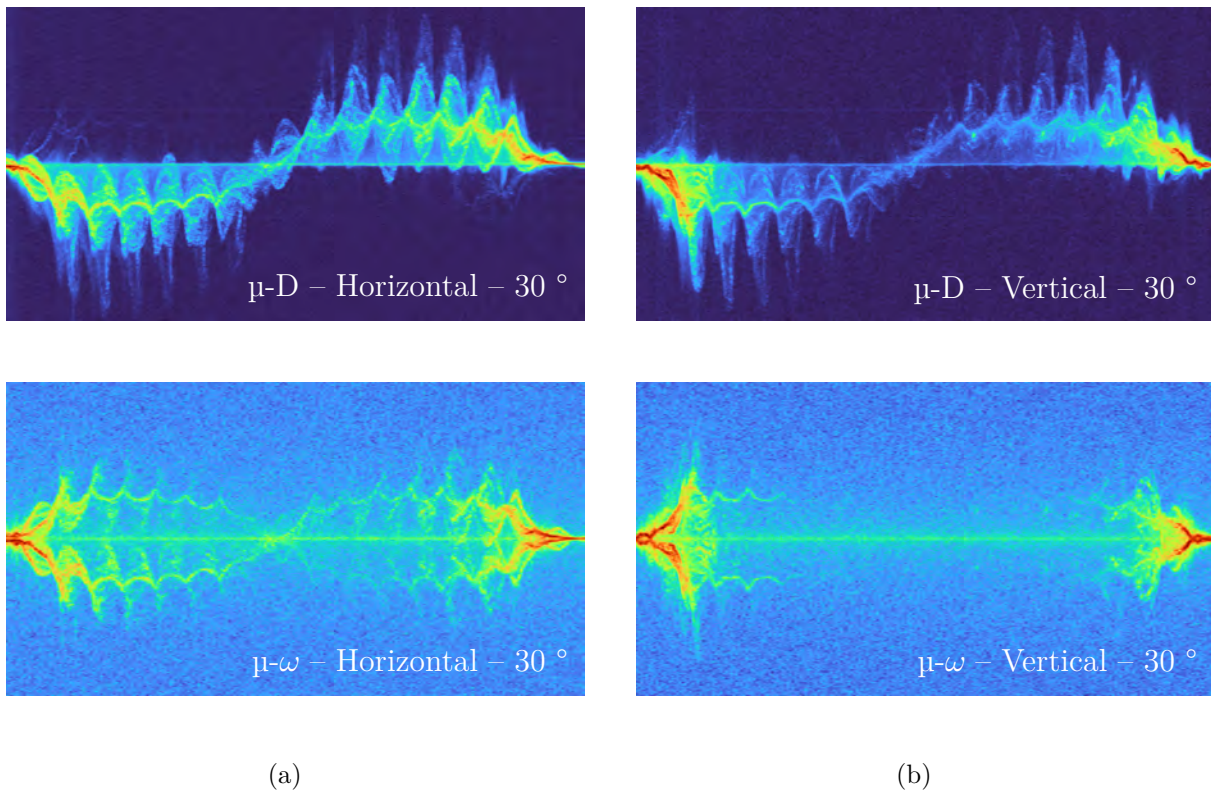


Figure 6.6: Comparison between vertical and horizontal radar polarization for walking in 30° .

between the vertical and horizontal orientations for both the LOS and the 30° from the LOS. As shown in Figures 6.5, for the LOS, the vertical orientation is showing a better SNR for both the μ -D and the μ - ω signatures, which is expected as most of the antenna power is utilized in the vertical scanning and reflected back from multiple points of the human body. On the other hand, the horizontal orientation is showing a better performance for the 30° direction as shown in Figure 6.6. Since, one of the main targets of the study is achieving an enhanced performance for the human-ID on multiple aspect angles, the horizontal orientation was selected.

The radar height was the second parameter that was taken into consideration, as most of the μ -D signatures is induced due to the lower part of the human body. This aspect is presented in Chapter 4 based on the results that we have presented in Abdulatif et al. 2017 and Abdulatif et al. 2018. To ensure that the captured micro motion signatures are reflecting the micro behavior of the whole body, the hand height is taken as a reference for the radar height. According to DIN 33402-2 and DIN EN ISO 14738, this is $\approx 0.75m$ for women and $\approx 0.82m$ for men as described in *Körpermasse – Überblick*. These defined

heights refer to people aging from 18-65 years old in Germany. Thus, the radar was set up at $\approx 0.7m$ to ensure meaningful captured micro motion signatures for the short people as long as the taller ones.

Rotating the antenna has also been tested to show a direct impact on the quality of the captured micro motion signatures as shown in Figure 6.7. Rotating the antenna upside-down has shown a better SNR performance than the upright one as can be seen in column (a), in which the lower body part is captured better. Moreover, the ghosting behavior of the torso's motion is more noticeable for the (a) column compared to the (b) column. Thus, the radar is set up for horizontal orientation at a height of $\approx 0.7m$ when rotated upside-down.

The radar is parameterized, in which the captured μ -D and μ - ω signatures are fully reflecting the micro motions behavior. As has been explained in Section 2.1.2, there is a trade-off between both the \mathbf{v}_{res} and the \mathbf{R}_{res} as shown in Figure 2.2. The radar is parameterized to be able to fully detect the induced velocities due to the human body, in which the walking velocity is estimated to be of $\approx 1.5\mathbf{m/s}$ as has been explained in Willen et al. 2013. However, the feet have been proved to induce a velocity component of 4 multiples of the walking velocity as has been explained in Chapter 4, and our study published in Abdulatif et al. 2017. Thus, \mathbf{T}_p has been calculated to ensure a reasonable \mathbf{v}_{max} and an \mathbf{N}_p has been tested to result a meaningful μ -D signature with a $\mathbf{v}_{\text{res}} \approx 0.5\mathbf{cm/s}$.

In order to support such fast chirp parametrization, the radar is parametrized with a small bandwidth of 250 MHz, in which a $\mathbf{R}_{\text{res}} \approx 60\mathbf{cm}$ is revealed that is reasonable for the average length of a single step. The last two parameters, which are the \mathbf{N}_s and the

Table 6.1: The parameterization for the utilized MIMO radar.

Radar Parametrization			Attributes
Start frequency	\mathbf{f}_{strt}	76.875 GHz	
Bandwidth	\mathbf{B}	250 MHz	$\mathbf{R}_{\text{res}} = 59.96 \text{ cm}$
Chirp repetition interval	\mathbf{T}_p	80 s	$\mathbf{v}_{\text{max}} = 6.08 \frac{\text{m}}{\text{s}}$
Number of samples per chirp	\mathbf{N}_s	128	$\mathbf{R}_{\text{max}} = 19.19 \text{ m}$
Number of Tx antennas	\mathbf{N}_{Tx}	2	$\mathbf{v}_{\text{res}} = 47.53 \frac{\text{mm}}{\text{s}}$
Number of Rx antennas	\mathbf{N}_{Rx}	16	$\theta_{\text{res}} = 0.0625^\circ$
(Ethernet data rate)			$\mathcal{R} = 409.6 \frac{\text{Mbit}}{\text{s}}$

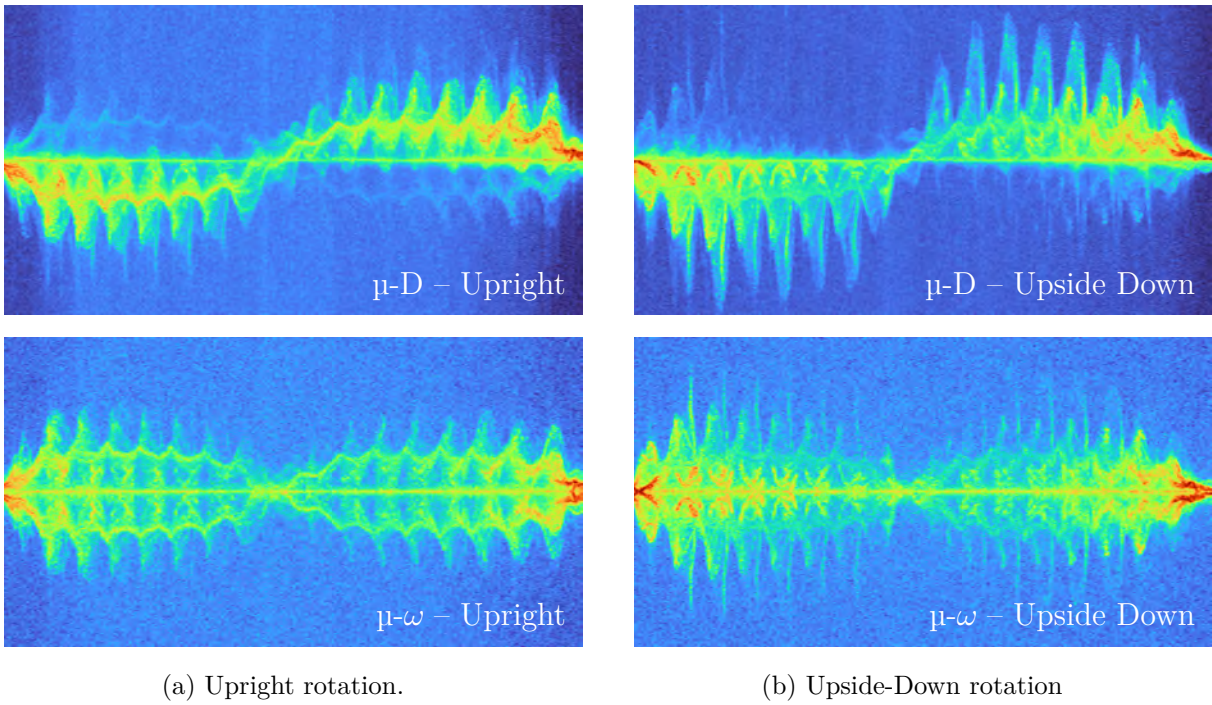


Figure 6.7: The effect of the rotation of the radar sensor frontend on the captured spectrograms.

$\mathbf{N}_{\mathbf{R}\mathbf{x}}$, were calculated carefully in order not to exceed the maximum possible data rate of the utilized radar module. The presented data path is based on Ethernet connection that is capable of data speed of 500 Mbit/s. However, the realistic data speed is tested to be $\approx 410\text{Mbit/s}$. The data rate \mathcal{R} is affected by both parameters the $\mathbf{N}_{\mathbf{s}}$ and the $\mathbf{N}_{\mathbf{R}\mathbf{x}}$ as the utilized radar is a MIMO-FMCW module. All the available 16 $\mathbf{N}_{\mathbf{R}\mathbf{x}}$ antennas are used to ensure capturing $\mu\text{-}\omega$ signature with the best SNR as has been explained in Section 2.1.3. Hence, an $\mathbf{N}_{\mathbf{s}}$ is calculated to fulfill the maximum achievable \mathcal{R} with an $\mathbf{R}_{\max} \approx 20\text{m}$, which is satisfactory for our indoor applications. The entire radar sensor parameterization is given in table 6.1. The left half states the parameters and their exact values, and the right half shows the resulting resolutions and maximum ratings. These are calculated with equations (2.16) through (2.15) and (2.31).

6.3.3 Experimental Setup

As described in Tahmoush & Silvius 2009, walking with an aspect angle to the radar LOS degrades the quality of the captured micro-motion signatures. In order to test the viability of the proposed triplet-loss method, walking at different aspect angles was considered.

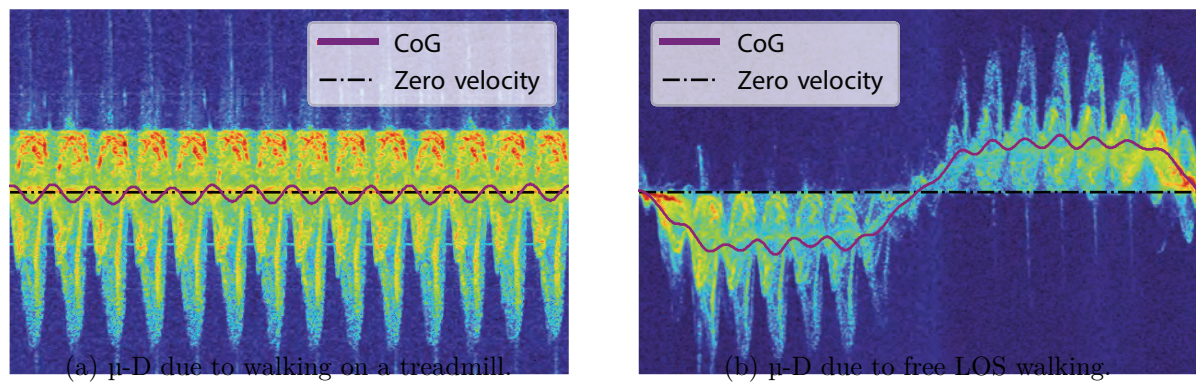


Figure 6.8: Comparison of the μ -D captured due to walking activity.

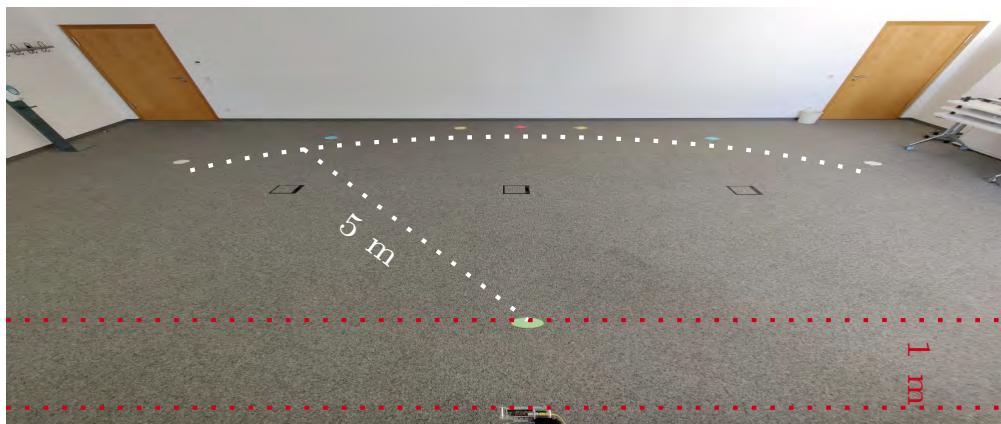


Figure 6.9: Experimental setup for the non-constrained walking.

Accordingly, the 22 subjects walked along predefined linear paths using their unrestricted, natural gait at the aspect angles $[\pm 50^\circ, \pm 30^\circ, \pm 10^\circ, 0^\circ]$ with regard to LOS. As shown in Figure 6.9, each path begins at a distance of 1 m from the radar and has a radial length of 5 m. The total recording duration per target is 6 min. The primary objective is to achieve human-ID on a larger FOV with improved unrestricted walking. The entire dataset is divided into 70% for training and 30% for validation. A separate test dataset consisting of 70 s per subject was also collected. The employed radar is parameterized according to Table 5.1. For comparison with our first feasibility study, which was published in Abdulatif et al. 2019 and described in Chapter 5, a treadmill-based dataset was collected on the same individuals using the same setup and aspect angles.

As depicted in Figure 6.8, the radar configuration is capable of capturing meaningful μ -D signatures. As depicted in Figure 6.8, relying solely on the treadmill configuration yields realistic signatures that will vary between individuals. However, the other major

effects that are induced by the free-walking scenarios, such as the SNR effect in relation to the distance from the radar, are not captured. Moreover, as depicted in Figure 6.8, the forced motion of the treadmill will induce a similar behavior for all the half gait cycles that occur during walking. The data for each subject are collected for the duration of time allotted, and then the STFT analysis is performed as a post-processing step. The adaptive slicing discussed in Section 3.3 is then applied.

6.3.4 Results and Discussion

Spectrogram Constellations

Human-ID has been primarily based on the μ -D signature. Departing from this, the presented method permits training on different spectrogram constellations. The unrestricted walking dataset is used for three training sessions. Initially, training on the μ -D and μ - ω are conducted independently. The feasibility of cohesive μ -D and μ - ω concatenation is subsequently evaluated. Each training includes all 22 classes. The respective error rates are 15.2% for the μ -D, 11.2% for the μ - ω , and 10.1% for concatenating both. The μ - ω spectrogram appears to capture more discriminative features than the well-explored μ -D spectrogram, despite the fact that both spectrograms display information that is closely related. The concatenated training reduces error relative to the μ -D and ω only cases by 33.6% and 9.8%, respectively. This demonstrates the effectiveness of spectrogram concatenation, which is employed in all subsequent experiments.

Dataset Complexity

According to previous research presented in Cao et al. 2018; Qiao et al. 2020, and Vandersmissen et al. 2018 the number of classes in the dataset has a significant effect on the accuracy of DCNN architectures that are based on conventional DL techniques. These difficulties arise from their fundamental working principle, which is to immediately classify a query sample into a particular class. Since FSL learns a distance function over given samples in an embedding space, it is possible to extract meaningful features as long as naturally related data is presented. Thus, the performance of classification is less affected by the dataset complexity. This behavior can be observed on the presented DCNN

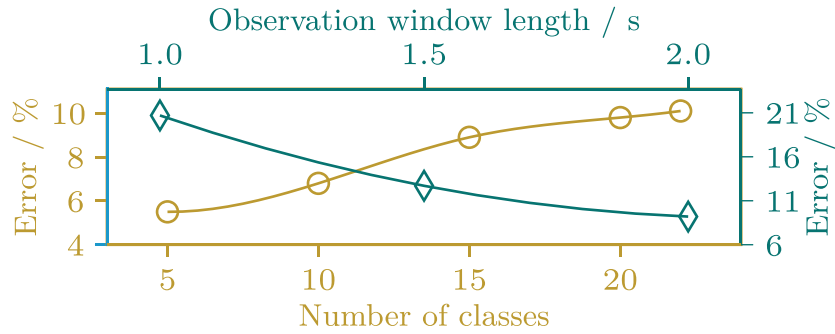


Figure 6.10: Error rate performance due to the dataset complexity and the observation window length.

architecture when given concatenations of coherent μ -D and μ - ω spectrograms from the dataset of unconstrained movement of 5 to 22 different targets, as depicted in Figure 6.10. The presented method produces an error of 5.5% for five different classes in the test set, which increases to 10.1% for 22 different classes. While the error rate still increases with increasing dataset complexity, the correlation between the two is less distinct than in previous research.

Dataset Size

Since FSL is primarily concerned with the limited availability of data, the presented DCNN architecture is evaluated on various data sizes. Recent studies collect on average ≈ 20 min of data per subject. Consequently, our collected data regarding unrestricted walking for 6 min per target is foreseen as limited. In order to test the system's capabilities, the dataset size is reduced further. Table 6.2 provides an overview of the training dataset sizes and associated error rates. Only 1 min of training data would result in an error rate of 24.9%, according to reports. By doubling this amount to 2 min, however, an error rate of 11.3% is achieved, which was not significantly impacted by larger training data sizes of 3.15 min and 4.20 min. This demonstrates the value of the proposed FSL method, which provides robust accuracy on unseen data despite a brief training period.

Table 6.2: Performance of the error rate due to training dataset sizes.

	Amount of data / min			
	1.05	2.10	3.15	4.20
Error rates / %	24.9	11.3	10.4	10.1

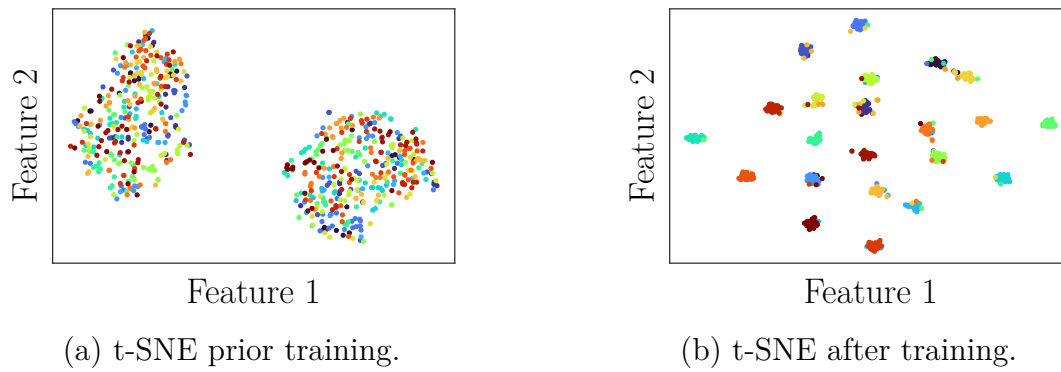


Figure 6.11: t-SNE visualizations of the embedding space.

Observation Window

To demonstrate the viability of half-gait-cycle-based human identification, a series of tests are conducted with varying observation window lengths. In order to minimize the impact of varying dataset sizes, the overlap between successive windows is modified throughout the experiments to maintain roughly the same size for each dataset. The results are shown in Figure 6.10. In addition to the proposed half gait cycle method, it can be observed that error rates decrease as the length of the observation window increases. This is because fixing the length of the observation window cannot account for the shape consistency of extracted signatures. This inconsistency in signatures increases the error rate for smaller observation windows. The utilized μ -D spectrogram adaptive segmentation results in the accurate extraction of the consistent shape of the captured gait cycles. As described previously, our proposed framework for the half gait cycle $\approx 0.5s$ yields a classification error comparable to the fixed observation window of 2 s.

