A Simulation-Based Framework for the Design of Direction-Independent Human Activity Recognition Systems Using Radar Sensors

Sahil Waqar

UiA University of Agder

A Simulation-Based Framework for the Design of Direction-Independent Human Activity Recognition Systems Using Radar Sensors

Sahil Waqar

A Simulation-Based Framework for the Design of Direction-Independent Human Activity Recognition Systems Using Radar Sensors

Doctoral Dissertation for the Degree *Philosophiae Doctor (Ph.D.)*
at the Faculty of Engineering and Science, Specialisation in Information and
Communication Technology

University of Agder
Faculty of Engineering and Science
2024

Preface

In the past decade, I have developed a strong interest in signal processing, machine learning, communications, radars, sensors, and the Internet of Things. This dissertation represents interdisciplinary research conducted at the Department of Information and Communication Technology (ICT), Faculty of Engineering and Science, University of Agder, Grimstad, Norway, under the supervision of Professor Dr.-Ing. habil. Matthias Uwe Pätzold.

The research presented in this dissertation will particularly pique your interest if we share the same enthusiasm for wireless channel modeling and simulation, machine learning, radar systems, and human-motion synthesis. While composing this dissertation, my goal was to elucidate the intricate relationships between radar and non-stationary objects in a wireless propagation environment. This research holds particular relevance for electrical, electronics, and computer engineers involved in motion synthesis and off-the-shelf commercial radar systems within the rapidly growing industries of healthcare, automotive, and smart homes. Furthermore, I stressed the importance of simulation models, which obviate the need for extensive measurement data acquisition for classifier training. For researchers, this dissertation unveils new research questions requiring further investigation.

I express my heartfelt gratitude to my principal supervisor, Prof. Matthias Pätzold, for his excellent guidance and supervision. I have learned a lot from his experience and knowledge. I am thankful to my co-supervisor Assoc. Prof. Bjørn Olav Hogstad from the Norwegian University of Science and Technology (NTNU). My gratitude goes to my colleagues at the mobile communications group (MCG), Dr. Muhammad Muaaz, Dr. Rym Hicheri, and Hasan Najjar. I thank the former members of the MCG, Assoc. Prof. Nurilla Avazov and Dr. Ahmed Abdelgawwad. I am grateful to Mrs. Katharina Pätzold and Ms. Emma Horneman for their administrative support during the past three years.

I am especially indebted to my parents, my wife, and my siblings for their constant support and love throughout this journey.

<div style="text-align: right">

Sahil Waqar
Grimstad, Norway
August 7, 2024

</div>

Abstract

Human activity recognition (HAR) systems play an important role in understanding and interpreting human movements across various domains, with applications ranging from automobiles to smart homes and health. This dissertation focuses on HAR within the realm of radio frequency (RF) sensing, with a primary focus on modeling the intricate influence of human motion on wireless channel characteristics, particularly in the context of frequency-modulated continuous wave (FMCW) radar systems. It presents a paradigm shift from experimental- to simulation-based approaches tailored for RF sensor-based HAR systems. The core innovation lies in a sophisticated channel model capable of transforming three-dimensional (3D) trajectories into high-fidelity simulated RF signals, offering substantial control over signal parameters for simulating diverse environmental conditions.

This research addresses two main challenges in HAR: accommodating multiple directions of human motion and tackling the scarcity of radar data for diverse scenarios. To overcome motion direction challenges, a distributed multiple-input multiple-output (MIMO) radar configuration is introduced, capturing multi-perspective radar signatures of multi-directional human activities. The configuration, complemented by the dynamic time warping (DTW) distance metric, facilitates the development of a direction-independent step counting system for multi-directional walking activities. To mitigate the problem of cross-channel interference, a novel range gating method is implemented, leveraging distinct RF delay lines within the distributed MIMO radar setup. This distributed MIMO radar configuration, providing complementary RF sensing, is well-suited for realizing direction-independent human activity recognition (DIHAR) systems. An experimental-based DIHAR system is developed, utilizing the multi-perspective MIMO radar configuration, to classify various multi-directional human activities. The system involves training a machine learning model with a large dataset of radar signatures, necessitating a comprehensive measurement campaign.

The dissertation highlights the limitations of experimental data-driven approaches, emphasizing the challenges of acquiring diverse and representative datasets for radar-based classifiers. It advocates simulation-based solutions, offering control over radar parameters, reducing training efforts, addressing user privacy concerns, and enabling the generation of varied training datasets tailored to specific conditions. An end-to-end simulation framework is introduced, incorporating an innovative channel model that transforms motion data into RF signals, alleviating the significant data scarcity challenge in radar systems. The simulation-centric approach eliminates the need for

resource-intensive measurement campaigns, showcasing a deep convolutional neural network (DCNN)-based HAR classifier with close to 100% accuracy. The simulation-centric HAR system's efficacy is validated using previously unseen experimental data from a physical FMCW radar system. The framework is further expanded to develop a DIHAR system, exclusively trained on simulated MIMO radar signatures, demonstrating its capability to simulate diverse radar datasets tailored to user-defined MIMO radar configurations.

The results presented in this dissertation showcase the successful mitigation of cross-channel interference, the development of robust experimental-based DIHAR systems, and the creation of a simulation framework with far-reaching implications for radar data generation. The simulation-based approach holds promise for advancements in various radar applications, marking a paradigm shift in radar-based classification and contributing to the elimination of resource-intensive, laborious and monotonous measurement campaigns.

Sammendrag

Systemer for gjenkjenning av menneskelig aktivitet (human activity recognition, HAR) spiller en viktig rolle når det gjelder å forstå og tolke menneskelig bevegelse i ulike domener, med bruksområder som spenner fra biler til smarte hjem og helse. Denne avhandlingen fokuserer på HAR innenfor sansing av radiofrekvens (radio frequency, RF), med hovedvekt på å modellere den intrikate påvirkningen av menneskelige bevegelser på trådløse kanalegenskaper, spesielt i forbindelse med av frekvensmodulerte kontinuerlige bølgeradar (frequency-modulated continuous wave, FMCW) systemer. Den representerer et paradigmeskifte fra eksperimentelle til simuleringsbaserte tilnærminger som er skreddersydd for RF-sensorbaserte HAR-systemer. Den viktigste innovasjonen ligger i en sofistikert kanalmodell som kan transformere tredimensjonale baner til simulerte RF-signaler med høy oppløsning, noe som gir betydelig kontroll over signalparametere for å simulere ulike miljøforhold.

Denne forskningen tar for seg to hovedutfordringer i HAR: hensynet til flere retninger for menneskelig bevegelse og mangelen på radardata for ulike scenarier. For å løse utfordringer knyttet til bevegelsesretningen introduseres radarkonfigurasjon med flere innganger og flere utganger (multiple-input multiple-output, MIMO) for å fange opp radarsignaturer med flere perspektiver fra aktiviteter i flere retninger. For å redusere problemet med interferens på tvers av kanaler implementeres en ny metode for avstandskontroll som utnytter ulike RF-forsinkelseslinjer i den distribuerte MIMO-radarkonfigurasjonen. Denne distribuerte MIMO radarkonfigurasjonen, som muliggjør komplementær RF-deteksjon, egner seg godt til å realisere retningsuavhengige systemer for gjenkjenning av menneskelig aktivitet (direction-independent human activity recognition, DIHAR). Vi har utviklet et eksperimentelt DIHAR-system som utnytter den multiperspektiviske MIMO-radarkonfigurasjonen til å klassifisere ulike multidireksjonelle aktiviteter. Systemet innebærer opplæring av en maskinlæringsmodell med et stort datasett av radarsignaturer, noe som krever en omfattende målekampanje.

Avhandlingen belyser begrensningene ved eksperimentelle databaserte tilnærminger, og understreker utfordringene med å skaffe ulike og representative datasett for radarklassifisering. Det argumenteres for simuleringsbaserte løsninger som gir kontroll over radarparametere, reduserer treningsinnsatsen, adresserer brukerens personvernhensyn, og gjør det mulig å generere ulike treningsdatasett som er skreddersydd til spesifikke forhold. Det introduseres et helhetlig rammeverk for simulering som inkluderer en innovativ kanalmodell som transformerer bevegelsesdata til RF-signaler og løser problemet med datamangel i radarsystemer. Den simu-

leringssentrerte tilnærmingen eliminerer behovet for ressurskrevende målekampanjer, og demonstrerer en HAR-klassifisering basert på et dypt konvolusjonalt nevralt nettverk med nesten 100% nøyaktighet. Effektiviteten til det simuleringssentrerte HAR-systemet valideres ved hjelp av tidligere usette eksperimentelle data fra et fysisk FMCW radarsystem. Rammeverket har blitt utvidet ytterligere for å utvikle et DIHAR-system som utelukkende er trent på simulerte MIMO radarsignaturer, og demonstrerer evnen til å simulere ulike radardatasett skreddersydd til brukerdefinerte MIMO radarkonfigurasjoner.

Resultatene i denne avhandlingen viser hvordan vi har lykkes med å redusere krysskanalinterferens, utvikling av robuste eksperimentelt baserte DIHAR-systemer, og skape et simuleringsrammeverk med vidtrekkende konsekvenser for radardatagenerering. Den simuleringsbaserte tilnærmingen er lovende for utvikling av ulike radarapplikasjoner, og markerer et paradigmeskifte innen radarbasert klassifisering og bidrar til å eliminere ressurskrevende, arbeidskrevende og monotone målekampanjer.

Publications

The dissertation incorporates five articles, presenting the outcomes of the author's research. Four of these articles have been published, while one (Paper E) is currently under review.

Paper A **S. Waqar** and M. Pätzold, "Interchannel interference and mitigation in distributed MIMO RF sensing," *Sensors*, vol. 21, no. 22, 2021, doi:10.3390/s21227496.

Paper B **S. Waqar**, M. Muaaz, and M. Pätzold, "Human activity signatures captured under different directions using SISO and MIMO radar systems," *Applied Sciences*, vol. 12, no. 4, 2022, doi:10.3390/app12041825.

Paper C **S. Waqar**, M. Muaaz, and M. Pätzold, "Direction-independent human activity recognition using a distributed MIMO radar system and deep learning," *IEEE Sensors Journal*, vol. 23, no. 20, pp. 24916-24929, 2023, doi:10.1109/JSEN.2023.3310620.

Paper D **S. Waqar** and M. Pätzold, "A simulation-based framework for the design of human activity recognition systems using radar sensors," *IEEE Internet of Things Journal*, 2023, doi:10.1109/JIOT.2023.3344179.

Paper E **S. Waqar**, M. Muaaz, S. Sigg, and M. Pätzold, "A paradigm shift from an experimental-based to a simulation-based framework using motion-capture driven MIMO radar data synthesis," *IEEE Sensors Journal*, 2024, doi:10.1109/JSEN.2024.3386221.

The papers listed below are relevant but are not included in the dissertation.

Paper F M. Muaaz, **S. Waqar**, and M. Pätzold, "Radar-based passive step counter and its comparison with a wrist-worn physical activity tracker," in *4th International Conference on Intelligent Technologies and Applications INTAP*, 2022, pp. 259–272.

Paper G   M. Muaaz, **S. Waqar**, and M. Pätzold, "Orientation-independent human activity recognition using complementary radio frequency sensing," *Sensors*, vol. 23, no. 13, 2023, doi:10.3390/s23135810.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
AAL	ambient assisted living
Adam	adaptive moment estimation
ADC	analog to digital converter
AIM	automatic identification of marker
ASD	active segment detection
CAE	convolutional auto encoder
CNN	convolutional neural network
CPI	coherent processing interval
CSI	channel state information
CW	continuous wave
DBF	digital beamforming
DCNN	deep convolutional neural network
DIHAR	direction-independent human activity recognition
DNN	deep neural network
DTW	dynamic time warping
FDMA	frequency division multiple access
FEN	feature extraction network
FFT	fast Fourier transform
FMCW	frequency-modulated continuous wave
FOV	field of view
fps	frames per second
FS-ADA	few-shot adversarial domain adaptation
GAN	generative adversarial network
HAR	human activity recognition
IoT	Internet of Things
IQ	in-phase and quadrature
JDS-TL	joint domain and semantic transfer learning
LiDAR	light detection and ranging
LSTM	long short-term memory
MBGAN	multibranch generative adversarial network

MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MLP	multilayer perceptron
mm-wave	millimeter wave
MoCap	motion capture
PCB	printed circuit board
PRF	pulse repetition frequency
PRI	pulse repetition interval
QTM	Qualisys track manager
ReLU	rectified linear unit
RF	radio frequency
RGB	red, green and blue
RNN	recurrent neural network
SAR	synthetic aperture radar
SD	standard deviation
SDR	software-defined radar
SIMO	single-input multiple-output
SISO	single-input single-output
STFT	short-time Fourier transform
SVM	support vector machine
TDMA	time division multiple access
TL	transfer learning
TV	time-variant
TV-CIR	time-variant channel impulse response
WRGAN-GP	Wasserstein refined generative adversarial network with gradient penalty

Chapter 1

Introduction

In the context of ambient assisted living (AAL), human activity recognition (HAR) systems allow real-time monitoring of human activities. These systems are instrumental in providing valuable insights into behavioral patterns, health and well-being of individuals. By accurately recognizing the type of human activities, HAR systems help the development of context-aware AAL systems [1, 2, 3, 4, 5]. Such AAL systems offer timely assistance, address anomalies, and ensure optimal day-to-day support. To date, various algorithms as well as sensing modalities have been investigated to develop HAR systems, including vision sensors, acoustic sensors, inertial sensors, and radio frequency (RF) sensors, as detailed in Figure 1.1. Vision-based sensors rely on visual data to perceive the environment, which may include ordinary red, green and blue (RGB) cameras, infrared cameras, thermal cameras, or depth sensors such as Kinect. Acoustic sensors, on the other hand, rely on sound signals coming from a microphone or a microphone array. Wearable inertial sensors like accelerometers, gyroscopes, and smartwatches are also widely accepted for capturing motion patterns. Moreover, for HAR systems, multi-modal sensor integration coupled with feature- and decision-fusion strategies, context-awareness, and ensemble learning approaches, plays a vital role in enhancing the accuracy and robustness of the overall system.

1.1 Radar-Based Classifiers

RF sensors, particularly those employing Wi-Fi and radar signals, have emerged as an effective modality for many applications including HAR. Other RF sensors such as Bluetooth, universal software radio peripherals and channel sounders are also relevant for HAR. An RF-based HAR system exploits the changes in signal propagation resulting from human movements to detect and recognize human activities. In other words, RF sensors leverage the variations in measured channel characteristics, which are amplitudes, phases, or frequencies of the received RF signals. For HAR systems, by virtue of these channel characteristics, RF sensors capture the signatures of moving body segments. Theoretically, body segments consist of an infinite number of point scatterers; however, we will see in this dissertation that

HAR system   **Sensing type**   **Techniques and sensors**

- Ambient RF sensing
  - SISO and MIMO radar systems [6, 7, 8, 9]
  - Wi-Fi systems [10, 11, 12]
  - Universal software radio peripherals [13]
  - Pattern recognition techniques [14]
  - Machine/deep learning models [15, 16, 17]

- Vision sensing
  - Cameras, infrared cameras, thermal cameras [18]
  - Depth sensors (e.g., Kinect) [19, 13]
  - Wearable cameras (e.g., smart glasses) [20]
  - Image processing techniques [21]
  - Object detection and tracking [22]
  - Pose estimation [23, 24]

- Acoustic sensing
  - Microphones [25]
  - Microphone arrays [26] (e.g., circular)
  - Audio signal processing [27]
  - Acoustic event detection [28, 27]

- Inertial sensing
  - Accelerometers [29, 30, 31, 32, 33]
  - Gyroscopes [34]
  - Magnetometers [32]
  - Wearable devices (e.g., smartwatches) [35]
  - Feature extraction algorithms [36]

- Multi-modal sensing
  - Multi-modal sensor integration [37, 26]
  - Feature fusion techniques [38, 39]
  - Context-aware approaches [40, 41, 42]
  - Ensemble learning methods [43, 44]
  - Decision fusion strategies [45]

HAR systems

Figure 1.1: HAR systems realized using different sensing modalities and techniques.

modeling each body segment with a limited number of point scatterers is adequate.

Within the realm of RF sensing, radar sensors are preferred because, unlike Wi-Fi sensors, radar sensors do not encounter carrier frequency offsets that lead to noisy phases in the channel frequency response. It is due to this factor and the recent advancements in the miniaturization and commercialization of single-input single-output (SISO) and multiple-input multiple-output (MIMO) radar systems that radar-based classifiers have garnered such a considerable interest among researchers. Generally, radar-based classifiers offer robust performance in classifying multiple classes, even in challenging situations and scenarios, e.g., classifying complex motion patterns based on their micro-Doppler signatures. Other advantages include detection and classification capabilities in non-line-of-sight environments and robustness against variable lighting conditions. Moreover, unlike vision-based approaches, users' privacy is not a disconcerting factor in radar-based classification systems, making them a compelling choice for various applications, including HAR in AAL. Radar-based systems offer privacy advantages over vision-based methods by capturing movement patterns and micro-Doppler signatures instead of visual images, thus reducing the risk of visual privacy intrusions. However, it is important to acknowledge that while radar-based systems offer enhanced privacy compared to cameras, they are not entirely free from privacy implications. With recent advancements, radar data can potentially reveal sensitive information about individuals' activities and routines, which could be misused if not properly safeguarded. Thus, ensuring data security, implementing access controls, and complying with privacy regulations remain essential.

1.2 Human Activity Recognition Systems

HAR systems play an important role in understanding and interpreting human movements in various contexts. These systems aim to automatically identify and categorize human activities based on data measured by different sensors (see Figure 1.1). Research in the field of HAR has demonstrated significant advancements in recent years [16, 46, 47, 48]. The constant interest in HAR classifiers stems from its wide-ranging applicability. Over time, HAR classifiers have demonstrated their utility in diverse areas, including automobiles [49], robotic systems for social interaction [50], smart homes [51], and health [52]. As outlined in Table 1.1, researchers have investigated various sensing modalities and features to realize HAR systems, including vision sensors, acoustic sensors, wearable motion capture (MoCap) sensors, and RF sensors. The table also indicates that a significant portion of the research in HAR systems relies on sensor data obtained through experiments. In this dissertation, we emphasize the constraints associated with experimental-data-driven HAR approaches and stress the significance of simulation-based methods in addressing these limitations (see Section 1.4).

Various approaches in the literature contribute to the field of HAR using radar and other sensing modalities (see Table 1.1). The authors of [53] integrated a re-

Table 1.1: Approaches towards HAR systems.

Sensors	Features	Experiments	Simulations	Papers
Wearable	Time-series	✓	–	[53, 64, 65]
Acoustic	Mel-spectrogram	✓	–	[26, 66]
Optical	Skeleton joints	✓	–	[67, 51]
Radar and Wi-Fi	Temporal features, micro-Doppler signatures, and/or range-Doppler maps	✓	–	[54, 55, 56, 57, 10, 11, 12, 58, 59]
Radar and Wi-Fi	Micro-Doppler signatures	–	✓	[60, 24, 61]
Radar and MoCap	Radial velocity distributions	–	✓	[62, 63]

current neural network (RNN) with attention mechanisms to extract time-series features for daily life activities. Recently, researchers have incorporated acoustic sensors for HAR utilizing Mel-spectrogram features [26]. In [51], Kim et al. presented a HAR method utilizing features extracted from joints of a skeleton model. Erol et al. [54] proposed a multilinear subspace HAR scheme leveraging slow-time features, fast-time features, and Doppler frequency. The authors of [55] conducted time-frequency analysis, both linear and bilinear, playing a crucial role in human motion classification. For HAR, Fioranelli et al. [56] analyzed and extracted valuable informaion from human micro-Doppler signatures, and Jokanovic et al. [57] integrated information from both time-frequency and range domains. Muaaz et al. [10, 11, 12] computed activity fingerprints from Wi-Fi channel state information (CSI). In [58, 59], we recently reported the use of complementary RF sensing for classifying multi-directional human activities, introducing a direction-independent human activity recognition (DIHAR) system. Chen et al. [60], Ahuja et al. [24], and Vishwakarma et al. [61] are among the few to focus on simulation-based solutions. In [62] and [63], Waqar et al. devised simulation-based HAR and DIHAR systems, respectively. Such diverse studies contribute significantly to the advancement of HAR methodologies, incorporating various sensing modalities and approaches.

Let us focus our attention to radar-based HAR systems, a topic central to this dissertation. In this context, two key research dimensions are considered: accommodating multiple directions of human motion and tackling the unavailability of radar data for diverse scenarios. To address the challenge posed by the direction of motion, we turn to the utilization of distributed MIMO radars, as mentioned in the subsequent section. This is the primary focus of our research in Paper B and C. In addressing the data scarcity challenge within radar systems, we present a novel simulation-based solution. The specifics of this solution are elaborated in Paper D and E, focusing on SISO and distributed MIMO radar systems, respectively.

1.3   Experimental-Based HAR systems

Contemporary HAR systems developed with monostatic radars exhibit good performance for unidirectional human motion, particularly when the motion is aligned with the radar boresight axis. However, a critical problem arises when a human moves at a right angle to the radar's boresight axis, impacting the performance of SISO or monostatic MIMO radar-based HAR systems. To address this limitation and consider diverse directions of motion, distributed MIMO radar systems are favored for their multi-perspective illumination within indoor environments. Figure 1.2 illustrate an example of such a distributed MIMO radar system, comprising two monostatic radar subsystems, designed for recognizing multi-directional activities of a human subject [58, 59, 68]. Such experimental setups provide complementary RF sensing, making them well-suited for realizing DIHAR systems.
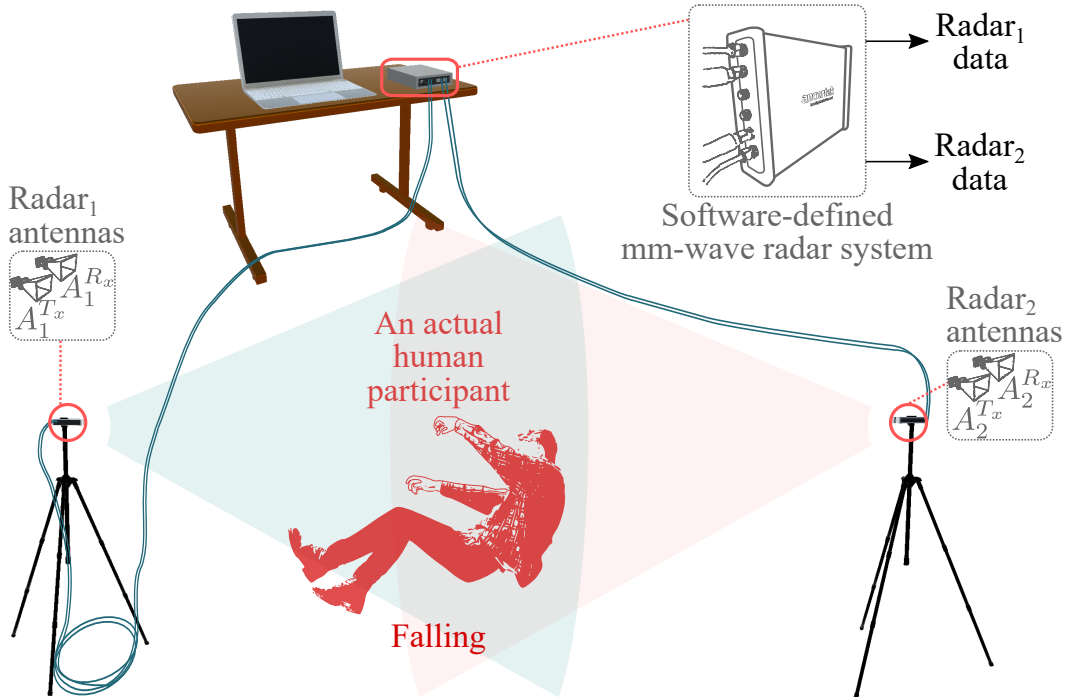
Figure 1.2: A measurement setup of a distributed 2×2 MIMO radar system to develop direction-independent HAR systems.

In the realization of a HAR system, a machine learning model is initially trained with a large dataset of radar signatures (human activity fingerprints). Figure 1.3 illustrates the design of an experimental-based DIHAR system utilizing multi-perspective radar signatures. From the activity fingerprints of Radar$_1$ and Radar$_2$, encompassing time-variant (TV) radial velocity distributions and/or TV mean radial velocities, features are extracted and fused to classify multi-directional human activities. This process is illustrated in Figure 1.3, showcasing the integration of information from both radar subsystems to realize a DIHAR system capable of recognizing diverse human motions. Note that relying solely on real radar signatures poses challenges, as acquiring them in large quantities demands significant resources and efforts. The time-consuming and resource-intensive nature of collecting radar signatures from
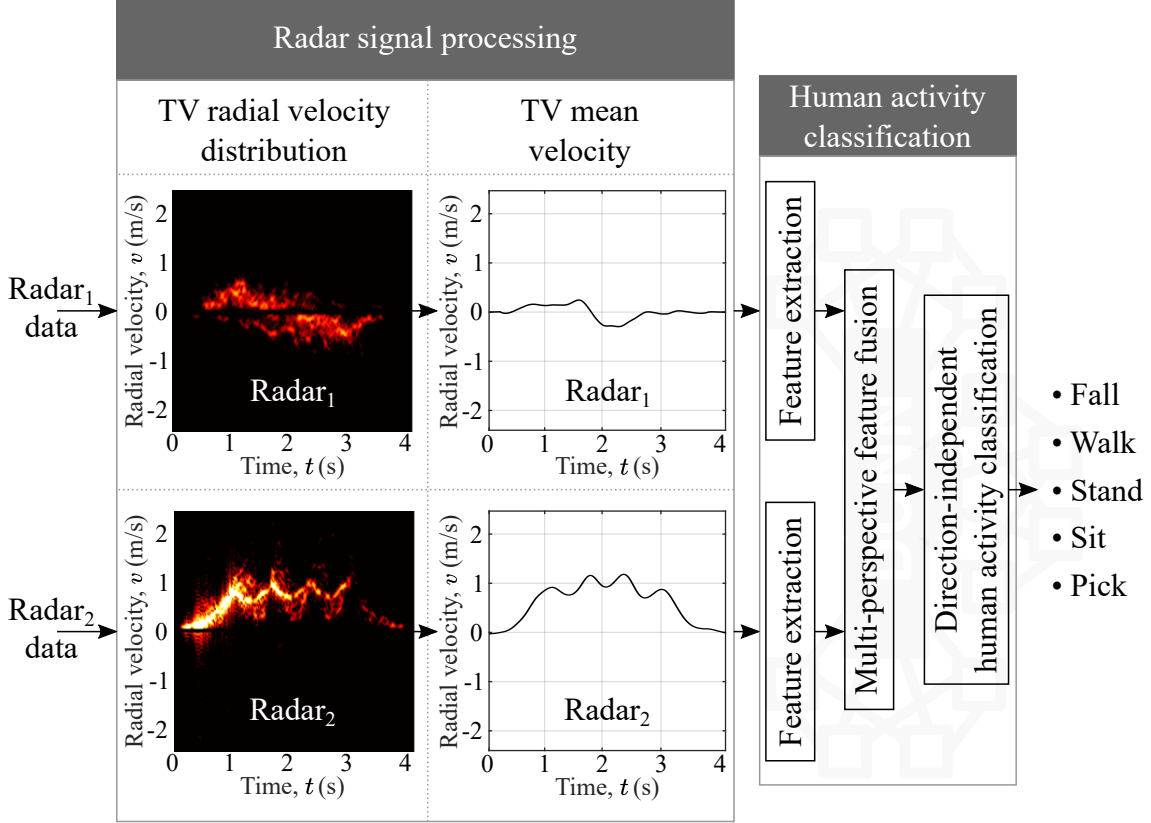
Figure 1.3: Design of an experimental-based direction-independent HAR system by using multi-perspective 2×2 MIMO radar signatures.

real individuals often results in datasets with limited scope, applicable only to specific scenarios and system parameters. In contrast to other sensing modalities, e.g., cameras, radar data is not readily available in various configurations and conditions, necessitating the development of fully customizable and flexible simulation-based data generation methods.

1.4   Simulation-Based HAR Systems

There is a need for a comprehensive simulation-based methodology to address the challenge of limited radar data. Several challenges may arise during the design and development phase of radar-based classifiers. One major challenge is the acquisition of diverse, high-fidelity, and representative datasets for radar-based classifiers. Most radar-based HAR systems found in the literature are realized using experimental data. In this context, simulation-centric approaches present numerous benefits over experimental methods. These benefits include addressing user privacy concerns, the capability to enhance training data, the flexibility to replicate radar datasets with user-defined target motion characteristics, and resource minimization. Simulation-centric approaches provide control over radar parameters for diverse operational and environmental conditions, allowing the generation of varied training datasets tailored to specific conditions. This significantly reduces the training effort for radar-based

classifiers. Thus, to facilitate the creation of diverse training datasets for radar-based classifiers, we suggest a simulation-centric HAR system relying solely on simulated radar data for training. We will see in Chapters 4 and 5 that the key benefit of the simulation approach lies in its ability to convert MoCap data into radar data, rendering it applicable to a broad range of real-world situations. Therefore, the availability of online MoCap data repositories, like Mixamo [69], further boosts the utility of such simulation-centric approaches across various domains such as healthcare, smart homes, sports, and more.

Our simulation-based design of HAR systems encompasses the conversion of human MoCap data into radar signatures using our channel model, as visually represented in Figure 1.4. Notably, the MoCap data can be incorporated from various sources, including the Mixamo online repository, optical/wearable MoCap systems and biomechanical modeling. To tailor the MoCap data to specific requirements, it can be retargeted and augmented using animation tools. Figure 1.4 illustrates the activity simulation module, which generates three-dimensional (3D) trajectories corresponding to the simulated motion. These spatial trajectories serve as input to the channel model that is responsible for simulating radar raw in-phase and quadrature (IQ) data. Among other things, the channel model takes into consideration antenna positions, radar type and its configuration, and radar operational parameters. This integrated approach ensures a realistic simulation of radar data, addressing the radar data scarcity challenge in a controlled and configurable manner. In order to prove the real-world efficacy of the simulation-centric HAR system, previously unseen experimental data from a physical frequency-modulated continuous wave (FMCW) radar system is used for the testing phase.

In Table 1.2, we compare the simulation-based HAR framework [62], originally proposed in Paper D, with state-of-the-art HAR methods, including joint domain and semantic transfer learning (JDS-TL) [70], hybrid convolutional neural network (CNN)-RNN [6], multi-view CNN-long short-term memory (LSTM) [71], few-shot adversarial domain adaptation (FS-ADA) [72], multibranch generative adversarial network (MBGAN) [73], Wasserstein refined generative adversarial network with gradient penalty (WRGAN-GP) [74], and Vid2Doppler [24]. The table provides a concise overview of their system description, the type of training dataset used (experimental, simulated, or both), the method of dataset simulation adopted (e.g., generative adversarial network (GAN) or other), and their corresponding classification accuracies. Clearly, the simulation-based HAR framework [62] surpasses other methods, obtaining an average (maximum) classification accuracy of 94% (98.4%). For more details on the state-of-the-art methods mentioned above, refer to Section D.8.

1.5 Research Questions

The list of research questions (RQs), addressed in the subsequent chapters, is provided as follows:

Table 1.2: State-of-the-art radar-based HAR methods and their classification accuracies (adapted from Paper D).

HAR method	System description	Training dataset	Dataset simulation method	Accuracy
JDS-TL [70]	Semisupervised transfer learning (TL) and domain adaptation using partially (10%) labeled radar data for HAR	Experimental	None	87.6%
Hybrid CNN-RNN [6]	Distributed radar network with spatial-temporal pattern extraction for HAR	Experimental	None	90.8%
Multi-view CNN-LSTM [71]	Convolutional auto encoder (CAE)-based unsupervised feature learning and multi-view data fusion via LSTM	Experimental	None	92%
FS-ADA [72]	Learning common feature space of pre-existing and newly acquired training datasets	Experimental	None	91.6%
MBGAN [73]	Physics-aware GAN-based systems that utilize limited real data to synthesize micro-Doppler signatures	Simulated and experimental	GAN-based data synthesis using limited real data	89.2%
WRGAN-GP [74]	Generation and refinement models for realistic spectrogram synthesis	Simulated and experimental	GAN-based data synthesis using limited real data	94.9%
Vid2Doppler [24]	Cross-domain translation: Creating synthetic Doppler signatures from videos of human activities	Simulated	Translation of video dataset to Doppler dataset	81.4%
Fully simulation-based HAR framework [62]	Simulating RF signals and radar signatures by using simulated 3D trajectories of virtual point scatterers	Simulated	Translating MoCap data to radar data using sensor modeling	Mean: 94% Max: 98.4%

RQ 1: How to resolve the cross-channel interference of a distributed MIMO radar system?

RQ 2: How to devise a distributed MIMO radar framework that overcomes the limitations posed by a monostatic SISO or a monostatic MIMO radar system in the context of the direction of motion?

RQ 3: How to realize a measurement-based DIHAR system? How to extract features from the activity fingerprints of a radar system and perform the feature fusion?

RQ 4: How to simulate realistic human activities, and generate 3D trajectories for different body segments?

RQ 5: How can a physical layer channel model be devised for millimeter wave (mm-wave) FMCW radar systems?

RQ 6: How can realistic radar signatures, including a TV range profile, TV radial velocity distribution (micro-Doppler signature), and TV mean radial velocity, be simulated for human activities?

RQ 7: How to devise a simulation-centric HAR system underlying a SISO radar model?
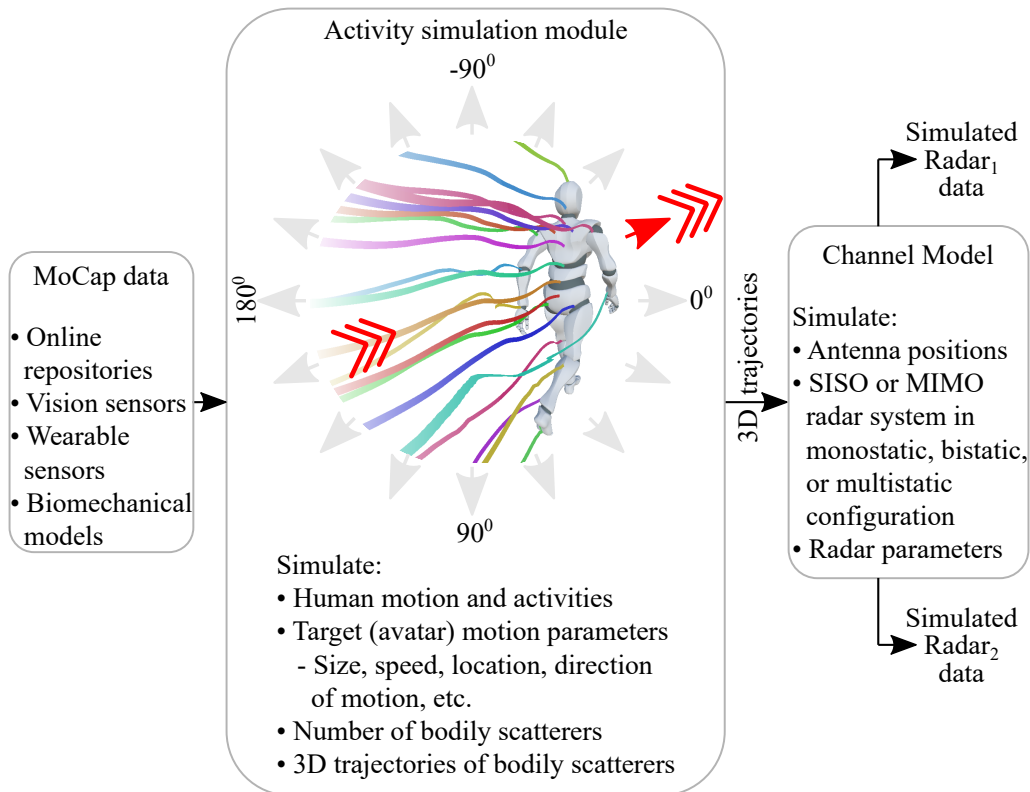
Figure 1.4: A simulation-based design to generate realistic radar signatures for developing real-world HAR systems.

RQ 8: How to devise a simulation-centric DIHAR system underlying a distributed MIMO radar model?

1.6 Dissertation Organization

The dissertation consists of six chapters and includes five papers. The six chapters provide a comprehensive summary of the contributions, while the five technical papers delve into the specific details of these contributions and present their respective results. The organization of the dissertation is as follows:

- Chapter 2 succinctly addresses RQs 1 and 2 by summarizing Papers A and B, respectively. A range gating method is introduced to mitigate the cross-channel interference in a distributed MIMO radar system, thereby addressing RQ 1. After mitigation, the measured TV radial velocity of a swinging pendulum are cross-validated using an extended analytical model. Furthermore, Chapter 2 addresses RQ 2 by investigating the impact of multi-directional human motion on measured channel characteristics. To overcome limitations inherent in monostatic radar systems, a multi-perspective distributed MIMO radar configuration is introduced. For quantitative assessment, its performance is compared with a conventional monostatic SISO radar system using the dynamic time warping (DTW) method.

- Chapter 3 answers RQ 3 by presenting the research findings from Paper C and addressing a physical layer challenge in contemporary radar-based HAR systems. To extend the application of HAR in RF sensing, we explore the intricacies of classifying multi-directional activities using a dataset obtained from a multi-view MIMO radar system. The development and analysis of three HAR systems—$SISO^{(1D)}$, $SISO^{(2D)}$, and $MIMO^{(2D)}$—are pivotal in tackling the identified physical layer challenge. $SISO^{(1D)}$ represents a conventional SISO radar-based HAR system. $SISO^{(2D)}$ is a conventional SISO radar-based DIHAR system, revealing a notable decline in classification performance. Lastly, $MIMO^{(2D)}$ is introduced, capable of classifying multi-directional human motion while significantly enhancing overall classification performance, approaching 100%.

- Chapter 4 provides a summary of the contributions in Paper D and addresses RQs 4–7. It introduces a comprehensive simulation framework, converting MoCap data into RF signals, to overcome the data scarcity issue in radar systems. The simulation framework includes synthesizing spatial trajectories, emulating a realistic mm-wave FMCW radar system through an innovative channel model, simulating radar signatures, and developing a simulation-centric HAR system. RQ 4 is answered by efficiently simulating diverse human activities and generating 3D trajectories for multiple markers on the avatar's body segments. A novel channel model is presented to address RQ 5, which takes

spatial trajectories as input and outputs simulated RF signals. RQ 6 is answered by simulating realistic radar signatures from the generated RF data. Finally, RQ 7 is addressed by training a HAR classifier exclusively with simulated radar signatures.

- Chapter 5 encapsulates the contributions of Paper E while addressing RQ 8. The chapter focuses on developing a DIHAR classifier exclusively trained with simulated MIMO radar signatures. It elucidates the data augmentation methods—motion-layer synthesis, physical-layer synthesis, and signal-layer synthesis—within the simulation-centric approach, enabling the simulation of diverse radar datasets for user-defined MIMO radar configurations in a resource-efficient manner. The motion-layer synthesis is explored, introducing random variations to avatar features for the synthesis of a broad spectrum of human motions. Additionally, the chapter details the physical-layer synthesis, facilitating the conversion of uni-directional movement into multi-directional movement and the emulation of user-defined antenna configurations (monostatic, multistatic) for SISO, single-input multiple-output (SIMO), multiple-input single-output (MISO), and MIMO radar systems. Lastly, the signal-layer synthesis stage is discussed, which allow the simulation of diverse radar signatures for individual MoCap examples.

- Chapter 6 outlines the key findings of this dissertation and offers insights into potential directions for future research.

Chapter 2

A Distributed MIMO Radar System

2.1 Introduction

Distributed multiple-input multiple-output (MIMO) radar systems have gained significant attention in recent years due to their ability to provide multi-perspective illumination of three-dimensional (3D) indoor wireless propagation environments. The multi-perspective illumination enables the capture of micro-Doppler characteristics irrespective of a target's direction of motion, thereby enhancing target detection and tracking capabilities. Nevertheless, in the operation of distributed MIMO radar systems, where multiple transmitters share temporal and spectral resources, the occurrence of cross-channel interference remains a pertinent concern. This interference can arise from various sources, including radio frequency (RF) leakage within the receivers' circuitry, mutual coupling between antennas, and crosstalk between transmitters and receivers. These interference factors collectively contribute to the degradation of system performance, as discussed in Paper A.

A critical problem with modern single-input single-output (SISO) or monostatic MIMO radar-based human activity recognition (HAR) systems is their inability to perform well when a human moves at a right angle to the radar's boresight axis. This rarely discussed crucial problem is addressed in Paper B. In Paper B, an investigation is carried out to assess the viability of capturing multi-directional human activities. This is achieved through the utilization of a multi-view distributed MIMO radar system, and its performance is compared to that of a conventional SISO radar. The central focus of Paper B is the computation of multi-view time-variant (TV) radial velocity distributions and mean radial velocities for multi-directional falling and walking activities. The results presented in Paper B underscore the efficacy of the proposed radar configuration, shedding light on the benefits of employing MIMO radar over its SISO counterpart in capturing complex human movement signatures.

Furthermore, Paper B delves into the practical applications of the multi-view distributed MIMO radar setup by showcasing its capability to function as a RF-based step counter for multi-directional human activities. We not only highlight the limitations of conventional SISO radar systems but also demonstrate the utility and versatility of the proposed MIMO radar configuration in addressing real-world challenges associated with human activity monitoring. In the following sections of this

chapter, we will look into the methodologies, experiments, and findings presented in Paper A and Paper B, providing a comprehensive understanding of the valuable contributions made in these studies.

2.2 Cross-Channel Interference

Cross-channel interference, which is commonly observed in multi-channel systems like MIMO radar systems, can arise from various factors such as electromagnetic coupling and RF energy leakage. It results in the contamination of signals among subchannels, potentially introducing spurious signals and significantly impairing system reliability and performance. We introduced a robust solution meticulously designed to alleviate cross-channel interference challenges originating from RF leakage within the context of a commercial-grade and cost-effective distributed MIMO radar system such as Ancortek SDR-KIT 2400T2R4. The ramifications of cross-channel interference extend beyond mere signal degradation. These ramifications may include the introduction of spurious or false targets, a reduction in the accuracy of target localization, and the system's compromised ability to detect weaker signals.

2.2.1 Problem Description

In the context of our research, we are confronted with the challenge of capturing the motion of scatterers from multiple directions, a task that necessitates the use of multiple pair of transmitter-receiver antennas. To address this, we configured the Ancortek SDR-KIT 2400T2R4 in a 2×2 MIMO radar setup, with a focus on a single moving scatterer, denoted as S^M, as depicted in Figure 2.1.

Within this setup, two transmitter antennas operate in different time slots, following a time division multiple access (TDMA) scheme. This scheduling strategy is employed to prevent interference between the subchannels or time-variant channel impulse responses (TV-CIRs) $h_{11}(\tau', t)$, $h_{12}(\tau', t)$, $h_{21}(\tau', t)$, and $h_{22}(\tau', t)$, as illustrated in Figure 2.1. However, even with the implementation of the TDMA scheme, the subchannels continue to encounter cross-channel interference due to the suboptimal RF isolation characteristic of the cost-effective commercial-grade MIMO
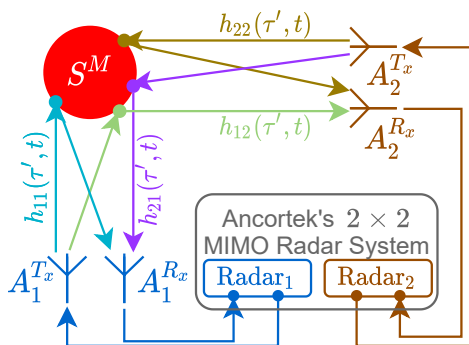
Figure 2.1: A 2×2 MIMO radar setup capturing the 3D movement of a non-stationary scatterer S^M (adapted from Paper A).

radar system. For comprehensive expressions and further details regarding these TV-CIRs, please consult Section A.2.

To elucidate the issue of cross-channel interference in the Ancortek radar, we conducted a pendulum experiment. In this experiment, a pendulum swung at a right angle to the boresight direction of the $A_2^{T_x}$–$A_2^{R_x}$ wireless link (see Figure 2.1). Figure 2.2 depicts the measured radial velocity profile associated with the wireless link between $A_2^{T_x}$ and $A_2^{R_x}$, corresponding to $h_{22}(\tau', t)$. Notably, this figure vividly demonstrates the presence of cross-channel interference from other subchannels. The consequences of this cross-channel interference become evident when we observe a discrepancy between the measured and analytical mean Doppler shifts, denoted as $B_{ij}^{(1)}(t)$. For an in-depth understanding of the mean Doppler shift $B_{ij}^{(1)}(t)$, please refer to the equation A.27. The mismatch between the measured and analytical mean Doppler shifts is visually illustrated in Figure 2.3.

Traditional interference mitigation techniques, such as TDMA, may exhibit limited effectiveness when confronted with the intricate challenges presented by distributed MIMO radar systems, as thoroughly discussed in Paper A. While frequency division multiple access (FDMA) can be an alternative, it imposes constraints on
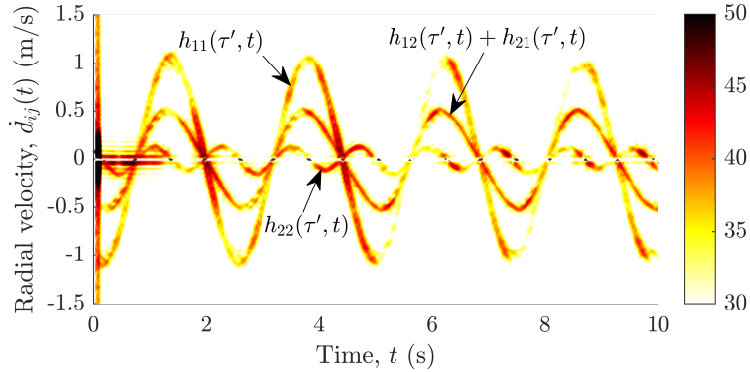
Figure 2.2: Measured radial velocity $\dot{d}_{ij}(t)$ of subchannel $h_{22}(\tau', t)$ with cross-channel interferences from other subchannels (reproduced from Paper A).
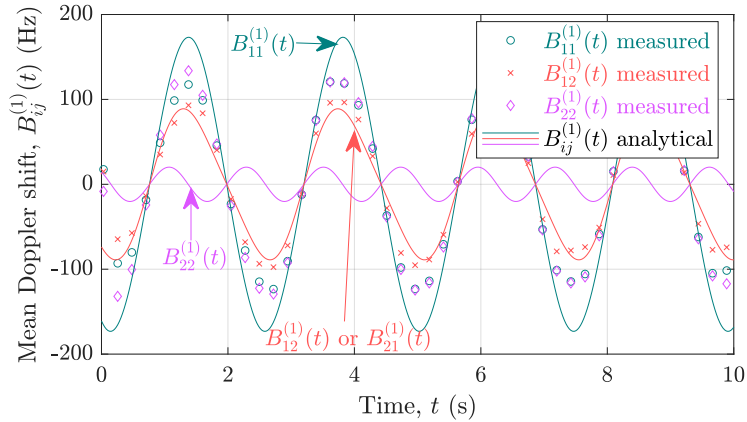
Figure 2.3: Comparison of the measured and analytical mean Doppler shifts $B_{ij}^{(1)}(t)$ in the absence of cross-channel interference mitigation scheme, revealing a notable mismatch (adapted from Paper A).

the instantaneous bandwidth of a frequency-modulated continuous wave (FMCW) radar, ultimately limiting range resolution. These inadequacies necessitate the development of novel solutions that are tailored to the specific interference scenarios encountered in these systems.

2.2.2 Proposed Interference Mitigation Approach

This subsection outlines our strategy to tackle the cross-channel interference, thereby offering a clear path to improved radar performance and signal fidelity. In our pursuit of mitigating the cross-channel interference issue, we introduce a straightforward yet effective solution. Within the context of the 2×2 MIMO radar system illustrated in Figure 2.1, we propose the utilization of either an RF delay line component or longer RF cables for one of the two transmitter-receiver antenna pairs, as illustrated in Figure A.4.

By employing this approach, the subchannel associated with the longer RF cables encounters a propagation delay commensurate to the cable length. In this manner, we achieve a thorough separation of all subchannels within the delay (range) domain. For additional details and to visually see the separation of subchannels within the delay (range) domain, please refer to Figure A.5(a) and A.5(b). Subsequently, we apply range gating to distinguish distinct signals from these individual TV-CIRs $h_{11}(\tau', t)$, $h_{12}(\tau', t)$, $h_{21}(\tau', t)$, and $h_{22}(\tau', t)$. The effectiveness of our approach becomes readily apparent when comparing the measured trajectories of the pendulum with their analytical counterparts. For comprehensive experimental results, please consult Section A.6. Figure 2.4 provides a visual representation of the close alignment between the measured and analytical mean Doppler shifts $B_{ij}^{(1)}(t)$.
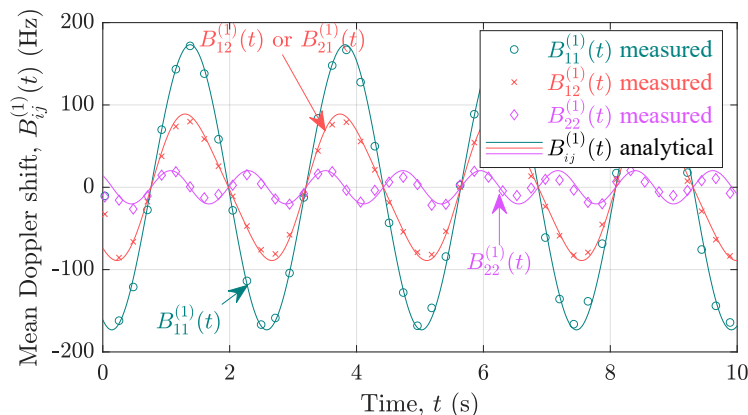
Figure 2.4: Comparison of the measured and analytical mean Doppler shifts $B_{ij}^{(1)}(t)$ following the implementation of the proposed cross-channel interference mitigation scheme, demonstrating a strong alignment (adapted from Paper A).

2.3 Multi-Directional Human Activities

In this section, we delve into the research findings from Paper B, which focus on addressing constraints observed in both SISO and monostatic MIMO radar systems. Our primary objective in Paper B was to introduce an innovative distributed MIMO radar configuration designed to tackle the complexities associated with multi-directional human motion. Furthermore, harnessing this distributed MIMO setup, we devised a versatile RF-based direction-independent step counting system capable of accurately tallying steps for multi-directional walking activities. First, let's examine the limitations intrinsic to SISO and monostatic MIMO radar systems, which hinder their ability to capture multi-directional human activities, such as falling and walking.

2.3.1 Limitations of SISO Radars

A rarely addressed yet critical issue with modern SISO or monostatic MIMO radar-based HAR systems is their inability to perform effectively when a human moves at a right angle to the radar's boresight axis. Various studies have reported classification accuracy well above 90% for radar-based HAR systems [75, 11, 10]. However, these studies typically assume human movement to be restricted to a single direction. In these studies, human subjects either move toward or away from the radar, resulting in significant changes in the radial information received by the radar. To illustrate this, consider the TV radial velocity distribution of a SISO radar shown in Figure 2.5(a). Such unidirectional movements, roughly parallel to a radar's boresight (angled at 0°), are easily detected and captured by the radar (see Scenario 1 in Figure 2.6).

Notably, in the context of the SISO radar, only $Radar_1$ is available in Figure 2.6; there is no $Radar_2$. The dashed lines in Figure 2.5(a), 2.5(b), and 2.5(c) represent the mean radial velocities of $Radar_1$ (calculated using equation B.15), derived from the TV radial velocity distributions corresponding to the three distinct walking scenarios depicted in Figure 2.6.
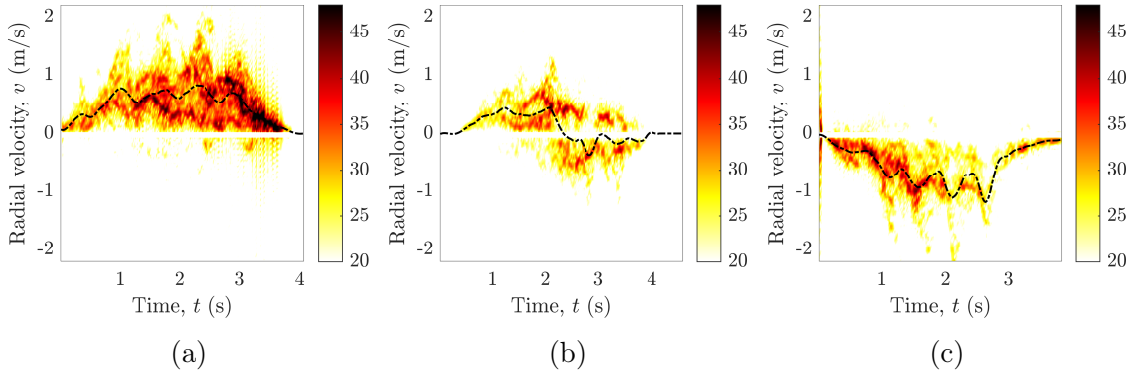
Figure 2.5: For $Radar_1$, the TV radial velocity distribution $p_{11}(v, t)$ and mean radial velocity $\bar{v}_{11}(t)$ of a human walking activity carried out at an angle of (a) 0°, (b) 90°, and (c) 45° with respect to the boresight axis of $Radar_1$ (reproduced from Paper B).
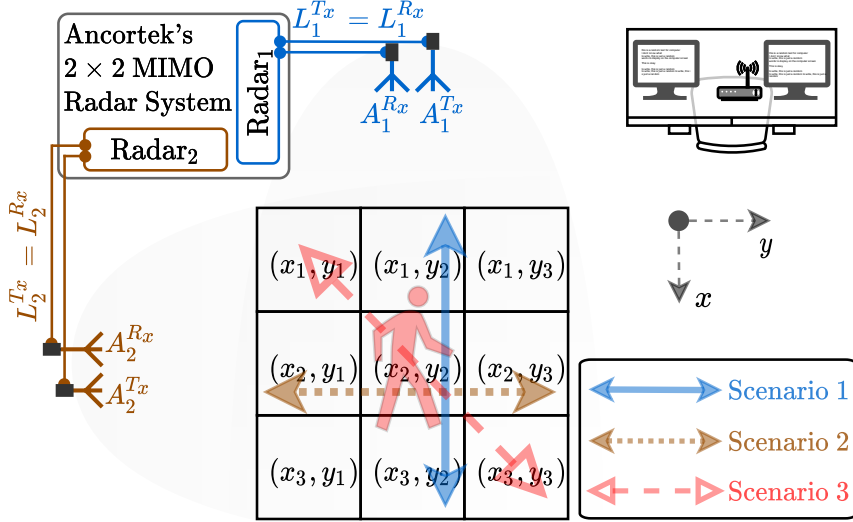
Figure 2.6: The proposed multi-perspective illumination strategy, where the radar subsystems (Radar$_1$ and Radar$_2$) of the 2×2 MIMO radar system capture multi-directional activities (adapted from Paper B).

A crucial limitation arises when the radial information captured by the radar does not change considerably. This occurs when a human moves roughly at a right angle or at a 90° angle to the radar's boresight axis, as exemplified in Scenario 2 in Figure 2.6. Such motion often leads to a low, insufficient, or undetectable micro-Doppler signature at the radar's signal preprocessor output, resulting in the suboptimal TV radial velocity distribution depicted in Figure 2.5(b). Additionally, consider the TV radial velocity distribution of Radar$_1$ in a scenario where a human moves diagonally (45°) to the boresight axis of Radar$_1$ (Scenario 3 in Figure 2.6). In this case, the SISO radar (Radar$_1$) captures an adequate walking pattern, as shown in Figure 2.5(c).

As a consequence of the aforementioned limitations of SISO radar systems, a step counter implemented using SISO radar (Radar$_1$) exhibits suboptimal performance when a walking activity is carried out approximately at a right angle or at a 90° angle to the SISO radar's boresight axis, as demonstrated in Figure 2.7(b). It's worth noting that the walking activities considered here consisted of four steps. For the other two scenarios where the walking activity is not perpendicular to the SISO radar's boresight, the step counter accurately tallies the number of steps, as illustrated in Figure 2.7(a) and 2.7(c). In the following subsection, we present an overview of our efforts to address the limitations associated with monostatic radar systems and introduce an adaptable solution that holds promise for effectively recognizing multi-directional human activities.

2.3.2 Proposed Solution

To overcome the challenges posed by perpendicular motion, we proposed a solution in Paper B that utilizes two collocated transmitter-receiver antenna pairs corresponding to Radar$_1$ and Radar$_2$, strategically positioned in an indoor environment.
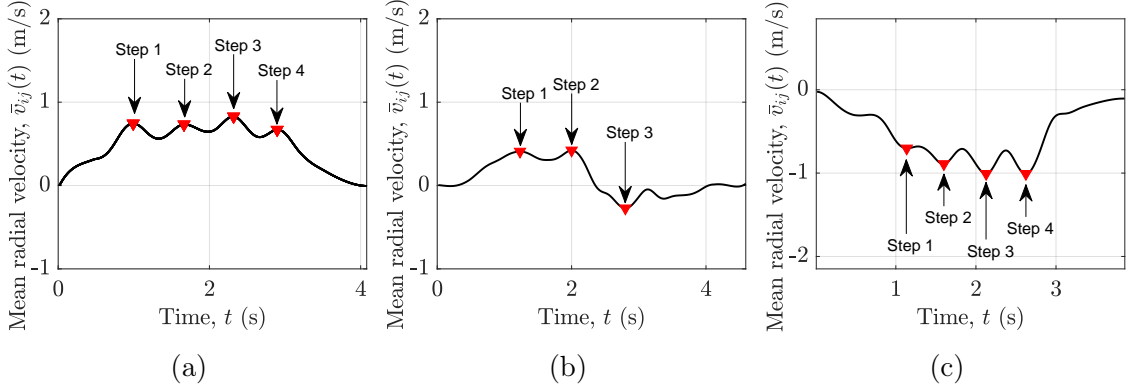
Figure 2.7: The detected number of steps based on the mean radial velocity $\bar{v}_{11}(t)$ of a human walking activity at different angles relative to the boresight axis of Radar$_1$: (a) 0°, (b) 90°, and (c) 45°. Note that the number of steps is not accurately detected by Radar$_1$ for the activity at a 90° angle (reproduced from Paper B).

The primary objective of this setup is to achieve multi-perspective illumination, as visually depicted in Figure 2.6. This approach effectively surmounts the limitations of monostatic radar systems, enabling comprehensive capture of multi-directional human activity signatures.