6.3.5 Embedding Space

The performance of the FSL architecture highly relies on the quality of its embedding space. Thus, a t-SNE representation is used for visualization as shown in Figure 6.11. The various hues represent the 22 distinct classes. Before training, two widely dispersed but well-separated clusters of randomly mixed classes can be observed. The direction of relative movement towards or away from the radar causes the two clusters. Intra-class variances are greater than inter-class variances, as indicated by the randomly mixed samples. Individual

classes are tightly clustered and well-separated in the embedding space following training, reflecting the classification error of 11.3%.

6.4 Activity Recognition task

Since the μ -D signature analysis is widely used in the research domain of human activity recognition, such task was investigated by us for testing the robustness of our proposed algorithm for free walking human identification. Our study including the experimental setup, the utilized neural network approach and the results are discussed and published in Aziz et al. 2022. Similar architecture is used, since both the human identification and activity recognition tasks suffer from the same limitations as both are sub topics under the main aspect of using the radar micro motion signatures for human detection. Accordingly, the same architecture is used, which can be summarized in the following points:

- A MIMO radar is used in order to be capable of capturing from multiple aspect angles and the study can be extended to include detecting humans simultaneously at the same time.
- The data were collected from different aspect angles and different persons were included to ensure data generalization.
- The triplet loss approach introduced in Section 6.1 is used.
- The μ - ω is combined with the μ -D signatures for enhanced activity recognition.
- The adaptive slicing technique, which is used for extracting single occurrences

The study addressed ten activities that are very common to occur in homes. The experimental setup is presented in Section 6.4.1. The results discussion is presented in Section 6.4.3, which is divided into three main sub sections discussing the main aspects that are addressed by us, which are the adaptive slicing, combining both the μ -D and μ - ω signatures, and the few-shot learning approach.

6.4.1 Experimental Setup

The presented study is based on ten activities that are essential for home applications and have been extensively studied in other studies such as Erol & Amin 2019 and Gurbuz &

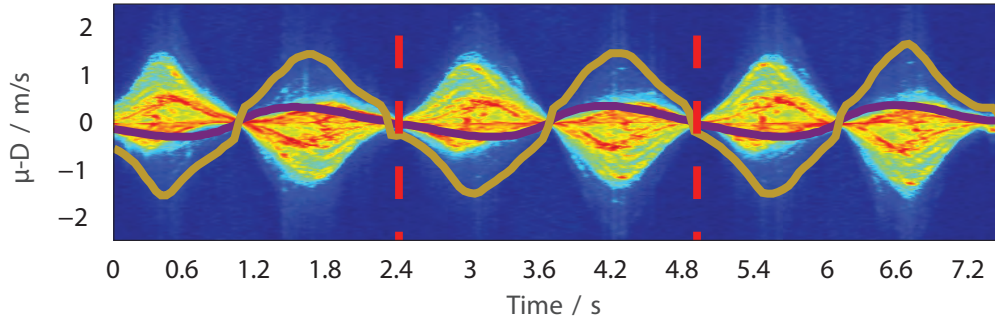


Figure 6.12: An example for the adaptive extraction of single instant of bending activity.

Amin 2019. Figure 6.12 displays both the μ -D and μ - ω signatures for the various activities, which can be classified as follows:

- Sitting, standing, and walking, which typically occur sequentially and involve the participation of the entire body.
- Gesture and reaching, which are essential for smart home applications, rely solely on the arm.
- Bending, kneeling, and jumping, all of which involve the lower body.
- As fall detection is vital for the elderly, the elderly must be able to detect and recover from falls.

To ensure generalization, signatures were collected from eight different human subjects ranging in height range 155-192 cm and weight range 70-105 Kg. Each subject performed each activity in front of the radar for 30 s at a distance range of 1-3 m, with the exception of walking, which was performed at a distance range of 1-5 m. Due to the fact that detection at side angles diminishes the radar signals as explained in Tahmoush & Silvius 2009, the data was collected at various side angles, including LOS, $\pm 30^\circ$ and $\pm 50^\circ$, to ensure ubiquitous recognition. To ensure data generalization and avoid redundancy, the detection range and aspect angles for each target were not fixed. Consequently, each subject was required to perform each task at four training locations and two test locations, with the line of sight and randomly selected aspect angles of $\pm 30^\circ$ and $\pm 50^\circ$. Both the μ -D and μ - ω signatures were recorded, and then the spectrograms were separated into single instances of each activity using various techniques. The sitting and standing activities were recorded sequentially, and the recording time was doubled to ensure that all classes received comparable data. Due to the complexity of the experiment, falling and standing

from falling data were collected sequentially and resulted in fewer data samples than other activities. The final two activities were not collected from different aspect angles; rather, each subject sequentially fell and stood on all four sides of the body (front, back, left and right sides). The duration of 30 s is the same for each record.

The measurements are based on the same MIMO radar module with the same parameterization that was presented in Section 6.3.2. Thus, the maximum achievable angular resolution is ensured, and activities such as gesture and reaching can be detected, which require high angle and range resolutions. The radar’s parameters are specified in Table+6.1. The final two activities, falling and getting up after falling, are collected to test the viability of the metric learning strategy for FSL. As shown in Figure 6.14, the full sequence of falling and standing lasts for $\approx 5s$, whereas a single occurrence of other activities lasts $\approx 1.5s$. Consequently, both the falling and standing from falling activities couldn’t be included in the training process from the beginning as they will induce much less samples compared to other activities.

6.4.2 Slicing Techniques

To extract a single occurrence of each activity from the captured spectrograms, three techniques were applied. First, a 1.5 s fixed-time window is employed, as this was determined to be the average duration of a single cycle for the majority of the activities. As the different activities are anticipated to occur sequentially and a single slice may contain half-cycles from two distinct activities or a period of inactivity, such a technique is not applicable to realistic situations. For improved data analysis, a 1.5 s sliding window was applied. This technique functions as a data augmentation technique in which the network is guaranteed to observe the full cycle of each activity from multiple perspectives, and the size of the training data is increased. The third technique is adaptive slicing, which accurately extracts a single instance of each activity, as depicted in Figure 6.12. It detects the behavior of the CoG in the captured spectrogram using the algorithm presented in Seifert et al. 2020, and described in Section 3.3. The CoG is used to determine the movement direction, which is then combined with the envelope of the motion cycles to determine the slicing locations. In addition, the frequency of slicing is adjustable and fixed for each activity to ensure generalization. The metric learning approach is highly

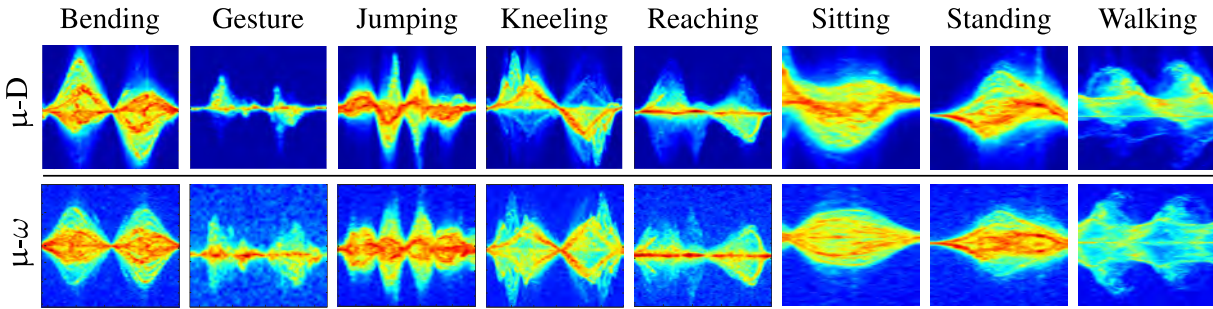


Figure 6.13: The μ -D and μ - ω signatures for the main 8 activities.

applicable to applications such as fall detection where only a small training dataset is available and FSL is required. For the purpose of this study, the model is initially trained on the primary eight activities that are shown in Figure 6.13 and can adapt to two unseen activities without additional training that are shown in Figure 6.14.

6.4.3 Results and Discussion

Slicing Techniques

The performance of the three slicing techniques is compared in terms of the size of the training data, the consistency of the data, and the classification precision. The proposed network is trained on the resulting μ -D sliced signatures, as this training method is the most prevalent. All reported classification accuracy for all activities is based on the test dataset. As mentioned in the previous section, the test dataset was collected independently. As shown in Table 6.3, the minimum number of samples is produced by the first technique of applying a constant window of 1.5 s. Despite the fact that 1.5 s is estimated to be the average duration of a single occurrence for all activities, this technique does not demonstrate a high level of consistency for the sliced samples, particularly for the sitting and standing that were collected sequentially. Consequently, this technique yields the lowest classification accuracy of 53.43%.

Using a sliding window of 1.5 s with an overlap of 80%, the second method yields a total of 6288 augmented samples. This technique has demonstrated an overall accuracy of 78.22%. This accuracy is nearly identical to that of the augmented dataset presented in Erol & Amin 2019 for comparable tasks and training dataset sizes. Moreover, the performance of our metric learning method is comparable to that of the study presented in

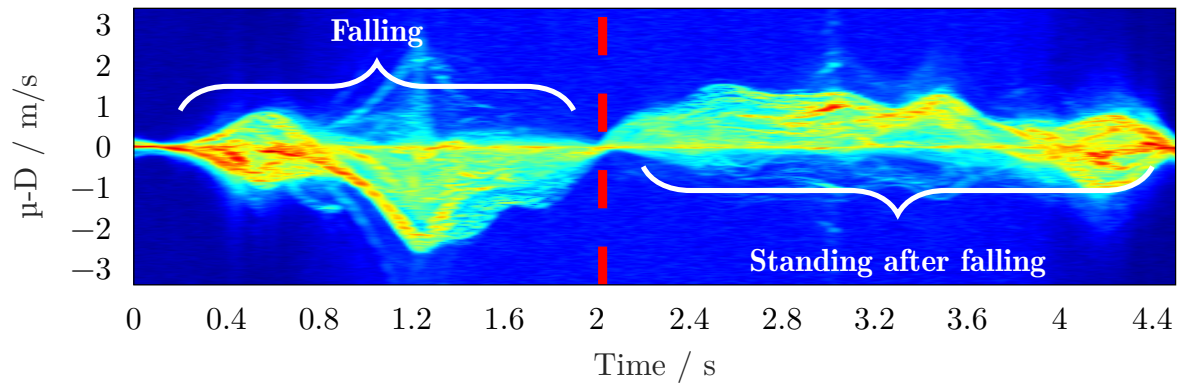
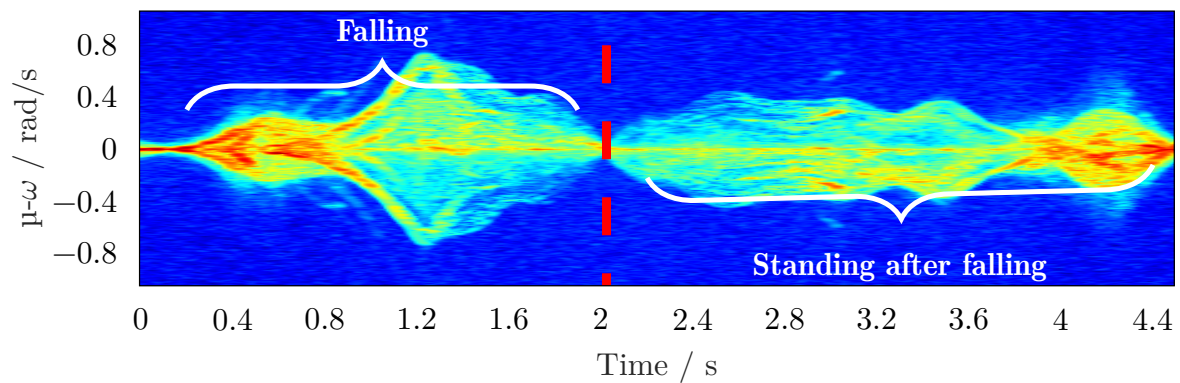
(a) μ -D signature.(b) μ - ω signature.

Figure 6.14: Falling/Standing from falling signatures.

Li et al. 2019, which employs SVM with a similar sliding window algorithm. Nonetheless, such a study has a real-time limitation due to the sliding window's 3-5 s duration and 70% overlap. Our study and the study presented in Li et al. 2019 indicate greater confusion between sitting and standing, which are captured sequentially as they are expected to occur sequentially. Consequently, adaptive slicing is necessary to reduce this classification ambiguity.

The third technique simulates the behavioral slicing by introducing the concept of adaptive slicing for each activity. This technique achieved the highest classification accuracy of 82.14% on a training dataset of 680 samples, which is comparable in size to the dataset utilized by the transfer learning approach described in Gurbuz & Amin 2019. Despite the fact that our study was not conducted on all 12 activities presented in Gurbuz & Amin 2019, our proposed metric learning approach provides a good indication based on the limited dataset available. In addition, the activities were captured from a variety of perspectives, as opposed to the majority of previous studies that only considered LOS.

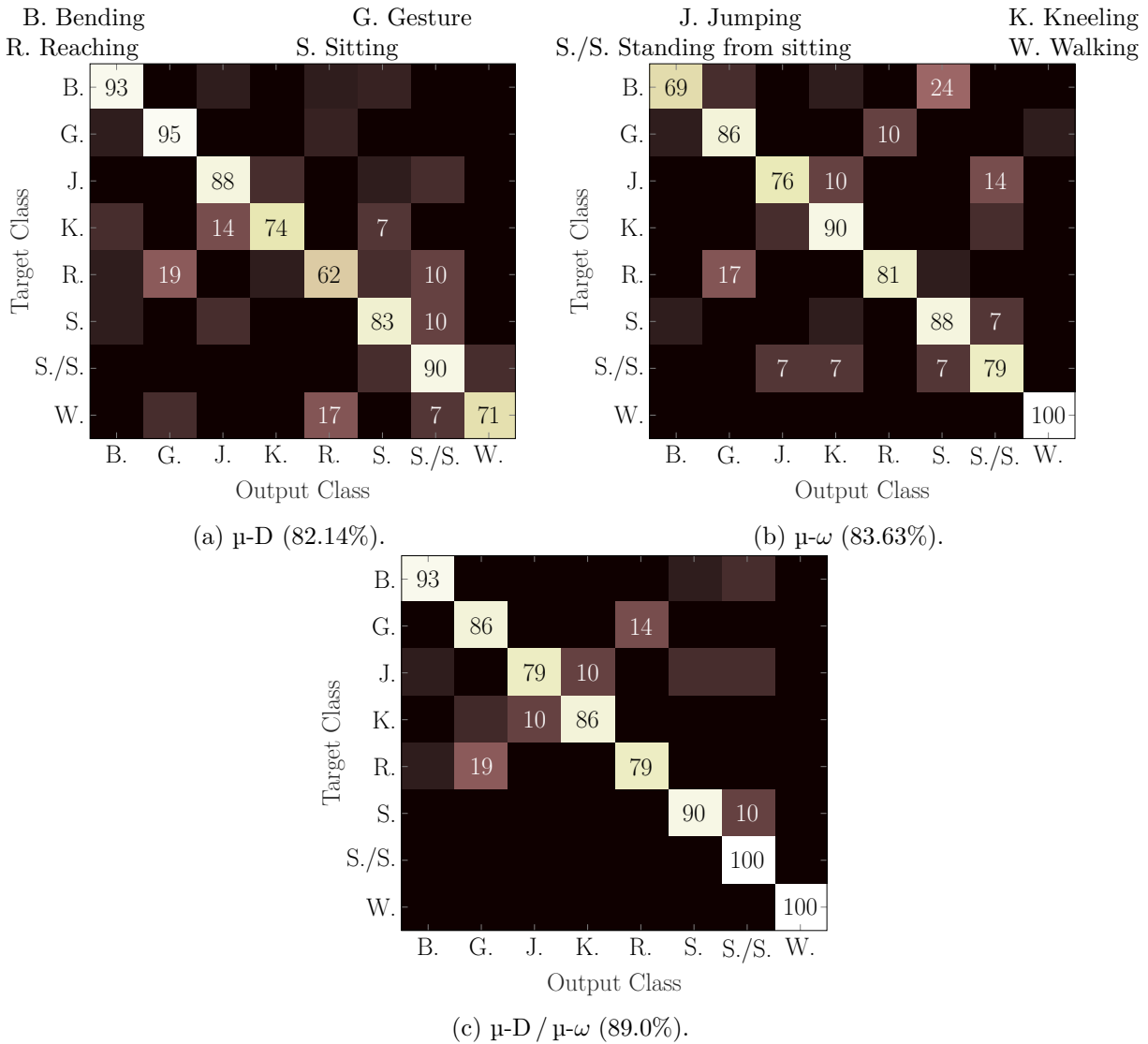


Figure 6.15: Confusion matrix for all the micro-motion signatures constellations.

Moreover, the adaptive slicing technique demonstrates a high degree of applicability for both standing and sitting activities, which can be extended to multiple activity streams.

Micro-motion Signatures Constellations

Adaptive slicing is used to slice both the μ -D and μ - ω signatures, as both are timely-synchronized. Training on the μ - ω yielded an overall accuracy of only 83.63%. As shown in Figure 6.15, relying solely on the μ -D or the μ - ω will result in a decrease in classification for certain activities. In the case of the μ -D, for instance, jumping is confused with kneeling, reaching with gesture, and standing with sitting. In contrast, the μ - ω confuses bending

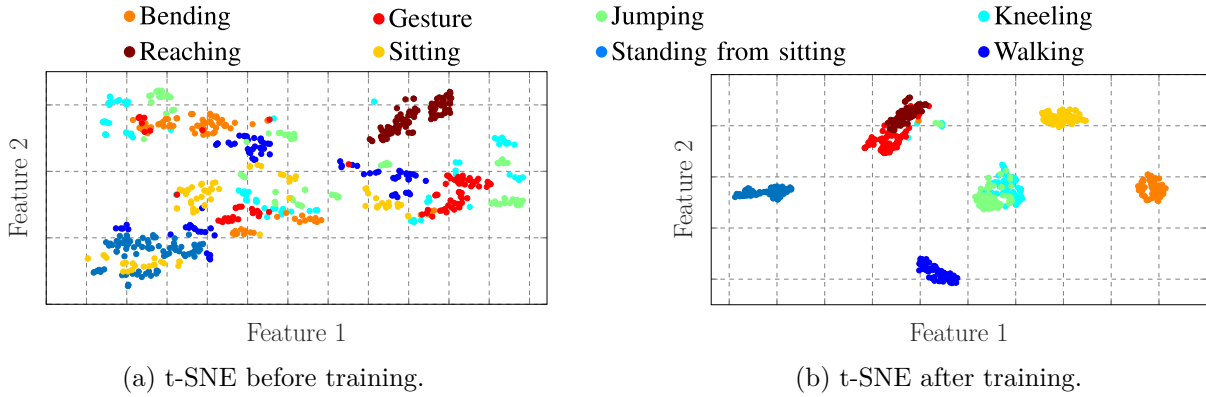


Figure 6.16: t-SNE for the adaptive-sliced and combined μ -D and μ - ω signatures.

with sitting, and jumping with both kneeling and standing from sitting. Accordingly, another training was conducted based on the combination of the μ -D and μ - ω . As depicted in Figure 6.15c, this training yielded the best performance of 89%, with classification accuracy for each class exceeding $\geq 80\%$.

The quality of the embedding space influences the effectiveness of metric learning. For the embedding space visualization illustrated in Figure 6.16, the t-SNE approach can be used. As depicted in Figure 6.16a, the dataset is complex because all the classes are confused with one another. As illustrated in Figure 6.16b, our proposed network resulted in well-separated classes. The t-SNE representation is consistent with the confusion matrix of the combined micro-motion signatures, as displayed in Figure 6.15c. The bending, walking, standing, and sitting form distinct clusters, whereas each gesture forms a cluster with reaching and each kneeling with jumping. Consequently, it is evident that the metric learning strategy could reduce the inter-class variances.

Table 6.3: The classification performance due to different slicing techniques.

	Slicing technique		
	Time (discrete)	Time (sliding)	Adaptive
Samples/Class	60	786	85
Total data	472	6288	680
Accuracy / %	53.43	78.22	82.14

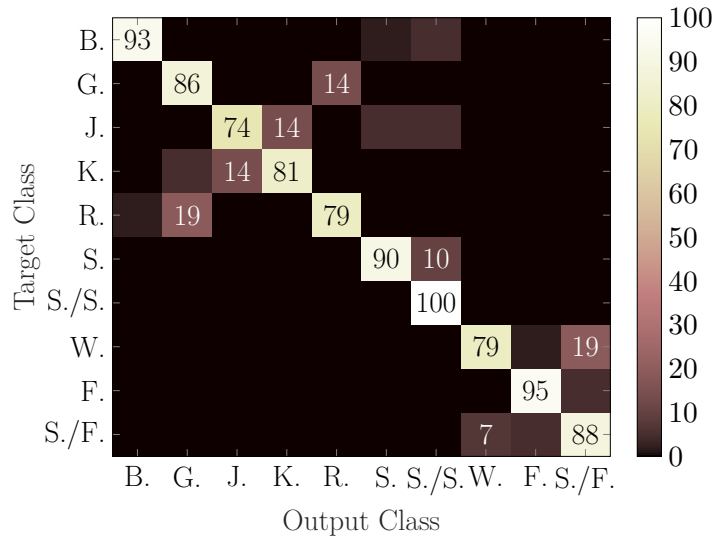


Figure 6.17: Confusion matrix after applying FSL including the 10 activities.