For Radar$_2$, Figure 2.8 displays the TV radial velocity distribution $p_{22}(v, t)$ and mean radial velocity $\bar{v}_{22}(t)$ of a human walking activity carried out at various angles with respect to the boresight axis of Radar$_1$: (a) 0°, (b) 90°, and (c) 45°. Subsequently, in Figure 2.9, we detect the number of steps by using the mean radial velocity $\bar{v}_{22}(t)$ of a human walking activity in three scenarios. Clearly, Radar$_1$ and Radar$_2$ complement each other in capturing multi-directional motion. In Scenario 1 and Scenario 2, Radar$_1$ and Radar$_2$ adequately capture the movement signatures, respectively.

Expanding upon this solution, we conducted further research to develop an RF-based step counting application that accurately tallies steps regardless of motion direction. Our proposed MIMO radar-based step counter leverages the dynamic time
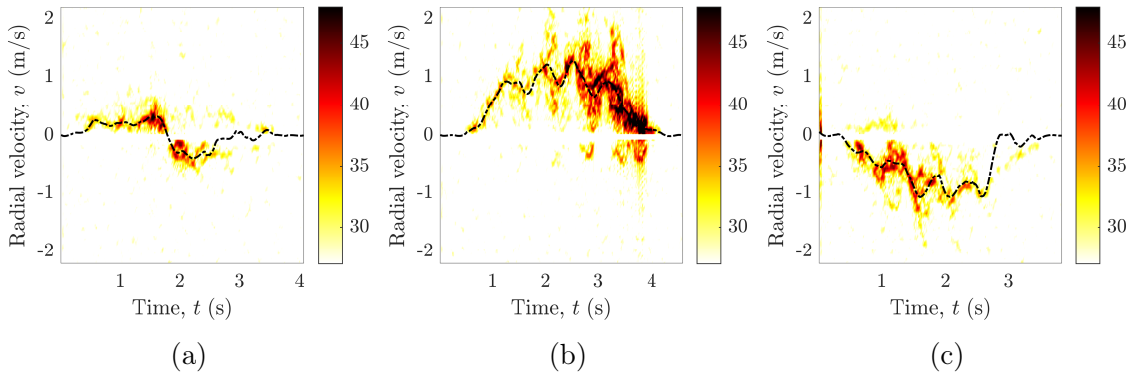
Figure 2.8: For Radar$_2$, the TV radial velocity distribution $p_{22}(v, t)$ and mean radial velocity $\bar{v}_{22}(t)$ of a human walking activity carried out at an angle of (a) 0°, (b) 90°, and (c) 45° with respect to the boresight axis of Radar$_1$ (reproduced from Paper B).
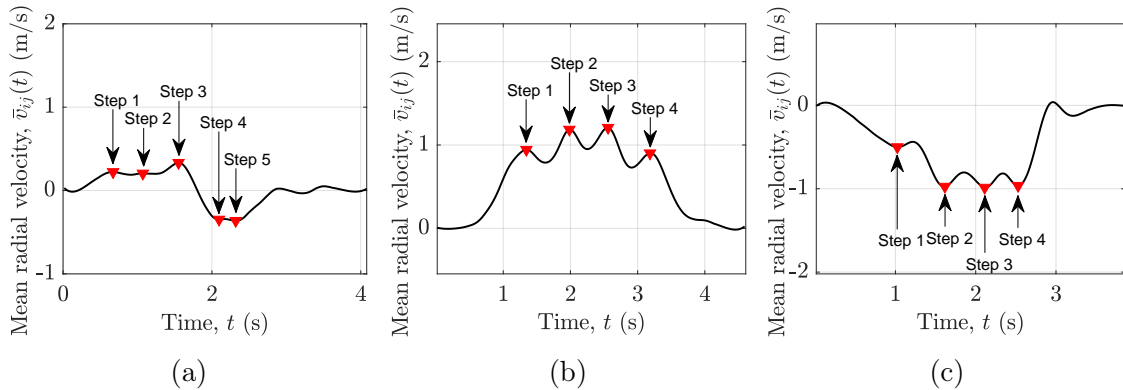
Figure 2.9: The detected number of steps based on the mean radial velocity $\bar{v}_{22}(t)$ of a human walking activity at different angles relative to the boresight axis of Radar_1: (a) 0°, (b) 90°, and (c) 45°. Note that the number of steps is not accurately detected by Radar_2 for the activity at a 0° angle or in Scenario 1 (reproduced from Paper B).

warping (DTW) distance metric to assess the similarity between a reference walking activity signature and those captured by the two subsystems of our distributed MIMO radar system, Radar_1 and Radar_2. Table 2.1 presents the DTW distances for three scenarios: (a) walking at 0° to Radar_1 boresight (Scenario 1), (b) walking at 90° to Radar_1 boresight or 0° to Radar_2 boresight (Scenario 2), and (c) walking at 45° to Radar_1 boresight (Scenario 3). For each radar, it is noteworthy that the walking activities with lower angles (Scenario 1 for Radar_1 and Scenario 2 for Radar_2) exhibited lower DTW distances (see Table 2.1), indicating better signature capture. Therefore, by selecting the radar subsystem with the lowest DTW distance, either Radar_1 or Radar_2, we ensure accurate step counting, regardless of motion direction.

Table 2.1: The DTW distances computed for three scenarios. The SISO radar system is assumed to be Radar_1. The MIMO radar system adopts the minimum DTW distance from Radar_1 and Radar_2 (adapted from Paper B).

Scenario #	Dist. of Radar_1	Dist. of Radar_2	Dist. of SISO Radar	Dist. of 2×2 MIMO Radar
1	19.6	81.2	19.6	19.6
2	73.8	12.0	73.8	12.0
3	7.3	9.1	7.3	7.3

2.4   Conclusion

In this chapter, along with the insights from Paper A and Paper B, we have addressed the inherent limitations in SISO and monostatic MIMO radar systems, particularly in the context of motion direction. Our solution revolves around a distributed MIMO radar configuration designed to comprehensively capture the micro-Doppler signatures of multi-directional human activities such as falling and walking. Building

upon this multistatic MIMO radar setup and leveraging the DTW distance metric, we have successfully developed a practical direction-independent step counting system capable of accurately tallying the number of steps in multi-directional walking activities.

Furthermore, the versatility of the proposed multistatic MIMO radar configuration opens the door to various radar-based applications. In our research, we have harnessed this configuration to realize a fully functional direction-independent HAR system, as discussed in more detail in the upcoming chapter of this dissertation.

Chapter 3

Measurement-Based HAR Systems Using MIMO Radars

3.1 Introduction

In the recent years, we have witnessed significant advancements in the realm of human activity recognition (HAR) [16, 46, 47, 48]. The constant interest in HAR is attributable to its applicability in diverse domains, including human–computer interaction [76] and health [52]. Numerous sensors have been employed for HAR, including radio frequency (RF) sensors. In contrast to alternative sensing modalities like vision and wearables, RF sensors present distinctive benefits, notably insensitivity to ambient lighting conditions and non-intrusiveness. Within RF sensing technologies, millimeter wave (mm-wave) radar sensors have lately emerged to be the preferred choice among many researchers.

For unidirectional human motion, particularly when the motion is aligned with the radar boresight axis, contemporary HAR systems developed with monostatic radars exhibit good performance [70]. Unfortunately, this is not true for the classification of multi-directional human activities. In this chapter, we elucidate the research findings of Paper C and primarily address the research question 3 that focuses on the realization of a measurement-based direction-independent human activity recognition (DIHAR) system. The measurement-based DIHAR system, employing a mm-wave distributed multiple-input multiple-output (MIMO) radar system (Ancortek SDR-KIT 2400T2R4), delves into the classification of five distinct kinds of multi-directional human activities.

Figure 3.1 depicts the overview of our proposed measurement-based multi-view 2×2 DIHAR system that incorporates two radar subsystems capturing multi-directional motion from different aspects. The radar signal processing module in Figure 3.1 computes the radar signatures for the multi-perspective radar data. Subsequently, the multi-view or multi-perspective radar signatures are used by two feature extraction networks (FENs), which compute the relevant features automatically. After combining or merging the computed features, our DIHAR system effectively classify a human activity regardless of its direction.
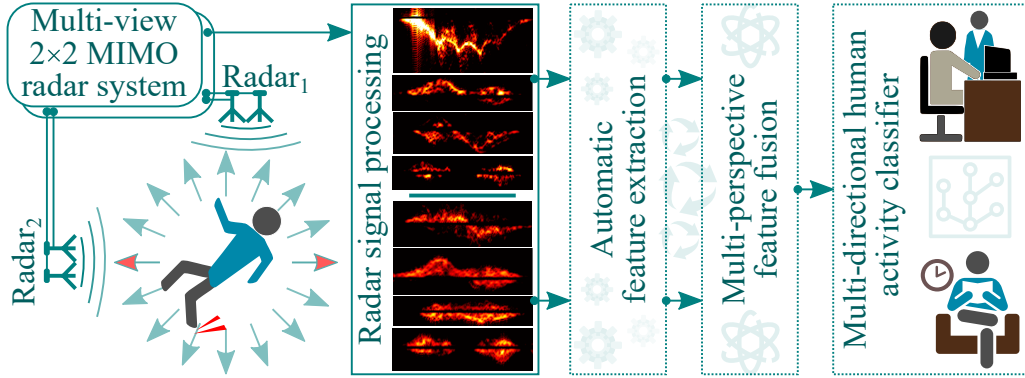
Figure 3.1: Our multi-perspective illumination strategy, where two radar subsystems (Radar$_1$ and Radar$_2$) of the multi-view MIMO radar system complement each other, thereby allowing for the multi-directional human activity classification (adapted from Paper C).

3.2 Direction-Independent Human Activity Recognition

In light of the constraints imposed by contemporary monostatic radar system, as detailed in the preceding chapter, current radar-based HAR systems encounter difficulties when confronted with the classification of multi-directional human activities. Traditional monostatic radar-based HAR systems are inherently insensitive to the direction of human movement, thus proving inadequate in categorizing activities that involve diverse motion patterns. This inherent limitation becomes evident when a human subject moves approximately perpendicularly, at a 90-degree angle, to the radar's boresight axis, leading to negligible changes in the radial information recorded by the radar.

In [77] Erol et al. presented findings related to the classification accuracy of human falling activities from multiple viewpoints. For instances where falls were aligned with the radar's primary boresight axis (0-degree aspect angle), they reported a 96% classification accuracy. However, at a 60-degree aspect angle, this accuracy declined to 85%, and when the falls occurred at a 90-degree aspect angle, the accuracy decreased drastically to 45%, thereby making the HAR system pointless. Furthermore, in a study by Ding et al. [78], involving multiple human activities, a decline in accuracy was observed, decreasing from around 96% to 87% as the viewing angle of the radar shifted from 15 degrees to 30 degrees. Additional research by Liu et al. [79, 80] demonstrated the detection of multi-directional human falls by deploying a radar on the ceiling, albeit with limited applicability in classifying more complex human activities.

To address DIHAR, there is an inclination towards the utilization of monostatic beamforming MIMO radar systems capable of angle measurement [81, 82]. However, as a consequence of hardware constraints, real radar systems tend to exhibit limited angular resolution. Therefore, for applications like gesture recognition, where angular accuracy is less critical at shorter ranges, researchers such as Molchanov et

al. [83] have rightly used the bearing information offered by a single-input multiple-output (SIMO) monopulse radar system. It's unfortunate that due to the limited angular accuracy, it is difficult to extend this method for developing DIHAR systems. Recent advances in HAR involve the use of three-dimensional (3D) point cloud data [84, 75] provided by monostatic frequency-modulated continuous wave (FMCW) MIMO radar systems, but such data also exhibit inadequate angular accuracy and cross-range resolution. A viable alternative is to develop a HAR system based on a radar in bistatic configuration [56]. Nevertheless, multi-view MIMO radar systems are the superior option for DIHAR. These systems provide the most comprehensive multi-perspective signatures of human activities, thereby improving classification performance for DIHAR, as elucidated in this chapter.

To circumvent the constraints imposed by the physical layer of monostatic radars, we present a DIHAR framework that employs a distributed mm-wave MIMO radar system (see Figure 3.1), providing a practical solution for multi-directional human activity classification. The physical layer design of our DIHAR system incorporates two independent FMCW radar subsystems, namely $Radar_1$ and $Radar_2$, each of which transmits the RF signal in distinct time slots and records multi-directional human movements from distinct vantage points. This approach yields two separate time-variant (TV) radial velocity distributions, effectively resolving the physical layer limitations associated with contemporary HAR systems that employ monostatic antenna configurations. Our solution takes the form of the multi-view 2×2 MIMO radar system (see Figure 3.1), facilitating the realization of DIHAR. In Section 3.3, we provide insights into our data collection campaign and elucidate the procedures employed to process the data acquired from the 2×2 MIMO radar system.

3.3 Data Collection and Processing

By using the multi-view 2×2 MIMO radar system, we recorded multiple types of multi-directional human activities from six individuals, which are: falling on a mattress, picking up an object from the floor, walking, sitting down on a chair, and standing up. The utilized 2×2 MIMO radar configuration in Figure 3.1 resembles the setup discussed in the previous chapter (see Figure 2.6). To see comprehensive details regarding the recording of the aforementioned activities from various angles, please refer to Section C.4, which explains three distinct movement scenarios that closely mirror what was presented in the preceding chapter. In Scenario 1 and 2, human motion aligns parallel to the boresight of $Radar_1$ and $Radar_2$, respectively. Scenario 3, on the other hand, involves human motion diagonal to the boresight of $Radar_1$ and $Radar_2$.

The dataset acquired by our multi-view MIMO radar system is referred to as $HAR^{(2D)}$, with the superscript "(2D)" signifying motion within the two-dimensional (2D) horizontal plane (see Figure C.1). Following a similar convention, specifically for the DIHAR task, we define $FEN^{(2D)}$ as a feature extraction network, $SISO^{(2D)}$ as

a single-input single-output (SISO) radar-based DIHAR classifier, and MIMO[(2D)] as a MIMO radar-based DIHAR classifier. Note that we employ the HAR[(2D)] dataset to develop our SISO and MIMO radar-based DIHAR systems, namely, SISO[(2D)] and MIMO[(2D)]. In Table C.2, we specify that the HAR[(2D)] dataset contains 1364 human activities from five individuals. We computed two TV radial velocity distributions denoted as $p_{11}(v,t)$ and $p_{22}(v,t)$ for all the recorded activities. These two distributions correspond to the respective radar subsystems Radar$_1$ and Radar$_2$.

We employed the active segment detection (ASD) technique to compute the variance of the in-phase component of a radar signal to determine the starting and ending of an activity. Note that the ASD technique was proposed by the authors of [85] in their work on smartphone-based gait recognition. We used this technique to obtain active segments from the radar's raw in-phase and quadrature (IQ) data of the human activities. An active segment corresponds to the actual duration of the activity in question, in which a human actively perform an activity. For the active segment of the raw data, the beat frequency function $S_{b,ik}(f_b,t)$ is obtained using the relation (C.5). Subsequently, we compute the square of the short-time Fourier transform (STFT) of the beat frequency function $S_{b,ik}(f_b,t)$ to obtain the TV micro-Doppler signature $S_{ik}(f,t)$ using (C.6). Let v, t, f_0, and c_0 represent the velocity, time, carrier frequency, and speed of light, respectively. We can then derive the TV radial velocity distribution, denoted as $p_{ik}(v,t)$, corresponding to the wireless link between the transmit antenna $A_i^{T_x}$ and the receive antenna $A_k^{R_x}$ using [68]

$$p_{ik}(v,t) = \frac{S_{ik}\left(\frac{2f_0}{c_0}v,t\right)}{\int_{-\infty}^{\infty} S_{ik}\left(\frac{2f_0}{c_0}v,t\right)dv} \tag{3.1}$$

where $i,k \in \{1,2\}$.

Figures 3.2 and 3.3 illustrate the radar signatures (TV radial velocity distribution $p_{ik}(v,t)$) corresponding to five distinct types of multi-directional human activities. These figures show the radar signatures associated with three distinct directions of human motion, labeled as Scenarios 1, 2, and 3, as defined in Figure C.1. Upon a closer examination of these figures, it becomes evident that employing a single monostatic radar system, which illuminates a subject from a single aspect angle, exhibits limitations. Using only a single monostatic radar, the acquired radar signatures are suboptimal, particularly when the direction of motion is perpendicular to the radar's boresight, as demonstrated in Figures 3.2 and 3.3. These limitations of monostatic radars were comprehensively addressed in the previous chapter. In the following section, we will primarily look into the classification performance of both monostatic SISO and distributed MIMO radar systems in the context of DIHAR.

3.4   Design Approaches for the DIHAR Systems

To show the applicability and efficacy of our proposed multi-view MIMO radar-based methodology, we have devised three distinct HAR systems, with two of them being DIHAR systems. In Paper C, we initially devise a SISO[(1D)] HAR system based
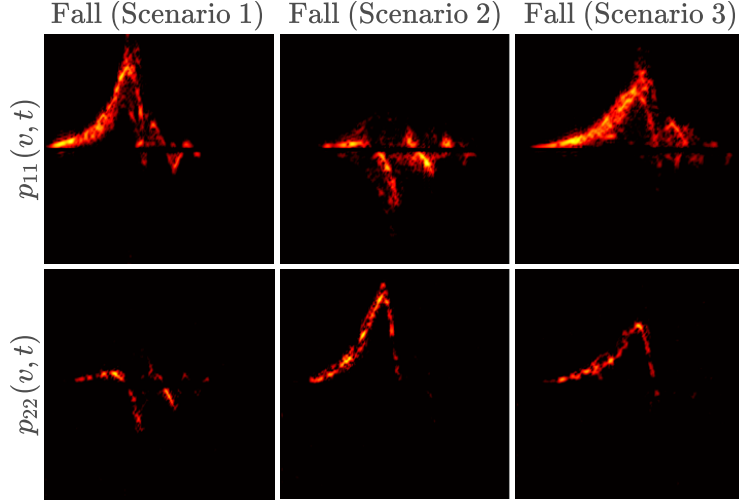
Figure 3.2: The heatmap of the TV radial velocity distributions $p_{ii}(v,t)$ (TV micro-Doppler signatures) in three distinct movement scenarios for the falling activity. The x- and y-axis of each distribution is the radial velocity v and time t, respectively. In Scenarios 1, 2, and 3, the motion is respectively parallel, perpendicular, and diagonal to the boresight of Radar$_1$ (reproduced from Paper C).
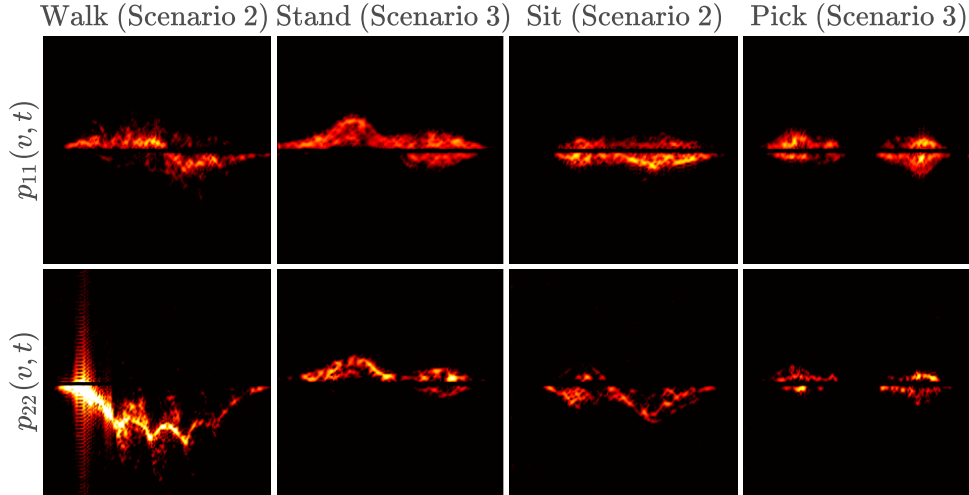
Figure 3.3: The heatmap of the TV radial velocity distributions $p_{ii}(v,t)$ (TV micro-Doppler signatures) in three distinct movement scenarios for multiple activities. In Scenarios 1, 2, and 3, the motion is respectively parallel, perpendicular, and diagonal to the boresight of Radar$_1$ (reproduced from Paper C).

on a monostatic SISO radar configuration. SISO$^{(1D)}$ HAR system is a conventional HAR system that only considers simple unidirectional human activities. These unidirectional activities are assumed to occur along the boresight axis of the SISO radar system. As a result, SISO$^{(1D)}$ focused on unidirectional human activities during both training and testing phases. This unidirectional HAR system demonstrated commendable performance, achieving a classification accuracy of approximately 97%. However, it's worth noting that this system exclusively dealt with unidirectional human activities, which have relatively uncomplicated radar signatures, leading to

a relatively straightforward classification task. For the sake of brevity, we omit the results of the SISO$^{(1D)}$ unidirectional HAR system in this chapter; please consult Paper C for a comprehensive presentation of those results.

As a second step, we aimed to underline the detrimental impact of introducing diverse movement directions on the classification performance of a HAR system. For this purpose, we designed a SISO radar-based DIHAR system known as SISO$^{(2D)}$. In contrast to SISO$^{(1D)}$, the SISO$^{(2D)}$ DIHAR system is tailored to classify multi-directional human activities. For multi-directional activities, it utilizes the HAR$^{(2D)}$ dataset for its training and testing. Note that in this chapter, we present exclusively the outcomes pertaining to the SISO$^{(2D)}$ DIHAR classifier, which was trained and tested using data from Radar$_1$. SISO$^{(2D)}$ DIHAR classifier realized using the data from Radar$_2$ are not discussed for the sake of brevity.

In our final design, we sought a substantial improvement in the classification performance of the SISO$^{(2D)}$ HAR system. To achieve this, we introduced MIMO$^{(2D)}$ system, which is a DIHAR system that employs the proposed 2×2 distributed MIMO radar antenna configuration (see Figure 3.1). The proposed MIMO$^{(2D)}$ DIHAR system, like the SISO$^{(2D)}$ DIHAR system, is designed to classify multi-directional human activities. Therefore, the MIMO$^{(2D)}$ DIHAR system also utilizes the HAR$^{(2D)}$ dataset during its training and testing phases. For both SISO$^{(2D)}$ and MIMO$^{(2D)}$ DIHAR classifiers, the recorded radar signatures or the TV radial velocity distributions were transformed into the images, as exemplified in Figures 3.2 and 3.3. These images were subsequently employed as input feature maps to the FEN$^{(2D)}$ (see Figure 3.4).

In the training and validation phases of the deep convolutional neural network (DCNN)-based SISO$^{(2D)}$ and MIMO$^{(2D)}$ DIHAR classifiers, we used data from two participants extracted from the HAR$^{(2D)}$ dataset. A significant portion, approximately eighty percent of this data, was dedicated to training the DIHAR classifiers, with the remaining twenty percent allocated for validation purposes. The data associated with the other four participants within the HAR$^{(2D)}$ dataset was exclusively reserved for testing the DCNN-based SISO$^{(2D)}$ and MIMO$^{(2D)}$ DIHAR classifiers. Now let's discuss the design and results of the SISO$^{(2D)}$ and MIMO$^{(2D)}$ DIHAR classifiers, individually.

The FEN$^{(2D)}$ consists of four convolutional layers, with each layer housing an array of filters. Each filter within these layers is characterized by a 5×5 kernel dimension, denoted as k_d. During the training of the SISO$^{(2D)}$ DIHAR classifier, the adaptive moment estimation (Adam) optimizer [86] is employed to minimize the empirical risk $R_J(\mathcal{C}_f)$ (see (C.8)). To mitigate the risk of overfitting, the SISO$^{(2D)}$ DIHAR classifier incorporates L2 regularization and utilizes dropout layers. It's noteworthy to mention that the training phase of the SISO$^{(2D)}$ classifier, depicted in Fig. C.10, shows no indications of overfitting.

The trained SISO$^{(2D)}$ DIHAR classifier underwent evaluation by using previously unseen data from the HAR$^{(2D)}$ dataset to determine its classification accuracy. The system's performance is depicted in the confusion matrix in Figure 3.5. It is evident that the SISO$^{(2D)}$ DIHAR system, with an overall accuracy of only 88.98%, falls

short as a suboptimal DIHAR system that misclassified numerous multi-directional human activities. Notably, multi-directional falling activities exhibited perfect precision and recall, whereas other activities with intricate radial velocity distributions (see Figures 3.2 and 3.3) encountered challenges, particularly the "Pick" activity that was misclassified 21 times. The decline in SISO$^{(2D)}$ DIHAR classifier's performance is attributed to the physical-layer constraints of monostatic radars. In the following discussion, we will see that these physical-layer constraints are addressed by employing the proposed multi-view MIMO$^{(2D)}$ DIHAR system.
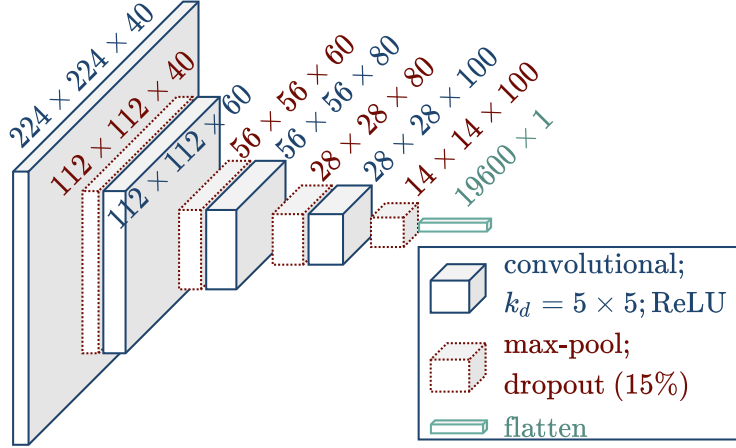
Figure 3.4: The architecture of the feature extraction network FEN$^{(2D)}$, employing four convolutional layers, designed for both SISO$^{(2D)}$ and MIMO$^{(2D)}$ DIHAR classifiers (reproduced from Paper C).

Figure 3.5: Confusion matrix quantitatively representing the summary of the results achieved by the SISO$^{(2D)}$ DIHAR classifier. The overall obtained accuracy is 88.98% for the five multi-directional activities (reproduced from Paper C).

While the neural network architectures of MIMO$^{(2D)}$ and SISO$^{(2D)}$ HAR systems differ (see Figures 3.6 and C.5), their basic components, hyperparameters, and training methods closely resemble each other. Both DIHAR classifiers share similar specifications and employ similar training methodologies, e.g., dropout layers, activation function, kernel dimension k_d, and Adam optimizer. For more details, see Section C.5.3. Furthermore, both DIHAR classifiers employ the same FEN$^{(2D)}$. Notably, MIMO$^{(2D)}$ DIHAR classifier employs two FEN$^{(2D)}$ modules, independently extracting features from the radar signatures $p_{ik}(v,t)$ (see Figure 3.6). In other words, in our multi-view MIMO$^{(2D)}$ DIHAR system, the TV radial velocity distributions $p_{11}(v,t)$ and $p_{22}(v,t)$ are obtained simultaneously, serving as inputs to FEN$^{(2D)}$. Subsequently, these features are merged before the multilayer perceptron (MLP) and softmax layers for the activity classification (see Figure 3.6).

The trained MIMO$^{(2D)}$ DIHAR classifier went through evaluation using unseen data from the HAR$^{(2D)}$ dataset to determine its classification accuracy. The system's performance is depicted in the confusion matrix in Figure 3.7. Evidently, the MIMO$^{(2D)}$ DIHAR classifier demonstrates remarkable classification accuracy of 98.52%, especially compared to the SISO$^{(2D)}$ DIHAR system. Our MIMO$^{(2D)}$ DIHAR classifier encountered merely seven misclassifications, as depicted by the confusion matrix in Figure 3.7. The enhanced classification accuracy of our MIMO$^{(2D)}$ DIHAR classifier stems from its multi-perspective illumination strategy.

3.5   Conclusion

In this chapter, we addressed a significant physical layer issue in contemporary radar-based HAR systems, specifically within the context of monostatic SISO and monostatic MIMO radar configurations. The development and analysis of three HAR systems, namely SISO$^{(1D)}$, SISO$^{(2D)}$, and MIMO$^{(2D)}$, were crucial in addressing these challenges. To expand the scope of HAR in RF sensing, we delved into the
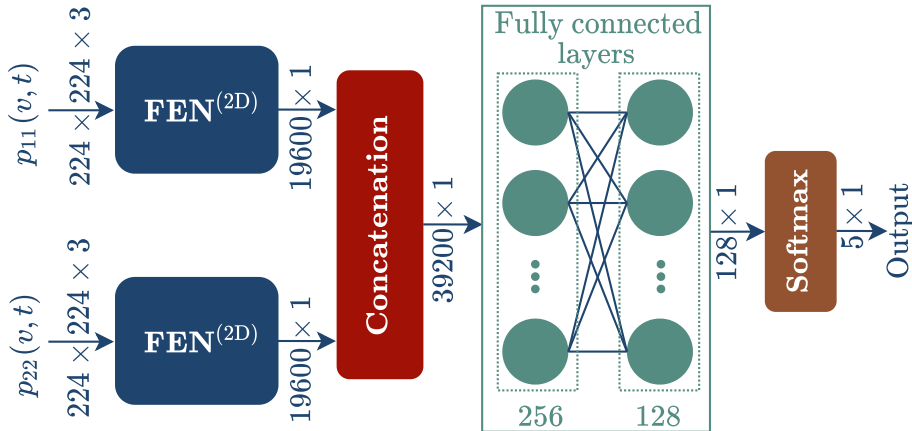
Figure 3.6: The DCNN architecture of the proposed MIMO$^{(2D)}$ DIHAR classifier. It merges two sets of distinct features that are computed by the two FEN$^{(2D)}$ modules (reproduced from Paper C).

	Fall	Walk	Stand	Sit	Pick	
Fall	36	0	0	0	0	100% 0%
Walk	0	125	0	0	0	100% 0%
Stand	0	1	107	0	0	99.07% 0.93%
Sit	0	0	0	97	0	100% 0%
Pick	0	0	5	1	100	94.34% 5.66%
	100% 0%	99.21% 0.79%	95.54% 4.46%	98.98% 1.02%	100% 0%	98.52% 1.48%
	Fall	Walk	Stand	Sit	Pick	

True Class / Predicted Class

Figure 3.7: Confusion matrix quantitatively representing the summary of the results achieved by our MIMO$^{(2D)}$ DIHAR classifier. The overall obtained accuracy is 98.52% for the five multi-directional activities (reproduced from Paper C).

complexity of classifying multi-directional activities utilizing the HAR$^{(2D)}$ dataset captured by a multi-view MIMO radar system.

While the SISO$^{(1D)}$ HAR classifier achieved a commendable 97.28% classification accuracy, its limitation in handling multi-directional activities prompted a deeper exploration. The examination of SISO$^{(2D)}$ DIHAR classifier, considering activities in the 2D xy-plane, revealed a significant accuracy drop, emphasizing the inherent constraints of monostatic radars. These limitations were mitigated by our proposed MIMO$^{(2D)}$ DIHAR classifier that employed a multi-view MIMO radar system, which illuminates the environment from multiple perspectives and combines the features computed by the convolutional layers of the DCNN. The MIMO$^{(2D)}$ DIHAR classifier demonstrated substantial superiority over the monostatic SISO$^{(2D)}$ by improving the classification accuracy from 88.98% to 98.52%.

In undertaking this investigation and implementing the radar-based DIHAR system, a comprehensive measurement campaign was undertaken, necessitating substantial allocation of human resources, time, and effort. Consider the potential ramifications of altering a single radar parameter—an adjustment that could potentially undermine the entire data acquisition endeavor. Such challenges are inherent in measurement-based methodologies. Consequently, a compelling need emerges for a refined simulation-based alternative capable of supplanting measurement-based approaches while assuring the preservation of radar signature fidelity. In the following chapters, we introduce advanced motion capture (MoCap)-data-driven simulation-based HAR systems in the context of RF sensing.

Chapter 4

A Simulation-Based Framework for SISO Radar-Based HAR Systems

4.1 Introduction

Due to the recent advent of artificial intelligence and machine learning techniques, the idea of generating synthetic data has never been more pertinent. The major obstacle to the development of systems based on modern artificial intelligence and machine learning methods is the acquisition of large, clean, and labeled datasets [87]. Usually, the amount of time, effort, and resources required to obtain a comprehensive dataset impede the development of such systems. Often, privacy and legal issues, especially in the areas of healthcare [88], finance [89], and social sciences [90], hinder the collection of real data. This results in an impasse that can only be overcome by methods for generating field-specific synthetic data. Therefore, the use of synthetic data is increasingly preferred in various research areas, such as medical imaging [91]. Here, our main focus is on Radio frequency (RF) sensors, chosen for their privacy protection, resilience to lighting conditions, and non-intrusiveness. In this chapter, we introduce an end-to-end simulation-centric framework for human activity recognition (HAR) using millimeter wave (mm-wave) radar systems. While the framework is generally applicable, our main focus is on a radar-based classifier for HAR. This research aims to devise a HAR system within this simulation-based framework and discusses the contributions of Paper D, primarily addressing research questions 4–7.

Collecting radar signatures from real individuals is a time-consuming and resource-intensive task, often resulting in datasets with limited scope. Typically, measurement-based radar datasets are applicable only to specific scenarios and specific system parameters. To overcome this limitation and facilitate the creation of diverse training datasets for radar-based classifiers, we present a simulation-centric HAR system relying solely on simulated radar data for both training and validation phases. In order to prove the real-world efficacy of our simulation-centric HAR system, we use previously unseen experimental data from a physical frequency-modulated continuous wave (FMCW) radar system for the testing phase.

The human motion synthesis tool, developed in Unity software [92], is integral to our framework, emulating various activities and simulating three-dimensional (3D) trajectories for virtual markers on an avatar. These virtual markers capture detailed human movements, eventually processed by a geometric model (see Sect. D.5) in relation to a software-defined antenna configuration. Additionally, scatterer-level modeling of radar signals is presented to enhance the realism of simulated micro-Doppler signatures. Our simulation encompasses the synthesis of in-phase and quadrature (IQ) signals and the generation of radar micro-Doppler signatures for various emulated activities, considering both the stationary and non-stationary aspects of the wireless channel. Note that the randomization feature of human motion, facilitated by our proposed simulation-based approach, ensures unbiased datasets for training radar-based classifiers.

The simulation-centric approach presents numerous benefits compared to experimental approaches. These benefits include the flexibility to replicate radar datasets with user-defined target motion characteristics, the capability to enhance training data, resource minimization, and addressing user privacy concerns. For diverse operational and environmental conditions, our approach offers control over radar parameters such as bandwidth, carrier frequency, pulse repetition frequency (PRF), and more. This capability enables the generation of diverse training datasets tailored to specific conditions, thereby significantly reducing the training effort for radar-based classifiers.

It's essential to note that the proposed end-to-end simulation-centric approach, initially devised for HAR, finds its utility in many domains. The efficacy of our proof-of-concept is demonstrated in Paper D by means of actual experimental data encompassing five human activities. Nevertheless, the main advantage of this proof-of-concept is its ability to convert motion capture (MoCap) data into radar data (see Section 4.4 and 4.5), which renders it applicable to a broad range of real-world situations. The accessibility of online MoCap data repositories, such as Mixamo [69], further boosts the utility of our approach across various domains such as sports, smart homes, healthcare, and more.

4.2 Overview of the Proposed Simulation-Based Approach

Traditionally acquired training datasets suffer from limited reusability due to their dependency on fixed operating parameters and fixed antenna configurations. Proposed in Paper D, our pragmatic solution involves a fully simulation-based approach for developing a practical HAR system. Our simulation-based approach, as detailed in Figure D.1(b), commences with six basic animations, which are then synthesized into five distinct human activities. Utilizing Unity software [92], the activity emulation block simulates the motion of virtual markers on human body segments and generate corresponding time-variant (TV) spatial trajectories (see Section D.4.3). Virtual markers in our simulation framework are simulated point scatterers that

represent actual scatterers on human body segments. These scatterers scatter the transmitted signal back to the radar system's receive antenna. Then, our channel model (see Section 4.4) transforms these trajectories into TV delays while taking into account the positions of radar antennas. The radar data synthesizer then simulates raw IQ data, considering scatterers' weights and positions (see Section 4.5).

The development of our simulation-centric HAR classifier is depicted in Figure 4.1 and its testing is shown in Figure 4.2. It is essential to clarify that the simulated RF signal in Figure 4.1 is generated by our channel model (refer to Section 4.4). On the other hand, the real RF signal in Figure 4.2 is the output of an actual radar. Notably, the structural similarity between the raw IQ data from the real and simulated radar allows us to employ the same radar signal processing block in Figures 4.1 and 4.2. From the simulated IQ data (RF signal), the radar signal processing block simulates the radar signatures for synthesized human activities. It is worth mentioning that the suggested simulation-based framework incorporates scatterer-level signal modeling, ensuring accuracy in radar modeling and comprehensive simulation of radar signatures, including TV radial range distribution, TV radial velocity distribution (micro-Doppler signature), and so forth. This simulation-centric framework allows for the simulation of realistic and varied radar signatures, offering the necessary training data for radar-based classifiers.

Our simulation-centric approach facilitates the generation of extensive training data without the need for real human subjects and radar systems, ensuring feasibility and practicality. The simulated radar signatures corresponding to synthesized human activities are stored in a dedicated radar dataset. Our simulation-based HAR system is trained exclusively on this dataset as shown in Figure 4.1. The recorded radar signatures contribute to a real radar dataset, as depicted in Figure 4.2. To
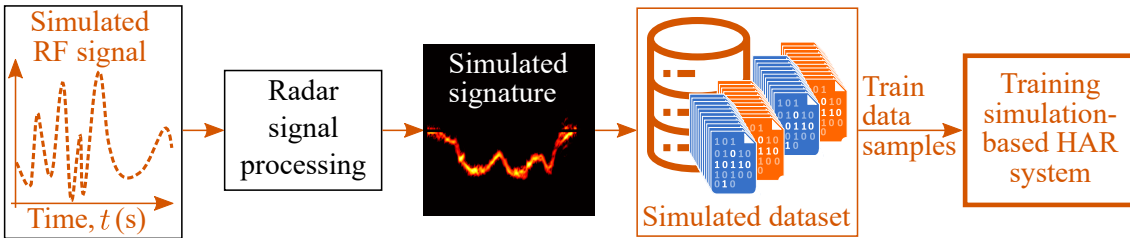
Figure 4.1: Design phase: The development of our HAR classifier relying exclusively on simulated radar signatures.
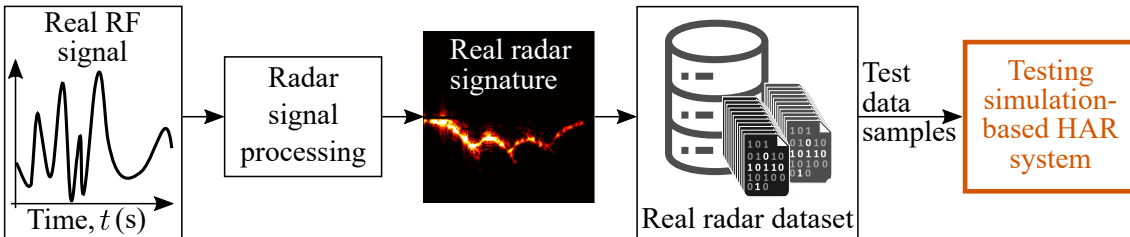
Figure 4.2: Testing phase: The trained HAR classifier evaluated with real radar signatures obtained from real human subjects.

validate the practical significance and applicability of our simulation-based HAR classifier, we assess its performance on real human activities recorded by a mm-wave FMCW radar system (see Figure 4.2). In the following sections, we will examine various approaches for generating MoCap data, elaborate on our simulation-based channel model, discuss the synthesis of RF data, and delve into the implementation of our simulation-centric HAR classifier.

4.3   Approaches for Generating 3D Trajectories

In this section, we delve into various approaches to capture and synthesize human motion, such as biomechanical modeling, wearable sensors, optical MoCap methods, and 3D animation tools, as illustrated in Figure 4.3. In the following, we first look into the biomechanical modeling.

4.3.1   Biomechanical Modeling

Biomechanical modeling [93] aids the synthesis of diverse human motions by using computational frameworks that integrate principles from biomechanics, employing mathematical models rooted in rigid body dynamics to capture nuanced interactions between body segments, joints, and muscles for varied activities. This approach enables the computation of spatial trajectories corresponding to the synthesized human motion, providing a comprehensive representation of movement patterns.

However, biomechanical modeling is inherently complex due to the intricate nature of human biomechanics. This complexity arises from the need to account for factors such as joint angles, segmental accelerations, and muscle forces, demanding a sophisticated integration of multidisciplinary principles. Several limitations exist in comparison to MoCap systems, notably in terms of precision and universality. Biomechanical models may lack the universal applicability of MoCap systems, facing challenges in accurately representing individual variations in anatomy and movement patterns. Additionally, the computational complexity of biomechanical modeling poses practical challenges in real-time applications.

4.3.2   Collecting 3D Trajectories Using Wearable Sensors

Wearable sensors demonstrate versatility in capturing human motion, providing a dynamic means to record and analyze movement patterns across diverse contexts. The integration of sensors into garments, exemplified by the Rokoko Smartsuit Pro [94], serves as a excellent choice for wearable modalities. Such sophisticated systems seamlessly combine comfort and functionality by encapsulating a comprehensive array of sensors within the fabric of the garment, as depicted in Figure 4.3. The integration of these sensors enables precise and real-time tracking of human motion, rendering it an effective solution for motion analysis across various applications.
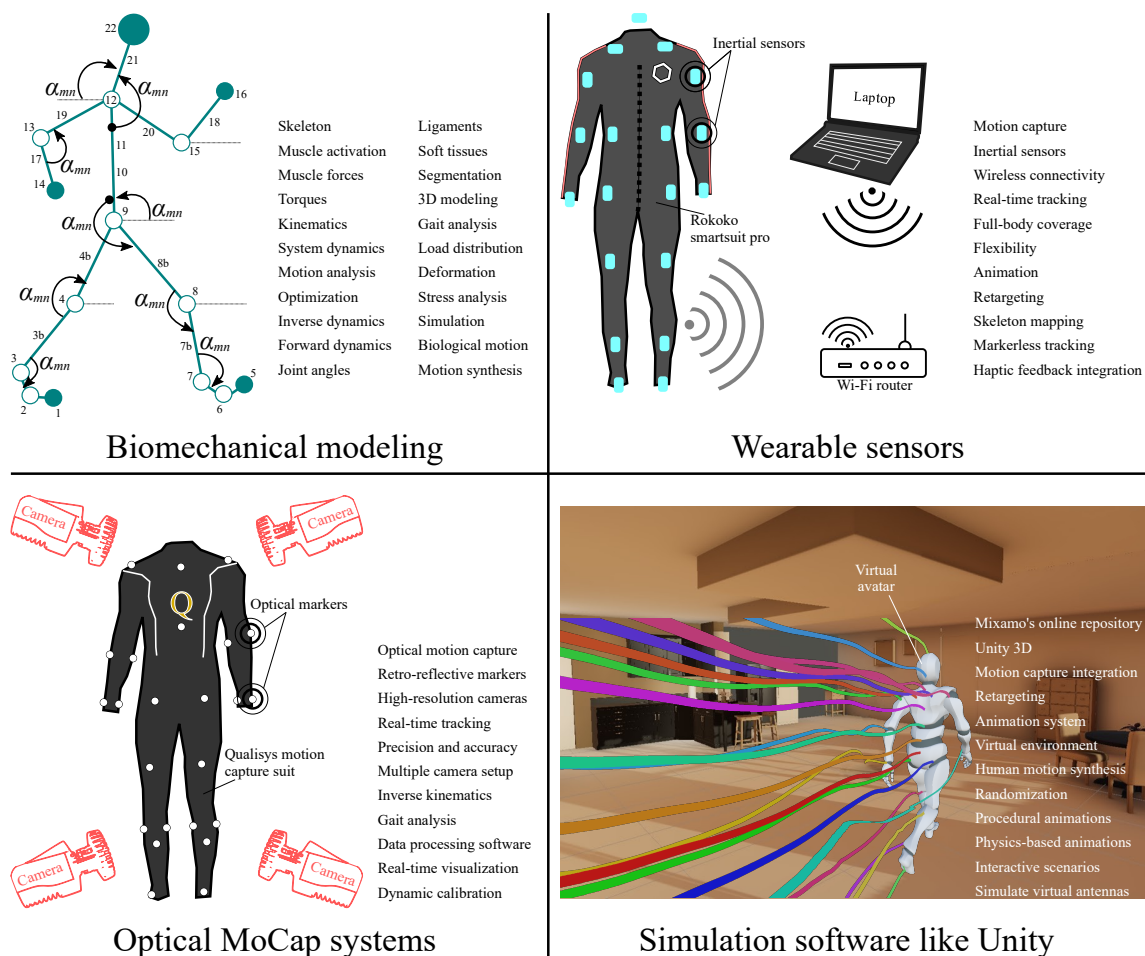
Figure 4.3: Four distinct approaches for capturing and synthesizing the human motion to generate 3D trajectories.

These wearable sensors find application across diverse domains, providing a portable and adaptable means for motion capture. Despite their advantages, wearable sensors exhibit limitations compared to optical MoCap systems, notably in terms of precision and spatial resolution. Additionally, susceptibility to magnetic interference poses a challenge, requiring careful consideration in environments where magnetic fields may impact data accuracy.

4.3.3 Collecting 3D Trajectories Using Optical Motion Capture Systems

In this study, our primary source of MoCap data was Mixamo [69], an online repository that includes optical MoCap data acquired from professional actors [95]. Mixamo serves as a valuable resource for MoCap data. However, the following chapter of this dissertation predominantly utilized MoCap data captured through the optical MoCap system, Qualisys [96]. In the Qualisys MoCap system, we employed a configuration of six cameras arranged in a daisy chain, tracking passive optical reflectors (markers) affixed to the human body. The Qualisys track manager (QTM) software facilitates marker identification and labeling through an automated proce-

dure. Camera calibration, as dictated by the QTM software, ensures the accurate tracking of marker positions and orientations. The skeleton solver functionality of Qualisys system further refines the skeleton calibration based on the positions of optical reflectors.

Human activities are captured using the Qualisys MoCap system, and the resulting spatial trajectories are subsequently imported into software platforms such as Unity and MotionBuilder for further data augmentation. For falling activities, exclusive use of Mixamo MoCap data is preferred due to the challenges and impracticality associated with executing falls while equipped with optical reflectors and costly MoCap apparatus attached to the body.

4.3.4   Simulation Tools for Generating 3D Trajectories

In the process of synthesizing diverse human movements, software tools like Unity and MotionBuilder play a crucial role. Motion data acquired from any MoCap system, including optical or wearable MoCap systems, can be imported into simulation software such as Unity and MotionBuilder. These software applications facilitate the dynamic generation of new motion data based on the fundamental MoCap data acquired from MoCap systems. This seamless integration efficiently augments the rudimentary MoCap data at the motion synthesis layer, thereby facilitating the creation of diverse and dynamic human motion datasets.

To simulate human activities, we utilize an avatar and six rudimentary animations. These animations, namely idle, walking, falling, standing, sitting, and picking, were sourced from the Mixamo repository. The MoCap data of the idle animation involves the avatar maintaining an innate standing pose with minimal in-place motion. The walking animation comprises two forward steps, and the falling animation mimics a sudden collapse as if experiencing a heart attack. In the standing animation, the avatar stands up from a sitting position, while in the sitting animation, it sits down. In the picking animation, it picks up an object. The MoCap data of these six rudimentary animations were imported into the Unity and MotionBuilder software.

Sometimes, it is necessary to switch between various types of animations. However, creating transitional animation data for all such cases can be quite daunting, particularly when working with a large number of animations. Moreover, for enhanced diversity, the existing animation data need to be augmented with slightly varied humanoid movements. To address these issues, a Unity's animation tool called blend tree can be utilized, which enables the dynamic generation of complex, varied, and entirely new sequences of human movements, and it also facilitates seamless transitions between multiple humanoid animations. The blend tree animation tool is quite pertinent to our virtual reality simulations as it significantly improves the expressiveness and realism of the avatar motion.

After synthesizing and augmenting human motion at the motion-layer synthesis, spatial (3D) trajectories for these motions can be simulated. To achieve this, 21 virtual markers (simulated point scatterers) are positioned on distinct body segments

of the avatar, as depicted in Figure D.3. These virtual markers mimic actual scatterers on human body segments. Assigned to specific body segments, these markers facilitate spatial tracking and recording of corresponding TV 3D trajectories for synthesized human activities. In the following sections, we will explore how these spatial trajectories are utilized to simulate volumes of radar signatures by using our channel model.

4.4 Channel Modeling and Simulation

In this section, we primarily discuss our channel model designed to simulate RF or raw IQ data utilizing the spatial trajectory information of virtual markers. Within this framework, we analyze and simulate the TV propagation delays associated with multipath components originating from these markers. This approach allows for the generation of innumerable radar signatures, including TV radial velocity distributions, offering substantial design control and flexibility. Before delving into further details, let us first briefly compare traditional wireless channels with our proposed channel model.

The input of a traditional wireless channel block is the transmitted RF signal, with the output being the received RF signal, as depicted in Figure 4.4. Traditional wireless channels [97, 98, 99, 100] take into account factors such as multipath fading, signal attenuation, noise, and interference during the RF signal's propagation through the wireless medium. The received RF signal is comprised of multipath components originating from various environmental objects. In Figure 4.4, it is evident that multipath components from both stationary and non-stationary objects contribute to the received RF signal. Contrary to the traditional wireless channel model, our proposed channel model introduces a paradigm shift in the wireless channel modeling. Unlike traditional wireless channels, it incorporates a more comprehensive set of inputs, as illustrated in Figure 4.5. These inputs include TV spatial trajectories from both stationary and non-stationary scatterers, offering a dynamic representation of spatial motion. Additionally, the model includes constant channel parameters, such as carrier frequency f_0, chirp's slope γ, ith transmitter antenna position $\mathcal{C}_i^{T_x}$, and kth receiver antenna position $\mathcal{C}_k^{R_x}$, where i and k range from 1 to the total number of antennas. The resulting output of our channel model is the simulated RF signal of the receiver, thereby essentially transforming the spatial trajectories of scatterers into realistic RF data.

The distinct advantages of our channel model, in comparison to the conventional wireless channel model, are manifold. Notably, our model offers the flexibility to simulate a myriad of scenarios. For instance, at the motion-synthesis layer, various scenarios of scatterers' motion can be simulated with precision. Moreover, the model enables the generation of diverse RF datasets by manipulating different environmental and operational parameters inherent to the wireless channel. This includes variations in the characteristics and motion of environmental objects, antenna locations, and the path gain of both stationary and non-stationary scatterers. This paradigm
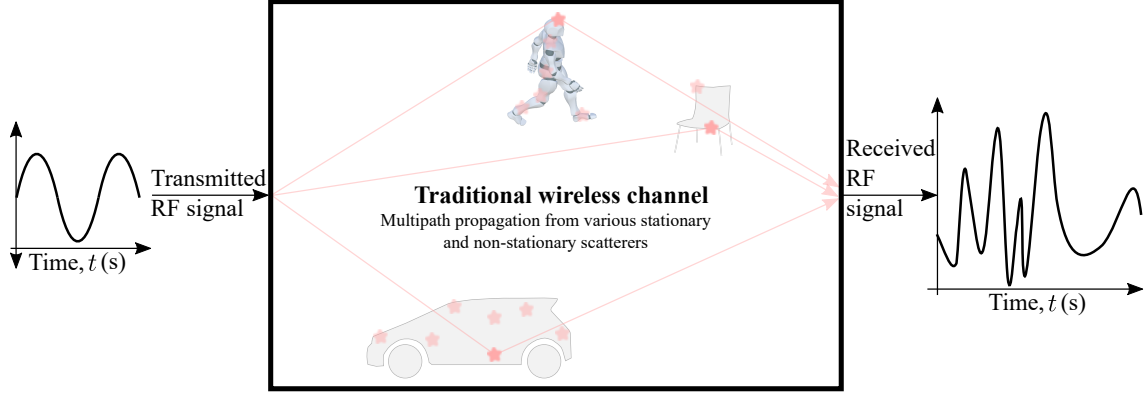
Figure 4.4: A traditional wireless channel model depicting the transformation of a transmitted signal into a composite signal. This composite signal encompasses multipath components originating from stationary/non-stationary dominant/non-dominant scatterers.
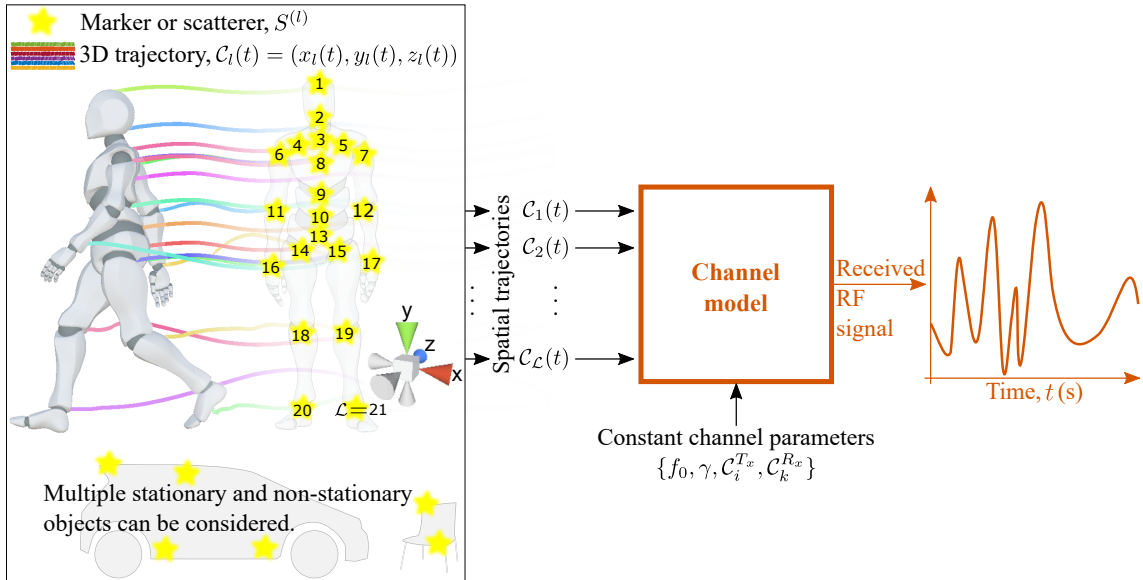
Figure 4.5: Our channel model translating the TV spatial trajectories $\mathcal{C}_l(t)$ of virtual markers into a realistic RF signal while taking into account the carrier frequency f_0, chirp's slope γ, ith transmitter antenna position $\mathcal{C}_i^{T_x}$ and kth receiver antenna position $\mathcal{C}_k^{R_x}$.

shift in wireless channel modeling signifies a departure from experimental-based approaches towards simulation-based methodologies. Leveraging our simulation-based channel model, an array of user-defined scenarios can be simulated, allowing for the generation of unlimited RF data. This capability proves invaluable for constructing RF-based classifiers utilizing simulated data. Let us delve deeper into the proposed channel modeling approach and consider its utility in realizing a simulation-based HAR system.

Multipath propagation of RF signals is common in indoor wireless channels. Within our simulation-centric approach, we exclusively take into account the multipath components emanating from the $\mathcal{L} = 21$ non-stationary scatterers. The fixed

scatterers, also known as clutter, are filtered out through signal preprocessing [101]. Thus, the multipath components from the fixed scatterers like walls, chairs, and tables are not included in analysis. Figure D.3 depicts the propagation of the RF signal from the transmitter to the receiver antennas of the radar. The signal traverses multiple virtual (actual) point scatterers on the humanoid (human) body segments, experiencing multipath propagation. In our simulation-centric approach, the $\mathcal{L} = 21$ virtual markers on the humanoid character represent actual or real bodily scatterers that scatter the electromagnetic energy back to the radar system's receiver antennas. The receiver antennas capture multiple replicas of the transmitted RF signal, with each replica characterized by a unique TV propagation delay denoted as $\tau^{(l)}(t)$ for the lth scatterer, where $l = 1, 2, \ldots, \mathcal{L}$. Let $d_l(t)$ represent the TV radial distance or range, and c_0 denote the speed of light. We can then derive the lth TV propagation delay $\tau^{(l)}(t)$ as $\tau^{(l)}(t) = 2d_l(t)/c_0$. So, in the context of radar sensing, the simulated TV propagation delays $\tau^{(l)}(t)$ fully characterize the synthesized motion.

Figure 4.6 depicts the TV propagation delays $\tau^{(l)}(t)$ simulated by our framework for the five synthesized human activities. The lth TV propagation delay, $\tau^{(l)}(t)$, is directly influenced by the 3D position of the lth marker. This is evident in Figure 4.6, where the rapid change in spatial trajectories during a person's fall is reflected in the TV propagation delays $\tau^{(l)}(t)$. The figure also illustrates the periodic nature of the walking activity, where four steps can be seen. On the other hand, the less mobile activities such as standing up, sitting down, and picking result in the TV propagation delays $\tau^{(l)}(t)$ with smaller fluctuations, indicating the reduced mobility of virtual markers during these activities. Having simulated the TV propagation delays $\tau^{(l)}(t)$ for various human activities, the subsequent step involves synthesizing radar data within our channel model, as elaborated below.

4.5   Radar Data Synthesis

The radar data synthesis module within our channel model emulates an FMCW radar system, simulating high-fidelity raw IQ data. This synthesis relies on the TV propagation delays $\tau^{(l)}(t)$ and TV path gains $a^{(l)}(t)$ of $\mathcal{L}$ virtual markers (see Section D.6.1). Note that, in the context of radar's coherent processing interval
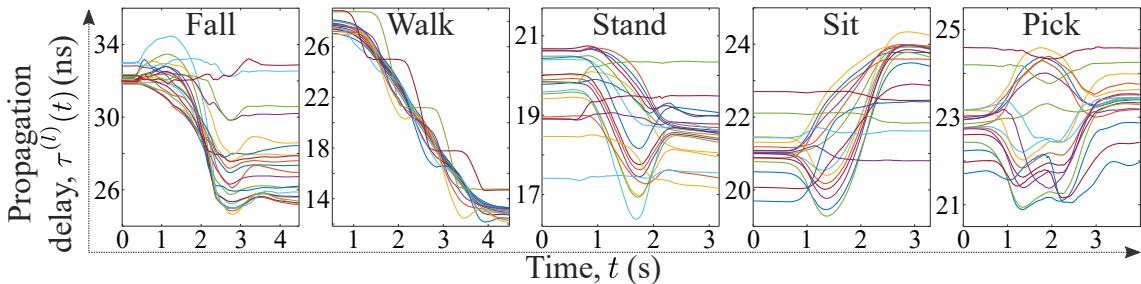
Figure 4.6: The simulated TV propagation delays $\tau^{(l)}(t)$ computed for the $\mathcal{L} = 21$ virtual markers and the five activities (adapted from Paper E).