Few-Shot Learning

FSL has demonstrated acceptable performance for human identification based on walking activity using μ -D as presented in Ni & Huang 2021 and Yang et al. 2019. The task of human identification necessitates a large amount of data, as classification must be performed on the same activity for numerous subjects. FSL and transfer learning were proposed as methods for collecting massive datasets. Similarly, distinguishing between multiple activities is plagued by the same issues. In our study, we selected falling and standing from falling to test the feasibility of FSL. For the elderly, fall detection is of utmost importance, and radar has been foreseen as a viable solution because it protects privacy and can be installed anywhere, such as in bathrooms. Other studies have included falling as a primary activity, necessitating the collection of more data than our method. The adaptive slicing method has demonstrated remarkable consistency in slicing both activities sequentially.

The eight-activity-trained network is fine-tuned to include the two new classes. Using a combination of the μ -D and μ - ω signatures, the training was conducted on only 15 samples for each of the two classes. The test dataset was evenly distributed across the ten classes, with each class containing 42 samples. The classification accuracy of the FSL procedure was 86.42%. As illustrated in Figure 6.17, both falling and standing from falling can be classified with overall accuracy of 95% and 88%, respectively. Moreover, the addition of

the last two activities had no effect on the classification accuracy of the eight primary classes.

6.4.4 Summary

In this section, the robustness of the discussed architecture in Section 6.3 for the human identification task is extended for a use-case of activity recognition. The performance of combining a newly-formulated $\mu\text{-}\omega$ signature with the commonly-used $\mu\text{-D}$ signature is tested using a MIMO radar. The presented study is published in Aziz et al. 2022, and addressed eight critical activities for home use. The experimental setup is based on capturing data on multiple aspect angles ranging and LOS. The data was collected in 30 s intervals for different subjects, and three slicing techniques were tested: a constant window of 1.5 s, a sliding window with an overlap of 80%, and adaptive slicing based on activity behavior. Both the $\mu\text{-D}$ and the $\mu\text{-}\omega$ produced comparable classification results of 82.14% for the $\mu\text{-D}$ and 83.63% for the $\mu\text{-}\omega$. The best accuracy of 89% is obtained by combining both micro-motion signatures. Finally, with only 15 samples for each activity, the FSL was trained on falling and standing from falling, yielding a classification accuracy of 86.42% for the ten activities.

7 Summary

In this thesis, the use of radar sensors for indoor human detection and identification is examined. The primary focus of the presented methods is a short-range analysis ($\leq 10m$). Investigating the μ -D signatures has started by developing a novel real-time μ -D limb decomposition technique for a walking human. A combination of two simulations is used to model a walking human target approaching a radar. The foundation of the initial simulation is an empirical mathematical model. In the second simulation, infrared cameras are used to track the positions of infrared-reflective markers placed on a real walking subject. Independent of a particular walking target range scenario, power-free μ -D and mean-free μ -R are deemed to be the most relevant classification features. The body is subsequently divided into four primary classes (base, arms, legs, and feet). Body parts are grouped into these four classes based on their relative velocity trajectories. Over time, the μ -D and μ -R are extracted from the employed model and then used as features for a decision tree classifier. The gathered characteristics are used to train the decision tree classifier, and a new data set is generated for validation. Implementing an outliers removal algorithm improves the true classification rates of the classifier. The full study is published in Abdulatif et al. 2017. The research outcome was examined on a human-robot classification use case. The CNN has shown the best performance compared to other learning techniques with a classification accuracy of $\approx 99\%$, and the results were published in Abdulatif et al. 2018

As the second main focus of this thesis, human fingerprinting while walking is examined. The identification of humans is considered the most difficult task, and the results can then be generalized to other applications of human recognition. Experimental research illustrating the effect of human body characteristics on the μ -D signatures of walking individuals is provided. Using a CW radar, the μ -D signatures of 22 participants of various genders and body types walking on a treadmill are measured. Consequently, the RCS and walking style of the moving subject are the primary factors that influence the measured μ -D signatures. The BMI is employed as a metric for the body volume of selected participants,

ranging from lean bodies between 18-20 kg/m² to obese levels at 37 kg². A CAE is utilized for encoding the μ -D signatures of the half gait into a latent space representation. To depict the latent space distribution in 2D, a nonlinear dimension reduction based on t-SNE is employed. On the basis of visual analysis, the BMI of the participants and their walking methods are found to have a direct effect on their μ -D signatures. A deep ResNet-50 architecture is trained on μ -D signatures for person identification. The presented approach accurately recognizes μ -D signatures with a 98 % accuracy at low SNR levels and an 84 % at varying SNR levels, with confusions occurring predominantly in participants with comparable BMIs. A full discussion, including the results of the feasibility study is published in Abdulatif et al. 2019

The results of the preliminary feasibility study elaborated on the key factors for increasing the number of included subjects, where the ResNet-50 network was presented as a remedy. In addition, it suggested that confusion may occur because of the similarity in body shape and demonstrated a direct correlation between body characteristics and walking style. However, this was constrained by the treadmill's motion. Consequently, the results and employed network architecture served as guidelines for the design of an experimental setup for unrestricted human recognition. The following study proposes a radar-based, inception-residual DCNN-based human identification framework. By leveraging FSL, the requirement for large training datasets is eliminated, and the triplet loss is successfully used to learn a discriminative embedding space on challenging data. In addition, the duration of individual data samples is reduced to half a gait cycle ($\approx 0.5s$). Each sample consists of time-synchronized μ -D and newly-measured μ - ω spectrograms, the latter of which reflects the micro-motion behavior of the angular velocity in the elevation plane. The half gait cycles of walking are segmented in an adaptive manner. Using only 2 min of training data, an error rate of 11.3% can be achieved. Implementing open-set capabilities and executing the framework in real-time will be the subject of additional investigation. Previous research identified $\pm 60^\circ$ as the dividing line between radial and tangential movements. This is considered a limitation of using micro-motion signatures for human identification. In these situations, a 3D point cloud radar can be useful for capturing more descriptive features that can aid in human identification in both static and dynamic scenarios.

The robustness and feasibility of the proposed metric learning approach are examined in human activity recognition. The performance of combining a newly-formulated $\mu\text{-}\omega$ signature with the commonly-used $\mu\text{-D}$ signature is evaluated using a MIMO radar. The research is conducted on eight essential activities that are highly probable to occur in homes. The experimental design is predicated on the collection of data for multiple aspect angles within the range of $[\pm 10^\circ\text{-}\pm 50^\circ]$ and LOS. Three slicing techniques, including a constant window of 1.5 s, a sliding window with an overlap of 80%, and adaptive slicing based on the behavior of each activity, were tested on the data collected for different subjects over periods of 30 s. Comparable classification performance was achieved by the $\mu\text{-D}$ and the $\mu\text{-}\omega$, where classification accuracy of 82.14% for the $\mu\text{-D}$ and 83.63% for the $\mu\text{-}\omega$ is achieved. Combining the two micro-motion signatures yields the highest accuracy of 89%. Finally, the FSL is trained on falling and standing from falling using only 15 samples for each activity, resulting in a classification accuracy of 86.4% for the ten activities.

7.1 Research Limitations and Future Work

The presented work is based on proposing a single walking step as the classification sample for all the presented studies, which introduces the results as feasible for real-time analysis since a single step lasts for $\approx 0.5\text{s}$. For all the analysis, the utilized $\mu\text{-D}$ signatures are demonstrated through post-processing, but it can be generated by concatenating the captured velocity profiles through frame-by-frame processing. However, the algorithm of adaptive slicing is not tested yet for such frame-by-frame processing. The feasibility of integrating the presented algorithms with a tracker is not yet tested. Demonstrating the $\mu\text{-D}$ and $\mu\text{-}\omega$ signatures through a tracker is expected to have a little different processing chain by considering only the bins that are expected to contain targets to save computational power. Other environmental conditions e.g., noise mitigation and the limited radar hardware capabilities, were not discussed in this dissertation as they are out of scope. Nevertheless, we have shared two studies about the $\mu\text{-D}$ de-noising and the radar super-resolution that were published in Abdulatif et al. 2019, and Armanious et al. 2019, respectively. Relying on the MIMO module enables the possibility of sensor fusion e.g., with any camera type for enhanced detection, and this can be the next step.

Bibliography

- Abdulatif et al. 2019** Abdulatif, Sherif; Armanious, Karim; Aziz, Fady; Schneider, Urs, Yang, Bin (eds.), 2019.
Towards adversarial denoising of radar micro-Doppler signatures.
IEEE International Radar Conference (RADAR).
ISBN: 1728126606
- Abdulatif et al. 2017** Abdulatif, Sherif; Aziz, Fady; Kleiner, Bernhard, Schneider, Urs (eds.), 2017.
Real-time capable micro-Doppler signature decomposition of walking human limbs.
IEEE Radar Conference (RadarConf).
ISBN: 1467388238
- Abdulatif et al. 2018** Abdulatif, Sherif; Wei, Qian; Aziz, Fady; Kleiner, Bernhard, Schneider, Urs (eds.), 2018.
Micro-doppler based human-robot classification using ensemble and deep learning approaches.
IEEE Radar Conference (RadarConf).
ISBN: 1538641674
- Addabbo et al. 2020** Addabbo, Pia; Bernardi, Mario Luca; Biondi, Filippo; Cimitile, Marta; Clemente, Carmine, Orlando, Danilo (eds.), 2020.
Gait recognition using fmcw radar and temporal convolutional deep neural networks.
IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace).
ISBN: 1728166365
- Ahmadi et al. 2015** Ahmadi, Amin; Destelle, Francois; Monaghan, David; Moran, Kieran; OConnor, Noel E.; Unzueta, Luis, Linaza, Maria Teresa (eds.), 2015.
Human gait monitoring using body-worn inertial sensors and kinematic modelling.
IEEE SENSORS.
ISBN: 1479982032
- Amin 2017** Amin, Moeness G., 2017.
Through-the-wall radar imaging. CRC press.
ISBN: 1439814775
- Amin & Erol 2018** Amin, Moeness G.; Erol, Baris (eds.), 2018.
Understanding deep neural networks performance for radar-based human motion recognition.
IEEE Radar Conference (RadarConf).
ISBN: 1538641674

- Anatomy 2016** Anatomy, Human, 2016,
URL: http://www.theodora.com/anatomy/the_pelvis.html
- Ankile et al. 2020** Ankile, Lars Lien; Heggland, Morgan Feet; Krange, Kjartan, 2020.
Deep Convolutional Neural Networks: A survey of the foundations, selected improvements, and some current applications.
arXiv preprint arXiv:2011.12960
- Armanious et al. 2019** Armanious, Karim; Abdulatif, Sherif; Aziz, Fady; Schneider, Urs, Yang, Bin (eds.), 2019.
An adversarial super-resolution remedy for radar design trade-offs.
IEEE European Signal Processing Conference (EUSIPCO).
ISBN: 9082797038
- Arrue et al. 2000** Arrue, Begoña C.; Ollero, Aníbal; de Dios JR, Matinez, 2000.
An intelligent system for false alarm reduction in infrared forest-fire detection.
IEEE Intelligent Systems and their Applications **15** (3), pp. 64–73
- Aziz et al. 2021** Aziz, Fady; Elmakhzangy, Bassam; Maufroy, Christophe; Schneider, Urs, Huber, Marco F. (eds.), 2021.
DimRad: A Radar-Based Perception System for Prosthetic Leg Barrier Traversing.
IEEE European Signal Processing Conference (EUSIPCO).
ISBN: 9082797062
- Aziz et al. 2022** Aziz, Fady; Metwally, Omar; Weller, Pascal; Schneider, Urs, Huber, Marco F. (eds.), 2022.
A MIMO Radar-Based Metric Learning Approach for Activity Recognition.
IEEE Radar Conference (RadarConf).
ISBN: 1728153689
- Bae et al. 2020** Bae, Youngseok; Yi, Minwoo; Shin, Jinwoo, Lee, Sang-Gug, 2020.
X-band photonic-based pulsed radar architecture with a high range resolution.
Applied Sciences **10** (18), pp. 6558
- Bartsch et al. 2012** Bartsch, A.; Fitzek, F.; Rasshofer, R. H., 2012.
Pedestrian recognition using automotive radar sensors.
Advances in Radio Science **10** (B. 2), pp. 45–55
- Basjaruddin et al. 2021** Basjaruddin, Noor Cholis; Syahbarudin, Febian; Sutjiredjeki, Ediana, 2021.
Measurement device for stress level and vital sign based on sensor fusion.
Healthcare Informatics Research **27** (1), pp. 11–18

- Bernard et al. 2013** Bernard, Vladan; Staffa, Erik; Mornstein, Vojtěch, Bourek, Aleš, 2013. Infrared camera assessment of skin surface temperature–effect of emissivity. *Physica Medica* **29** (6), pp. 583–591
- Bialer et al. 2021** Bialer, Oded; Jonas, Amnon; Tirer, Tom, 2021. Code optimization for fast chirp FMCW automotive MIMO radar. *IEEE Transactions on Vehicular Technology* **70** (8), pp. 7582–7593
- Błaszczuk et al. 2011** Błaszczuk, Janusz W.; Plewa, Michał; Cieślinska-Swider, Joanna; Bacik, Bogdan; Zahorska-Markiewicz, Barbara, Markiewicz, Andrzej, 2011. Impact of excess body weight on walking at the preferred speed. *Acta neurobiologiae experimentalis* **71** (4), pp. 528–540
- Boulic et al. 1990** Boulic, Ronan; Thalmann, Nadia Magnenat; Thalmann, Daniel, 1990. A global human walking model with real-time kinematic personification. *The visual computer* **6** (6), pp. 344–358
- Boye et al. 2014** Boye, Jenny Katrine; Thomsen, Maj Halling; Pfau, Thilo, Olsen, Emil, 2014. Accuracy and precision of gait events derived from motion capture in horses during walk and trot. *Journal of Biomechanics* **47** (5), pp. 1220–1224
- Campbell & Ahmad 2020** Campbell, Christopher; Ahmad, Fauzia (eds.), 2020. *Attention-augmented convolutional autoencoder for radar-based human activity recognition*. IEEE International Radar Conference (RADAR). ISBN: 1728168139
- Cao et al. 2018** Cao, Peibei; Xia, Weijie; Ye, Ming; Zhang, Jutong, Zhou, Jianjiang, 2018. Radar-ID: Human identification based on radar micro-Doppler signatures using deep convolutional neural networks. *IET Radar, Sonar & Navigation* **12** (7), pp. 729–734
- Cao et al. 2019** Cao, Peibei; Xia, Weijie; Li, Yi, 2019. Heart-ID: Human identification based on radar micro-doppler signatures of the heart using deep Learning. *Remote Sensing* **11** (10), pp. 1220

- Caruana et al. 2008** Caruana, Rich; Karampatziakis, Nikos; Yessenalina, Ainur, 2008.
An empirical evaluation of supervised learning in high dimensions,
Pp. 96–103
- Chen & Ling 2002** Chen, Victor C.; Ling, Hao, 2002.
Time-frequency transforms for radar imaging and signal analysis. Artech house.
ISBN: 1580535496
- Chen et al. 2006** Chen, Victor C.; Li, Fayin; Ho, S-S, Wechsler, Harry, 2006.
Micro-Doppler effect in radar: phenomenon, model, and simulation study.
IEEE Transactions on Aerospace and Electronic Systems **42** (1), pp. 2–21
- Chen 2008** Chen, Victor C., 2008.
Doppler signatures of radar backscattering from objects with micro-motions.
IET signal processing **2** (3), pp. 291–300
- Chen 2019** Chen, Victor C., 2019.
The micro-Doppler effect in radar. Artech house.
ISBN: 1630815489
- Chen et al. 2014** Chen, Xiaoli; Leung, Henry; Tian, Mao, 2014.
Multitarget detection and tracking for through-the-wall radars.
IEEE Transactions on Aerospace and Electronic Systems **50** (2), pp. 1403–1415
- Cheng et al. 2020** Cheng, Can; Ling, Fei; Guo, Shisheng; Cui, Guolong; Jian, Qiang; Jia, Chao, Ran, Qingxin (eds.), 2020.
A real-time human activity recognition method for through-the-wall radar.
IEEE Radar Conference (RadarConf).
ISBN: 172818942X
- Chetty et al. 2017** Chetty, Kevin; Chen, Qingchao; Ritchie, Matthew, Woodbridge, Karl, 2017.
A low-cost through-the-wall FMCW radar for stand-off operation and activity detection.
10188, pp. 65–73
- Chiao et al. 1997** Chiao, Richard Y.; Thomas, Lewis J.; Silverstein, Seth D. (eds.), 1997.
Sparse array imaging with spatially-encoded transmits.
IEEE Ultrasonics Symposium Proceedings. An International Symposium.
ISBN: 0780341538

- Cho et al. 2014** Cho, Hyunggi; Seo, Young-Woo; Kumar, B. VijayaV.K., Rajkumar, Ragunathan Raj (eds.), 2014.
A multi-sensor fusion system for moving object detection and tracking in urban driving environments.
IEEE International Conference on Robotics and Automation (ICRA).
ISBN: 1479936855
- Cho et al. 2007** Cho, Namjun; Yoo, Jerald; Song, Seong-Jun; Lee, Jeabin; Jeon, Seonghyun, Yoo, Hoi-Jun, 2007.
The human body characteristics as a signal transmission medium for intrabody communication.
IEEE transactions on microwave theory and techniques **55** (5), pp. 1080–1086
- Choi et al. 2011** Choi, JunHyeok; Park, JeongHo; Yeom, DongJin (eds.), 2011.
High angular resolution estimation methods for vehicle FMCW radar.
IEEE CIE International Conference on Radar.
ISBN: 142448443X
- Cui & Dahnoun 2021** Cui, Han; Dahnoun, Naim, 2021.
High precision human detection and tracking using millimeter-wave radars.
IEEE Aerospace and Electronic Systems Magazine **36** (1), pp. 22–32
- Cutting & Kozlowski 1977** Cutting, James E.; Kozlowski, Lynn T., 1977.
Recognizing friends by their walk: Gait perception without familiarity cues.
Bulletin of the psychonomic society **9** (5), pp. 353–356
- Da Andrade 1959** Da Andrade, E. C. N., 1959.
Doppler and the Doppler effect
- De Jesus Romero-Troncoso 2016** De Jesus Romero-Troncoso, Rene, 2016.
Multirate signal processing to improve FFT-based analysis for detecting faults in induction motors.
IEEE Transactions on industrial informatics **13** (3), pp. 1291–1300
- Denavit & Hartenberg 1955** Denavit, Jacques; Hartenberg, Richard S., 1955.
A kinematic notation for lower-pair mechanisms based on matrices
- Dickmann et al. 2015** Dickmann, Juergen; Appenrodt, Nils; Klappstein, Jens; Bloecher, Hans-Ludwig; Muntzinger, Marc; Sailer, Alfons; Hahn, Markus, Brenk, Carsten, 2015.
Making berth see even more: Radar contribution.
IEEE Access **3**, pp. 1233–1247

- Dong et al. 2020** Dong, Shiqi; Xia, Weijie; Li, Yi; Zhang, Qi, Tu, Dehao, 2020. Radar-based human identification using deep neural network for long-term stability. *IET Radar, Sonar & Navigation* **14** (10), pp. 1521–1527
- Du et al. 2013** Du, Lan; Wang, Baoshuai; Li, Yanbing, Liu, Hongwei, 2013. Robust classification scheme for airplane targets with low resolution radar based on EMD-CLEAN feature extraction method. *IEEE sensors journal* **13** (12), pp. 4648–4662
- Du et al. 2015** Du, Lan; Wang, Baoshuai; Wang, Penghui; Ma, Yanyan, Liu, Hongwei, 2015. Noise reduction method based on principal component analysis with beta process for micro-Doppler radar signatures. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **8** (8), pp. 4028–4040
- Dubois & Charpillet 2013** Dubois, Amandine; Charpillet, Francois (eds.), 2013. *Human activities recognition with RGB-Depth camera using HMM*. IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). ISBN: 1457702169
- Eden 1992** Eden, Alec (ed.), 1992. *The Search for Christian Doppler*. Springer
- Erol & Amin 2019** Erol, Baris; Amin, Moeness G., 2019. Radar data cube processing for human activity recognition using multisubspace learning. *IEEE Transactions on Aerospace and Electronic Systems* **55** (6), pp. 3617–3628
- Erol et al. 2017** Erol, Baris; Amin, Moeness G.; Boashash, Boualem (eds.), 2017. *Range-Doppler radar sensor fusion for fall detection*. IEEE Radar Conference (RadarConf). ISBN: 1467388238
- Erol et al. 2019** Erol, Baris; Gurbuz, Sevgi Z.; Amin, Moeness G. (eds.), 2019. *GAN-based synthetic radar micro-Doppler augmentations for improved human activity recognition*. IEEE Radar Conference (RadarConf). ISBN: 1728116791
- Erol & Gurbuz 2015** Erol, Baris; Gurbuz, Sevgi Zubeyde, 2015. A kinect-based human micro-doppler simulator. *IEEE Aerospace and Electronic Systems Magazine* **30** (5), pp. 6–17