(CPI), t' and t denote the fast and slow time, respectively. The TV propagation delays $\tau^{(l)}(t)$ dictates both the beat frequency $f_b^{(l)}(t)$ and phase $\phi^{(l)}(t)$ of the beat signal $s_b^{(l)}(t', t)$. For comprehensive details regarding the simulation of beat frequency $f_b^{(l)}(t)$, phase $\phi^{(l)}(t)$, and eventually the beat signal $s_b^{(l)}(t', t)$, refer to Section D.6. Let the function $\delta(\cdot)$ denote the Dirac delta function, and let T_n symbolize the discrete slow time, which is linked to the chirp duration T_{sw} as $T_n = nT_{sw}$ for $n = 0, 1, \ldots$ Then, the beat signal $s_b^{(l)}(t', t)$ can be simulated using the expression given as [62]

$$s_b^{(l)}(t', t) = \sum_{n=0}^{\infty} a^{(l)}(t) \exp \left[j \left(2\pi f_b^{(l)}(t)t' + \phi^{(l)}(t) \right) \right] \times \delta(t - \tau^{(l)}(t) - T_n). \quad (4.1)$$

The TV path gain $a^{(l)}(t)$ represents the electromagnetic energy received by the radar system from the lth scatterer, thereby simulating the strength of the lth virtual marker in synthesizing the corresponding beat signal $s_b^{(l)}(t', t)$. In this proof-of-concept study, we have employed $\mathcal{L} = 21$ constant path gains, denoted as $a^{(l)}$, and for the five simulated human activities, five distinct sets of path gains were employed. The path gains can be meticulously designed using the body surface area [102] and real TV radial velocity distributions (refer to Section D.7). Notably, different sets of path gains can be designed to simulate various TV radial velocity distributions for a single synthesized activity.

We present the simulated and real radar signatures (TV radial velocity distributions $p(v, t)$) corresponding to the simulated and real activities in Figures 4.7 and 4.8, respectively. Visual similarities are evident between the simulated and real radar signatures. It is essential to highlight that these images will be utilized to realize our simulation-based HAR system as detailed in the following section.

4.6 Simulation-Based HAR Systems

In this section, we elaborate the implementation of our HAR system, which utilizes a deep convolutional neural network (DCNN). Simulated radar data is employed for the realization of this system. Different architectures and parameters of the DCNN
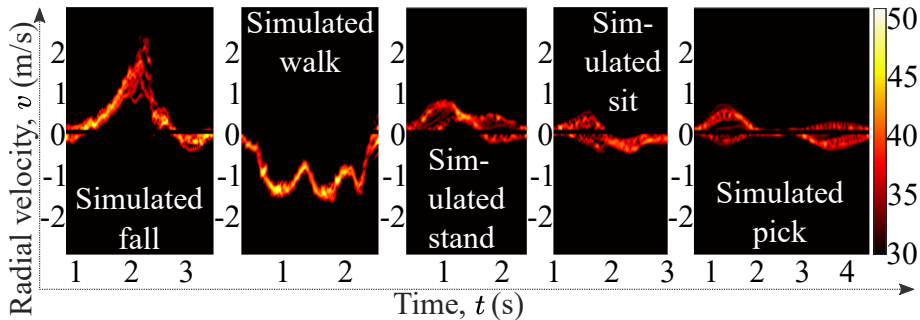
Figure 4.7: Simulated radar signatures for the five emulated activities (reproduced from Paper D).
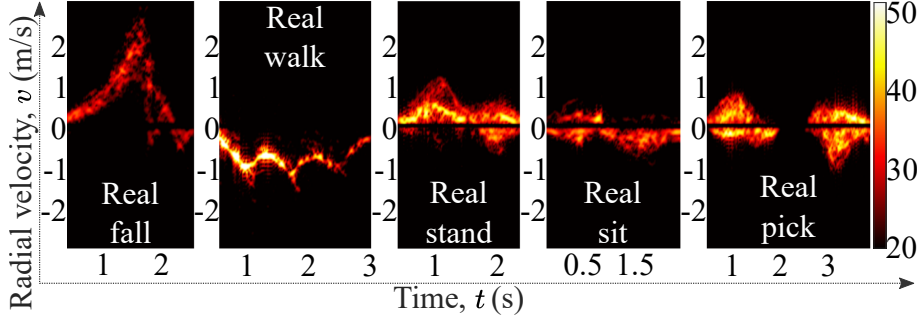
Figure 4.8: Real radar signatures recorded from real human subjects (reproduced from Paper D).

classifier are investigated to select a suitable model. Finally, we test the effectiveness of our HAR classifier (best-performing DCNN model) on real radar data.

The initial step involves simulating a HAR dataset, which is essential for the training phase. This dataset is created by synthesizing human motion within the Unity software. For each of the five activity types, ten distinct activity samples are generated in Unity by varying parameters such as location, speed, acceleration, and more. The 3D trajectories of these fifty activity samples are then exported to MATLAB. In MATLAB, for the sake of data augmentation, eight marginally different radar locations $\{\mathcal{C}^{T_x}, \mathcal{C}^{R_x}\}$ and three signal strengths are simulated for each activity sample. Consequently, a simulation-based HAR dataset is obtained, comprising a total of 1200 simulated TV radial velocity distributions $p(v,t)$. This dataset encompasses five activities, ten emulations per activity, eight radar locations, and three signal strengths. The resulting simulated dataset is utilized for training the DCNN-based HAR classifier.

The simulated radar signatures (TV radial velocity distributions $p(v,t)$) are transformed into images, serving as input feature maps for the DCNN-based HAR classifier, illustrated in Figure D.9. For training and validation, an 80 : 20 split ratio is applied to the simulation dataset. The adaptive moment estimation (Adam) optimizer [86] is employed to optimize the parameters of the DCNN classifier using the training examples from the simulated dataset. Real human activities were recorded by an off-the-shelf FMCW radar system to evaluate the performance of our trained HAR classifier. Both the real and simulated radar systems share same operating parameters, such as the chirp interval, and utilize a monostatic configuration for antennas.

Several DCNN network configurations were carefully investigated to implement the HAR classifier. These network configurations differ in the structure of their hidden layers, each with a unique depth and complexity, as detailed in Table 4.1. The classification performance of these networks (HAR classifiers) was evaluated utilizing the recorded TV radial velocity distributions $p(v,t)$. Models 4–6 exhibited mean accuracies exceeding 86.2% and a standard deviation (SD) of less than 4.8%, thereby showing superior average classification performance compared to models with fewer and more trainable parameters. Consequently, Models 4–6 were further investigated and the optimal values were determined for their hyperparameters. For our HAR

system, Model 6 came out to be the best-performing classifier with optimized hyperparameters like kernel dimension k_d and learning rate l_r. Figure 4.9 illustrates the average accuracies of Model 6 over several values of these hyperparameters. The mean performance of Model 6 is shown in Figures 4.9(a) and 4.9(b) through a graph and a heatmap, respectively, for various hyperparameter combinations.

By utilizing the optimal hyperparameters and the best-performing model identified through the aforementioned mean performance analysis, we were able to achieve the classification accuracy of 98.4%, as illustrated in Figure 4.10. It is important to highlight that the classification performance of our simulation-driven HAR classifier is comparable to real-data-driven HAR systems [7, 103, 58, 59]. Nevertheless, the uniqueness of our simulation-driven framework lies in its ability to easily produce large amounts of realistic simulation data for training purposes.

4.7 Conclusion

The evolution of RF-based HAR systems encounters challenges attributed to restricted and incomplete datasets. These challenges primarily arise from difficulties in procuring radar data for human subjects. In this chapter, we have addressed data scarcity issues by introducing a comprehensive simulation framework. This framework includes the synthesis of human motion and the simulation of realistic mm-wave FMCW radar signatures, thereby significantly facilitating the generation of training data for radar-centric HAR classifiers. Furthermore, the versatility of the simulation framework is notable, providing precise control over various radar and target parameters. This capability enables the generation of diverse radar datasets corresponding to various scenarios, thereby enhancing the applicability of the proposed

Table 4.1: Details and the mean performance of the DCNN architectures used to realize the simulation-based HAR system (reproduced from Paper D).

Model	Convolutional neural network (CNN) layers	multilayer perceptron (MLP) layers	Trainable parameters	Mean accuracy $\pm$ SD (%)
0	$[16, 32]$	$[32, 16]$	$3,232,117$	80.5 ± 7.5
1	$[16, 32, 48]$	$[48, 16]$	$1,882,805$	81.9 ± 5.7
2	$[16, 16, 32, 32, 48, 48]$	$[48, 32, 16]$	$658,485$	77.4 ± 16.5
3	$[32, 48, 64, 80]$	$[128, 64, 32]$	$2,371,557$	85.3 ± 4.7
4	$[32, 64, 72, 80]$	$[256, 128, 32]$	$4,502,205$	87 ± 3.3
5	$[32, 64, 96, 128]$	$[256, 128, 32]$	$7,201,029$	86.2 ± 4.7
6	$[32, 64, 128, 256]$	$[256, 128, 32]$	$14,434,725$	86.7 ± 4.8
7	$[48, 128, 256, 512]$	$[256, 128, 32]$	$31,853,109$	47.3 ± 31.1
8	$[48, 128, 512, 512]$	$[256, 128]$	$37,747,957$	66.4 ± 30.8
9	$[48, 128, 256, 256, 512, 512]$	$[256, 128, 64]$	$43,654,645$	19.7 ± 5.5

(a)

(b)

Figure 4.9: (a) Mean classification accuracies and (b) mean accuracy heatmap of Model 6 for different combinations of kernel dimensions k_d and learning rates l_r (reproduced from Paper D).

Figure 4.10: Confusion matrix of simulation-centric HAR system, realized using Model 6, demonstrating 98.4% accuracy (reproduced from Paper D).

proof-of-concept. The proposed simulated-data-driven HAR system showcased an impressive classification accuracy of 98.4% on previously unseen real radar data, underscoring its effectiveness and potential for future radar-driven HAR classifiers.

In this proof-of-concept, we presented scatterer-level signal modeling for radar signals, improving simulated micro-Doppler signatures and opening new possibilities for radar-centric classifiers. For example, one can explore optimization techniques to improve the realism of simulated radar signals and signatures. Another potential extension involves considering various human activities beyond those studied in this research. The proposed approach holds promise for future applications, such as a sign language detector realized using FMCW radar systems. In the next chapter, we will extend this framework to multiple-input multiple-output (MIMO) radar systems focusing on direction-independent human activity recognition (DIHAR).

Chapter 5

A Simulation-Based Framework for MIMO Radar-Based DIHAR Systems

5.1 Introduction

In this chapter, we present a simulation-centric framework tailored for direction-independent human activity recognition (DIHAR) applications, with a particular emphasis on utilizing millimeter wave (mm-wave) multiple-input multiple-output (MIMO) radar systems. Building on the groundwork laid in earlier chapters, where we highlighted the significance of motion direction in human activity recognition (HAR) (Chapter 2 and Chapter 3) and addressed data scarcity issues in single-input single-output (SISO) radar systems (Chapter 4), our focus now shifts to exploring the potential of MIMO radar systems for developing simulation-centric DIHAR classifiers. This investigation basically extends the simulation-based framework introduced in Paper D, with a specific focus on incorporating MIMO radar systems and multi-directional human activities. The chapter outlines contributions from Paper E, specifically addressing research question 8. In this chapter, we mainly concentrate on aspects of the DIHAR problem and innovative multi-stage data augmentation techniques enabled by our simulation-based framework that haven't been extensively covered in the preceding chapter.

We propose a simulation-based approach that fundamentally differs from existing machine learning-based simulation methods for HAR, such as few-shot learning [15] and domain translation methods [8]. While these state-of-the-art HAR methods address the challenge of limited radar training data in specific scenarios, they still require some radar training data from those scenarios, relying on basic experimental data. In contrast, our fully simulation-based approach does not require any experimental data from an actual radar system for training. Our approach stands out by simulating radar datasets entirely from scratch, utilizing simulated three-dimensional (3D) trajectories of motion capture (MoCap) and incorporating radar and wireless channel modeling. This eliminates the need for new recordings and

enables the simulation of radar signatures for various scenarios, a unique aspect not found even in recent generative adversarial network (GAN)-based HAR approaches. The novelty of our simulation model lies in motion synthesis, physical-layer radar modeling, and radar signal synthesis. By incorporating these three stages, we augment the radar dataset at the motion-layer, physical-layer, and signal-layer syntheses (see Figure 5.1). This level of control in generating realistic MIMO radar data is unprecedented, setting our approach apart from other simulation techniques.

5.2 Our DIHAR Approach

In Paper E, we introduce an approach that addresses the challenges associated with the scarcity of MIMO radar data, enables the simulation of extensive datasets, and provides substantial control and adaptability in designing MIMO radar-based classifiers. Our approach involves simulating a 2×2 MIMO radar system within our framework to implement a simulation-centric DIHAR system. Analogous to our approach in the preceding chapter, we initially emulate diverse human activities in a virtual setting using the 3D animation features of Unity [92] and MotionBuilder [104] software. To animate and retarget recorded MoCap data through these software platforms (refer to Section E.3), we employ a suitable avatar with virtual markers
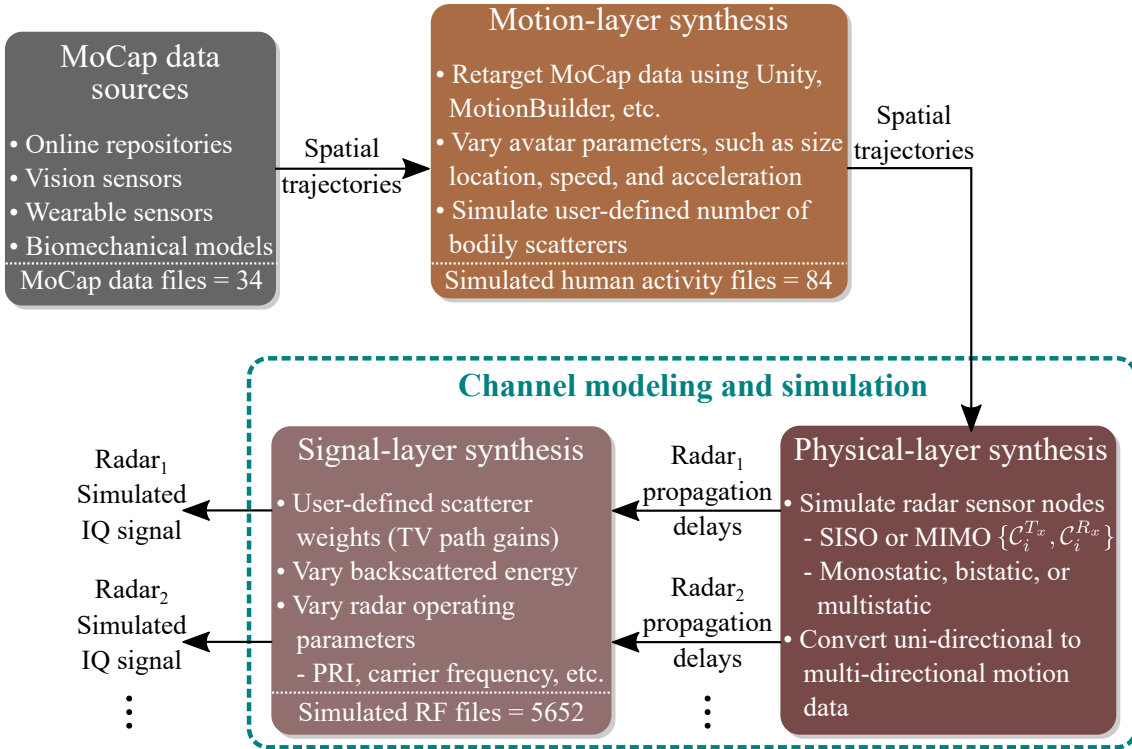
Figure 5.1: The data augmentation possibilities across various stages within our simulation-centric framework. Our channel model facilitating effective data augmentation at the signal- and physical-layer syntheses, enabling the conversion of MoCap data from any source into realistic MIMO radar data (adapted from Paper E).

on its body segments. For characterizing the motion of the avatar, we generate 3D trajectories of virtual markers.

While taking into account user-defined antenna locations, our channel model processes the 3D trajectories from all virtual markers, thereby simulating the composite baseband radio frequency (RF) signal or the raw in-phase and quadrature (IQ) signal of a mm-wave radar system. When generating the simulated raw IQ signal, the channel model considers multipath propagation components primarily emanating from the avatar's virtual markers. These markers have distinct time-variant (TV) propagation delays, as detailed in Section E.4. Moreover, to train our DIHAR system, we generated multi-directional activities by changing the positions of the transmitter and receiver antennas at the physical-layer synthesis stage (refer to Section 5.3.2).

In contrast to measurement-based DIHAR classifiers, our simulation-centric methodology exhibits exceptional adaptability and confers numerous benefits, including the capability to simulate varied training datasets tailored to diverse radar applications and operational needs. For both monostatic and multistatic configurations, encompassing SISO, single-input multiple-output (SIMO), multiple-input single-output (MISO) and MIMO radar systems, our simulation-centric methodology opens avenues for advanced research in refining radar signatures. These signatures may include simulated TV micro-Doppler signatures, simulated TV range distributions, and more (refer to Section E.4 and Section E.6 for more details). Furthermore, we incorporate multi-stage data augmentation methods, as detailed in Section 5.3, enabling the generation of varied radar datasets for both SISO and MIMO configurations in a resource-efficient way. Notably, our approach significantly reduces resources required for classifier training. Given its versatility, our approach can realize various other classifiers stemming from either SISO, SIMO, MISO or MIMO radar systems, such as gesture recognition [105].

Within our proposed simulation-centric framework, we introduce a series of diverse data augmentation strategies encompassing motion-layer synthesis, physical-layer synthesis, and signal-layer synthesis, as outlined in Figure 5.1. In the motion-layer synthesis, we vary the target motion data, and in the physical-layer synthesis, we expand our simulation-based dataset by introducing variations in physical layer parameters, such as the radar's aspect angle or the angle from which the radar illuminates the environment. Finally, at the signal-layer synthesis stage, we employ different sets of simulated TV path gains to simulate different power levels. The adoption of these multi-layered, simulation-centric data augmentation strategies enables the manipulation of target features and antenna positions, the simulation of several sensor nodes, and the conversion of uni-directional movement into multi-directional movement. In the subsequent section, we present the details on the multi-stage data augmentation techniques in the context of our simulation-based DIHAR system.

5.3 Proposed Multi-Stage Data Augmentation Techniques

Now we delve into the data augmentation methods presented by our simulation-centric methodology, as illustrated in Figure 5.1. These techniques empower us to generate extensive, diverse and realistic radar signatures such as TV radial velocity distributions. Initially, we explore the motion-layer synthesis, wherein a range of avatar features, such as the acceleration, deceleration and height of the avatar, undergo random variations. This process allows for the synthesis of a broad spectrum of human motions. Following that, we detail the the physical-layer synthesis. It enables the conversion of uni-directional movement into multi-directional movement as well as the emulation of user-defined antenna configurations, such as monostatic and multistatic, for SISO, SIMO, MISO and MIMO radar systems. Finally, we explore the signal-layer synthesis stage.

5.3.1 Motion-Layer Synthesis

In the preceding chapter, our primary source of MoCap data was Mixamo [69], an online repository that includes optical MoCap data acquired from professional actors [95]. In this chapter, we leverage both the Mixamo repository and the optical MoCap system Qualisys [96] to acquire a limited and fundamental MoCap dataset (refer to Chapter 4). This dataset encompasses five distinct human activities: falling on the ground, walking in an indoor environment, standing up from and sitting down on a seat, and picking up an item. These activities were performed multiple times, and the initial MoCap dataset consisted of just 34 files. Employing the motion-layer synthesis within our approach, we aim to generate an extensive dataset portraying realistic human movement. Note that, the Unity software platform offers a comprehensive suite of features, particularly in animation. One prominent tool within Unity is the blend tree, which is highly pertinent at the motion-layer synthesis. This is due to the fact that blend trees enable the creation of new animation data, contributing to the enhancement of motion data at the motion-synthesis layer. In Unity and MotionBuilder software platforms, we visualized the MoCap data in a virtual 3D environment for the five activities. Subsequently, by modifying the features of the avatar in these software, specifically altering the avatar's height to 5ft and 6ft, we augmented the rudimentary MoCap data, resulting in a total of 84 simulated human activities, as depicted in Figure 5.1.

5.3.2 Physical-Layer Synthesis

Our simulation-centric framework provides the flexibility to fine-tune physical layer configurations according to specific requirements, including adjusting parameters such as pulse repetition frequency (PRF), chirp rate, bandwidth, and antenna positions $\{\mathcal{C}_i^{T_x}, \mathcal{C}_i^{R_x}\}$, where $i = 1, 2$. This flexibility serves to expand radar datasets and model new conditions or scenarios. In the physical-layer synthesis, we selected

appropriate antenna configurations $\{\mathcal{C}_i^{T_x}, \mathcal{C}_i^{R_x}\}$ to emulate two radar subsystems, Radar$_i$, in monostatic configuration, as depicted in Figure 5.2. In order to be consistent with the real Ancortek 2×2 mm-wave radar system (see Figure E.1(a)), we maintained the same simulation parameters for the two monostatic radar subsystems, Radar$_1$ and Radar$_2$, including PRF, chirp rate, bandwidth, and antenna positions $\{\mathcal{C}_i^{T_x}, \mathcal{C}_i^{R_x}\}$.

We emulated different positions of Radar$_i$ using a rotation matrix $R_y(\theta_{Ri})$, where the angle θ_{Ri} is the clockwise rotation with respect to the y-axis, with $i = 1, 2$. For detailed information on the rotation matrix $R_y(\theta_{Ri})$, refer to (E.4). As depicted in Figure 5.2, Radar$_1$ is located at $\theta_{R1} = 0°$, and Radar$_2$ is positioned at $\theta_{R2} = -90°$. Within our framework, Radar$_2$ can be simulated by rotating Radar$_1$ by 90 degrees. Thus, the physical-layer simulation enables the emulation of a 2×2 MIMO radar system and its signatures, akin to the actual radar system illustrated in Figure E.1(a). Importantly, the physical-layer synthesis facilitated by our framework allows us to simulate various radar antenna configurations, including monostatic, bistatic, and multistatic setups for SISO, SIMO, MISO, and MIMO radar systems.

The captured MoCap data originally represented activities in a single direction. To simulate multi-directional human activities for DIHAR, our framework utilizes the physical-layer synthesis. In our simulated MIMO radar system with two subsystems (depicted in Figure 5.2), instead of performing activities in all directions, we rotate both radar subsystems simultaneously, creating 18 different aspect angles. For instance, in Direction 1, Radar$_1$ (Radar$_2$) is at $0°$ ($-90°$), and by subsequently rotating the simulated MIMO radar system by $20°$ clockwise, we simulate Direction 2. This process converts basic uni-directional MoCap data into multi-directional data with respect to the simulated MIMO radar system.
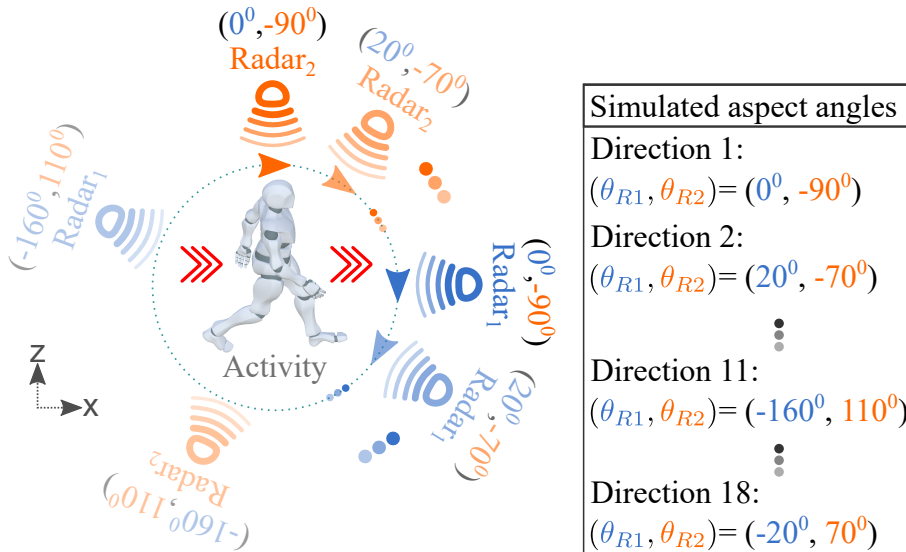
Figure 5.2: Rotating the transmitter and receiver antennas of the 2×2 MIMO radar system to simulate an activity in all directions (reproduced from Paper E).

5.3.3 Signal-Layer Synthesis

Within our framework, the signal-layer synthesis enables the simulation of diverse radar signatures for individual MoCap sample. For the simulated baseband signal $s_{b,i}(t', t)$ (refer to (E.2)), we generate multipath signals corresponding to the TV propagation delays of virtual markers. In this study, we do not simulate the signals reflected from stationary objects as they are filtered out. Nevertheless, the signal-layer synthesis is useful in emulating various wireless conditions, including those with or without radar clutter. Additionally, it empowers us to control all aspects of the simulated baseband signal $s_{b,i}(t', t)$, including the TV path gains $a_i^{(l)}(t)$, beat frequencies $f_{b,i}^{(l)}(t)$, and phases $\phi_i^{(l)}(t)$. For instance, for the simulated activities, we fine-tuned the time-invariant path gains $a_i^{(l)}$ using actual radar signatures and body surface area [102]. To further augment the radar data, we varied power levels by employing distinct sets of time-invariant path gains $a_i^{(l)}$.

We explored data augmentation methods embedded at various stages of our simulation-centric framework for DIHAR: motion-, physical-, and signal-layer syntheses. For each radar subsystem in our simulated MIMO radar system, we generated 2826 radar signatures by employing these methods. In total, 5652 radar signatures were generated for our two radar subsystems, demonstrating the utility and practicality of our simulation-centric framework. By converting the rudimentary MoCap data, initially comprising only 34 samples, into 5652 multi-perspective radar signatures, the data augmentation techniques within our framework played a crucial role in producing comprehensive and realistic radar data.

5.4 Simulation-Based DIHAR Systems

In this section, we detail the implementation and evaluation of our simulation-driven DIHAR system and assess its real-world applicability using a real 2×2 MIMO radar dataset. To train our deep convolutional neural network (DCNN)-based DIHAR system, we simulated a radar dataset with 5652 simulated TV radial velocity distributions. The DIHAR system consists of two identical feature extraction networks (FENs) and a multilayer perceptron (MLP), as illustrated in Figure E.9. During training and testing, the FENs process radar signatures corresponding to each radar subsystem, and the feature fusion block (see Figure E.9) merges features from the two monostatic radar subsystems, Radar$_1$ and Radar$_2$. The MLP network then learns to classify multi-directional human activities, making the radar-based HAR classifier direction-independent.

The scarcity of publicly available datasets containing both raw IQ data and clean radar signatures led us to utilize our own recorded dataset, which included the five types of human activities. While simulating new complex activities presents minimal challenges, acquiring real data remains a significant obstacle. Our simulation-centric approach can adeptly generate radar signatures for numerous activities, thereby providing an efficient and streamlined process. However, assessing the performance of the simulation-based HAR system is constrained by the lack of real radar data for

a wider range of activities. To determine the classification accuracy of our trained DCNN-based DIHAR classifier, we utilized our real dataset acquired with a mm-wave MIMO radar system (see Figure E.2).

The real dataset comprised 875 multi-directional human activities from six subjects, resulting in a dataset with 1750 TV radial velocity distributions. Our simulation-centric DIHAR classifier demonstrated excellent real-world performance, yielding an accuracy of 97.83% as indicated by the confusion matrix in Figure 5.3. Note that our well-balanced test dataset led to a macro average F1-score [106] of approximately 97.6%, aligning closely with the overall accuracy. The simulation-driven DIHAR system achieves accuracy similar to contemporary radar-based HAR systems [103, 7]. However, it uniquely considers diverse human movement directions, it enables the generation of unbounded MIMO radar signatures, and it can be extended to realize various radar-based classifiers like gesture recognition systems.

5.5 Conclusion

The advancement of classifiers, whether utilizing SISO or MIMO radar systems, is primarily impeded by a shortage of labeled datasets. This study introduces a simulation-centric framework to mitigate data scarcity for mm-wave frequency-modulated continuous wave (FMCW) MIMO radar systems. While the emphasis of this study is on a DIHAR system utilizing a MIMO radar system, its application goes beyond DIHAR classifiers.

Our simulation framework generates software-defined multi-directional human

	Fall	Walk	Stand	Sit	Pick	
Fall	94	2	0	1	0	96.91% / 3.09%
Walk	2	151	0	0	0	98.69% / 1.31%
Stand	2	0	163	0	0	98.79% / 1.21%
Sit	0	1	3	223	4	96.54% / 3.46%
Pick	4	0	0	0	225	98.25% / 1.75%
	92.16% / 7.84%	98.05% / 1.95%	98.19% / 1.81%	99.55% / 0.45%	98.25% / 1.75%	97.83% / 2.17%
	Fall	Walk	Stand	Sit	Pick	

Figure 5.3: Confusion matrix quantitatively representing the summary of the results achieved by our simulation-based DIHAR classifier. The overall obtained accuracy is 97.83% for the five multi-directional activities (reproduced from Paper E).

motions from basic MoCap data. It converts simulated spatial trajectories into raw IQ data, while taking into account software-defined physical-layer and signal-layer parameters. For instance, these parameters include the placement of transmitters and receivers, characteristics of multipath components, and more. The simulated raw IQ data undergoes radar signal processing to simulate multiperspective TV radial velocity distributions, used for DIHAR system training.

To diversify radar-based HAR training data, our simulation framework employs multi-stage data augmentation at motion, physical, and signal levels. This enables precise control over avatar properties, motion characteristics, PRF, antenna setup, and more. By augmenting rudimentary MoCap data (34 examples) to 5652 radar signatures, our multi-stage data augmentation techniques remarkably reduces training efforts and proves effective for developing radar-based classifiers. Our trained DIHAR system achieves 97.83% accuracy on real test samples. By eliminating the necessity for human participation and a real radar system, our simulation-based approach holds promise for training future radar-data-driven classifiers.

Chapter 6

Summary and Outlook

This dissertation addresses the challenge of human activity recognition (HAR) within the domain of radio frequency (RF) sensing, primarily focusing on modeling the impact of human motion on wireless channel characteristics, with a specific emphasis on millimeter wave (mm-wave) frequency-modulated continuous wave (FMCW) radar systems. The subsequent sections concisely outline the key findings of this dissertation and offer insights into potential directions for future research.

6.1   Major Contributions

This research signifies a paradigm shift from experimental-based approaches to simulation-based approaches, specifically tailored for HAR systems employing RF sensors. The pivotal innovation lies in the introduction of a sophisticated channel model that inputs three-dimensional (3D) trajectories and outputs high-fidelity simulated RF signals. Notably versatile, this proposed channel model grants substantial control over signal parameters, enabling the simulation of diverse environmental conditions. The key contributions of this dissertation are succinctly delineated as follows:

- In a distributed mm-wave FMCW multiple-input multiple-output (MIMO) radar system, the challenge of cross-channel interference is mitigated through the introduction of a range gating method. This method involves the physical implementation of different RF delay lines, effectively separating distinct RF links.

- This research investigates the limitations of single-input single-output (SISO) and monostatic MIMO radar systems, particularly in relation to motion direction. It introduces a distributed MIMO radar configuration designed to capture multi-perspective time-variant (TV) radial velocity distributions of multi-directional human activities. Leveraging this multistatic MIMO radar setup and the dynamic time warping (DTW) distance metric, a direction-independent step counting system is developed for accurately tallying the number of steps for multi-directional walking activities.

- An experimental-based direction-independent human activity recognition (DIHAR) system is devised using a multi-perspective MIMO radar configuration to address physical layer challenges in current monostatic radar-based HAR systems and classify various multi-directional human activities. A comprehensive measurement campaign is conducted to acquire a dataset for training and testing the experimental-based DIHAR system. Three HAR systems are developed and analyzed: a monostatic radar-based HAR system, a monostatic radar-based DIHAR system, and a distributed MIMO radar-based DIHAR system. Performance comparisons are made to underscore the physical layer challenges present in contemporary monostatic radar-based HAR systems.

- A comprehensive end-to-end simulation framework is introduced that addresses a significant data scarcity challenge in radar systems by transforming motion data into RF signals. It efficiently simulates diverse human activities, generating 3D trajectories for multiple markers on the avatar's body segments through adjustments to avatar's parameters like position, velocity, and size. An innovative channel model is introduced that utilizes spatial trajectories to produce simulated raw in-phase and quadrature (IQ) data or baseband RF signals while considering the long- and short-time stationarity properties of the indoor wireless channel. High-fidelity radar signatures are generated from the raw IQ data. A close match between simulated and real radar signatures is shown by computing normalized DTW distances, thereby demonstrating the framework's effectiveness in capturing the kinematic features of diverse human activities. This approach enables the simulation of numerous realistic radar signatures for diverse operational and environmental conditions, eliminating the need for resource-intensive measurement campaigns. Simulated radar signatures exclusively train a deep convolutional neural network (DCNN)-based HAR classifier, achieving close to 100% classification accuracy when tested with real radar signatures.

- A simulation framework, driven by motion capture (MoCap) data, is introduced, focusing on the development of a DIHAR system that incorporates a MIMO channel model for classifying multi-directional activities. The DIHAR classifier within this simulation-centric framework is exclusively trained using simulated MIMO radar signatures. Data augmentation methods, including motion-layer synthesis, physical-layer synthesis, and signal-layer synthesis, are presented within the simulation-centric approach, enabling the simulation of diverse radar datasets tailored to user-defined MIMO radar configurations. Motion-layer synthesis introduces random variations to avatar features to synthesize a broad spectrum of human motions. Physical-layer synthesis details the conversion of uni-directional movement into multi-directional movement and the emulation of user-defined antenna configurations. Signal-layer synthesis allow the simulation of diverse radar signatures for individual MoCap samples.

56

6.2 Future Work

While this study focuses on simulation-centric HAR systems, the applications of the presented proof-of-concept extend beyond HAR classifiers. The dissertation introduces a paradigm shift in radar sensing, transitioning from experimental-based approaches to simulation-centric methodologies. The future directions are envisioned as follows:

- The scatterer-level signal modeling, presented in the simulation approach, allows for the optimization of scatterers' weights. Optimization techniques can be employed to improve the realism of simulated radar data, incorporating intricate environmental factors and propagation effects.

- A thorough investigation can be conducted into adjusting simulation parameters to ensure the generated data is representative across a broader spectrum of scenarios and conditions, thereby creating a more comprehensive training dataset.

- Investigate the customization of the simulation framework for emerging radar applications, such as autonomous vehicles, smart cities, and other innovative use cases. Tailor the framework to meet the specific requirements of these applications, expanding its capabilities to generate realistic radar data for a diverse range of entities, including animals, vehicles, aircraft, drones, and more. Furthermore, explore the integration of dynamic operational conditions within the simulation, enabling the radar system to adapt to scenarios involving changes in interference and other environmental factors like moving clutter. Additionally, consider extending simulated radar signatures to encompass a broader spectrum of human activities beyond those explored in this research, facilitating the development of radar-based classifiers trained using simulated data, such as sign language detectors.

- Develop methodologies for the validation and benchmarking of simulated radar data, comparing its performance against real-world datasets and identifying areas for improvement.

- The feasibility of real-time simulation capabilities within the framework can be explored to enable on-the-fly generation of radar data. This functionality caters to applications requiring continuous adaptation and learning.

- Investigate the potential for collaborative simulation environments, facilitating interactions among multiple radar systems within a simulated space. This offers a platform for studying complex scenarios involving multiple sensors.

Bibliography

[1] G. A. Oguntala, Y. F. Hu, A. A. S. Alabdullah, R. A. Abd-Alhameed, M. Ali, and D. K. Luong, "Passive RFID module with LSTM recurrent neural network activity classification algorithm for ambient-assisted living," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10953–10962, 2021.

[2] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. De Munari, "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8553–8562, 2019.

[3] G. Cicirelli, R. Marani, A. Petitti, A. Milella, and T. D'Orazio, "Ambient assisted living: A review of technologies, methodologies and future perspectives for healthy aging of population," *Sensors*, vol. 21, no. 10, p. 3549, 2021.

[4] A. Patel and J. Shah, "Sensor-based activity recognition in the context of ambient assisted living systems: A review," *J. Ambient Intell. Smart Environ.*, vol. 11, no. 4, pp. 301–322, 2019.

[5] C. Voigtmann, M. Söllner, K. David, and J. M. Leimeister, "Support-U: Designing an ambient assisted living system using interdisciplinary development patterns," *Socio-technical Des. Ubiquitous Comput. Syst.*, pp. 277–294, 2014.

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

[7] W.-Y. Kim and D.-H. Seo, "Radar-based human activity recognition combining range–time–Doppler maps and range-distributed-convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[8] Y. Yang, Y. Zhang, H. Ji, B. Li, and C. Song, "Radar-based human activity recognition under the limited measurement data support using domain translation," *IEEE Signal Process. Lett.*, vol. 29, pp. 1993–1997, 2022.

[9] N. Avazov, R. Hicheri, M. Muaaz, F. Sanfilippo, and M. Pätzold, "A trajectory-driven 3D non-stationary mm-wave MIMO channel model for a single moving point scatterer," *IEEE Access*, vol. 9, pp. 115990–116001, 2021.

[10] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-Sense: a passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems," *Ann. Telecommun.*, vol. 77, no. 3-4, pp. 163–175, 2022.

[11] M. Muaaz, A. Chelli, and M. Patzold, "WiHAR: From Wi-Fi channel state information to unobtrusive human activity recognition," in *2020 IEEE 91st Veh. Technol. Conf.*, pp. 1–7, 2020.

[12] M. Muaaz, A. Chelli, A. A. Abdelgawwad, A. C. Mallofre, and M. Pätzold, "WiWeHAR: Multimodal human activity recognition using Wi-Fi and wearable sensing modalities," *IEEE Access*, vol. 8, pp. 164453–164470, 2020.

[13] M. J. Bocus, W. Li, S. Vishwakarma, R. Kou, C. Tang, K. Woodbridge, I. Craddock, R. McConville, R. Santos-Rodriguez, K. Chetty, and R. Piechocki, "OPERAnet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors," *Sci. Data*, vol. 9, no. 1, 2022. doi:10.1038/s41597-022-01573-2.

[14] S.-H. Kim, Z. W. Geem, and G.-T. Han, "A novel human respiration pattern recognition using signals of ultra-wideband radar sensor," *Sensors*, vol. 19, no. 15, p. 3340, 2019.

[15] Z. Liu, C. Wu, and W. Ye, "Category-extensible human activity recognition based on Doppler radar by few-shot learning," *IEEE Sens. J.*, vol. 22, no. 22, pp. 21952–21960, 2022.

[16] F. Gu, M. H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu, "A survey on deep learning for human activity recognition," *ACM Comput. Surv.*, vol. 54, no. 8, 2022.

[17] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 669–673, 2018.

[18] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions," *Knowledge-Based Syst.*, vol. 223, p. 106970, 2021.

[19] W. Qi, N. Wang, H. Su, and A. Aliverti, "DCNN based human activity recognition framework with depth vision guiding," *Neurocomputing*, vol. 486, pp. 261–271, 2022.

[20] J. Meyer, A. Frank, T. Schlebusch, and E. Kasneci, "U-har: A convolutional approach to human activity recognition combining head and eye movements for context-aware smart glasses," *Proc. ACM Human-Computer Interact.*, vol. 6, no. ETRA, pp. 1–19, 2022.

[21] A. P. Ismail, M. A. B. Azahar, N. M. Tahir, K. Daud, and N. M. Kasim, "Human action recognition (HAR) using image processing on deep learning," in *2023 IEEE 13th Int. Conf. Control Syst. Comput. Eng.*, pp. 179–183, 2023.

[22] A. S. Patel, R. Vyas, O. P. Vyas, M. Ojha, and V. Tiwari, "Motion-compensated online object tracking for activity detection and crowd behavior analysis," *Vis. Comput.*, vol. 39, no. 5, pp. 2127–2147, 2023.

[23] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep learning-based human pose estimation: A survey," *ACM Comput. Surv.*, vol. 56, no. 1, pp. 1–37, 2023.

[24] K. Ahuja, Y. Jiang, M. Goel, and C. Harrison, "Vid2Doppler: Synthesizing Doppler radar data from videos for training privacy-preserving activity recognition," in *Proc. 2021 CHI Conf. Hum. Factors Comput. Syst.*, pp. 1–10, may 2021.

[25] N. Siddiqui and R. H. M. Chan, "Hand gesture recognition using multiple acoustic measurements at wrist," *IEEE Trans. Human-Machine Syst.*, vol. 51, no. 1, pp. 56–62, 2021.

[26] M. Mohtadifar, M. Cheffena, and A. Pourafzal, "Acoustic- and radio-frequency-based human activity recognition," *Sensors*, vol. 22, no. 9, p. 3125, 2022.

[27] H. Purwins, B. Li, T. Virtanen, J. Schluter, S.-Y. Chang, and T. Sainath, "Deep learning for audio signal processing," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 2, pp. 206–219, 2019.

[28] S. Pandya and H. Ghayvat, "Ambient acoustic event assistive framework for identification, detection, and recognition of unknown acoustic events of a residence," *Adv. Eng. Informatics*, vol. 47, p. 101238, 2021.

[29] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," *Appl. Soft Comput.*, vol. 62, pp. 915–922, 2018.

[30] I. Suarez, A. Jahn, C. Anderson, and K. David, "Improved activity recognition by using enriched acceleration data," in *Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 1011–1015, 2015.

[31] B. N. Klein, S. L. Lau, and K. David, "Evaluation of the influence of time synchronisation on classification learning based movement detection with accelerometers," in *2011 IEEE/IPSJ Int. Symp. Appl. Internet*, pp. 56–64, 2011.

[32] S. L. Lau, I. Konig, K. David, B. Parandian, C. Carius-Dussel, and M. Schultz, "Supporting patient monitoring using activity recognition with a smartphone," in *2010 7th Int. Symp. Wirel. Commun. Syst.*, pp. 810–814, 2010.

[33] S. L. Lau and K. David, "Movement recognition using the accelerometer in smartphones," in *2010 Futur. Netw. Mob. Summit*, pp. 1–9, 2010.

[34] M. Webber and R. F. Rojas, "Human activity recognition with accelerometer and gyroscope: A data fusion approach," *IEEE Sens. J.*, vol. 21, no. 15, pp. 16979–16989, 2021.

Bibliography

[35] S. Mekruksavanich and A. Jitpattanakul, "RNN-based deep learning for physical activity recognition using smartwatch sensors: A case study of simple and complex activity recognition," *Math. Biosci. Eng.*, vol. 19, no. 6, pp. 5671–5698, 2022.

[36] Y. Tang, L. Zhang, F. Min, and J. He, "Multiscale deep feature learning for human activity recognition using wearable sensors," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 2106–2116, 2023.

[37] F. Lin, Z. Wang, H. Zhao, S. Qiu, X. Shi, L. Wu, R. Gravina, and G. Fortino, "Adaptive multi-modal fusion framework for activity monitoring of people with mobility disability," *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 8, pp. 4314–4324, 2022.

[38] X. Feng, Y. Weng, W. Li, P. Chen, and H. Zheng, "DAMUN: A domain adaptive human activity recognition network based on multimodal feature fusion," *IEEE Sens. J.*, vol. 23, no. 18, pp. 22019–22030, 2023.

[39] Y. Zhang, Q. Liu, Y. Wang, and G. Yu, "CSI-based location-independent human activity recognition using feature fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[40] L. Arrotta, C. Bettini, and G. Civitarese, "MICAR: Multi-inhabitant context-aware activity recognition in home environments," *Distrib. Parallel Databases*, vol. 41, no. 4, pp. 571–602, 2023.

[41] M. Morold, M. Bachmann, and K. David, "Toward context awareness for cooperative vulnerable road user collision avoidance: Incorporating related contextual information," *IEEE Veh. Technol. Mag.*, vol. 17, no. 3, pp. 75–83, 2022.

[42] M. Morold, Q.-H. Nguyen, M. Bachmann, K. David, and F. Dressler, "Requirements on delay of VRU context detection for cooperative collision avoidance," in *2020 IEEE 92nd Veh. Technol. Conf.*, pp. 1–5, 2020.

[43] S. Deng, J. Chen, D. Teng, C. Yang, D. Chen, T. Jia, and H. Wang, "LHAR: Lightweight human activity recognition on knowledge distillation," *IEEE J. Biomed. Heal. Informatics*, pp. 1–10, 2023.

[44] R. Kan, H. Qiu, X. Liu, P. Zhang, Y. Wang, M. Huang, and M. Wang, "Indoor human action recognition based on dual Kinect V2 and improved ensemble learning method," *Sensors*, vol. 23, no. 21, p. 8921, 2023.

[45] Y. Zhang, X. Yao, Q. Fei, and Z. Chen, "Smartphone sensors-based human activity recognition using feature selection and deep decision fusion," *IET Cyber-Physical Syst. Theory Appl.*, vol. 8, no. 2, pp. 76–90, 2023.

[46] M. Straczkiewicz, P. James, and J.-P. Onnela, "A systematic review of smartphone-based human activity recognition methods for health research," *npj Digit. Med.*, vol. 4, no. 1, p. 148, 2021.

[47] Y. Guo, Y. Chu, B. Jiao, J. Cheng, Z. Yu, N. Cui, and L. Ma, "Evolutionary Dual-Ensemble Class Imbalance Learning for Human Activity Recognition," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 4, pp. 728–739, 2022.

[48] Z. Ma, L. T. Yang, M. Lin, Q. Zhang, and C. Dai, "Weighted Support Tensor Machines for Human Activity Recognition with Smartphone Sensors," *IEEE Trans. Ind. Informatics*, 2021.

[49] M. Tammvee and G. Anbarjafari, "Human activity recognition-based path planning for autonomous vehicles," *Signal, Image Video Process.*, vol. 15, no. 4, pp. 809–816, 2021.

[50] F. J. Rodriguez Lera, F. Martín Rico, A. M. Guerrero Higueras, and V. M. Olivera, "A context-awareness model for activity recognition in robot-assisted scenarios," *Expert Syst.*, vol. 37, no. 2, 2020.

[51] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Electr. Eng. Technol.*, vol. 14, no. 6, pp. 2567–2573, 2019.

[52] M. Al-khafajiy, T. Baker, C. Chalmers, M. Asim, H. Kolivand, M. Fahim, and A. Waraich, "Remote health monitoring of elderly through wearable sensors," *Multimed. Tools Appl.*, vol. 78, no. 17, pp. 24681–24706, 2019.

[53] M. A. A. Al-qaness, A. Dahou, M. A. Elaziz, and A. M. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," *IEEE Trans. Ind. Informatics*, vol. 19, no. 1, pp. 144–152, 2023.

[54] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3617–3628, 2019.

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

[56] F. Fioranelli, M. Ritchie, and H. Griffiths, "Bistatic human micro-Doppler signatures for classification of indoor activities," in *2017 IEEE Radar Conf.*, pp. 610–615, 2017.

[57] B. Jokanovic and M. Amin, "Fall detection using deep learning in range-Doppler radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, 2018.

[58] S. Waqar, M. Muaaz, and M. Pätzold, "Direction-independent human activity recognition using a distributed MIMO radar system and deep learning," *IEEE Sens. J.*, vol. 23, no. 20, pp. 24916–24929, 2023.

Bibliography

[59] M. Muaaz, S. Waqar, and M. Pätzold, "Orientation-independent human activity recognition using complementary radio frequency sensing," *Sensors*, vol. 23, no. 13, p. 5810, 2023.

[60] V. C. Chen, *The micro-Doppler effect in radar.* Artech House, 2019.

[61] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, "SimHumalator: An open-source end-to-end radar simulator for human activity recognition," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 3, pp. 6–22, 2022.

[62] S. Waqar and M. Pätzold, "A simulation-based framework for the design of human activity recognition systems using radar sensors," *IEEE Internet Things J.*, 2023. doi:10.1109/JIOT.2023.3344179.

[63] S. Waqar, M. Muaaz, S. Sigg, and M. Pätzold, "A paradigm shift from an experimental-based to a simulation-based framework using motionCapture driven MIMO radar data synthesis," *IEEE Sens. J.*, 2023 (under review).

[64] E. Ramanujam, T. Perumal, and S. Padmavathi, "Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review," *IEEE Sens. J.*, vol. 21, pp. 13029–13040, jun 2021.

[65] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K. R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," *Inf. Fusion*, vol. 53, pp. 80–87, 2020.

[66] A. H. Yuh and S. J. Kang, "Real-time sound event classification for human activity of daily living using deep neural network," in *2021 IEEE Int. Conf. Internet Things IEEE Green Comput. Commun. IEEE Cyber, Phys. Soc. Comput. IEEE Smart Data IEEE Congr. Cybermatics*, pp. 83–88, 2021.

[67] N. Tasnim, M. K. Islam, and J.-H. Baek, "Deep learning based human activity recognition using spatio-temporal image formation of skeleton joints," *Appl. Sci.*, vol. 11, no. 6, p. 2675, 2021.

[68] S. Waqar, M. Muaaz, and M. Pätzold, "Human activity signatures captured under different directions using SISO and MIMO radar systems," *Appl. Sci.*, vol. 12, no. 4, 2022.

[69] "Get animated." Mixamo, accessed Feb. 20, 2023. [Online]. Available: https://www.mixamo.com/#/.

[70] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022.

[71] H.-U.-R. Khalid, A. Gorji, A. Bourdoux, S. Pollin, and H. Sahli, "Multi-view CNN-LSTM architecture for radar-based human activity recognition," *IEEE Access*, vol. 10, pp. 24509–24519, 2022.

[72] X. Li, Y. He, J. A. Zhang, and X. Jing, "Supervised domain adaptation for few-shot radar-based human activity recognition," *IEEE Sens. J.*, vol. 21, no. 22, pp. 25880–25890, 2021.

[73] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, "Physics-aware generative adversarial networks for radar-based human activity recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 3, pp. 2994–3008, 2023.

[74] L. Qu, Y. Wang, T. Yang, and Y. Sun, "Human activity recognition based on WRGAN-GP-synthesized micro-Doppler spectrograms," *IEEE Sens. J.*, vol. 22, no. 9, pp. 8960–8973, 2022.

[75] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proc. 3rd ACM Work. Millimeter-wave Networks Sens. Syst.*, pp. 51–56, 2019.

[76] F. Gurcan, N. E. Cagiltay, and K. Cagiltay, "Mapping human–computer interaction research themes and trends from its existence to today: A topic modeling-based review of past 60 years," *Int. J. Hum. Comput. Interact.*, vol. 37, no. 3, pp. 267–280, 2021.

[77] B. Erol and M. Amin, "Effects of range spread and aspect angle on radar fall detection," in *2016 IEEE Sens. Array Multichannel Signal Process. Work.*, pp. 1–5, 2016.

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

[79] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on Doppler radar motion signature," in *2011 5th Int. Conf. Pervasive Comput. Technol. Healthc. Work.*, pp. 222–225, 2011.

[80] L. Liu, M. Popescu, K. C. Ho, M. Skubic, and M. Rantz, "Doppler radar sensor positioning in a fall detection system," in *2012 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 256–259, 2012.

[81] P. Nallabolu, L. Zhang, H. Hong, and C. Li, "Human presence sensing and gesture recognition for smart home applications with moving and stationary clutter suppression using a 60-GHz digital beamforming FMCW radar," *IEEE Access*, vol. 9, pp. 72857–72866, 2021.

[82] S. Sana, S. Waqar, H. Yusaf, M. Waqas, and F. A. Siddiqui, "Software defined digital beam forming processor," in *2016 13th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 671–676, 2016.

[83] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *2015 IEEE Radar Conf.*, pp. 1491–1496, 2015.

[84] C. Yu, Z. Xu, K. Yan, Y.-R. Chien, S.-H. Fang, and H.-C. Wu, "Noninvasive human activity recognition using millimeter-wave radar," *IEEE Syst. J.*, vol. 16, no. 2, pp. 3036–3047, 2022.

[85] M. Muaaz and R. Mayrhofer, "Smartphone-based gait recognition: from authentication to imitation," *IEEE Trans. Mob. Comput.*, vol. 16, no. 11, pp. 3209–3221, 2017.

[86] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv Prepr.*, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980.

[87] H. Shang, Z. Sun, W. Yang, X. Fu, H. Zheng, J. Chang, and J. Huang, "Leveraging other datasets for medical imaging classification: Evaluation of transfer, multi-task and semi-supervised learning," in *Med. Image Comput. Comput. Assist. Interv. – MICCAI 2019*, pp. 431–439, Springer International Publishing, 2019.

[88] J. Shahid, R. Ahmad, A. K. Kiani, T. Ahmad, S. Saeed, and A. M. Almuhaideb, "Data protection and privacy of the internet of healthcare things (IoHTs)," *Appl. Sci.*, vol. 12, no. 4, p. 1927, 2022.

[89] C. C. Nicholls, "Open banking and the rise of FinTech: Innovative finance and functional regulation," *Bank. Financ. Law Rev.*, vol. 35, no. 1, pp. 121–151, 2019.

[90] J. R. Saura, D. Ribeiro-Soriano, and D. Palacios-Marqués, "Assessing behavioral data science privacy issues in government artificial intelligence deployment," *Gov. Inf. Q.*, vol. 39, no. 4, p. 101679, 2022.

[91] V. Thambawita, P. Salehi, S. A. Sheshkal, S. A. Hicks, H. L. Hammer, S. Parasa, T. de Lange, P. Halvorsen, and M. A. Riegler, "SinGAN-Seg: Synthetic training data generation for medical image segmentation," *PLoS One*, vol. 17, no. 5, 2022.

[92] "Unity Real-Time Development Platform | 3D, 2D VR & AR Engine." Unity, accessed Feb. 20, 2023. [Online]. Available: https://www.unity.com/.

[93] I. Roupa, M. R. da Silva, F. Marques, S. B. Gonçalves, P. Flores, and M. T. da Silva, "On the modeling of biomechanical systems for human movement analysis: A narrative review," *Arch. Comput. Methods Eng.*, vol. 29, no. 7, pp. 4915–4958, 2022.

[94] "Smartsuit Pro II - Quality body motion capture in one simple mobile mocap suit." Rokoko, accessed Jul. 10, 2023. [Online]. Available: https://www.rokoko.com/products/smartsuit-pro.

[95] Mixamo, "Mixamo motion capture (behind the scenes)." YouTube, Aug. 24, 2012. [Online]. Available: https://www.youtube.com/watch?v=q7jWMd7ix98.

[96] "Qualisys | Motion Capture Systems." Qualisys, accessed Jun. 21, 2023. [Online]. Available: https://www.qualisys.com/.

[97] A. L. Imoize, A. E. Ibhaze, A. A. Atayero, and K. V. N. Kavitha, "Standard propagation channel models for MIMO communication systems," *Wirel. Commun. Mob. Comput.*, vol. 2021, pp. 1–36, 2021.

[98] Y. Ma, B. O. Hogstad, M. Pätzold, and P. M. Crespo, "Statistical modeling, simulation, and experimental verification of wideband indoor mobile radio channels," *Wirel. Commun. Mob. Comput.*, vol. 2018, pp. 1–13, 2018.

[99] Y. Ai, B. O. Hogstad, M. Cheffena, and M. Patzold, "Geometry-based modeling of wideband industrial indoor radio propagation channels," in *IECON 2015 - 41st Annu. Conf. IEEE Ind. Electron. Soc.*, pp. 004299–004304, 2015.

[100] M. Pätzold and B. O. Hogstad, "A wideband MIMO channel model derived from the geometric elliptical scattering model," *Wirel. Commun. Mob. Comput.*, vol. 8, no. 5, pp. 597–605, 2008.

[101] S. Waqar, H. Yusaf, S. Sana, M. Waqas, and F. A. Siddiqui, "Reconfigurable monopulse radar tracking processor," in *2018 15th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 805–809, 2018.

[102] M. Mance, M. Prutki, A. Dujmovic, M. Miloševic, V. Vrbanovic-Mijatovic, and D. Mijatovic, "Changes in total body surface area and the distribution of skin surfaces in relation to body mass index," *Burns*, vol. 46, no. 4, pp. 868–875, 2020.

[103] Y. Zhao, H. Zhou, S. Lu, Y. Liu, X. An, and Q. Liu, "Human activity recognition based on non-contact radar data and improved PCA method," *Appl. Sci.*, vol. 12, no. 14, p. 7124, 2022.

[104] "Autodesk MotionBuilder Software." Autodesk, accessed Jun. 21, 2023. [Online]. Available: https://www.autodesk.com/products/motionbuilder/overview?term=1-YEAR&tab=subscription.

[105] S. Ahmed, W. Kim, J. Park, and S. H. Cho, "Radar-based air-writing gesture recognition using a novel multistream CNN approach," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 23869–23880, 2022.

[106] H. Schutze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval.* Cambridge University Press, 2008.

Appendix A

Paper A

Interchannel Interference and Mitigation in Distributed MIMO RF Sensing

Sahil Waqar and Matthias Pätzold

Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway

E-mails: {sahil.waqar, matthias.paetzold}@uia.no

Abstract— In this paper, we analyze and mitigate the cross-channel interference, which is found in multiple-input multiple-output (MIMO) radio frequency (RF) sensing systems. For a millimeter wave (mm-Wave) MIMO system, we present a geometrical three-dimensional (3D) channel model to simulate the time-variant (TV) trajectories of a moving scatterer. We collected RF data using a state-of-the-art radar known as Ancortek SDR-KIT 2400T2R4, which is a frequency-modulated continuous wave (FMCW) MIMO radar system operating in the K-band. The Ancortek radar is currently the only K-band MIMO commercial radar system that offers customized antenna configurations. It is shown that this radar system encounters the problem of interference between the various subchannels. We propose an optimal approach to mitigate the problem of cross-channel interference by inducing a propagation delay in one of the channels and apply range gating. The measurement results prove the effectiveness of the proposed approach by demonstrating a complete elimination of the interference problem. The application of the proposed solution on Ancortek's SDR-KIT 2400T2R4 allows resolving all subchannel links in a distributed MIMO configuration. This allows using MIMO RF sensing techniques to track a moving scatterer (target) regardless of its direction of motion.

Keywords— Interchannel interference; distributed MIMO; 3D channel model; sensor network; millimeter wave (mm-Wave); FMCW; micro-Doppler signatures; RF sensing.