- Falanga et al. 2020** Falanga, Davide; Kleber, Kevin; Scaramuzza, Davide, 2020. Dynamic obstacle avoidance for quadrotors with event cameras.
Science Robotics **5** (40), pp. eaaz9712
- Fernández-Delgado et al. 2014** Fernández-Delgado, Manuel; Cernadas, Eva; Barro, Senén, Amorim, Dinani, 2014. Do we need hundreds of classifiers to solve real world classification problems.
Journal of Machine Learning Research **15** (1), pp. 3133–3181
- Filippeschi et al. 2017** Filippeschi, Alessandro; Schmitz, Norbert; Miezal, Markus; Bleser, Gabriele; Ruffaldi, Emanuele, Stricker, Didier, 2017. Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion.
Sensors **17** (6), pp. 1257
- Fischer et al. 2018** Fischer, Bettina; Krauthäuser, Horst; Krebs, Christian; Gütgemann, Sabine; Nüssler, Dirk, Antoine, Claus-Peter, 2018. Radar Solutions for Harsh Environmental Conditions.
BHM Berg-Und Hüttenmännische Monatshefte **163** (3), pp. 84–89
- Foued et al. 2017** Foued, Cherchour; Ammar, Mesloub; Arezki, Younsi (eds.), 2017. *Detection and classification of ground targets using a Doppler radar*.
IEEE Seminar on Detection Systems Architectures and Technologies (DAT).
ISBN: 1509045082
- Futatsumori et al. 2013** Futatsumori, Shunichi; Kohmura, Akiko; Yonemoto, Naruto, 2013. Performance measurement of compact and high-range resolution 76GHz millimeter-wave radar system for autonomous unmanned helicopters.
IEICE transactions on electronics **96** (4), pp. 586–594
- Geibig et al. 2016** Geibig, Thomas; Shoykhetbrod, Alex; Hommes, Alexander; Herschel, Reinhold, Pohl, Nils (eds.), 2016. *Compact 3D imaging radar based on FMCW driven frequency-scanning antennas*.
IEEE Radar Conference (RadarConf).
ISBN: 1509008632
- Geiss & Hardin 2020** Geiss, Andrew; Hardin, Joseph C., 2020. Radar super resolution using a deep convolutional neural network.
Journal of Atmospheric and Oceanic Technology **37** (12), pp. 2197–2207

- Gonzalez 2009** Gonzalez, Rafael C., 2009.
Digital image processing. Pearson education india.
ISBN: 8131726959
- Goodfellow et al. 2016** Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron, 2016.
Deep learning. MIT press.
ISBN: 0262337371
- Greenberg & Kingsbury 1997** Greenberg, Steven; Kingsbury, Brian E. D. (eds.), 1997.
The modulation spectrogram: In pursuit of an invariant representation of speech.
IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).
ISBN: 0818679190
- Guldogan et al. 2011** Guldogan, Mehmet Burak; Gustafsson, Fredrik; Orguner, Umut; Bjoerklund, Svante; Petersson, Henrik, Nezirovic, Amer (eds.), 2011.
Human gait parameter estimation based on micro-doppler signatures using particle filters.
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
ISBN: 1457705397
- Guo et al. 2004** Guo, Hongwei; He, Haitao; Chen, Mingyi, 2004.
Gamma correction for digital fringe projection profilometry. *Applied optics* **43** (14), pp. 2906–2914
- Gurbuz et al. 2020** Gurbuz, Sevgi Z; Gurbuz, Ali Cafer; Malaia, Evie A; Griffin, Darrin J; Crawford, Chris S; Rahman, Mohammad Mahbubur; Kurtoglu, Emre; Aksu, Ridvan; Macks, Trevor, Mdrafi, Robiulhossain, 2020.
American sign language recognition using rf sensing. *IEEE Sensors Journal* **21** (3), pp. 3763–3775
- Gurbuz et al. 2007** Gurbuz, Sevgi Z.; Melvin, William L.; Williams, Douglas B. (eds.), 2007.
Detection and identification of human targets in radar data. SPIE
- Gurbuz et al. 2017** Gurbuz, Sevgi Zubeyde; Clemente, Carmine; Balleri, Alessio, Soraghan, John J., 2017.
Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems. *IET Radar, Sonar & Navigation* **11** (1), pp. 107–115
- Gurbuz & Amin 2019** Gurbuz, Sevgi Zubeyde; Amin, Moeness G., 2019.
Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring. *IEEE Signal Processing Magazine* **36** (4), pp. 16–28
- Habash 2001** Habash, Riadh WY, 2001.
Electromagnetic fields and radiation: Human bioeffects and safety. CRC Press

- Hardeman et al. 2019** Hardeman, AM; Serra Bragança, FM; Swagemakers, Jan-Hein; Van Weeren, PR, Roepstorff, Lars, 2019.
Variation in gait parameters used for objective lameness assessment in sound horses at the trot on the straight line and the lunge.
Equine veterinary journal **51** (6), pp. 831–839
- Hartenberg & Danavit 1964** Hartenberg, Richard; Danavit, Jacques, 1964.
Kinematic synthesis of linkages. New York: McGraw-Hill
- He et al. 2016** He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing, Sun, Jian (eds.), 2016.
Deep residual learning for image recognition
- Hecht-Nielsen 1992** Hecht-Nielsen, Robert, 1992.
Theory of the backpropagation neural network.
In: *Neural networks for perception*.
Elsevier, pp. 65–93
- Hermans et al. 2017** Hermans, Alexander; Beyrer, Lucas; Leibe, Bastian, 2017.
In defense of the triplet loss for person re-identification.
arXiv preprint arXiv:1703.07737
- Ho et al. 2005** Ho, Clifford K.; Robinson, Alex; Miller, David R., Davis, Mary J., 2005.
Overview of sensors and needs for environmental monitoring.
Sensors **5** (1), pp. 4–37
- Ho & Gader 2002** Ho, K. C.; Gader, Paul D., 2002.
A linear prediction land mine detection algorithm for hand held ground penetrating radar.
IEEE transactions on geoscience and remote sensing **40** (6), pp. 1374–1384
- Hongwei 2009** Hongwei, Wang, 2009.
Evaluation of Various Window Functions using Multi-Instrument.
Hornak, JP (2010). The Basics of MRI. Retrieved January 24, pp. 2010
- Hornsteiner & Detlefsen 2008** Hornsteiner, Christoph; Detlefsen, Jürgen, 2008.
Characterisation of human gait using a continuous-wave radar at 24 GHz.
Advances in Radio Science **6** (B. 2), pp. 67–70
- Hu et al. 2018** Hu, Jie; Shen, Li; Sun, Gang, 2018.
Squeeze-and-excitation networks
- Huang et al. 2017** Huang, Gao; Liu, Zhuang; van der Maaten, Laurens, Weinberger, Kilian Q., 2017.
Densely connected convolutional networks

- Hughes & Cowley 2017** Hughes, Stephen W.; Cowley, Michael, 2017. Teaching the Doppler effect in astrophysics. *European Journal of Physics* **38** (2), pp. 025603
- Hwang et al. 2017** Hwang, Jea-Woong; Lee, Su-Kyoung; Park, Jin-Seong; Ahn, Su-Hong; Lee, Kwnag-Jun, Lee, Sung-Jae, 2017. The effects of ankle weight loading on the walking factors of adults without symptoms. *Journal of exercise rehabilitation* **13** (4), pp. 425
- Iatsenko et al. 2016** Iatsenko, Dmytro; McClintock, Peter V. E.; Stefanovska, Aneta, 2016. Extraction of instantaneous frequencies from ridges in time–frequency representations of signals. *Signal Processing* **125**, pp. 290–303
- Instruments 2018** Instruments, Texas, 2018. Mimo radar. *Online*], <https://www.ti.com/lit/an/swra554a/swra554a.pdf>
- Ioffe & Szegedy 2015** Ioffe, Sergey; Szegedy, Christian (eds.), 2015. *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. PMLR
- Jalalvand et al. 2019** Jalalvand, Azarakhsh; Vandersmissen, Baptist; de Neve, Wesley, Mannens, Erik (eds.), 2019. *Radar signal processing for human identification by means of reservoir computing networks*. IEEE Radar Conference (RadarConf). ISBN: 1728116791
- Janakaraj et al. 2019** Janakaraj, Prabhu; Jakkala, Kalvik; Bhuyan, Arupjyoti; Sun, Zhi; Wang, Pu, Lee, Minwoo (eds.), 2019. *STAR: Simultaneous tracking and recognition through millimeter waves and deep learning*. IEEE IFIP Wireless and Mobile Networking Conference (WMNC). ISBN: 3903176184
- Javier & Kim 2014** Javier, Rios Jesus; Kim, Youngwook, 2014. Application of linear predictive coding for human activity classification based on micro-Doppler signatures. *IEEE Geoscience and Remote Sensing Letters* **11** (10), pp. 1831–1834
- Jokanovic & Amin 2017** Jokanovic, Branka; Amin, Moeness, 2017. Fall detection using deep learning in range-Doppler radars. *IEEE Transactions on Aerospace and Electronic Systems* **54** (1), pp. 180–189

- Joyce 2011** Joyce, James M., 2011.
Kullback-leibler divergence.
In: Joyce, James M. (ed.): *International encyclopedia of statistical science*.
Springer, pp. 720–722
- Kaya & Bilge 2019** Kaya, Mahmut; Bilge, Hasan Şakir, 2019.
Deep metric learning: A survey.
Symmetry **11** (9), pp. 1066
- Kim & Ling 2009** Kim, Youngwook; Ling, Hao, 2009.
Human activity classification based on micro-Doppler signatures using a support vector machine.
IEEE transactions on geoscience and remote sensing **47** (5), pp. 1328–1337
- Kim & Ling 2008** Kim, Youngwook; Ling, Hao (eds.), 2008.
Human activity classification based on micro-Doppler signatures using an artificial neural network.
IEEE Antennas and Propagation Society International Symposium.
ISBN: 1424420415
- Kingma & Ba 2014** Kingma, Diederik P.; Ba, Jimmy, 2014.
Adam: A method for stochastic optimization.
arXiv preprint arXiv:1412.6980
- Kingsbury et al. 1998** Kingsbury, Brian E. D.; Morgan, Nelson; Greenberg, Steven, 1998.
Robust speech recognition using the modulation spectrogram.
Speech communication **25** (1-3), pp. 117–132
- Klauder 1960** Klauder, John R., 1960.
The design of radar signals having both high range resolution and high velocity resolution.
Bell System Technical Journal **39** (4), pp. 809–820
- Koay et al. 2014** Koay, Kheng Lee; Syrdal, Dag Sverre; Ashgari-Oskoei, Mohammadreza; Walters, Michael L., Dautenhahn, Kerstin, 2014.
Social roles and baseline proxemic preferences for a domestic service robot.
International Journal of Social Robotics **6** (4), pp. 469–488
- Kozlowski & James 1977** Kozlowski, Lynn T.; James, E., 1977.
Cutting. 1977. Recognizing the sex of a walker from a dynamic point-light display.
Perception & psychophysics **21** (6), pp. 575–580
- Krizhevsky et al. 2012** Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E., 2012.
Imagenet classification with deep convolutional neural networks.
Advances in neural information processing systems **25**

- Krogh & Vedelsby 1994** Krogh, Anders; Vedelsby, Jesper, 1994.
Neural network ensembles, cross validation, and active learning.
Advances in neural information processing systems **7**
- Kronauge & Rohling 2013** Kronauge, Matthias; Rohling, Hermann, 2013.
Fast two-dimensional CFAR procedure.
IEEE Transactions on Aerospace and Electronic Systems **49** (3), pp. 1817–1823
- Kueppers et al. 2020** Kueppers, Simon; Cetinkaya, Harun; Herschel, Reinhold, Pohl, Nils, 2020.
A compact 24×24 channel MIMO FMCW radar system using a substrate integrated waveguide-based reference distribution backplane.
IEEE transactions on microwave theory and techniques **68** (6), pp. 2124–2133
- Laakso et al. 2012** Laakso, Ilkka; Tsuchida, Shogo; Hirata, Akimasa, Kamimura, Yoshitsugu, 2012.
Evaluation of SAR in a human body model due to wireless power transmission in the 10 MHz band.
Physics in Medicine & Biology **57** (15), pp. 4991
- Lang et al. 2019** Lang, Yue; Wang, Qing; Yang, Yang; Hou, Chunping; Liu, Haiping, He, Yuan, 2019.
Joint motion classification and person identification via multitask learning for smart homes.
IEEE Internet of Things Journal **6** (6), pp. 9596–9605
- Le 2013** Le, Quoc V. (ed.), 2013.
Building high-level features using large scale unsupervised learning.
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
ISBN: 1479903566
- LeCun et al. 1998** LeCun, Yann; Bottou, Léon; Bengio, Yoshua, Haffner, Patrick, 1998.
Gradient-based learning applied to document recognition.
Proceedings of the IEEE **86** (11), pp. 2278–2324
- Li et al. 2017** Li, Haobo; Shrestha, Aman; Fioranelli, Francesco; Le Kerneec, Julien; Heidari, Hadi; Pepa, Matteo; Cippitelli, Enea; Gambi, Ennio, Spinsante, Susanna (eds.), 2017.
Multisensor data fusion for human activities classification and fall detection.
IEEE SENSORS.
ISBN: 1509010122
- Li & Stoica 2008** Li, Jian; Stoica, Petre, 2008.
MIMO radar signal processing. John Wiley & Sons.
ISBN: 047039143X

- Li et al. 2012** Li, Wei; Zhao, Rui; Wang, Xiaogang (eds.), 2012.
Human reidentification with transferred metric learning.
Springer
- Li et al. 2019** Li, Xinyu; He, Yuan; Jing, Xiaojun, 2019.
A survey of deep learning-based human activity recognition
in radar.
Remote Sensing **11** (9), pp. 1068
- Liao et al. 2001** Liao, Ping-Sung; Chen, Tse-Sheng; Chung, Pau-Choo, 2001.
A fast algorithm for multilevel thresholding.
J. Inf. Sci. Eng. **17** (5), pp. 713–727
- Lin & Ling 2006** Lin, A.; Ling, H., 2006.
Three-dimensional tracking of humans using very
low-complexity radar.
Electronics Letters **42** (18), pp. 1062–1064
- Lipa & Barrick 1990** Lipa, Belinda J; Barrick, Donald E, 1990.
FMCW signal processing.
FMCW signal processing report for Mirage Systems
- Liu et al. 2020** Liu, Haiping; Yang, Ruixia; Yang, Yang; Hou, Chunping;
Hu, Zhiqi, Jiang, Tianli, 2020.
Human–human interaction recognition based on
ultra-wideband radar.
Signal, Image and Video Processing **14** (6), pp. 1181–1188
- Ma et al. 2022** Ma, Yugang; Zeng, Yonghong; Sun, Sumei, 2022.
A deep learning based super resolution DOA estimator with
single snapshot mimo radar data.
IEEE Transactions on Vehicular Technology **71** (4), pp.
4142–4155
- Mahafza 2005** Mahafza, Bassem R., 2005.
Radar systems analysis and design using MATLAB.
Chapman and Hall/CRC.
ISBN: 0429125410
- Mahafza 2017** Mahafza, Bassem R., 2017.
Introduction to radar analysis. Chapman and Hall/CRC.
ISBN: 131511934X
- Majumder et al. 2006** Majumder, Uttam; Minardi, Michael; Blasch, Erik;
Gorham, LeRoy; Naidu, Kiranmai; Lewis, Thomas,
Williams, Robert, 2006.
Radar signals dismount data production.
In: *Algorithms for Synthetic Aperture Radar Imagery XIII*,
Pp. 21–24

- Martin et al. 2006** Martin, Christian; Schaffernicht, Erik; Scheidig, Andrea, Gross, H-M, 2006.
Multi-modal sensor fusion using a probabilistic aggregation scheme for people detection and tracking.
Robotics and Autonomous Systems **54** (9), pp. 721–728
- Michniewicz & Reinhart 2016** Michniewicz, Joachim; Reinhart, Gunther, 2016.
Cyber-Physical-Robotics–Modelling of modular robot cells for automated planning and execution of assembly tasks.
Mechatronics **34**, pp. 170–180
- Milligan 2005** Milligan, Thomas A., 2005.
Modern antenna design. John Wiley & Sons.
ISBN: 0471720607
- Mitra & Kuo 2006** Mitra, Sanjit Kumar; Kuo, Yonghong, 2006.
Digital signal processing: a computer-based approach. McGraw-Hill New York
- Moindrot 2018** Moindrot, Olivier, 2018.
tensorflow-triplet-loss,
URL: <https://github.com/omindrot/tensorflow-triplet-loss>
- Moyer & Stork 1977** Moyer, Elmer A.; Stork, Ronald F. (eds.), 1977.
The zoom FFT using complex modulation.
IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
- Mummolo et al. 2013** Mummolo, Carlotta; Mangialardi, Luigi; Kim, Joo H, 2013.
Quantifying dynamic characteristics of human walking for comprehensive gait cycle.
Journal of biomechanical engineering **135** (9)
- Nanzer 2010** Nanzer, Jeffrey A., 2010.
Millimeter-wave interferometric angular velocity detection.
IEEE transactions on microwave theory and techniques **58** (12), pp. 4128–4136
- Nanzer & Zilevu 2014** Nanzer, Jeffrey A.; Zilevu, Kojo S., 2014.
Dual interferometric-Doppler measurements of the radial and angular velocity of humans.
IEEE Transactions on Antennas and Propagation **62** (3), pp. 1513–1517
- Nanzer 2016** Nanzer, Jeffrey A. (ed.), 2016.
Micro-motion signatures in radar angular velocity measurements.
IEEE radar conference (RadarConf).
ISBN: 1509008632
- Narayanan & Zenaldin 2015** Narayanan, Ram M.; Zenaldin, Matthew, 2015.
Radar micro–Doppler signatures of various human activities.
IET Radar, Sonar & Navigation **9** (9), pp. 1205–1215

- Ni & Huang 2020** Ni, Zhongfei; Huang, Binke, 2020.
Human identification based on natural gait micro-Doppler signatures using deep transfer learning.
IET Radar, Sonar & Navigation **14** (10), pp. 1640–1646
- Ni & Huang 2021** Ni, Zhongfei; Huang, Binke, 2021.
Open-set human identification based on gait radar micro-Doppler signatures.
IEEE sensors journal **21** (6), pp. 8226–8233
- Niazi et al. 2021** Niazi, Usman; Hazra, Souvik; Santra, Avik, Weigel, Robert (eds.), 2021.
Radar-based efficient gait classification using Gaussian prototypical networks.
IEEE Radar Conference (RadarConf21).
ISBN: 1728176093
- Nixon & Carter 2006** Nixon, Mark S.; Carter, John N., 2006.
Automatic recognition by gait.
Proceedings of the IEEE **94** (11), pp. 2013–2024
- Nuttall 2015** Nuttall, Frank Q., 2015.
Body mass index: obesity, BMI, and health: a critical review.
Nutrition today **50** (3), pp. 117
- Otsu 1979** Otsu, Nobuyuki, 1979.
A threshold selection method from gray-level histograms.
IEEE transactions on systems, man, and cybernetics **9** (1), pp. 62–66
- Papanastasiou et al. 2021** Papanastasiou, V. S.; Trommel, R. P.; Harmanny, R. I.A., Yarovoy, Alexander (eds.), 2021.
Deep Learning-based identification of human gait by radar micro-Doppler measurements.
IEEE 17th European Radar Conference (EuRAD).
ISBN: 2874870617
- Parker 2011** Parker, Michael, 2011.
Radar Basics-Part 2: Pulse Doppler Radar.
Electronics Engineering Times. Available online at:
<http://www.eetimes.com/document.asp>
- Pegoraro et al. 2020** Pegoraro, Jacopo; Meneghello, Francesca; Rossi, Michele, 2020.
Multiperson continuous tracking and identification from mm-wave micro-Doppler signatures.
IEEE transactions on geoscience and remote sensing **59** (4), pp. 2994–3009
- Pegoraro & Rossi 2021** Pegoraro, Jacopo; Rossi, Michele, 2021.
Real-time people tracking and identification from sparse mm-wave radar point-clouds.
IEEE Access **9**, pp. 78504–78520