A.1   Introduction

The birth of radar in the first half of the last century led to numerous research studies and advances in the field. Although radar systems were originally developed for military surveillance tasks, modern radars have found many applications in our daily lives due to their continuous development over the decades. Conventionally, radar systems were limited to official or governmental entities, but now their smaller form factor, lower cost, higher precision, and easier handling have led to more general utilization. Conventional applications of radars have been aerial [1] and terrestrial [2] traffic control, missile and aerial defense [3], altimetry [4], naval surveillance [5], weather surveillance [6], and astronomy [7], whereas the contemporary radar systems have also been employed in modern medicine [8], autonomous vehicles [9, 10, 11], geology [12], building security, human activity recognition systems [13, 14, 15, 16], and even in consumer electronics such as mobile phones [17] (serving as a gesture recognition system). It is now safe to assert that the idea of radar sensors being ubiquitous is not far-fetched anymore due to their miniaturization, affordability, and utility. For a non-trivial problem such as autonomous driving in automotive engineering, several types of radar systems (short-range, medium-range, and long-range) [18] are typically integrated to achieve the desired performance, especially under adverse lighting conditions, where other sensing modalities do not perform as required.

A radar system transmits electromagnetic waves and processes the received backscattered waves to estimate one or more parameters of an object present in the environment. Depending on the type of radar, it may measure the range, Doppler (or micro-Doppler) signature, and angular information of a target within certain limitations. Depending on the problem, a radar may be designed and deployed as a continuous wave (CW) radar [19], frequency-modulated continuous wave (FMCW) radar [20], pulsed radar [21], bistatic radar, monopulse radar [22, 23], synthetic aperture radar (SAR) [24], digital beamforming (DBF) multiple-input multiple-output (MIMO) radar in a monostatic configuration [25, 26], or distributed MIMO radar [27, 28, 29]. Recently, short- to medium-range FMCW radars have been gaining increasing attention for commercial indoor and outdoor applications. For instance, the authors of [30, 31] have used a K-band FMCW radar system in indoor settings to monitor human vital functions. More recently, FMCW radar systems operating in the W-band have been adopted for more sophisticated applications, such as sign language recognition [32], multimodal traffic monitoring [33], and skeletal posture estimation [34].

Generally, radar systems suppress the static clutter by filtering out the zero-Doppler frequency components from the received signal, which prevents detection and tracking of the scatterer's motion perpendicular to its boresight. Thus, to acquire the scatterer's motion information from multiple aspect angles, the deployment of a single-input single-output (SISO) radar or a monostatic MIMO radar is not a suitable choice. Instead, with the idea of macrodiversity, a distributed MIMO radar system or a multistatic radar network is preferred to circumvent the shortcomings

of the aforementioned radar configurations. It is in this context that we will focus our attention on the deployment of a distributed MIMO radar system in indoor environments. For different application areas, researchers are investigating different target–antenna configurations while leaning towards multistatic radar networks. For example, the authors of [35] deployed a network known as NetRAD for the detection of armed/unarmed personnel, and the authors of [36] report the use of a commercial DWM1000 ultra wideband wireless transceiver module in a multistatic configuration to track a moving person in a cluttered indoor or outdoor environment.

The probability of mutual interference between radar systems is increasing gradually as commercial radars become more widely used. In distributed MIMO radar systems, cross-channel interference exists between the different nodes of a multistatic radar network. For this research, we chose a radar system that uses the time division multiple access (TDMA) scheme to avoid cross-channel interference. In the TDMA mode, the transmitters of a MIMO radar system operate in different time slots. As part of the physical channel characteristics, it is also imperative for the system performance to consider the interchannel radio frequency (RF) isolation inside the RF circuitry. In case of RF leakage in MIMO radar subchannels, the received signals are of the same order of magnitude for all receiver channels. For a consumer grade hardware that undergoes such RF leakage, the signal from one receiver leaks into the other receiver, and vice versa, making it impossible to separate the subchannels from each other. The problem is then to distinguish the received signal once impaired by RF-leakage from the co-channel signals. The interference problems arising due to the RF leakage between the RF chains cannot be resolved by the TDMA scheme, because the TDMA scheme is only effective against cross-channel interference if good RF-isolation is ensured beforehand. Thus, for such consumer grade MIMO radar systems, we propose a robust approach in this paper to solve the interference problem.

To estimate the trajectories of a non-stationary scatterer from different aspect angles in a cluttered indoor environment, we adopt Ancortek's commercial MIMO radar system SDR-KIT 2400T2R4, which operates in the 24–26 GHz frequency band. It has in aggregate six independent physical RF chains: two transmitter chains and four receiver chains. For this research, we utilize Ancortek's 2×4 MIMO radar system in a 2×2 configuration for simplicity. Ancortek's radar system is currently the only commercially available MIMO radar system that offers the flexibility to distribute its antennas and to process all eight MIMO subchannel links individually. We distribute two pairs of collocated transmitter–receiver antennas in an indoor setting to illuminate a non-stationary scatterer from different aspect angles. The problem of cross-channel interference arises in Ancortek's MIMO radar system even with the utilization of the TDMA scheme. Furthermore, we will point out that Ancortek's SDR-KIT 2400T2R4 has a very poor interchannel RF isolation, which leads to incorrect measurements of the mean Doppler shift. Thus, without any hardware or firmware alteration, there is no known optimal solution to effectively isolate the different RF MIMO channel links. The problem of interference in the Ancortek radar has also been reported by the authors of [37], where they have subtracted the

spectrograms to alleviate the interference problem. The solution proposed in [37] is suboptimal and non-robust; it works when the interference component is smaller than the subchannel's main component and fails when the interference component is of the order of the magnitude of the main component of the subchannel. Therefore, in this paper, we propose an optimal and robust solution that completely eradicates the problem of cross-channel interference. The proposed solution performs effectively even when the interference component is stronger than the subchannel's main component. Although our focus is on Ancortek's radar, similar interference problems may also persist for future commercially available MIMO radar sensors. Thus, for such MIMO radar systems, the proposed solution can be adopted without entailing any hardware or firmware modifications. Additionally, the proposed solution also helps alleviate the maximum measurable velocity or the pulse repetition frequency (PRF) of the radar by completely avoiding the TDMA scheme, and still being able to segregate the MIMO channel links.

The principal contributions of this paper are as follows:

1. For a MIMO radar system whose antennas are distributed in an indoor cluttered environment, we present a system-theoretical approach to simulate the time-variant (TV) trajectories of a scatterer with arbitrary antenna placements.

2. We illuminate a non-stationary scatterer from different aspect angles (by deploying two pairs of collocated transmitter-receiver antennas) to analyze the TV micro-Doppler spectrogram, TV radial velocity profile, and TV mean Doppler shift.

3. For Ancortek's SDR-KIT 2400T2R4 distributed MIMO radar system, we highlight the problem of cross-channel interference. We propose an optimal and robust solution to completely eradicate the interference components without modifying the hardware or firmware of the MIMO radar system.

4. We conduct experiments to verify the effectiveness of the proposed solution by successfully segregating the measured MIMO subchannels' data.

5. We cross-validate the analytical model and the proposed solution of the interference problem by comparing the simulation results with the measurement results.

The organization of the paper is as follows. Section A.2 formulates the interference problems that persist in Ancortek's SDR-KIT 2400T2R4 distributed MIMO radar system. The geometrical 3D indoor channel model and the radar system model are presented in Sections A.3 and A.4, respectively. Section A.5 elucidates the proposed solution to the interference problem. The simulation results and the measurement results are discussed in Section A.6. Finally, Section A.7 summarizes our results and draws the conclusions.

A.2 Problem Description

Capturing and tracking nonlinear trajectories of moving scatterers indoors by means of RF-sensing modalities presents a number of challenges. One major challenge is to detect the scatterer trajectories regardless of the radar's aspect angle, which requires multiple RF sensors. Therefore, for our experiments, a software-defined radar (SDR) known as Ancortek SDR-KIT 2400T2R4 has been configured in a 2×2 MIMO radar setup in the presence of a single moving scatterer S^M as illustrated in Figure A.1. The 2×2 MIMO radar system is composed of two radar subsystems, denoted as Radar$_1$ and Radar$_2$. The first subsystem (Radar$_1$) is equipped with the transmitter antenna $A_1^{T_x}$ and the receiver antenna $A_1^{R_x}$, whereas the second subsystem (Radar$_2$) is composed of the transmitter antenna $A_2^{T_x}$ and the receiver antenna $A_2^{R_x}$. Although the two radar subsystems are part of the same Ancortek system, they have identical but completely separate signal processing units.

The wireless channel link from the transmitter antenna $A_i^{T_x}$ to the receiver antenna $A_j^{R_x}$ via the scatterer S^M is denoted by $A_i^{T_x}$–$A_j^{R_x}$, where $i, j \in \{1, 2\}$. The time-variant channel impulse response (TV-CIR) $h_{ij}(\tau', t)$ corresponds to the link $A_i^{T_x}$–$A_j^{R_x}$ as illustrated in Figure A.1. Moreover, the two subradars operate in the same frequency range but in different time slots. Each subradar is assigned a different time slot according to the TDMA scheme to avoid cross-channel interference between the two subradars. In TDMA mode, the TV-CIRs $h_{21}(\tau', t)$ and $h_{12}(\tau', t)$ do not interfere with $h_{11}(\tau', t)$ and $h_{22}(\tau', t)$, respectively, but this is not true for the Ancortek SDR-KIT 2400T2R4 MIMO radar. The commercially available Ancortek MIMO radar system poses the problem of cross-channel interference even in TDMA mode due to its poor interchannel RF isolation. It is vital for system designers to ensure a good RF-isolation in the MIMO radar RF-circuitry, but such insurance is hard to realize for miniaturized and cost-effective RF circuits. Here, this phenomenon of RF leakage between the physical RF channels has been first investigated for the Ancortek radar because it is currently the only commercially available K-band radar that allows to distribute its antennas. However, the same problem may persist in future commercial MIMO radar systems. Note that this analysis provides guidelines for radar system designers to avoid cross-channel interference
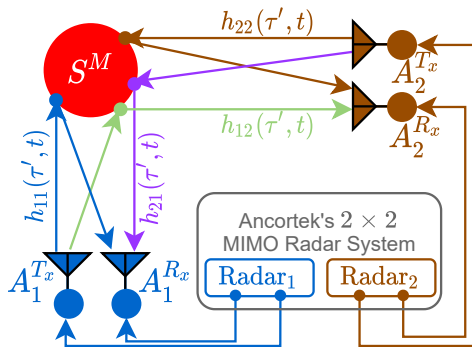
Figure A.1: Two radar subsystems forming a 2×2 MIMO radar system in the presence of a single moving scatterer S^M.

in their future designs. In addition, the analysis provides a performance criterion for the test and evaluation of the future FMCW MIMO radar systems. Note that the frequency division multiple access (FDMA) scheme is generally not preferred in commercial FMCW MIMO radar systems because of the associated complexity and cost. The FDMA approach limits the instantaneous bandwidth of an FMCW radar, which in turn limits the range resolution of the radar (see Section A.4).

The TV-CIRs $h_{11}(\tau', t)$ and $h_{22}(\tau', t)$ are related to Radar$_1$ and Radar$_2$, respectively. Under ideal circumstances, Radar$_1$ would only receive the signal corresponding to the wireless channel link $A_1^{T_x}$–$A_1^{R_x}$, and Radar$_2$ would only receive the signal corresponding to the link $A_2^{T_x}$–$A_2^{R_x}$. However, due to the poor interchannel RF isolation of the Ancortek radar system, the receivers of the two radars strongly interfere with each other. This problem is independent of the channel impulse response length. The system is paused between switching from Radar$_1$ to Radar$_2$, but the two subsystems, i.e., Radar$_1$ and Radar$_2$, are part of one and the same MIMO radar system having a single RF printed circuit board (PCB). This RF circuit has poor RF isolation, due to which we encounter the problems of RF-leakage and cross-channel interference. The actual measured TV-CIRs $\tilde{h}_{11}(\tau', t)$ and $\tilde{h}_{22}(\tau', t)$ incorporating the problem of cross-channel interference are

$$\tilde{h}_{11}(\tau', t) = h_{11}(\tau', t) + \alpha_{22}^{11} h_{22}(\tau', t) + \alpha_{12}^{11} h_{12}(\tau', t) + \alpha_{21}^{11} h_{21}(\tau', t) \qquad \text{(A.1)}$$

and

$$\tilde{h}_{22}(\tau', t) = h_{22}(\tau', t) + \alpha_{11}^{22} h_{11}(\tau', t) + \alpha_{12}^{22} h_{12}(\tau', t) + \alpha_{21}^{22} h_{21}(\tau', t) \qquad \text{(A.2)}$$

respectively, where α_{ij}^{kk} is the weight corresponding to the TV-CIR of the interfering link for $i, j, k \in \{1, 2\}$. The system model described by (A.1) takes into account that the measured TV-CIR $\tilde{h}_{11}(\tau', t)$ comprises the desired component $h_{11}(\tau', t)$ and the three undesired cross-channel interference components $\alpha_{22}^{11} h_{22}(\tau', t)$, $\alpha_{12}^{11} h_{12}(\tau', t)$, and $\alpha_{21}^{11} h_{21}(\tau', t)$. Equation (A.2) presents an analogous system model for the cross-channel inference impairing the actual measured TV-CIR $\tilde{h}_{22}(\tau', t)$. The weights α_{ij}^{kk} depend on the RF isolation between the subchannels of the MIMO radar system. An ideal MIMO radar system fulfills the condition $\alpha_{ij}^{kk} = 0$, implying that $\tilde{h}_{ii}(\tau', t) = h_{ii}(\tau', t)$, but in practice, we have $\alpha_{ij}^{kk} \neq 0 \ \forall \ i, j, k \in \{1, 2\}$.

To demonstrate the practical relevance of the described problem, we study the cross-channel interference of the Ancortek MIMO radar. Therefore, we measure the nonlinear trajectories of a swinging pendulum in a 2×2 MIMO radar setup. Let us consider a swinging pendulum as a physical model for a moving scatterer S^M as shown in Figure A.1. The choice of a pendulum as a moving scatterer S^M is appropriate as the trajectory of S^M can be described by a mathematical reference model as shown in Section A.6, which is important for the cross-validation of the experimental results. The two subradars are positioned on the two-dimensional orthogonal axes (x, y). This arrangement of subradars enables the overall system to capture the scatterer's motion in the horizontal plane, which is not possible with a SISO radar system. For instance, if the scatterer moves in the direction of the

boresight of $Radar_1$, then $Radar_1$ will detect the motion, while $Radar_2$ may not. Conversely, $Radar_2$ will obtain a relatively much stronger movement signature if the scatterer moves in the direction of the boresight of $Radar_2$.

The pendulum is set to swing in a direction parallel to the boresight of $Radar_1$. The pendulum's trajectories are recorded simultaneously by two subradars. Then, the recorded raw data are processed and the spectrogram is computed individually for each radar unit. Section A.4 provides the details on the computation of the spectrogram from the radar's raw data. Subsequently, the radial velocity profile is computed from the spectrogram (see Section A.4). The radial velocity profile of the measured TV-CIR $\tilde{h}_{22}(\tau', t)$ in the presence of the swinging pendulum is shown in Figure A.2a. Figure A.2b shows the motion of the pendulum in terms of the radial velocity $\dot{d}_{ij}(t)$ and range $d_{ij}(t)$. Although both subradars experience interferences, for brevity, only the measurements from $Radar_2$ are shown here in Figure A.2.

Evidently, the radial velocity profile in Figure A.2a not only contains the pendulum's trajectories from the desired wireless link $A_2^{T_x}$–$A_2^{R_x}$, but also the undesired trajectories from the interfering links $A_1^{T_x}$–$A_1^{R_x}$, $A_1^{T_x}$–$A_2^{R_x}$, and $A_2^{T_x}$–$A_1^{R_x}$. Similarly, Figure A.2b also aids unmasking the problem of interference by depicting the three separate curves corresponding to the links $A_i^{T_x}$–$A_j^{R_x}$. As expected, the radial ve-

(a)

(b)
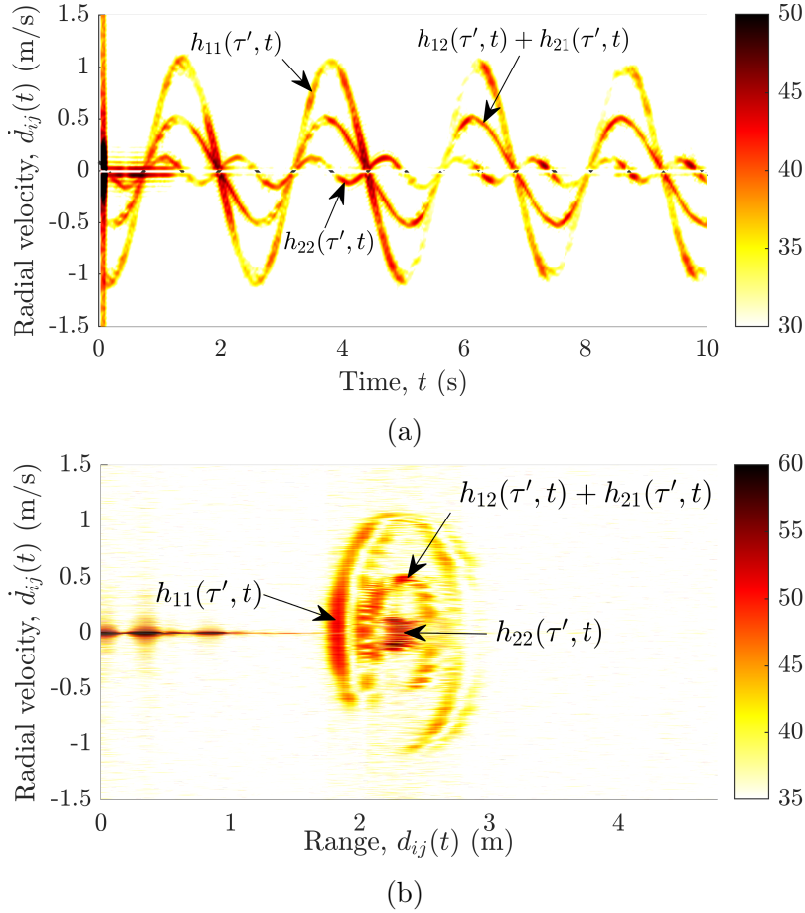
Figure A.2: Radial velocity $\dot{d}_{ij}(t)$ of the pendulum vs. (a) time t and (b) range $d_{ij}(t)$ for the measured subchannel $\tilde{h}_{22}(\tau', t)$.

locities of the pendulum in Figure A.2a,b are identical for the links $A_1^{T_x}$–$A_2^{R_x}$ and $A_2^{T_x}$–$A_1^{R_x}$. In Figure A.2a,b, the three different components of the swinging pendulum are labeled with the corresponding TV-CIRs $h_{ij}(\tau', t)$. Furthermore, we have confirmed and validated this observed phenomenon of cross-channel interference by simulating the different wireless links $A_i^{T_x}$–$A_j^{R_x}$. The geometrical 3D indoor channel model and the extended pendulum model have been presented in Sections A.3 and A.6, respectively, enabling the simulation of the wireless links $A_i^{T_x}$–$A_j^{R_x}$.

The aforementioned interferences encountered by the MIMO radar system hinder us to track the scatterer's motion. To efficiently compute the radial range and radial velocity of the scatterer at each radar, we must first eradicate the interferences shown in Figure A.2. This impels us to propound a solution to the problem of cross-channel interferences, which is presented in Section A.5. For a better understanding of the proposed solution, we first describe the underlying geometrical 3D indoor model and the radar system model in Sections A.3 and A.4, respectively.

A.3   Geometrical 3D Indoor Channel Model

In this section, we consider a 2×2 MIMO system deployed in an indoor 3D propagation scenario as depicted in Figure A.3. The transmitter antenna $A_i^{T_x}$ is placed at a fixed position $(x_i^{T_x}, y_i^{T_x}, z_i^{T_x})$ for $i = 1, 2$. Similarly, the receiver antenna $A_j^{R_x}$ is fixed at the position $(x_j^{R_x}, y_j^{R_x}, z_j^{R_x})$ for $j = 1, 2$. The RF cable of length $L_i^{T_x}$ ($L_j^{R_x}$) connects the ith transmitter (jth receiver) antenna to the SDR as illustrated in Figure A.3. The 3D propagation scenario consists of a single moving object, which is modeled as a scatterer S^M with the TV coordinates $(x(t), y(t), z(t))$ as shown in Figure A.3. In addition, the propagation environment consists of K fixed objects S_k^F ($k = 1, 2, \ldots, K$), such as walls, furniture, and decoration items. As the fixed scatterers S_k^F are of no interest, they are eliminated from the spectrogram by radar signal preprocessing techniques.

The TV trajectory $\mathcal{C}(t)$ of the moving scatterer S^M, the position $\mathcal{C}_i^{T_x}$ of the transmitter antenna $A_i^{T_x}$, and the position $\mathcal{C}_j^{R_x}$ of the receiver antenna $A_j^{R_x}$ are defined as

$$\mathcal{C}(t) = \begin{bmatrix} x(t) & y(t) & z(t) \end{bmatrix}^T \tag{A.3}$$

$$\mathcal{C}_i^{T_x} = \begin{bmatrix} x_i^{T_x} & y_i^{T_x} & z_i^{T_x} \end{bmatrix}^T \tag{A.4}$$

and

$$\mathcal{C}_j^{R_x} = \begin{bmatrix} x_j^{R_x} & y_j^{R_x} & z_j^{R_x} \end{bmatrix}^T \tag{A.5}$$

respectively. The Euclidean distance between the ith transmitter (jth receiver) antenna and the non-stationary scatterer S^M is denoted by $d_i^{T_x}(t)$ and $d_j^{R_x}(t)$, which can be expressed as

$$d_i^{T_x}(t) = \left\| \mathcal{C}(t) - \mathcal{C}_i^{T_x} \right\| \tag{A.6}$$

and

$$d_j^{R_x}(t) = \left\| \mathcal{C}(t) - \mathcal{C}_j^{R_x} \right\| \tag{A.7}$$
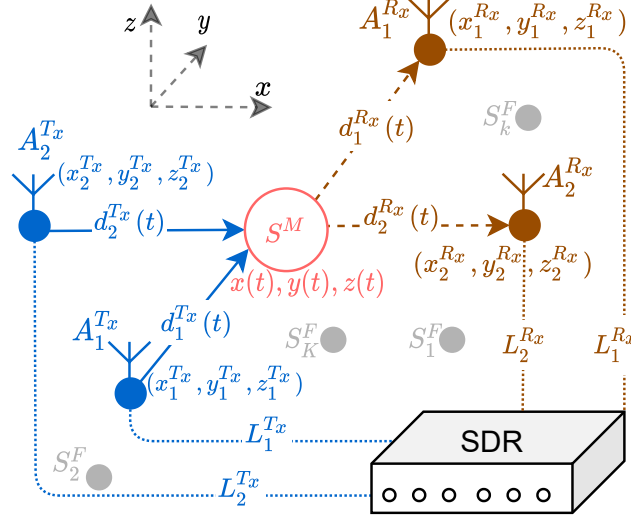
Figure A.3: Geometrical 3D model for a 2×2 MIMO system with a single moving scatterer S^M and K fixed scatterers S_k^F $(k = 1, 2, \ldots, K)$.

respectively, where $\|x\|$ denotes the Euclidean norm of x. The TV radial velocity components $\dot{d}_i^{T_x}(t)$ and $\dot{d}_j^{R_x}(t)$ can be represented as

$$\dot{d}_i^{T_x}(t) = \frac{1}{d_i^{T_x}(t)} \left[\dot{\mathcal{C}}(t) \right]^T \left[\mathcal{C}(t) - \mathcal{C}_i^{T_x} \right] \tag{A.8}$$

and

$$\dot{d}_j^{R_x}(t) = \frac{1}{d_j^{R_x}(t)} \left[\dot{\mathcal{C}}(t) \right]^T \left[\mathcal{C}(t) - \mathcal{C}_j^{R_x} \right] \tag{A.9}$$

respectively. The radar's radial range $d_{ij}(t)$ of the moving scatterer S^M is given by $1/2$ of the total propagation distance, i.e,

$$d_{ij}(t) = \frac{1}{2} \left[d_i^{T_x}(t) + d_j^{R_x}(t) + L_i^{T_x} + L_j^{R_x} \right]. \tag{A.10}$$

Finally, the composite radial velocity $\dot{d}_{ij}(t)$ can be expressed as

$$\dot{d}_{ij}(t) = \frac{1}{2} \left[\dot{d}_i^{T_x}(t) + \dot{d}_j^{R_x}(t) \right]. \tag{A.11}$$

A.4   Radar System Model

For a 2×2 MIMO TDMA FMCW radar system, the transmitter signal $s_i(t')$ is defined as

$$s_i(t') = \exp \left[j\phi_i + j2\pi \left(\frac{c_r}{2} t'^2 + f_0 t' \right) \right] \tag{A.12}$$

for $i = 1, 2$, where ϕ_i is the initial phase, c_r is the chirp rate, and f_0 is the start frequency. The chirp rate c_r is defined as $c_r = (f_1 - f_0)/T_{sw}$, where f_1 is the stop frequency, and T_{sw} is the sweep time of the periodic up-chirp signal being transmitted. In the TDMA mode, both transmitters operate in different time slots but use the same waveform as in (A.12). The time slots for the ith transmitter are defined as $(2n + i - 1)T_{sw} \leq t' < (2n + i)T_{sw}$ for $n = 0, 1, \ldots$.

Paper A

The transmitted signal $s_i(t')$ is reflected to the radar receiver antennas due to stationary and non-stationary scatterers present in the indoor environment. Therefore, each multipath component associated with the link A_i^{Tx}–A_j^{Rx} experiences a propagation delay $\tau_{ij}'^{(l)}$ for $l = 1, 2, \ldots, \mathcal{L}$, where $\mathcal{L}$ denotes the total number of scatterers, which is given by $\mathcal{L} = K + 1$. The received signal, which is modeled as a weighted sum of $\mathcal{L}$ back-scattered multipath components, is then passed through the quadrature mixer stage of the radar. At the output of the mixer, we obtain the so-called beat (also known as deramped, dechirped or intermediate frequency) signal. The beat signal $s_{b,ij}^{(l)}(t')$ corresponding to the channel link A_i^{Tx}–A_j^{Rx} in the presence of a particular scatterer $S^{(l)}$ is given as [38]

$$s_{b,ij}^{(l)}(t') = a_{ij}^{(l)} \exp\left(j2\pi f_{b,ij}^{(l)} t' + j\phi_{ij}^{(l)}\right) \tag{A.13}$$

where

$$f_{b,ij}^{(l)} = \frac{2d_{ij}^{(l)} c_r}{c_0} \tag{A.14}$$

is the beat frequency, and

$$\phi_{ij}^{(l)} = \frac{4\pi d_{ij}^{(l)}}{\lambda} \tag{A.15}$$

is the phase corresponding to the range $d_{ij}^{(l)} = c_0 \tau_{ij}'^{(l)}/2$, where c_0 is the speed of light, and λ is the radar's wavelength. The symbol $a_{ij}^{(l)}$ in (A.13) represents the net amplitude attenuation, which is related to the radar cross section of the lth scatterer, antenna gains, and transmission losses. In the presence of $\mathcal{L}$ scatterers in the radar's field of view (FOV), the composite beat signal $s_{b,ij}(t')$ is simply the sum of all beat signals, i.e.,

$$s_{b,ij}(t') = \sum_{l=1}^{\mathcal{L}} s_{b,ij}^{(l)}(t'). \tag{A.16}$$

Furthermore, note that according to the authors of [39], the complex conjugate of the composite beat signal $s_{b,ij}^*(t')$ is equal to the Fourier transform of the TV-CIR $h_{ij}(\tau', t)$, i.e.,

$$s_{b,ij}^*(t') = \mathcal{F}\{h_{ij}(\tau', t)\} \tag{A.17}$$

where $\mathcal{F}$ represents the Fourier transform. The time delay τ' in (A.17) is related to the dual value of t' denoted by f_b as $\tau' = f_b/c_r$. Due to relation (A.17) and $\mathcal{F}\{.\}$ being a linear operator, the interference components in (A.1) and (A.2) also affect the measured composite beat signal $s_{b,ij}(t')$.

The composite beat signal $s_{b,ij}(t')$ is sampled by an analog to digital converter (ADC) module with sampling frequency $F_s = 1/T_s$, where T_s is the sampling interval. Let N_s denote the number of samples taken from $s_{b,ij}(t')$ with the sampling interval T_{sw}, and let N_c denote the number of chirps within a frame of the FMCW radar. Then, for a single frame duration of $T_f = N_c \times N_s \times T_s$, the sampled beat

signal $s_{b,ij}(nT_s)$ can be arranged in a raw data matrix $\mathcal{D}_{ij}$ as

$$
\mathcal{D}_{ij} =
\begin{bmatrix}
s_{b,ij}(0) & s_{b,ij}(T_s) & \dots & s_{b,ij}(T_{sw} - T_s) \\
s_{b,ij}(T_{sw}) & s_{b,ij}(T_{sw} + T_s) & \dots & s_{b,ij}(2T_{sw} - T_s) \\
\vdots & \vdots & \vdots & \vdots \\
s_{b,ij}((N_c - 1)T_{sw}) & s_{b,ij}((N_c - 1)T_{sw} + T_s) & \dots & s_{b,ij}(N_c T_{sw} - T_s)
\end{bmatrix}
\tag{A.18}
$$

where $T_{sw} = N_s T_s$. Note that the dimension of the raw data matrix is $N_c \times N_s$. Each row of $\mathcal{D}_{ij}$ contains the fast-time data that has been sampled with the sampling interval T_s, and each column of $\mathcal{D}_{ij}$ contains the slow-time data sampled with the sampling interval T_{sw}.