- Pisa et al. 2016** Pisa, Stefano; Pittella, Erika; PiuZZi, Emanuele, 2016. A survey of radar systems for medical applications. *IEEE Aerospace and Electronic Systems Magazine* **31** (11), pp. 64–81
- Qi et al. 2015** Qi, Yongbin; Soh, Cheong Boon; Gunawan, Erry; Low, Kay-Soon, Thomas, Rijil, 2015. Assessment of foot trajectory for human gait phase detection using wireless ultrasonic sensor network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **24** (1), pp. 88–97
- Qian et al. 2021** Qian, Jiang; Huang, Shaoyin; Wang, Lu; Bi, Guoan, Yang, Xiaobo, 2021. Super-resolution ISAR imaging for maneuvering target based on deep-learning-assisted time–frequency analysis. *IEEE transactions on geoscience and remote sensing* **60**, pp. 1–14
- Qiao et al. 2020** Qiao, Xingshuai; Shan, Tao; Tao, Ran, 2020. Human identification based on radar micro–Doppler signatures separation. *Electronics Letters* **56** (4), pp. 195–196
- Qu et al. 2019** Qu, Zhiyu; Wang, Wenyang; Hou, Changbo, Hou, Chenfan, 2019. Radar signal intra-pulse modulation recognition based on convolutional denoising autoencoder and deep convolutional neural network. *IEEE Access* **7**, pp. 112339–112347
- Qualisys 2015** Qualisys, 2015, URL: <http://www.qualisys.com/>
- Radzicki et al. 2016** Radzicki, Vincent R.; Boutte, David; Taylor, Paul V., Lee, Hua (eds.), 2016. *Standoff CW radar for through-the-wall detection of human heartbeat signatures*. IEEE Radar Conference (RadarConf). ISBN: 1509008632
- Raemer 1996** Raemer, Harold R., 1996. *Radar systems principles*. CRC press. ISBN: 0849394813
- Raj et al. 2010** Raj, R. G.; Chen, V. C.; Lipps, R., 2010. Analysis of radar human gait signatures. *IET signal processing* **4** (3), pp. 234–244
- Raj et al. 2009** Raj, Raghu G.; Chen, Victor C.; Lipps, Ronald (eds.), 2009. *Analysis of radar dismount signatures via non-parametric and parametric methods*. IEEE Radar Conference. ISBN: 142442870X

- Ram & Ling 2008** Ram, Shobha Sundar; Ling, Hao, 2008.
Through-wall tracking of human movers using joint Doppler and array processing.
IEEE Geoscience and Remote Sensing Letters **5** (3), pp. 537–541
- Ram et al. 2007** Ram, Shobha Sundar; Li, Yang; Lin, Adrian, Ling, Hao (eds.), 2007.
Human tracking using doppler processing and spatial beamforming.
IEEE Radar Conference.
ISBN: 1424402832
- Reich et al. 2020** Reich, Olaf; Huebner, Erik; Ghita, Bogdan; Wagner, Matthias F., Schaefer, Joerg (eds.), 2020.
A Survey Investigating the Combination and Number of IMUs on the Human Body Used for Detecting Activities and Human Tracking.
IEEE Radar Conference.
ISBN: 1728197384
- Rhodin et al. 2018** Rhodin, Marie; Persson-Sjodin, E; Egenvall, Agneta; Serra Bragança, FM; Pfau, Thilo; Roepstorff, Lars; Weishaupt, Michael A; Thomsen, Maj H; van Weeren, P René, Hernlund, Elin, 2018.
Vertical movement symmetry of the withers in horses with induced forelimb and hindlimb lameness at trot.
Equine veterinary journal **50** (6), pp. 818–824
- Richards et al. 2010** Richards, Mark A.; Scheer, Jim; Holm, William A., Melvin, William L., 2010.
Principles of modern radar. Citeseer
- Richards 2014** Richards, Mark A., 2014.
Fundamentals of radar signal processing. McGraw-Hill Education.
ISBN: 0071798323
- Robey et al. 2004** Robey, Frank C.; Coutts, Scott; Weikle, Dennis; McHarg, Jeffrey C., Cuomo, Kevin (eds.), 2004.
MIMO radar theory and experimental results.
IEEE Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers.
ISBN: 0780386221
- Robots 2017** Robots, Staubli, 2017,
URL: <http://www.staubli.com/robotics/>

- Rock et al. 2019** Rock, Johanna; Toth, Mate; Messner, Elmar; Meissner, Paul, Pernkopf, Franz (eds.), 2019.
Complex signal denoising and interference mitigation for automotive radar using convolutional neural networks.
IEEE 22th International Conference on Information Fusion (FUSION).
ISBN: 0996452788
- Rohling & Moller 2008** Rohling, Hermann; Moller, Christof (eds.), 2008.
Radar waveform for automotive radar systems and applications.
IEEE Radar Conference.
ISBN: 1424415381
- Russakovsky et al. 2015** Russakovsky, Olga; Deng, Jia; Su, Hao; Krause, Jonathan; Satheesh, Sanjeev; Ma, Sean; Huang, Zhiheng; Karpathy, Andrej; Khosla, Aditya, Bernstein, Michael, 2015.
Imagenet large scale visual recognition challenge.
International journal of computer vision **115** (3), pp. 211–252
- Saha 2018** Saha, Sumit, 2018.
A Comprehensive Guide to Convolutional Neural Networks – the ELI5 way,
URL:
<https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- Saho et al. 2017** Saho, Kenshi; Fujimoto, Masahiro; Masugi, Masao, Chou, Li-Shan, 2017.
Gait classification of young adults, elderly non-fallers, and elderly fallers using micro-Doppler radar signals: Simulation study.
IEEE sensors journal **17** (8), pp. 2320–2321
- Santra et al. 2018** Santra, Avik; Ulaganathan, Raghavendran Vagarappan; Finke, Thomas, 2018.
Short-range millimetric-wave radar system for occupancy sensing application.
IEEE sensors letters **2** (3), pp. 1–4
- Schapire 1990** Schapire, Robert E., 1990.
The strength of weak learnability.
Machine learning **5** (2), pp. 197–227
- Schmid 2012** Schmid, Hanspeter, 2012.
How to use the FFT and Matlab’s pwelch function for signal and noise simulations and measurements.
FHNW/IME, pp. 2–13
- Schroff et al. 2015** Schroff, Florian; Kalenichenko, Dmitry; Philbin, James (eds.), 2015.
Facenet: A unified embedding for face recognition and clustering

- Seifert et al. 2019a** Seifert, Ann-Kathrin; Amin, Moeness G.; Zoubir, Abdelhak M., 2019. Toward unobtrusive in-home gait analysis based on radar micro-Doppler signatures. *IEEE Transactions on Biomedical Engineering* **66** (9), pp. 2629–2640
- Seifert et al. 2020** Seifert, Ann-Kathrin; Grimmer, Martin; Zoubir, Abdelhak M., 2020. Doppler radar for the extraction of biomechanical parameters in gait analysis. *IEEE Journal of Biomedical and Health Informatics* **25** (2), pp. 547–558
- Seifert et al. 2018a** Seifert, Ann-Kathrin; Schaefer, Lukas; Amin, Moeness G., Zoubir, Abdelhak M. (eds.), 2018. *Subspace classification of human gait using radar micro-Doppler signatures*. IEEE European Signal Processing Conference (EUSIPCO). ISBN: 9082797011
- Seifert et al. 2018b** Seifert, Ann-Kathrin; Zoubir, Abdelhak M.; Amin, Moeness G. (eds.), 2018. *Radar classification of human gait abnormality based on sum-of-harmonics analysis*. IEEE Radar Conference (RadarConf). ISBN: 1538641674
- Seifert et al. 2019b** Seifert, Ann-Kathrin; Zoubir, Abdelhak M.; Amin, Moeness G. (eds.), 2019. *Detection of gait asymmetry using indoor Doppler radar*. IEEE Radar Conference (RadarConf). ISBN: 1728116791
- Seyfioglu et al. 2018** Seyfioglu, Mehmet Saygin; Erol, Baris; Gurbuz, Sevgi Zubeyde, Amin, Moeness G., 2018. DNN transfer learning from diversified micro-Doppler for motion classification. *IEEE Transactions on Aerospace and Electronic Systems* **55** (5), pp. 2164–2180
- Shah & Fioranelli 2019** Shah, Syed Aziz; Fioranelli, Francesco (eds.), 2019. *Human activity recognition: Preliminary results for dataset portability using FMCW radar*. IEEE International Radar Conference (RADAR). ISBN: 1728126606
- Siliciano et al. 2010** Siliciano, Bruno; Sciavicco, Lorenzo; Villani, Luigi, Oriolo, Giuseppe, 2010. Robotics: modelling, planning and control. *New York, NY, USA: Springer*, pp. 415–418

- Simonyan & Zisserman 2014** Simonyan, Karen; Zisserman, Andrew, 2014.
Very deep convolutional networks for large-scale image recognition.
arXiv preprint arXiv:1409.1556
- Singh et al. 2019** Singh, Akash Deep; Sandha, Sandeep Singh; Garcia, Luis, Srivastava, Mani (eds.), 2019.
Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar
- Singhal 2012** Singhal, Kunal, 2012.
Walsh Codes, PN Sequences and their role in CDMA Technology.
Term paper, EEL 201, pp. 1–4
- Skolnik 2001** Skolnik, M., 2001.
Introduction to Radar systems.
IEEE Aerospace and Electronic Systems Magazine **16** (10), pp. 19
- Smith et al. 2018** Smith, Karly A.; Csech, Clément; Murdoch, David, Shaker, George, 2018.
Gesture recognition using mm-wave sensor for human-car interface.
IEEE sensors letters **2** (2), pp. 1–4
- Spline 2016** Spline, Cubic, 2016,
URL: <http://mathworld.wolfram.com/CubicSpline.html>
- Srivastava et al. 2014** Srivastava, Nitish; Hinton, Geoffrey; Krizhevsky, Alex; Sutskever, Ilya, Salakhutdinov, Ruslan, 2014.
Dropout: a simple way to prevent neural networks from overfitting.
Journal of machine learning research **15** (1), pp. 1929–1958
- Starr & Lattimer 2014** Starr, Joseph W.; Lattimer, B. Y., 2014.
Evaluation of navigation sensors in fire smoke environments.
Fire Technology **50** (6), pp. 1459–1481
- Suhr & Jung 2017** Suhr, Jae Kyu; Jung, Ho Gi, 2017.
Rearview camera-based backover warning system exploiting a combination of pose-specific pedestrian recognitions.
IEEE transactions on intelligent transportation systems **19** (4), pp. 1122–1129
- Sun et al. 2014** Sun, Hongbo; Brigrui, Frederic; Lesturgie, Marc (eds.), 2014.
Analysis and comparison of MIMO radar waveforms.
IEEE International Radar Conference.
ISBN: 1479941956
- Sundararajan 2017** Sundararajan, Duraisamy, 2017.
Digital image processing: a signal processing and algorithmic approach. Springer.
ISBN: 9811061130

- Szegedy et al. 2015** Szegedy, Christian; Liu, Wei; Jia, Yangqing; Sermanet, Pierre; Reed, Scott; Anguelov, Dragomir; Erhan, Dumitru; Vanhoucke, Vincent, Rabinovich, Andrew (eds.), 2015.
Going deeper with convolutions
- Tahmoush 2015** Tahmoush, Dave, 2015.
Review of micro-Doppler signatures.
IET Radar, Sonar & Navigation **9** (9), pp. 1140–1146
- Tahmoush & Silvius 2009** Tahmoush, Dave; Silvius, Jerry (eds.), 2009.
Angle, elevation, PRF, and illumination in radar microDoppler for security applications.
IEEE Antennas and Propagation Society International Symposium.
ISBN: 1424436478
- Taipalus & Ahtiainen 2011** Taipalus, Tapio; Ahtiainen, Juhana (eds.), 2011.
Human detection and tracking with knee-high mobile 2D LIDAR.
IEEE International Conference on Robotics and Biomimetics.
ISBN: 1457721384
- Takeda et al. 2013** Takeda, Takahiro; Kuramoto, Kei; Kobashi, Syoji, Hata, Yutaka (eds.), 2013.
A fuzzy human detection for security system using infrared laser camera.
IEEE International Symposium on Multiple-Valued Logic.
ISBN: 1467360678
- Teixeira et al. 2010** Teixeira, Thiago; Dublon, Gershon; Savvides, Andreas, 2010.
A survey of human-sensing: Methods for detecting presence, count, location, track, and identity.
ACM Computing Surveys **5** (1), pp. 59–69
- Ulrich et al. 2018** Ulrich, Michael; Hess, Thomas; Abdulatif, Sherif, Yang, Bin (eds.), 2018.
Person recognition based on micro-Doppler and thermal infrared camera fusion for firefighting.
IEEE International Conference on Information Fusion (FUSION).
ISBN: 0996452761
- Und Arbeitswelt** Und Arbeitswelt, Industrieverband Büro.
Körpermasse – Überblick,
URL: <https://iba.online/raeume-planen/flachenplanung/koerpermasse/>
- van der Maaten & Hinton 2008** van der Maaten, Laurens; Hinton, Geoffrey, 2008.
Visualizing data using t-SNE.
Journal of machine learning research **9** (11)

- van Dorp & Groen 2003** van Dorp, Ph; Groen, F. C.A., 2003.
Human walking estimation with radar.
IEEE Proceedings-Radar, Sonar and Navigation **150** (5), pp. 356–365
- Vandersmissen et al. 2018** Vandersmissen, Baptist; Knudde, Nicolas;
Jalalvand, Azarakhsh; Couckuyt, Ivo; Bourdoux, Andre;
de Neve, Wesley, Dhaene, Tom, 2018.
Indoor person identification using a low-power FMCW radar.
IEEE transactions on geoscience and remote sensing **56** (7),
pp. 3941–3952
- Venture 2010** Venture, Gentiane (ed.), 2010.
Human characterization and emotion characterization from gait.
IEEE Annual International Conference of the IEEE
Engineering in Medicine and Biology.
ISBN: 1424441234
- Wagner et al. 2017** Wagner, Thomas; Feger, Reinhard; Stelzer, Andreas, 2017.
Radar signal processing for jointly estimating tracks and
micro-Doppler signatures.
IEEE Access **5**, pp. 1220–1238
- Wang et al. 2016** Wang, H.; Lei, Z.; Zhang, X.; Zhou, B., Peng, J., 2016.
Machine learning basics.
Deep learning, pp. 98–164
- Wang et al. 2019** Wang, Siying; Kueppers, Simon; Cetinkaya, Harun,
Herschel, Reinhold (eds.), 2019.
*3D localization and vital sign detection of human subjects
with a 120 GHz MIMO radar.*
IEEE 20th International Radar Symposium (IRS).
ISBN: 3736998600
- Wang & Fathy 2011** Wang, Yazhou; Fathy, Aly E. (eds.), 2011.
*Micro-Doppler signatures for intelligent human gait
recognition using a UWB impulse radar.*
IEEE International Symposium on Antennas and
Propagation (APSURSI).
ISBN: 1424495636
- Wang et al. 2020** Wang, Yong; Ren, Aihu; Zhou, Mu; Wang, Wen,
Yang, Xiaobo, 2020.
A novel detection and recognition method for continuous
hand gesture using FMCW radar.
IEEE Access **8**, pp. 167264–167275

- Wang et al. 2021** Wang, Yuheng; Liu, Haipeng; Cui, Kening; Zhou, Anfu; Li, Wensheng, Ma, Huadong (eds.), 2021.
m-activity: Accurate and real-time human activity recognition via millimeter wave radar.
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
ISBN: 1728176050
- Weller et al. 2021** Weller, Pascal; Aziz, Fady; Abdulatif, Sherif; Schneider, Urs, Huber, Marco F., 2021.
A mimo radar-based few-shot learning approach for human-id.
arXiv preprint arXiv:2110.08595
- Willen et al. 2013** Willen, Carin; Lehmann, Kirsten; Sunnerhagen, Katharina, 2013.
Walking speed indoors and outdoors in healthy persons and in persons with late effects of polio.
Journal of Neurology Research **3** (2), pp. 62–67
- Winkler 2007** Winkler, Volker (ed.), 2007.
Range Doppler detection for automotive FMCW radars.
IEEE European Radar Conference.
ISBN: 2874870048
- Winter 2009** Winter, David A., 2009.
Biomechanics and motor control of human movement. John Wiley & Sons.
ISBN: 0470398183
- Wu et al. 2017** Wu, Chao-Yuan; Manmatha, R.; Smola, Alexander J., Krahenbuhl, Philipp (eds.), 2017.
Sampling matters in deep embedding learning
- Yang et al. 2020** Yang, Xin; Liu, Jian; Chen, Yingying; Guo, Xiaonan, Xie, Yucheng (eds.), 2020.
MU-ID: Multi-user identification through gaits using millimeter wave radios.
IEEE International Conference on Machine Learning and Applications (ICMLA).
ISBN: 1728164125
- Yang et al. 2017** Yang, Xiuzhu; Yin, Wenfeng; Zhang, Lin, 2017.
People counting based on CNN using IR-UWB radar.
In: *2017 IEEE/CIC International Conference on Communications in China (ICCC)*,
Pp. 1–5

- Yang et al. 2019** Yang, Yang; Hou, Chunping; Lang, Yue; Yue, Guanghui; He, Yuan, Xiang, Wei, 2019.
Person identification using micro-Doppler signatures of human motions and UWB radar.
IEEE Microwave and Wireless Components Letters **29** (5), pp. 366–368
- Yu et al. 2020** Yu, Myoungseok; Kim, Narae; Jung, Yunho, Lee, Seongjoo, 2020.
A frame detection method for real-time hand gesture recognition systems using CW-radar.
Sensors **20** (8), pp. 2321
- Zenaldin & Narayanan 2016** Zenaldin, Matthew; Narayanan, Ram M. (eds.), 2016.
Radar micro-Doppler based human activity classification for indoor and outdoor environments.
SPIE
- Zetik et al. 2006** Zetik, Rudolf; Crabbe, Stephen; Krajnak, Jozef; Peyerl, Peter; Sachs, Juergen, Thoma, Reiner (eds.), 2006.
Detection and localization of persons behind obstacles using M-sequence through-the-wall radar.
SPIE
- Zhang et al. 2019** Zhang, Yonghua; Zhang, Jiawan; Guo, Xiaojie (eds.), 2019.
Kindling the darkness: A practical low-light image enhancer
- Zhang et al. 2016** Zhang, Zhili; Zhang, Renli; Sheng, Weixing; Han, Yubing, Ma, Xiaofeng (eds.), 2016.
Feature extraction and classification of human motions with LFM CW radar.
IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM).
ISBN: 1467381942
- Zhao et al. 2006** Zhao, Guoying; Liu, Guoyi; Li, Hua, Pietikainen, Matti (eds.), 2006.
3D gait recognition using multiple cameras.
IEEE.
ISBN: 0769525032
- Zhao et al. 2019** Zhao, Peijun; Lu, Chris Xiaoxuan; Wang, Jianan; Chen, Changhao; Wang, Wei; Trigoni, Niki, Markham, Andrew (eds.), 2019.
mid: Tracking and identifying people with millimeter wave radar.
IEEE International Conference on Automatic Face and Gesture Recognition (FGRO6).
ISBN: 1728105706

- Zhao et al. 2020** Zhao, Xiangmo; Sun, Pengpeng; Xu, Zhigang; Min, Haigen, Yu, Hongkai, 2020.
Fusion of 3D LIDAR and camera data for object detection in autonomous vehicle applications.
IEEE sensors journal **20** (9), pp. 4901–4913
- Zhou & Hu 2008** Zhou, Huiyu; Hu, Huosheng, 2008.
Human motion tracking for rehabilitation—A survey.
Biomedical signal processing and control **3** (1), pp. 1–18
- Zhou 2021** Zhou, Zhi-Hua, 2021.
Machine learning. Springer Nature.
ISBN: 9811519676
- Zhu et al. 2018** Zhu, Shangyue; Xu, Junhong; Guo, Hanqing; Liu, Qiwei; Wu, Shaoen, Wang, Honggang (eds.), 2018.
Indoor human activity recognition based on ambient radar with signal processing and machine learning.
IEEE International Conference on Communications (ICC).
ISBN: 1538631806

A Appendix

A.1 Description of the Limb Decomposition Process

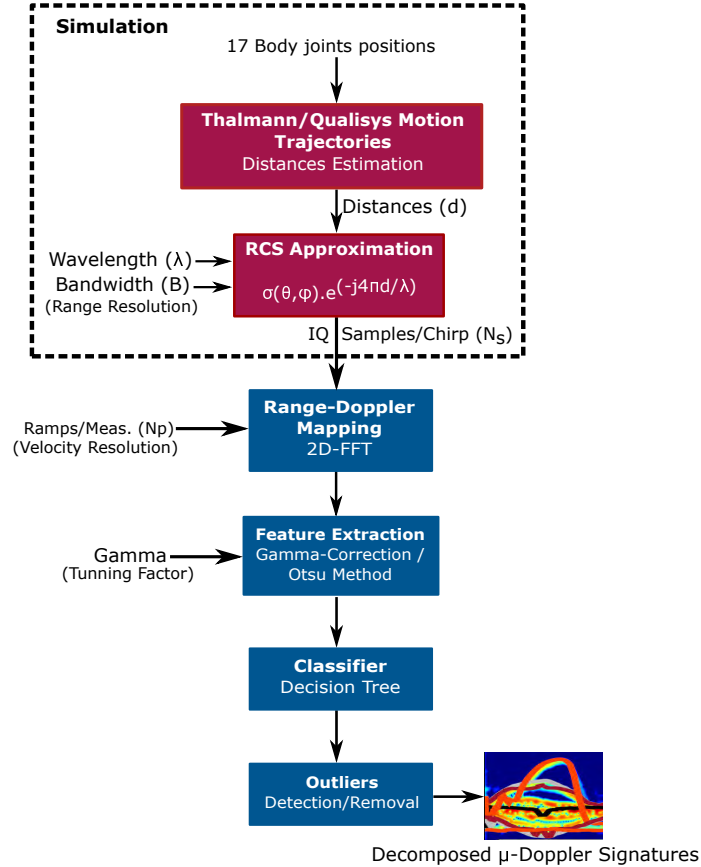


Figure A.1: Algorithm of limbs decomposition presented in Section 4.3 and published in Abdulatif et al. 2017

A.2 Datasheet of Ancortek Radar

In this section, the user manual for the Ancortek radar is presented. This radar module is used in this thesis work in two phases. First, it was used for the human-robot classification use-case that is discussed briefly in Section 4.4, and published in Abdulatif et al. 2018. Second, it was used by us for conducting a feasibility study on human identification that is presented in Chapter 5, and published in Abdulatif et al. 2019.

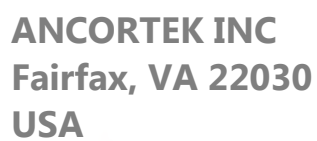
The logo for ANCORTEK, featuring the company name in a bold, blue, sans-serif font. The background of the top half of the page is a decorative pattern of overlapping blue triangles in various shades, creating a geometric, crystalline effect.

ANCORTEK

SDR SYSTEM EVALUATION KIT

USER MANUAL

**ANCORTEK INC
Fairfax, VA 22030
USA**

The contact information for ANCORTEK INC, including the company name, address (Fairfax, VA 22030), and country (USA). The text is in a bold, black, sans-serif font. The bottom half of the page features a continuation of the blue geometric triangle pattern seen at the top.

SDR SYSTEM EVALUATION KIT

USER MANUAL

Rev. 1.3

Edition 2016-05-01

Published by

ANCORTEK INC, USA

© 2016 ANCORTEK INC

All Rights Reserved.

Legal Disclaimer

The information given in this document shall in no event be regarded as a guarantee of conditions or characteristics. With respect to any examples or hints given herein, any typical values stated herein and/or any information regarding the application of the device, ANCORTEK INC hereby disclaims any and all warranties and liabilities of any kind, including without limitation, warranties of non-infringement of intellectual property rights of any third party.

Information

For further information on technology, delivery terms and conditions and prices, please contact the ANCORTEK INC at www.ancortek.com.

Important Notice

This evaluation kit is intended for use for ENGINEERING DEVELOPMENT, DEMONSTRATION, OR EVALUATION PURPOSES ONLY and is not considered to be a finished end-product fit for general consumer use. Persons handling the product(s) must have electronics training and observe good engineering practice standards. As such, the goods being provided are not intended to be complete in terms of required design-, marketing-, and/or manufacturing-related protective considerations, including product safety and environmental measures typically found in end products that incorporate such semiconductor components or circuit boards. This evaluation kit does not fall within the scope of the European Union directives regarding electromagnetic compatibility, restricted substances (RoHS), recycling (WEEE), FCC, CE or UL, and therefore may not meet the technical requirements of these directives or other related directives.

TABLE OF CONTENTS

1	System Specification.....	3
1.1	Scope.....	3
1.2	Features.....	3
1.3	Parameters.....	3
2	Operation Instruction.....	10
2.1	Equipment List.....	10
2.2	System Description.....	11
2.3	Getting Started.....	12
2.4	Graphical User Interface.....	15

1 System Specification

1.1 Scope

Ancortek Inc has built compact-size, light-weight and low-power software defined RF modules and FPGA-based processor modules in S, C, X, and K bands operating at 2.4 GHz, 5.8 GHz, 6.2 GHz, 9.8 GHz and 25 GHz. Our SDR Evaluation Kits offer the ability of integrating transmitter-receiver systems for applications to industry automation, medical monitoring, public safety and security and academic research.

1.2 Features

- Support FMCW, FSK, and CW Signal Waveforms.
- Expandable Bandwidth.
- Selectable Operation Parameters.
- Selectable Stream Filtering.
- Selectable Display Mode.
- Recording of Complex (I & Q) Data.
- USB 2.0 Interface to Host Computer.
- Multifunctional Graphical User Interface (GUI).

1.3 Parameters

- An overview of the current SDR-KITs is shown in Table 1.
- The parameters of 2.4 GHz, 5.8 GHz, 6.2 GHz, 9.8 GHz and 25 GHz SDR-KITs are illustrated in Table 2-6.
- The parameters of the FPGA-based SDR-PM processor module are listed in Table 7.

Table 1 Overview of the SDR-KITs

SDR Evaluation Kits	SDR-KIT 240B	SDR-KIT 580B	SDR-KIT 620B	SDR-KIT 980B	SDR-KIT 2500B
Waveforms	FMCW/FSK/CW	FMCW/FSK/CW	FMCW/FSK/CW	FMCW/FSK/CW	FMCW/FSK/CW
Frequency Range	2.25 - 2.65 GHz	5.6 - 6.0 GHz	6.0 - 6.4 GHz	9.6 – 10.0 GHz	24 – 26 GHz
Expandable Frequency Range	2.05 - 2.65 GHz	5.2 - 6.0 GHz	5.6 – 6.4 GHz	9.2 – 10.2 GHz	23 – 26 GHz
Beam Width Horz/Vert	Patched Antenna 38°/38°	Patched Antenna 40°/20°	Patched Antenna 40°/20°	Patched Antenna 40°/20°	Horn Antenna 30°/20°
Bandwidth	100/150/300/400 MHz	100/150/300/400 MHz	100/150/300/400 MHz	100/150/300/400 MHz	0.5/0.75/1.5/2 GHz
Expandable Bandwidth	600 MHz	800 MHz	800 MHz	1 GHz	3 GHz
Power Output	22 dBm	19 dBm	19 dBm	18 dBm	16 dBm
Noise Figure	2.8 dB	3.4 dB	3.4 dB	3.4 dB	6.4 dB
Application	Through-Wall, Ground Penetration	Industry, Medical, Security, Through-Wall	Industry, Medical, Security, Through-Wall	Industry, Medical, Security	Industry, Medical, Security

Table 2 SDR-RF 240 Module Parameters

Parameter	Min.	Typ.	Max.	units
<i>Frequency Range</i>	2.25		2.65	GHz
<i>Expandable Frequency Range</i>	2.05		2.65	GHz
<i>Tune Voltage</i>	0		5	V
<i>Tuning Sensitivity @RF Port</i>		0.1		GHz/V
<i>Power Output</i>	21	22	23	dBm
<i>SSB Phase Noise @10KHz offset</i>		-80		dBc
<i>SSB Phase Noise @1MHz offset</i>		-130		dBc
<i>Conversion Gain Over Rx Channel</i>	28	30	32	dB
<i>Noise Figure over Rx channel</i>	2.6	2.8	3.0	dB
<i>Maximum input power</i>		10		dBm
<i>Input 1dB power compression</i>		-15		dBm
<i>Supply voltage</i>	4.75	5	5.25	V
<i>Supply current</i>	650	670	700	mA
<i>Operating temperature</i>	-40		85	C°
<i>Storage temperature</i>	-65		150	C°
<i>Dimensions</i>	L=79 W=56 H=13			mm

Table 3 SDR-RF 580 Module Parameters

Parameter	Min.	Typ.	Max.	units
<i>Frequency Range</i>	5.6		6.0	GHz
<i>Expandable Frequency Range</i>	5.2		6.0	GHz
<i>Tune Voltage</i>	0		5	V
<i>Tuning Sensitivity @RF Port</i>		0.13		GHz/V
<i>Power Output</i>	18	19	20	dBm
<i>SSB Phase Noise @10KHz offset</i>		-80		dBc
<i>SSB Phase Noise @1MHz offset</i>		-130		dBc
<i>Conversion Gain Over Rx Channel</i>	26	28	30	dB
<i>Noise Figure over Rx channel</i>	3.2	3.4	3.6	dB
<i>Maximum input power</i>		10		dBm
<i>Input 1dB power compression</i>		-11		dBm
<i>Supply voltage</i>	4.75	5	5.25	V
<i>Supply current</i>	650	670	700	mA
<i>Operating temperature</i>	-40		85	C°
<i>Storage temperature</i>	-65		150	C°
<i>Dimensions</i>	L=79 W=56 H=13			mm

Table 4 SDR-RF 620 Module Parameters

Parameter	Min.	Typ.	Max.	units
<i>Frequency Range</i>	6.0		6.4	GHz
<i>Expandable Frequency Range</i>	5.6		6.4	GHz
<i>Tune Voltage</i>	0		5	V
<i>Tuning Sensitivity @RF Port</i>		0.13		GHz/V
<i>Power Output</i>	18	19	20	dBm
<i>SSB Phase Noise @10KHz offset</i>		-80		dBc
<i>SSB Phase Noise @1MHz offset</i>		-130		dBc
<i>Conversion Gain Over Rx Channel</i>	26	28	30	dB
<i>Noise Figure over Rx channel</i>	3.2	3.4	3.6	dB
<i>Maximum input power</i>		10		dBm
<i>Input 1dB power compression</i>		-11		dBm
<i>Supply voltage</i>	4.75	5	5.25	V
<i>Supply current</i>	650	670	700	mA
<i>Operating temperature</i>	-40		85	C°
<i>Storage temperature</i>	-65		150	C°
<i>Dimensions</i>	L=79	W=56	H=13	mm

Table 5 SDR-RF 980 Module Parameters

Parameter	Min.	Typ.	Max.	units
<i>Frequency Range</i>	9.6		10.0	GHz
<i>Expandable Frequency Range</i>	9.2		10.2	GHz
<i>Tune Voltage</i>	0		5	V
<i>Tuning Sensitivity @RF Port</i>		0.2		GHz/V
<i>Power Output</i>	17	18	19	dBm
<i>SSB Phase Noise @10KHz offset</i>		-80		dBc
<i>SSB Phase Noise @1MHz offset</i>		-130		dBc
<i>Conversion Gain Over Rx Channel</i>	26	28	30	dB
<i>Noise Figure over Rx channel</i>	3.2	3.4	3.6	dB
<i>Maximum input power</i>		10		dBm
<i>Input 1dB power compression</i>		-13		dBm
<i>Supply voltage</i>	4.75	5	5.25	V
<i>Supply current</i>	650	670	700	mA
<i>Operating temperature</i>	-40		85	C°
<i>Storage temperature</i>	-65		150	C°
<i>Dimensions</i>	L=79 W=56 H=13			mm

Table 6 SDR-RF 2500 Module Parameters

Parameter	Min.	Typ.	Max.	units
<i>Frequency Range</i>	24		26	GHz
<i>Expandable Frequency Range</i>	23		26	GHz
<i>Tune Voltage</i>	0		5	V
<i>Tuning Sensitivity @RF Port</i>		1.5		GHz/V
<i>Power Output</i>	14	16	18	dBm
<i>SSB Phase Noise @1MHz offset</i>		-100		dBc
<i>Conversion Gain Over Rx Channel</i>	18	26	31	dB
<i>Noise Figure over Rx channel</i>		6.4		dB
<i>Maximum input power</i>		0		dBm
<i>Input 1dB power compression</i>		-12		dBm
<i>Supply voltage</i>	4.75	5	5.25	V
<i>Supply current</i>	750	800	850	mA
<i>Operating temperature</i>	-40		85	C°
<i>Storage temperature</i>	-65		150	C°
<i>Dimensions</i>	L=79 W=56 H=13			mm

Table 7 SDR-PM 402 Module Parameters

Parameter	Min.	Typ.	Max.	units
<i>Supply voltage</i>	4.75	5	5.25	V
<i>Supply current</i>	180	200	220	mA
<i>Operating temperature</i>	-40 to +85			C°
<i>Storage temperature</i>	-65 to +150			C°
<i>Dimensions</i>	L=79	W=76	H=13	mm

2 Operation Instruction

2.1 Equipment List

The Ancortek SDR Evaluation Kit comes with SDR-RF module and SDR-PM processor module.

For complete operation and data collection, you will also need

1. AC/DC power adapter +5V.
2. USB 2.0 A to Mini-B cable.
3. SMA-M to SMA-M cables.
4. Transmitting and receiving antennas.
5. Ancortek Graphical User Interface: SDR-GUI or MATLAB-based SDR-GUI.
6. PC Windows Operating System.



Figure 1 - Equipment List.

2.2 System Description

Figure 2 is a high level block diagram of the SDR Evaluation Kit. A graphical user interface (SDR-GUI) is used to control the configuration of the SDR via a USB 2.0 cable. Digital samples of control voltage are generated by the FPGA firmware. After D/A converter, an analog control voltage is directly sent to the voltage controlled oscillator (VCO). The output of the mixer in the receive chain is digitized and streamed to host computer for further processing.

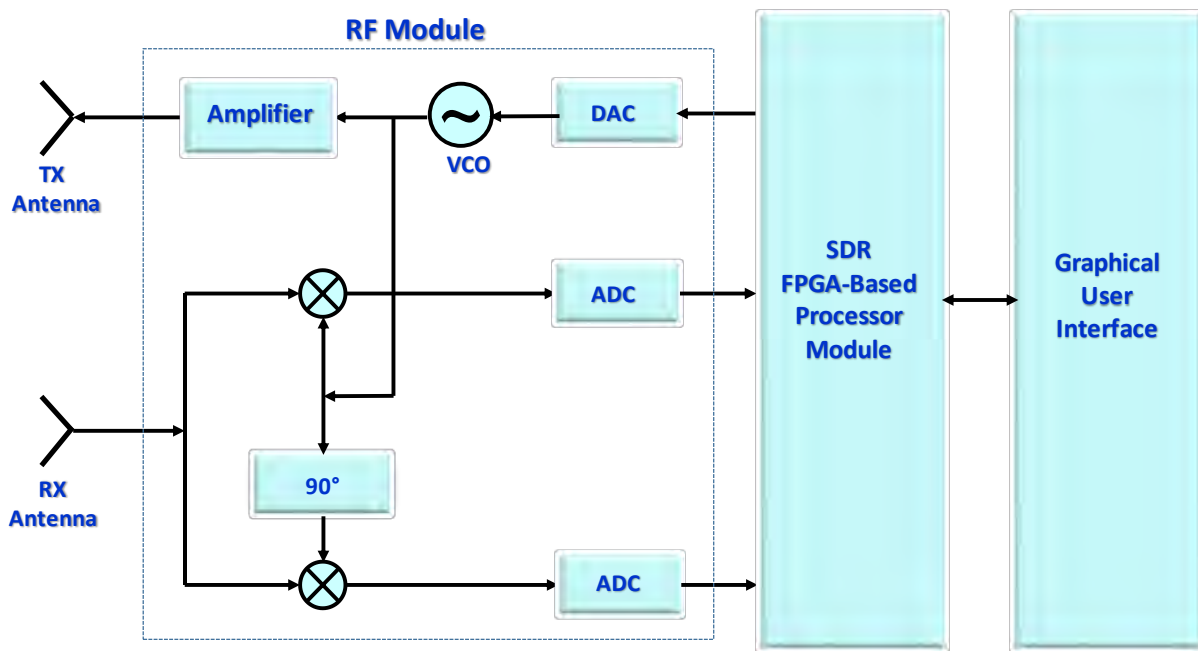


Figure 2 - System Block Diagram.

2.3 Getting Started

1. Run setup.exe to install the Ancortek SDR-GUI.
2. Plug in the Ancortek transceiver and go to the "Device Manager" (see Figure 3), right click "Unknown Device".

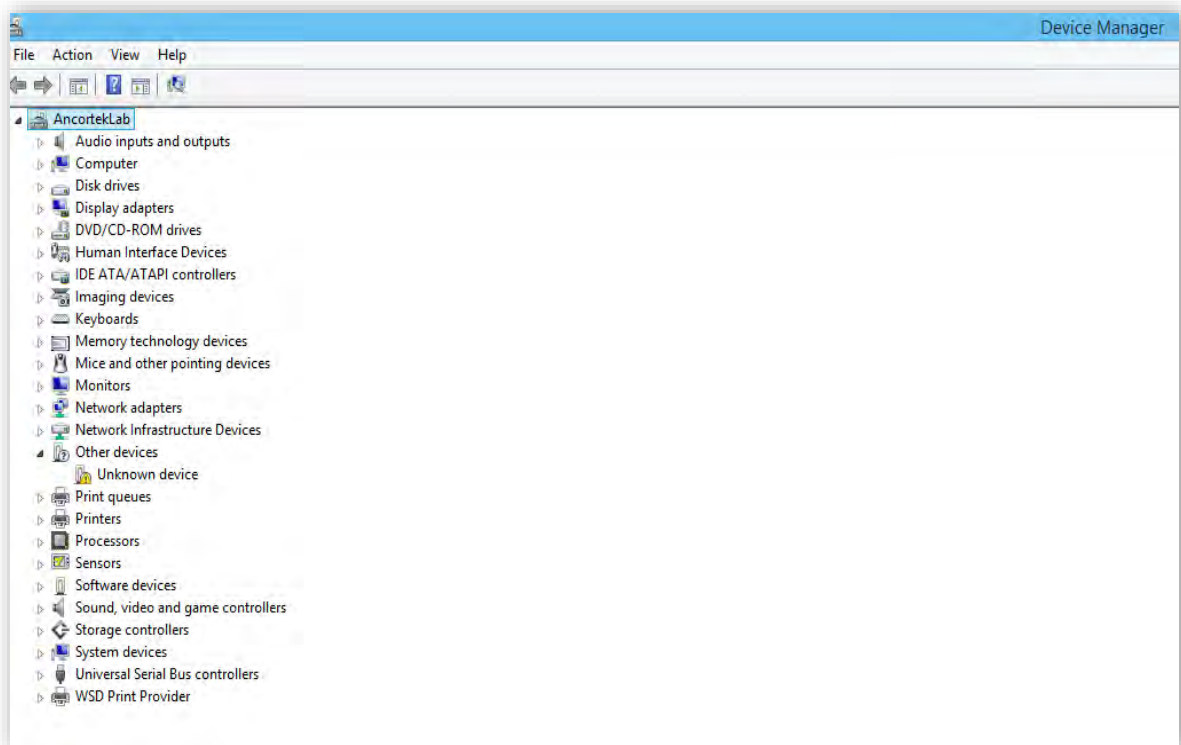


Figure 3 - Device Manager Window.

3. Click "Update Driver" (see Figure 4).

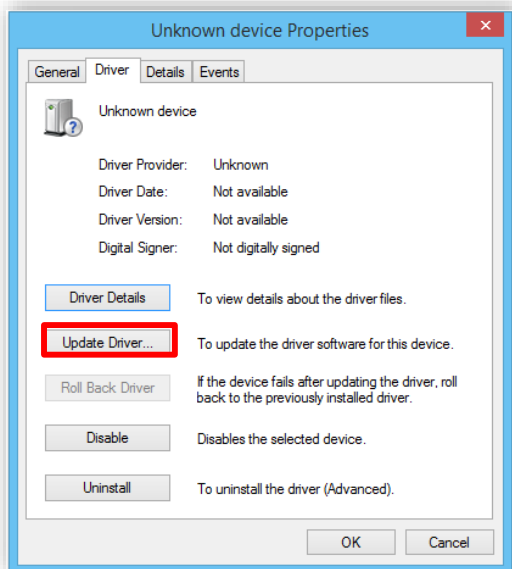


Figure 4 - Update Driver.

4. Click "Browse my computer for driver software" (see Figure 5).

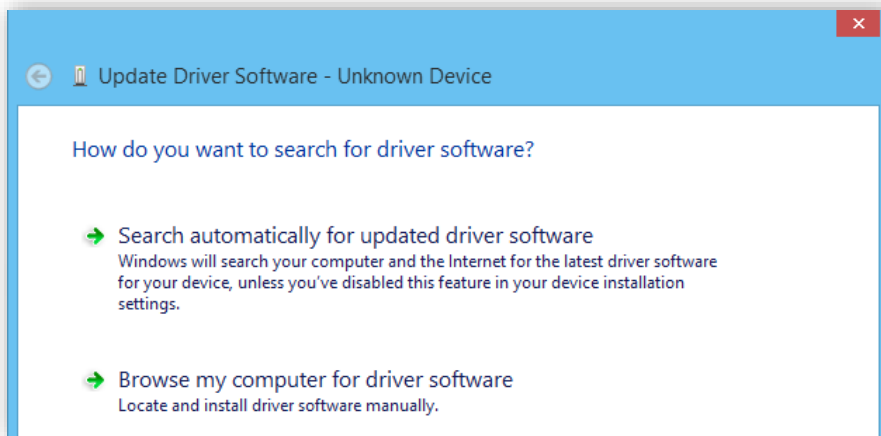


Figure 5 - Browse Computer for Driver Software.

5. Select the driver in the installation directory "Ancortek SDR/driver" according to Figure 6 below.

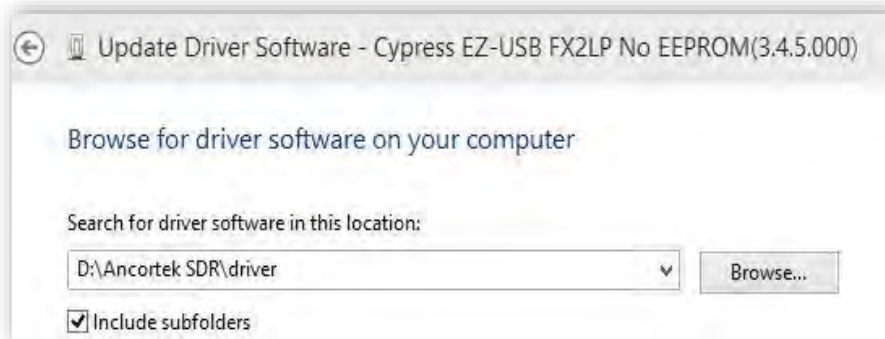


Figure 6 - Update Driver Software.

6. Driver software installed successfully.

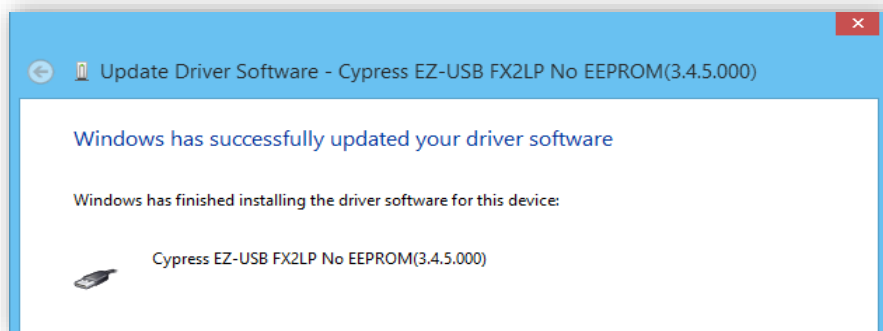


Figure 7 - Driver Software Installed Successfully.

7. Open the SDR-GUI.
8. MATLAB-based SDR-GUI is available for academic research groups upon request.

2.4 Graphical User Interface

2.4.1 Ancortek SDR-GUI

After running setup.exe and updating the driver software, Ancortek SDR-GUI, as shown in Figure 8, will appear. The GUI gives users access to selecting of waveforms, operating parameters, filtering types, and recording. It is capable of showing graphical representation of signals in time and frequency domain.

1. Control Panel

The Control Panel includes selecting of waveforms, operating parameters, filtering types, and recording.

- **Active Device**

- 1) The upper-left pop-up menu: List of detected Ancortek devices.
- 2) The upper-right pop-up menu: Center frequency or channel selection. Please choose the corresponding center frequency of the SDR-KIT before activating it. When using SDR-2400AD2, please choose CH1 or CH2 to select which channel to display.
- 3) Activate: Start the data stream and plotting. This button is also used to stop the data stream or plotting.
- 4) Refresh: Find Ancortek devices and list them in the upper-left pop-up menu.

- **Operating Parameters: Operating Mode**

- 1) FMCW Sawtooth: Range and Doppler measurements.
- 2) FSK: Range and Doppler measurements of moving targets.
- 3) CW: Doppler measurement.

- **Operating Parameters: Signal Parameters**

- 1) Bandwidth: Sweep bandwidth for FMCW Sawtooth waveform.
- 2) Samples per Sweep: Number of time samples per sweep.
- 3) Sweep Time: Length of one sweep. Changing the Sweep Time and Samples per Sweep will change the sampling rate.

- **Operating Parameters: Display Mode**

- 1) Range & Time Scope: Range obtained by FFT of the beat signal in FMCW Sawtooth mode.
- 2) Velocity Scope: Velocity obtained by FFT of the complex signal in FSK & CW modes.
- 3) Range-Velocity Map: Range-Doppler map in FMCW Sawtooth mode.
- 4) Range Waterfall: Waterfall of Range & Time Scope in FMCW Sawtooth mode and waterfall of detected range in FSK mode.
- 5) Velocity Waterfall: Waterfall of Velocity in FMCW Sawtooth, FSK, and CW modes.

- **Stream Filtering**

- 1) DC Subtraction: Remove the mean value of the signal waveform.
- 2) Amplitude Correction: Normalize the amplitude of signal waveform.
- 3) Hamming Window: Using Hamming Window before taking FFTs.
- 4) Direct Clutter Cancellation: Background subtraction for FMCW-Sawtooth mode.

Please make sure that there is no target of interest on the scene at the very beginning to collect clutter data when Direct Clutter Cancellation is enabled.

- 5) Outdoor Range Weighting: Correct range-based signal attenuation via radar equation.
- 6) Range Windowing: Select range to be displayed.
- 7) Velocity Windowing: Select velocity to be displayed.

- **Display Parameters: Dynamic Range**

- 1) Dynamic Range Auto: Automatic calculate the color limits using minimum and maximum data values.
- 2) Dynamic Range: Sets the color limits to specified minimum and maximum values when Auto is unchecked.

Note: Data values less than minimum or greater than maximum map to minimum and maximum, respectively. Values between minimum and maximum linearly map to the current color map.

- **Display Parameters: Update Rate**

Update Rate: Set update rate of graphical plots. Data processing time will affect the actual update rate.

- **Display Parameters: Doppler Parameters**

Sweep Count: Number of pulses collected for signal processing. Decrease this value will ease processing burden, thus, may increase update rate of graphical plots.

- **Display Parameters: Waterfall Parameters**

- 1) History Size: Number of rows in data for waterfall display.
- 2) Display Method: Resampling algorithm used for matrix data.
- 3) Color Scheme: Change the color map.

Note: In FSK mode, we suggest using Standard or B&W color scheme for Range Waterfall.

- **Export**

- 1) Screen Capture: Save the Top or Bottom View of graphical results into a file.
- 2) Duration: Raw data length to record.
- 3) Record Data Stream: Start recording. The raw data will be saved into .dat file for post-processing. Below listed is MATLAB sample codes for reading the *.dat file. Data is the raw data in (I + j Q) format.

```
filename = '2015-03-25-10-08-12.dat'; % File name
fileID = fopen(filename, 'r');
dataArray = textscan(fileID, '%f');
fclose(fileID);
radarData = dataArray{1};
clearvars fileID dataArray ans;

fc = radarData(1); % Center frequency
Tswep = radarData(2); % Sweep time in ms
NTS = radarData(3); % Number of time samples per sweep
Bw = radarData(4); % FMCW Bandwidth. For FSK, it is frequency step;
For CW, it is 0.
Data = radarData(5:end); % raw data in I+j*Q format
```

For SDR-KIT 2400AD2, please use the following codes:

```
% 24GHz two channel radar data read
filename = '2016-04-25-16-22-16.dat'; % File name
fileID = fopen(filename, 'r');
dataArray = textscan(fileID, '%f');
fclose(fileID);
radarData = dataArray{1};
clearvars fileID dataArray ans;

fc = radarData(1); % Center frequency
Tswep = radarData(2); % Sweep time in ms
NTS = radarData(3); % Number of time samples per sweep
Bw = radarData(4); % FMCW Bandwidth. For FSK, it is frequency step; For
CW, it is 0.
```

```
Data = radarData(5:end); % raw data in I+j*Q format
Data_1 = Data(1:2:end); % Data of channel 1
Data_2 = Data(2:2:end); % Data of channel 2
```

4) Record Status: When recording and saving are done, success information will appear.

- **Camera Controls**

1) Current Camera: Refresh Devices to show available cameras, and choose one from the listed cameras.

2) Camera Status: Press DISABLED to activate the camera. Press ENABLED to deactivate.

NOTE : Please resize the Camera window to your preferred size at start up. You could drag the Camera window or Radar Controls window to a different position.

2. Graphic Panel

The Graphic Panel shows the graphical representation of signals in time and frequency domain.

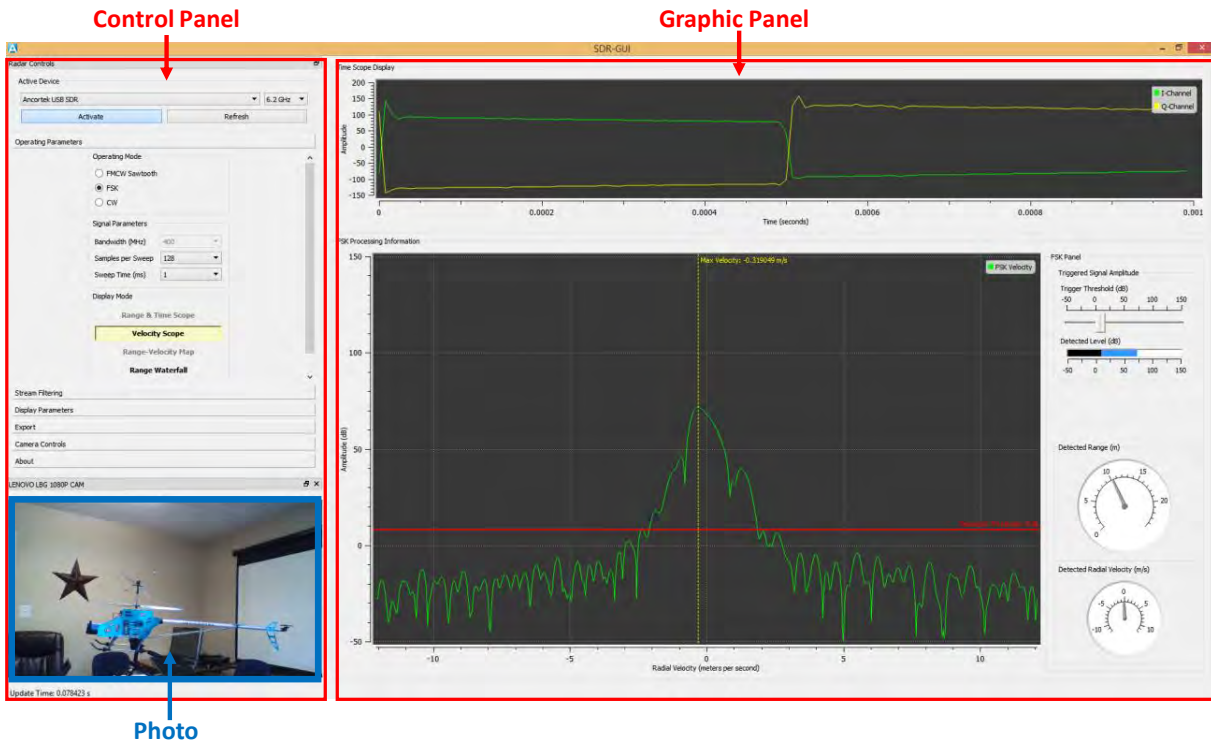


Figure 8 - Ancortek SDR-GUI.

2.4.2 MATLAB-based SDR-GUI

MATLAB-based SDR-GUI is available upon request. It is shown in Figure 9.

1. Control Panel

The Control Panel includes selecting of waveforms, operating parameters, filtering types, and recording.

- **Start/Stop**

Start: Start the data stream and plotting. Please remember to choose the right center frequency from the pop-up menu on the right. When using SDR-2400AD2, please choose CH1 or CH2 to select which channel to display.

Stop: Stop the data stream or plotting.

- **Waveforms**

- 1) FMCW_Sawtooth: Range and Doppler estimation.
- 2) FMCW_Triangle: Range and Doppler estimation without range-Doppler coupling effect.
- 3) FSK: Range and Doppler estimation of moving targets.
- 4) CW: Doppler estimation.

- **Parameters**

- 1) Bandwidth: Sweep bandwidth for FMCW.
- 2) Sweep Time: Length of one sweep or pulse.
- 3) Sampling Number: Number of time samples per sweep. Changing the Sweep Time and Sampling Number will change the sampling rate.
- 4) Frequency: Frequency of transmitted signal.

- **Filtering**

- 1) Direct Clutter Cancellation: Background subtraction for FMCW-Sawtooth and FMCW-Triangle. Please make sure that there is no target at the scene at the very beginning to collect clutter data when Direct Clutter Cancellation is enabled. Just after a few milliseconds, the target could enter the scene.
- 2) Range Notch Filter: Filter out the clutter nearby when the target of interest is far from the transceiver for FMCW-Sawtooth and FMCW-Triangle. Please change notch width according to the distance of target of interest.
- 3) Doppler Notch Filter: Filter out unwanted Doppler for range-Doppler map of FMCW-Sawtooth. Please change notch width according to the velocity of target of interest.

- **RawData**

- 1) Record: Record up to 8 seconds worth of raw data. Micro-Doppler analysis of the recorded raw data will appear in the Graphic Panel.
- 2) Save: Save the recorded raw data into .mat file for post-processing. Below is Matlab sample codes for reading the .mat file. Data is the raw data in (I+jQ) format.

```
raw = load('fmcw.mat');  
Data = raw.DATA;  
SweepTime = raw.SWEEPTIME;  
NTS = raw.samplingnumberpersweep;  
BandWidth = raw.BANDWIDTH;
```

```
samp_rate = 1/SweepTime*NTS*1000;
```

For SDR-KIT 2400AD2, please use the following codes:

```
% 24GHz two channel radar data read  
raw = load('two_channel.mat');  
Data_1 = raw.DATA1; % data of channel 1  
Data_2 = raw.DATA2; % data of channel 2  
SweepTime = raw.SWEEPTIME;  
NTS = raw.sampnumberpersweep;  
BandWidth = raw.BANDWIDTH;  
samp_rate = 1/SweepTime*NTS*1000;
```

3) Replay: Reshow the micro-Doppler analysis of the selected raw data.

2. Graphic Panel

The Graphic Panel includes selecting of graphical representation of signals in time and frequency domain.

• FMCW_Sawtooth graphical results

- 1) Stream: I & Q data.
- 2) Waterfall: Waterfall of range profile.
- 3) Range Profile: Range obtained from FFT of the beat signal.
- 4) Range Doppler: Range-Doppler map.

• FMCW_Triangle graphical results

- 1) Stream: I & Q data.
- 2) Range History: History of detected range.
- 3) Velocity History: History of detected velocity.

• FSK graphical results

- 1) Stream: I & Q data.
- 2) Velocity History: History of detected velocity.
- 3) Range History: History of detected range.

• CW graphical results

- 1) Stream: I & Q data.
- 2) Waterfall: Waterfall of velocity profile.
- 3) Velocity Profile: Velocity obtained from the FFT of the (I + jQ)

4) Velocity History: History of detected velocity.

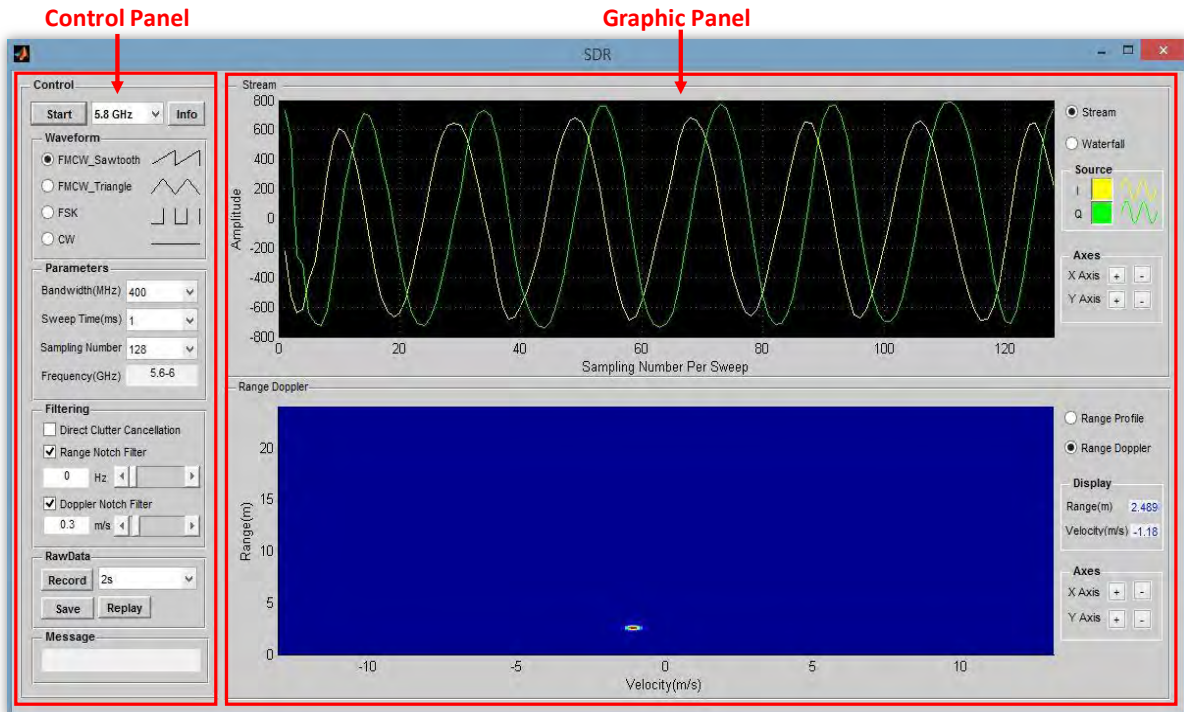


Figure 9 - MATLAB version of the SDR-GUI.

A.3 Histogram-based Thresholding

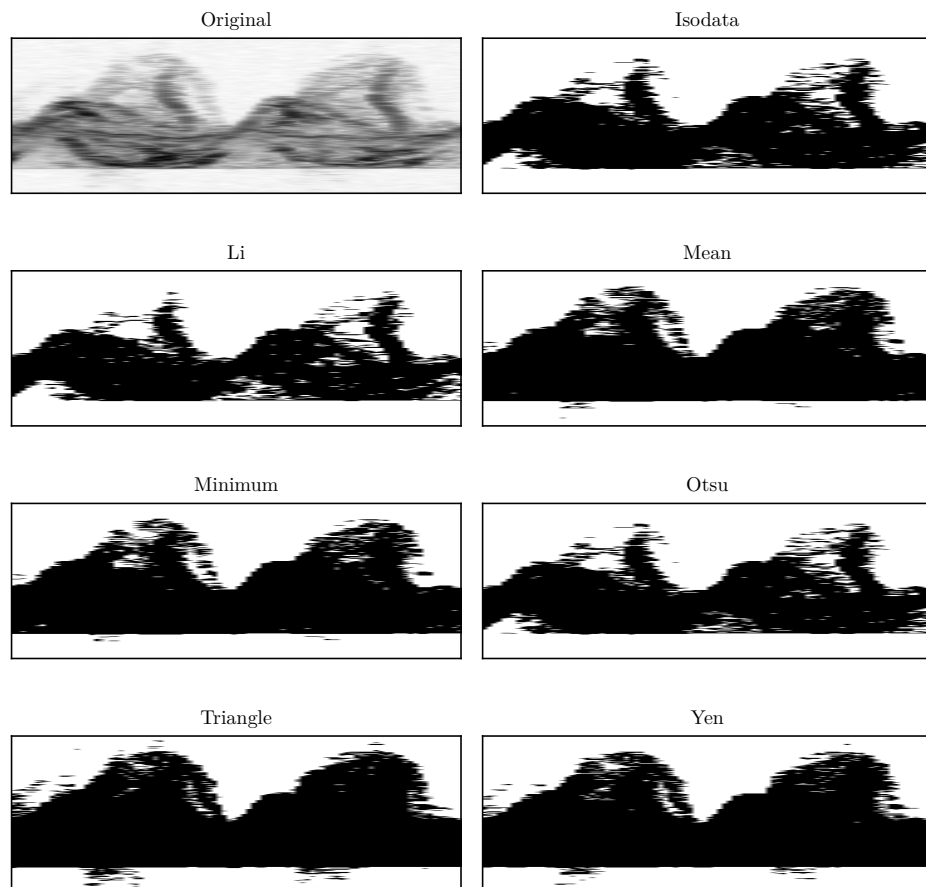


Figure A.2: Histogram-based thresholding techniques of Python library *skimage.filters*.

A.4 Datasheet of Radarbook2 Radar

In this section, the datasheet for the Radarbook2 module is presented. This radar module is used in this thesis work for two use-cases. First, the free walking human-ID task presented in Section 6.3, and published in Weller et al. 2021. Second, the activity recognition task presented in Section 6.4, and published in Aziz et al. 2022.



AN77-02 FMCW Basics

Radarbook2

Inras GmbH
Altenbergerstraße 69
4040 Linz, Austria
Email: office@inras.at
Phone: +43 732 2468 6384

Linz, July 2016

1 FMCW Basics

This application note shows how to perform FMCW measurements with the Radarbook2 (RBK2) equipped with a 77-GHZ MIMO frontend.

Goal of The Application Note

- Configure RF frontend and collect FMCW raw data.

To configure the RF board the `Rbk2IfxTx2Rx16` class design is provided. The implementation is inherited from the `Rbk2` class. Therefore it offers all methods specified in the base class and in addition methods for configuring the frontend are provided. The `Rbk2` class implements methods to configure the baseband board and the signal processing inside the Arria 10 SoC. The `Rbk2` class is inherited from the `Connection` class. The `Connection` class implements the TCP/IP communication to the board and it also can be used to communicate with the RadServe software.

In Fig. 1 a block diagram of the frontend is shown. In the following paragraph the configuration of the frontend and the transceiver chips with the RBK2 are described in detail. This application note highlights how to perform FMCW measurements without the need to handle the configuration of all the registers of the transceiver chips.

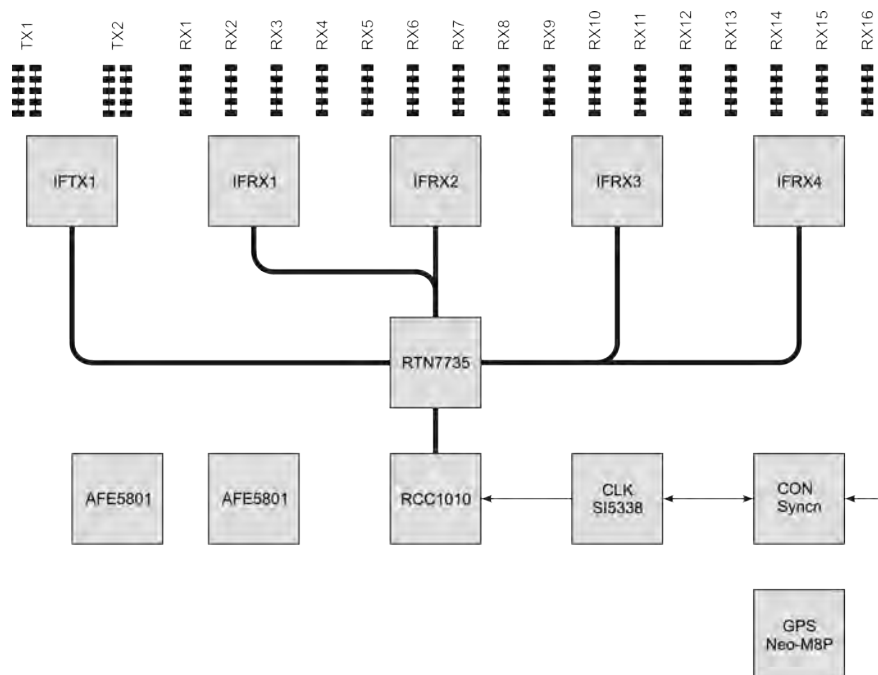


Figure 1: Block diagram of the frontend.

In the first step the RBK2 with the MIMO frontend must be powered on and connected to the PC. Check the status LEDs on the RBK2 to verify that the board has booted successfully and that the HPS system is running and the application loaded. In addition the RBK2, SeqTrig, and

Rbk2IfxTx2Rx16 classes must be added to the Matlab path. After generating a class object

```
Brd = Rbk2IfxTx2Rx16('PNet', '192.168.1.1');
```

the frontend and the sampling chain can be configured. In this application note we show how to perform FMCW measurements. The desired timing for the instantaneous transmit frequency $f_t(t)$ is shown in Fig. 2. In the current example, upchirps with a start frequency f_{Start} , a stop frequency f_{Stop} , a chirp duration T_{RampUp} are generated and repeated in an interval T_{Int} . The sampling chain

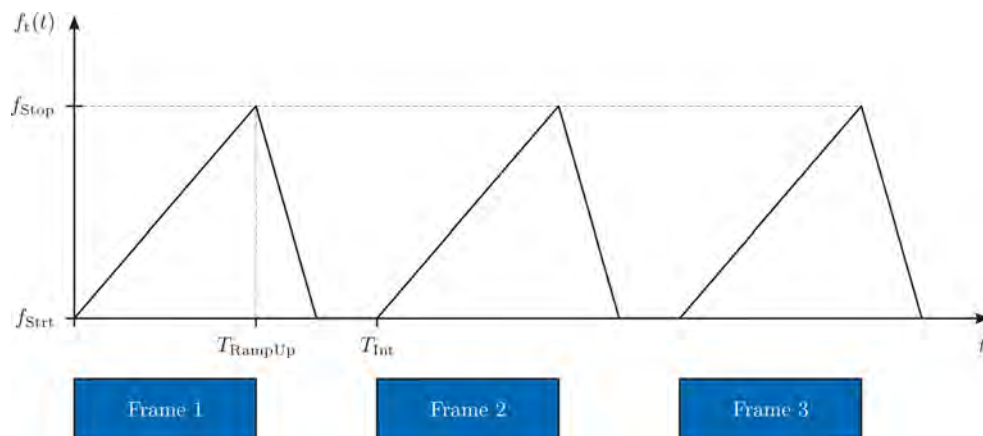


Figure 2: Measurement timing with instantaneous transmit frequency $f_t(t)$.

is configured to record the IF signals of all receive channels during the upchirp as shown in Fig. 2. At first the timing unit in the FPGA is put into reset.

```
Brd.BrdRst();
```

This is required to disable ongoing measurements, in case a previous measurement script was interrupted unexpectedly. In order to configure the frontend it is required to enable the frontend supply on the RBK2.

```
Brd.BrdPwrEna();
```

The supply LEDs on the baseboard indicate that the frontend supply is turned on, as shown in Fig. 3. The 77-GHz frontend requires both supplies and hence both LEDs are turned on after enabling the supply.

The frontends for the RBK2 support different clocking and trigger sources in order to enable radar networks. In this example we operate the board as master (MS) and we use its own clock signal. In addition the clock signal for the ADCs is set to 40 MHz. Currently 20 and 40 MHz are supported for the ADC clock f_{Adc} . The clock should be set to 40 MHz because this ensures that the sampling is synchronous to the programmed frequency ramps.

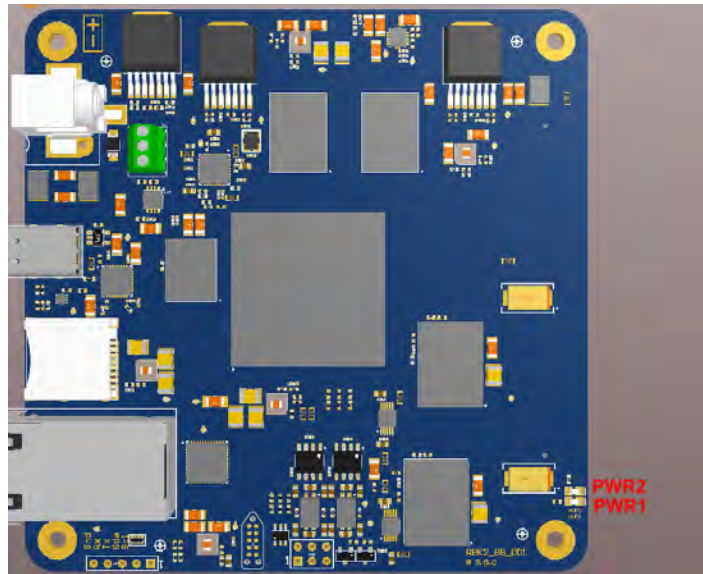


Figure 3: Power LEDs indicate the state of the frontend supply.

```
Brd.BrdSetRole('Ms', 40e6);
```

In the next step the receivers (IFRX1 - IFRX4) are enabled. The method

```
Brd.RfRxEna();
```

configures all IF channels of the four receive chips. In the next step the transmit chip is enabled

```
Brd.RfTxEna(1, 63);
```

The first argument specifies the transmit antenna to be activated and the second argument specifies the output power. The argument for the transmit antenna can be in the range from 0 to 2. The transmit path is deactivated by specifying 0 for the transmit antenna. If the value for the transmit antenna is non zero, then the specified transmit antenna is turned on permanently. The second argument is used to define the output power for the activated transmit channel. According to the datasheet a value in the range from 0 to 63 is allowed. Thereafter the timing of the transmit signal is specified with the following structure.

```
Cfg.fStrt = 76e9;  
Cfg.fStop = 77e9;  
Cfg.TRampUp = 512e-6;  
Cfg.TRampDo = 64e-6;  
Cfg.TInt = 200e-3;  
Cfg.N = 1000;
```

The parameter `TRampUp` defines the duration of the upchirp and the chirp repetition interval is

specified with the parameter `TInt`. The number of samples during a single upchirp is defined by the field `N`. Sampling is only performed during the upchirp sequence. The duration of the downchirp is specified by `TRampDo`. To configure the synthesizer and program the timing unit in the FPGA the method

```
Brd.RfMeas('ExtTrigUp', Cfg);
```

is called with the first argument `'ExtTrigUp'` and the second argument being the previously defined configuration structure. The `'ExtTrigUp'` measurement mode generates a trigger for the RCC1010 in the defined interval `TInt`. In addition, the sampling chain of the RBK2 is triggered to record the data during the upchirp. The method automatically configures the sampling frequency and the CIC filters in the sampling chain. The `Get` method `Brd.Get('fs')` can be used to read the configured sampling frequency. The desired sampling rate

$$f_{s,des} = \frac{N}{T_{RampUp}}$$

is defined by the chirp duration and the number of samples. Only if the f_{Adc} is an integer multiple of the desired sampling rate, then the actual sampling rate is the desired one. If not the firmware selects the sampling rate reduction

$$R = \text{floor}\left(\frac{f_{Adc}T_{RampUp}}{N}\right)$$

by truncating the fractional part and hence the actual sampling rate

$$f_s = \frac{f_{Adc}}{R}$$

is higher than the desired one. In this case only the first part of the ramp is sampled. The actual sampling rate can be readback with `Brd.Get('fs')`.

After the configuration is programmed, the method

```
Data = Brd.BrdGetData();
```

can be used to read back the sampled IF signals. Every call of `Brd.BrdGetData()` returns the IF signals for a single upchirp and all receive channel. In the default configuration the first values of every column contains the frame counter, which numbers the adjacent measurements. The frame counter numbers all frames recorded and if one or multiple values are missing, this indicates that the data transfer could not forward all the measured frames. In Fig. 4 the sampled signals for a single frame are plotted.

After finishing measurements the power supply of the frontend can be disabled again.

```
Brd.BrdPwrDi();
```

Moreover the timing unit can be put into reset with the method

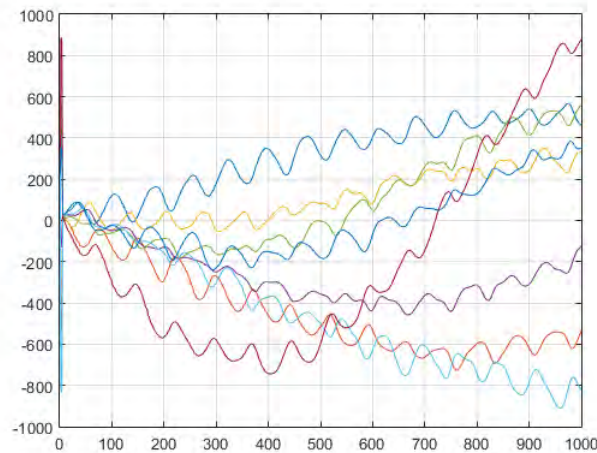


Figure 4: Sampled IF signals for a single upchirp and eight data channels (plot(Data)).

```
Brd.BrdRst();
```

Measurement Mode: ExtTrigUp

The 'ExtTrigUp' measurement mode is a basic mode, where the timing is generated by the timing unit (SeqTrig) in the FPGA. The timing unit generates a trigger signal for the RF synthesizer and for the ADC interface that collects the ADC data. The configuration structure provides additional entries that can be used to configure the timing.

```
Cfg.IniEve = 0;
Cfg.IniTim = 100e-3;
Cfg.CfgTim = 40e-6;
Cfg.ExtEve = 0;
Cfg.Strt = 1;
```

These entries are optional and need not be specified before calling RfMeas. The parameters IniEve and IniTim can be used to delay the start of the measurement. If IniEve is zero then the IniTim is the time before the measurement sequence is started as shown in Fig. 5. At the end of the RfMeas method the function BrdSampStrt() is executed. If Cfg.Strt is zero, then BrdSampStrt() is not called at the end of RfMeas and the user can start the timing manually. The BrdSampStrt() function starts the timing unit in the FPGA. After the period specified by IniTim the measurement sequence is started. At first in the configuration interval (mcodeCfg.CfgTim) the synthesizer is configured with an SPI command and after the configuration time a trigger is generated to start the upchirp and the sampling of the IF data. The duration of the configuration interval can be altered with Cfg.CfgTim. The measurement time is selected so that the configuration time and the measurement time add up to TInt.

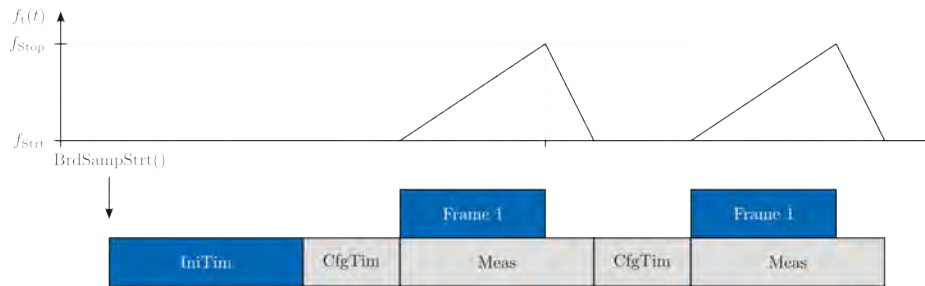


Figure 5: Measurement timing with `IniEve` and `ExtEve` set to zero.

If `IniEve` is set, then the timing unit waits for an event before measurements are started. In the master mode the event source is set to the firmware function `BrdSampTrig()`. Hence, the event can be triggered by the application script. In Fig. 6 the timing is started with an event after the selected `IniTim`. The event is only recognized at the end of the initialization interval. If the initialization interval is not finished an event is ignored. The user has to ensure that the event is generated after `IniTim`.

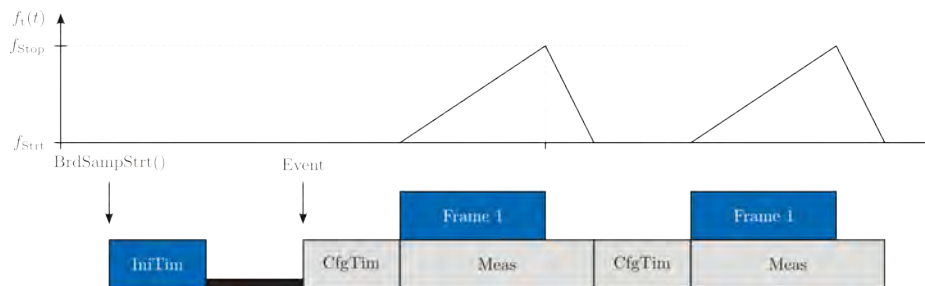


Figure 6: Measurement timing with `IniEve` set and `ExtEve` cleared.

If `ExtEve` is set, then the timing unit waits for an event at the end of every ramp as shown in Fig. 7. In the master mode the event for the measurements can be generated with the firmware function `BrdSampTrig()`.

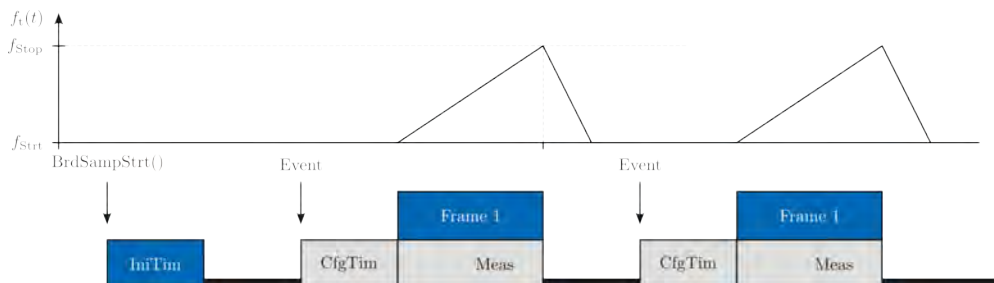


Figure 7: Measurement timing with `IniEve` and `ExtEve` set.

Kurzfassung

Das Studium von Radar zur menschlichen Erkennung basierend auf Deep Learning wird immer populärer. Es wurde gezeigt, dass der Mikro-Doppler (μ -D) Spektrogrammeffekt den Gang eines gehenden Menschen durch Erfassung der periodischen Mikrobewegungen der Gliedmaßen widerspiegeln kann. Der Forschungsbereich wurde erweitert, um die Erkennung von Menschen bei variablen Aktivitäten einzubeziehen, wodurch eine Vielzahl von Anwendungen untersucht wurde, wie z.B. Sturzerkennung. Zudem gibt es zwei Hauptfaktoren, die das Radar als leistungsstarken Sensor für solche Anwendungen etabliert haben. Erstens, die Radarerkennungsfähigkeit, die nicht durch Umweltbedingungen beeinträchtigt wird. Zweitens, das Mehrfach-Eingabe-Mehrfach-Ausgabe Übertragungsprotokoll, das die Radarerfassung und -verfolgung mehrerer Menschen ermöglicht.

In dieser Arbeit werden Radarsensoren für die indoor Erkennung und Identifikation von Menschen untersucht. Eine Kurzstreckenanalyse ($\leq 10m$) steht im Mittelpunkt aller vorgeschlagenen Ansätze. Zu Beginn der Forschung werden die Hauptmerkmale der erfassten gehenden μ -D Signatur extrahiert und analysiert. Erste Untersuchungen werden mit einem kombinierten halb-simulierten Signalmodell durchgeführt, das wir entwickelt haben. Das Modell simuliert ein typisches Einzel-Eingabe-Einzel-Ausgabe Radar zur Überwachung eines einzelnen gehenden Ziels. Reichweite und Geschwindigkeit sind die ersten beiden Radarmerkmale, die untersucht werden. Die Klassifizierung von Menschen und Robotern dient als Fallstudie zur Überprüfung der Ergebnisse. Das Reichweitenmerkmal wird bewertet, um das Mikrobewegungsverhalten zu reflektieren und so eine verbesserte Erkennungsfähigkeit zu gewährleisten. Aufgrund des Verhaltens von Körperteilen innerhalb der μ -D Signaturen bildeten die Ergebnisse der ersten Machbarkeitsstudien die Grundlage für das Konzept der menschlichen Erkennung. Zudem lenkte es unsere Aufmerksamkeit auf die Bedeutung der Nutzung mehrerer Radarmerkmale zur verbesserten Erkennung. Gehen ist die Hauptaktivität in all unseren Untersuchungen und wird als die wichtigste Aktivität für das Studium des μ -D Effekts im Radar betrachtet.

Die menschliche Identifikation ist das zweite Hauptforschungsziel aufgrund ihrer kritischen Anforderung in den meisten Sicherheitssystemen. Die Gehaktivität ist festgelegt, und daher werden ähnliche μ -D Signaturen unter Menschen erwartet. Daher gilt diese Aufgabe als die schwierigste und allgemeinste, die später auf andere Szenarien verallgemeinert werden kann. Das entwickelte Verfahren kann daher leicht auf andere Aufgaben zur menschlichen Erkennung angewendet werden. Für eine einzige Entscheidung besteht immer ein Kompromiss zwischen der Anzahl der einbezogenen Subjekte und der Länge des für die Identifikationsaufgabe erfassten Spektrogramms. Daher besteht das Forschungsziel darin, diese Einschränkungen durch Kombination mehrerer Radarmerkmale zu überwinden, um die Klassifizierungsambiguität zu reduzieren und die Klassenanzahl zu erhöhen, während eine Echtzeitklassifikation erreicht wird. Dafür wird der geschätzte Einfallswinkel des Radars untersucht, um ein neu gemessenes Mikro-Winkelgeschwindigkeitsspektrogramm zu erstellen, das das Verhalten der Mikrobewegungen des Körpers im Einfallswinkelmerkmal widerspiegelt.

Es wird auch berücksichtigt, dass die Qualität des Spektrogramms erheblich abnimmt, wenn es aus verschiedenen Betrachtungswinkeln in Bezug auf das Radar erfasst wird. Der Hauptgedanke ist, einen machbaren Deep Learning Ansatz zu untersuchen, um eine adaptive Lösung zu gewährleisten. Die Generalisierung und Robustheit des endgültigen Ansatzes werden durch seine Anwendung auf eine Aktivitätsklassifikationsaufgabe bewertet, die die Sturzerkennung einschließt. In diesem Zusammenhang sieht die Arbeit zwei Hauptaspekte vor. Erstens, Verringerung des erworbenen Trainingsdatensatzes. Zweitens wird der Echtzeitaspekt berücksichtigt, sodass alle Klassifizierungen bei einem einzigen Vorkommnis der Aktivität erfolgen, wobei ein einzelner Gehschritt $\approx 0.5s$ dauert.