The fast Fourier transform (FFT) of the fast-time data is known as the range FFT. The range FFT is applied to the rows of the raw data matrix $\mathcal{D}_{ij}$ to acquire the beat frequencies $f_{b,ij}^{(l)}$ of the composite beat signal $s_{b,ij}(t')$ (see (A.13)). Subsequently, the range maps or the range $d_{ij}^{(l)}$ for each scatterer can be computed using the relation in (A.14). As the observation interval of the range FFT is T_{sw}, the frequency resolution f_{res} of the range FFT is limited to $f_{\text{res}} = 1/T_{sw}$. Therefore, it can be shown [40] that the spectral components caused by two different moving scatterers at different ranges can be resolved in the spectrum of (A.16) provided that the scatterers are at least

$$
d_{\text{res}} = \frac{c_0}{2B}
\tag{A.19}
$$

apart in range, where d_{res} is the range resolution, and B is the bandwidth of the radar. Furthermore, from the Nyquist criterion, it can be shown [41] that the radar's maximum unambiguous range is $d_{\max} = F_s c_0 / 2 c_r$.

Let us define $\Delta d_{ij}^{(l)}$, $\Delta \tau_{ij}'^{(l)}$, $\Delta \phi_{ij}^{(l)}$, and $\Delta f_{b,ij}^{(l)}$ as the net change in $d_{ij}^{(l)}$, $\tau_{ij}'^{(l)}$, $\phi_{ij}^{(l)}$, and $f_{b,ij}^{(l)}$, respectively, over the period of one sweep interval T_{sw}. Note that a moving scatterer is fixed over an observation window T_{sw}, because $\Delta d_{ij}^{(l)} \ll d_{\text{res}}$. Therefore, a small change in the displacement $\Delta d_{ij}^{(l)}$ results in a small change in the frequency of the beat signal, denoted by $\Delta f_{b,ij}^{(l)}$. This frequency change $\Delta f_{b,ij}^{(l)}$ is not discernible in the spectrum of (A.16) because $\Delta f_{b,ij}^{(l)} < f_{\text{res}}$. In order to capture $\Delta d_{ij}^{(l)}$, we need to observe the phase of the beat signal $\phi_{ij}^{(l)}$ over multiple sweep intervals T_{sw}. The phase of the beat signal is very sensitive and changes significantly from sweep to sweep even for slight displacements of the scatterer. In analogy to (A.15), the relation of the phase change $\Delta \phi_{ij}^{(l)}$ and the displacement Δd_{ij} is given as

$$
\Delta \phi_{ij}^{(l)} = \frac{4\pi \Delta d_{ij}^{(l)}}{\lambda}.
\tag{A.20}
$$

Therefore, the phase change $\Delta \phi_{ij}^{(l)}$ of the beat signal can be observed over two sweeps to determine the radial velocity by means of

$$
v_{ij}^{(l)} = \frac{\lambda \Delta \phi_{ij}^{(l)}}{4\pi T_{sw}}.
\tag{A.21}
$$

However, two or more equidistant scatterers with different radial velocities cannot be resolved using the phase difference observed only over two chirps. To capture all the different phase changes $\Delta\phi_{ij}^{(l)}$ corresponding to the equidistant non-stationary scatterers, the Doppler FFT is applied to the columns of the radar range maps to obtain the micro-Doppler frequencies $f_{d,ij}^{(l)}(t)$. From the micro-Doppler frequencies $f_{d,ij}^{(l)}(t)$, the radial velocities $v_{ij}^{(l)}(t)$ can be computed as

$$v_{ij}^{(l)}(t) = \frac{f_{d,ij}^{(l)}(t)c_0}{2f_0}. \tag{A.22}$$

Furthermore, the radar velocity resolution is given as $v_{\text{res}} = \lambda/2T_f$. The maximum unambiguous radial velocity can be derived as $v_{\text{max}} = \lambda/4T_{sw}$.

The components of the radar signal processing of the raw data matrix $\mathcal{D}_{ij}$ are delineated here. First, the Hanning window function

$$w_H(t') = \begin{cases} \frac{1}{2}\left[1 - \cos\left(\frac{2\pi t'}{T_{sw}}\right)\right], & 0 \leq t' \leq T_{sw} \\ 0, & \text{otherwise} \end{cases} \tag{A.23}$$

is applied to the fast-time data of the frame, where the window length is equal to the chirp duration T_{sw}. Then, the range maps are computed by applying the range FFT to the windowed data. To acquire the range evolution of the scatterers over time, the slow-time data can be agglomerated to obtain the processing gain.

After the application of the range FFT, the slow-time data are split into many overlapping or consecutive disjoint segments. Then, for each segment and each range-bin, the short-time Doppler FFT is computed to obtain the local micro-Doppler information of the scatterers. A further processing gain can be achieved by agglomerating the range maps. In other words, for a particular range, the slow-time non-stationary data are composed of the TV micro-Doppler frequencies of the scatterers, which can be obtained by the spectrogram defined as [42]

$$S_{ij}(f,t) = \left| \int_{-\infty}^{\infty} x_{ij}(t'',t)e^{-j2\pi ft''}dt'' \right|^2 \tag{A.24}$$

where

$$x_{ij}(t'',t) = s_{b,ij}(t'')w_R(t''-t) \tag{A.25}$$

in which t is the local time, and t'' represents the running time. In (A.25), $w_R(t'')$ denotes a window function, which is in our case a rectangular function defined as

$$w_R(t'') = \begin{cases} 1, & 0 \leq t'' < N_c T_{sw} \\ 0, & \text{otherwise.} \end{cases} \tag{A.26}$$

Finally, from the spectrogram $S_{ij}(f,t)$, we can compute the TV mean Doppler shift as

$$B_{ij}^{(1)}(t) = \frac{\int_{-\infty}^{\infty} f S_{ij}(f,t)df}{\int_{-\infty}^{\infty} S_{ij}(f,t)df}. \tag{A.27}$$

The measured mean Doppler shift $B_{ij}^{(1)}(t)$ will be compared with the mean Doppler shift of the analytical model in Section A.6 for the cross-validation of the experimental results and the analytical results.

A.5 Proposed Solution

In this section, we propose a solution to mitigate the problem of the cross-channel interferences described in Section A.2. The proposed approach is to induce a controlled propagation delay in one of the subchannels, so that the desired channel links $A_1^{T_x}$–$A_1^{R_x}$ and $A_2^{T_x}$–$A_2^{R_x}$ can be separated in the range domain of the MIMO radar. To this end, we can use an RF delay line component as a tool for increasing the propagation delay in one of the subradars of the 2×2 MIMO radar system shown in Figure A.4a. More conveniently, a pair of RF cables with different lengths can be used instead of the RF delay line component to induce a fixed propagation delay in the channel of interest as shown in Figure A.4b. As illustrated in Figure A.3, a cable of length $L_i^{T_x}$ connects the SDR to the ith transmitter antenna $A_i^{T_x}$, and a cable of length $L_j^{R_x}$ connects the SDR to the jth receiver antenna $A_j^{R_x}$.

For each subradar, the cables of the same length are used for the transmitter and the receiver antennas, i.e., $L_i^{T_x} = L_j^{R_x}$ for $i = j$. To obtain a virtual propagation delay in the link $A_2^{T_x}$–$A_2^{R_x}$, we choose the cable lengths $L_2^{T_x}$ and $L_2^{R_x}$ depending on the dimensions of the indoor environment or the desired coverage area of the MIMO radar system. We deploy connector cables with lengths $L_2^{T_x}$ and $L_2^{R_x}$ according to

(a)

(b)
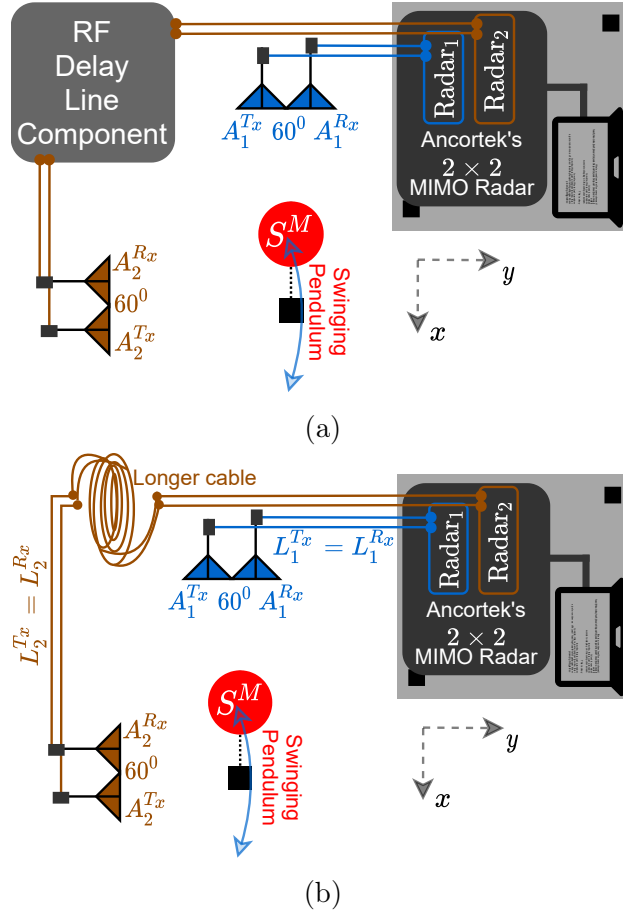
Figure A.4: Setup to induce a fixed propagation delay by either using (a) an RF delay line component or (b) different cable lengths, i.e., $(L_1^{T_x}, L_1^{R_x}) \neq (L_2^{T_x}, L_2^{R_x})$.

the relations

$$L_2^{T_x} \geq 2L_a + L_1^{T_x} \tag{A.28}$$

and

$$L_2^{R_x} \geq 2L_a + L_1^{R_x} \tag{A.29}$$

respectively, where L_a represents the length of the area of interest, which is essentially the square area covered by the MIMO radar system. Using (A.28) and (A.29), the channel links $A_i^{T_x}$–$A_j^{R_x}$ are guaranteed to be separable for the scatterers in the square area $A_{sq} = L_a \cdot L_a$. Therefore, the radar range $d_{ij}(t)$ in (A.10) is controlled using a longer pair of cables for the link $A_2^{T_x}$–$A_2^{R_x}$. Then, the radial ranges of the channel links $A_i^{T_x}$–$A_j^{R_x}$ follow the inequality $d_{11}(t) < d_{12}(t) < d_{22}(t)$. Furthermore, the links $A_1^{T_x}$–$A_2^{R_x}$ and $A_2^{T_x}$–$A_1^{R_x}$ have identical radial distances, i.e., $d_{12}(t) = d_{21}(t)$.

Finally, an additional range gating module is implemented after the range FFT module in the radar signal processing chain described in Section A.4. The range profile of the MIMO radar system (obtained by the range FFT module) is partitioned by the range gating module to acquire $d_{11}(t)$, $d_{22}(t)$, and $d_{12}(t)$. In other words, the range gating module segregates the independent trajectories of the scatterers for each channel link $A_i^{T_x}$–$A_j^{R_x}$. Subsequently, each channel link can now be further processed without the problem of cross-channel interferences. The results of the proposed approach are presented in the subsequent section.

Note that the proposed approach can also be adopted to completely avoid the use of the TDMA scheme. The TDMA scheme limits the PRF of the MIMO radar system, which in turn limits the system's maximum measurable unambiguous radial velocity $v_{\max}$. The PRF and the maximum radial velocity $v_{\max}$ decrease by the same factor as the number of subradars of the MIMO system increases. On the other hand, the proposed approach allows multiple RF delay lines to be used for different channel links $A_i^{T_x}$–$A_j^{R_x}$ so that all the subradars can operate simultaneously without effecting the PRF and $v_{\max}$ of the MIMO radar system. For instance, for an $N \times N$ MIMO radar system, the cable difference for different channel links $A_i^{T_x}$–$A_j^{R_x}$ must follow the inequality $\min[L_i^{T_x/R_x} - L_j^{T_x/R_x}] \geq 2L_a$ for $i \neq j$, where $i, j \in \{1, 2, \ldots, N\}$.

A.6 Experimental Results

In this section, we elaborate our measurement campaign carried out using an FMCW-based MIMO radar system (Ancortek SDR-KIT 2400T2R4) operating in the K-band. The detailed analytical model for a swinging pendulum is laid out in this section for the validation of the experimental results. The efficacy of the proposed solution against the interferences of Ancortek's MIMO radar system is also highlighted by the measurement results.

The measurements were carried out in a semi-controlled environment, a laboratory with the dimensions of 11.5 m × 6 m. The laboratory was equipped with many stationary objects such as chairs, tables, boards, and computers. The pendulum bob weighing 3 kg was suspended from the ceiling of the laboratory by means of a rope of length L. The pendulum bob acted as a single non-stationary scatterer

($\mathcal{L} = K + 1$) initially resting at the coordinates $(0, 0, 1.07)$ m. The Ancortek radar was placed inside the laboratory and configured as a 2×2 MIMO radar system in FMCW mode. The transmitter antennas $A_1^{T_x}$ and $A_2^{T_x}$, and the receiver antennas $A_1^{R_x}$ and $A_2^{R_x}$ were positioned in a monostatic configuration according to Table A.1. The length of the RF cables, $L_i^{T_x}$ and $L_j^{R_x}$, the maximum displacement $x_{\max}$ and the length L of the pendulum, and the MIMO radar operating parameters f_c, BW, T_{sw}, and PRF were fixed according to the values listed in Table A.1. The two subradars of the MIMO system were configured to share the time according to the TDMA scheme, but even so, the Ancortek system experienced cross-channel interference as stated in Section A.2. Needless to say, due to the TDMA mode of operation, the PRF of the subradars was reduced to half, i.e., PRF $= 1/2T_{sw}$, as listed in Table A.1.

We now present the analytical model for the pendulum swinging in xz-plane, so that we are able to cross-validate the experimental results with the analytical results. The pendulum is displaced by $x_{\max}$ to set it in a swinging motion. The TV nonlinear trajectories of the pendulum can be obtained as [43]

$$x(t) = L \sin \left\{ \arcsin \left(\frac{x_{\max}}{L} \right) \cos \left(\sqrt{\frac{g}{L}} t \right) \right\} \tag{A.30}$$

$$y(t) = 0 \tag{A.31}$$

$$z(t) = L \left[1 - \cos \left\{ \arcsin \left(\frac{x(t)}{L} \right) \right\} \right] \tag{A.32}$$

where g represents the gravitational field strength. The above model for the pendulum's trajectories is valid for an ideal pendulum, which swings only in the xz-plane. The model can readily be used for a pendulum swinging in the yz-plane by interchanging the right-hand side of the expressions in (A.30) and (A.31). To analytically determine the radial range of the scatterer, the pendulum model expressed by (A.30)–(A.32) can be used with (A.10) of the geometrical 3D indoor channel model

Table A.1: MIMO experimental setup.

Description	Parameters	Values
$A_1^{T_x}$ position	$(x_1^{T_x}, y_1^{T_x}, z_1^{T_x})$	$(1.56, 0.01, 1.195)$ m
$A_1^{R_x}$ position	$(x_1^{R_x}, y_1^{R_x}, z_1^{R_x})$	$(1.56, -0.01, 1.185)$ m
$A_2^{T_x}$ position	$(x_2^{T_x}, y_2^{T_x}, z_2^{T_x})$	$(-0.01, 1.56, 1.195)$ m
$A_2^{R_x}$ position	$(x_2^{R_x}, y_2^{R_x}, z_2^{R_x})$	$(1.0, 1.56, 1.185)$ m
RF cable lengths	$(L_1^{T_x}, L_1^{R_x}, L_2^{T_x}, L_2^{R_x})$	$(0.3, 0.3, 3.5, 3.5)$ m
Length of pendulum	L	1.48 m
Max. displacement	$x_{\max}$	0.4 m
Carrier frequency	f_c	25 GHz
Radar's bandwidth	BW	2 GHz
Sweep time	T_{sw}	1 ms
Pulse repetition freq.	PRF	500 Hz

Paper A

introduced in Section A.3. On the other hand, to obtain the radial velocity using
(A.11), we must first derive the expressions for $\dot{x}(t)$, $\dot{y}(t)$, and $\dot{z}(t)$, which results in

$$\dot{x}(t) = -\sqrt{Lg}\cos(\phi')\arcsin\left(\frac{x_{\max}}{L}\right)\sin\left(\sqrt{\frac{g}{L}}t\right) \tag{A.33}$$

$$\dot{y}(t) = 0 \tag{A.34}$$

and

$$\dot{z}(t) = \frac{x(t)\dot{x}(t)}{\sqrt{L^2 - x^2(t)}}, \quad |x(t)| \leq L \tag{A.35}$$

respectively, where $\phi' = \arcsin(x_{\max}/L)\cos(\sqrt{g/L}\cdot t)$. By making use of the extended pendulum model (A.30)–(A.35) combined with the geometrical 3D indoor channel model, we can compute analytically the TV radial range components $d_{ij}(t)$ and the radial velocity components $\dot{d}_{ij}(t)$ for all wireless channel links $A_i^{T_x}$–$A_j^{R_x}$ shown in Figure A.1.

For the experimental setup from Table A.1, the measured radial range profile is shown in Figure A.5a and the measured radial velocity profile is plotted against the measured range in Figure A.5b. The two subradars capture and process the
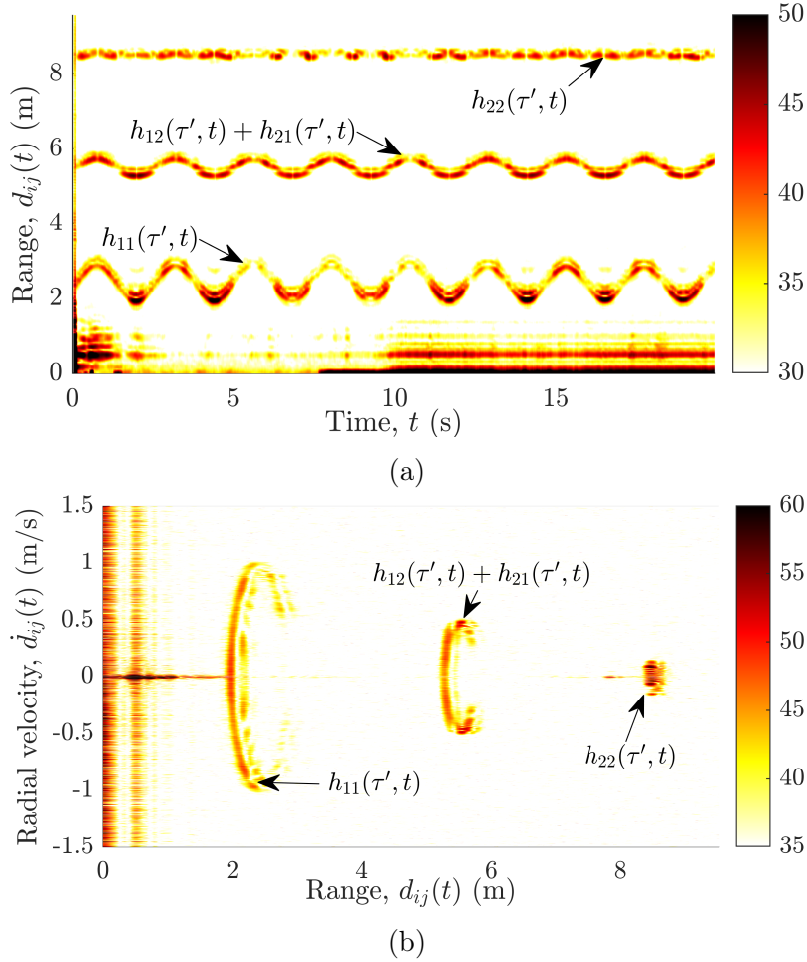
Figure A.5: Different cable lengths, i.e., $(L_1^{T_x}, L_1^{R_x}) \neq (L_2^{T_x}, L_2^{R_x})$, result in the segregation of (a) measured range profiles and (b) measured range–velocity profiles.

nonlinear trajectories of the pendulum by means of the radar signal preprocessing described in Section A.4. We obtain the processing gain in the radial range profile by agglomerating the slow-time data, whereas the radial velocity profile is acquired by integrating over the range maps. The radial range profile is obtained from the measured beat frequency profile by using (A.14). On the other hand, the radial velocity profile is mapped from the measured micro-Doppler frequency profile by utilizing the relation in (A.22). The two subradars adopt the proposed solution (see Section A.5) for the mitigation of the cross-channel interferences encountered by the Ancortek MIMO radar system. Figure A.5a,b illustrates the effect of different cable lengths on the measured range profile for a pendulum swinging in the xz-plane. Due to the deployment of different cables, three distinct curves can be observed in Figure A.5a,b that can be segregated by means of the range gating module (see Section A.5). While the pendulum swings in the xz-plane, the radial range $d_{11}(t)$ in Figure A.5a changes to a much greater extent than the radial ranges $d_{22}(t)$ and $d_{12}(t)$ or $d_{21}(t)$. A similar inference can be drawn regarding the radial velocities $\dot{d}_{ij}(t)$ in Figure A.5b.

After the application of the proposed interference mitigation approach, we obtain the distinct radial velocity components $\dot{d}_{11}(t)$, $\dot{d}_{12}(t)$ or $\dot{d}_{21}(t)$, and $\dot{d}_{22}(t)$ as illustrated in Figure A.6a–c, respectively, where $\dot{d}_{12}(t) = \dot{d}_{21}(t)$. The MIMO radar system captures the pendulum trajectories in the x-axis and y-axis, which signifies the importance of the deployment of multiple RF sensors in an indoor environment. Figure A.6a,c depicts the radial velocities corresponding to Radar$_1$ and Radar$_2$, respectively, whereas Figure A.6b shows the radial velocities corresponding to the channel link $A_1^{T_x}$–$A_2^{R_x}$ or $A_2^{T_x}$–$A_1^{R_x}$. The pendulum is swinging in the xz-plane (parallel to the boresight of Radar$_1$), consequently, one can observe that the radial velocity is much higher in Figure A.6a compared to Figure A.6c. Furthermore, as anticipated, the number of crests and troughs in the radial velocity profile of Radar$_2$ is twice as high. Note that the radial velocities $\dot{d}_{11}(t)$ and $\dot{d}_{22}(t)$ captured by Radar$_1$ and Radar$_2$, respectively, are independent and unique, which cannot be achieved with a SISO system. Moreover, the measured radial velocities are validated by the analytical model that comprises the geometrical 3D indoor model for the distributed MIMO system (see Section A.3) and the extended pendulum model described by (A.30)–(A.35). A good match between the measurements and the analytical model is shown in Figure A.6, which confirms the validity of the geometrical 3D indoor model and the extended pendulum model. The efficacy of the proposed approach against the interferences can be apprehended by comparing Figure A.6 with Figure A.2a. Evidently, the proposed approach eliminates the cross-channel interferences altogether by separating the measured trajectories for each radar of the MIMO system. Therefore, although the radial velocity components in Figure A.6 are identical to the radial velocity components of Figure A.2a, they are without any interferences.

Figures A.7–A.9 show the reference curves for the nonlinear trajectories of the pendulum, which are used to cross-validate the measurement results obtained for all subchannel links $A_i^{T_x}$–$A_j^{R_x}$ of the 2×2 MIMO system. Figures A.7–A.9 illustrate

(a)

(b)

(c)
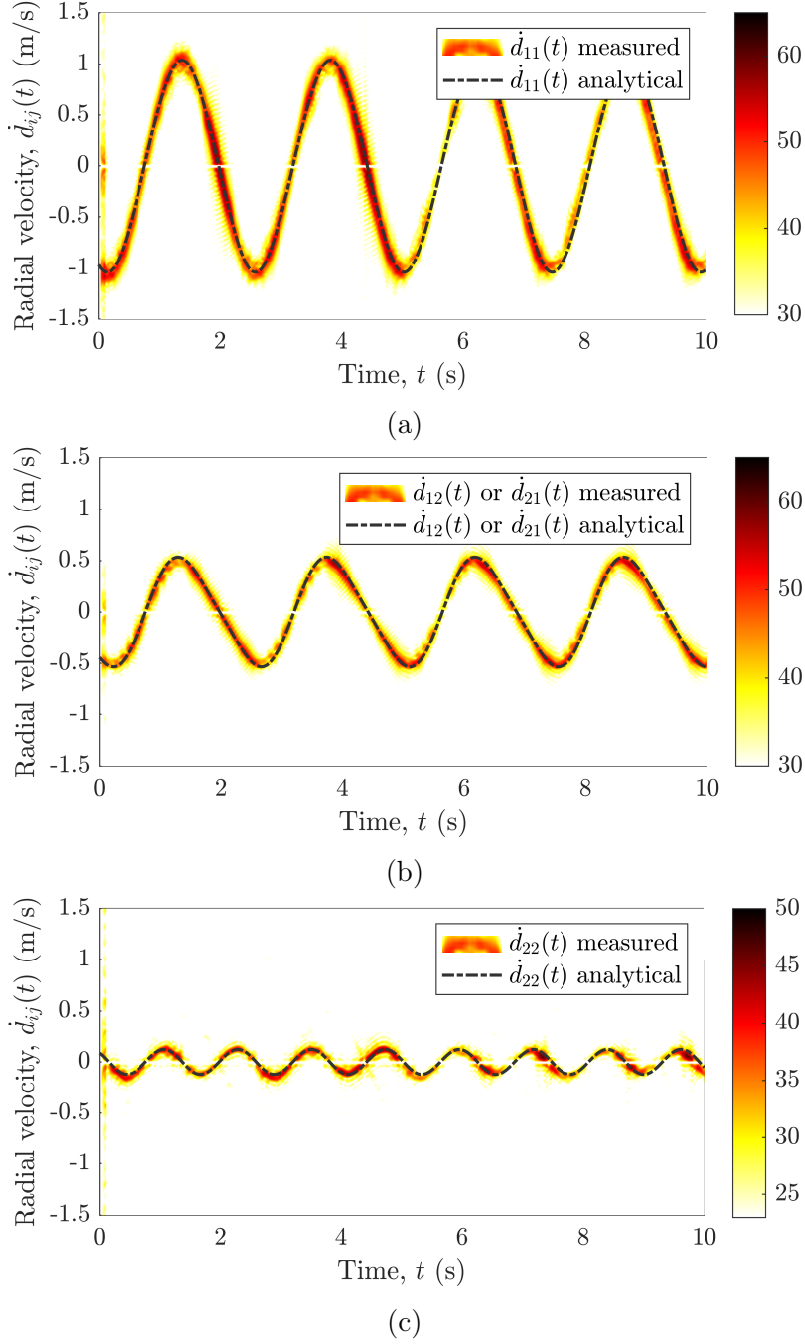
Figure A.6: Application of the proposed interference mitigation scheme results in segregated measured radial velocity components $\dot{d}_{ij}(t)$ for the channel links: (a) $A_1^{T_x}$–$A_1^{R_x}$ (Radar$_1$), (b) $A_1^{T_x}$–$A_2^{R_x}$ (or $A_2^{T_x}$–$A_1^{R_x}$), and (c) $A_2^{T_x}$–$A_2^{R_x}$ (Radar$_2$).

the trajectories of the pendulum swinging in the xz-plane (parallel to the boresight of Radar$_1$) within the FOV of the two subradars. Figure A.7 illustrates the analytical radial velocity components $\dot{d}_{ij}(t)$ that do not depend on the deployment of longer cables. Figure A.8a,b shows the scenario when the two subradars of the MIMO system use the same cable lengths, i.e., $(L_1^{T_x}, L_1^{R_x}) = (L_2^{T_x}, L_2^{R_x})$, whereas Figures A.9a,b shows the case when the two subradars use different cable lengths, i.e., $(L_1^{T_x}, L_1^{R_x}) \neq (L_2^{T_x}, L_2^{R_x})$. Figure A.9, analogous to Figure A.5, shows the effect
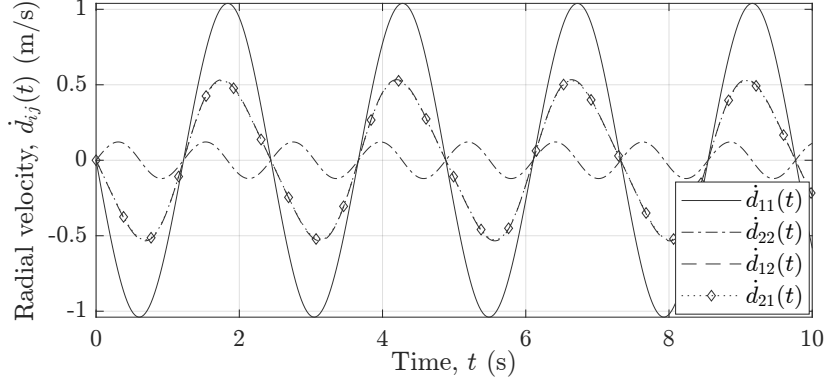
Figure A.7: Analytical radial velocity components $\dot{d}_{ij}(t)$ for the channel links $A_1^{T_x}$–$A_1^{R_x}$, $A_1^{T_x}$–$A_2^{R_x}$, $A_2^{T_x}$–$A_1^{R_x}$, and $A_2^{T_x}$–$A_2^{R_x}$.

(a)

(b)
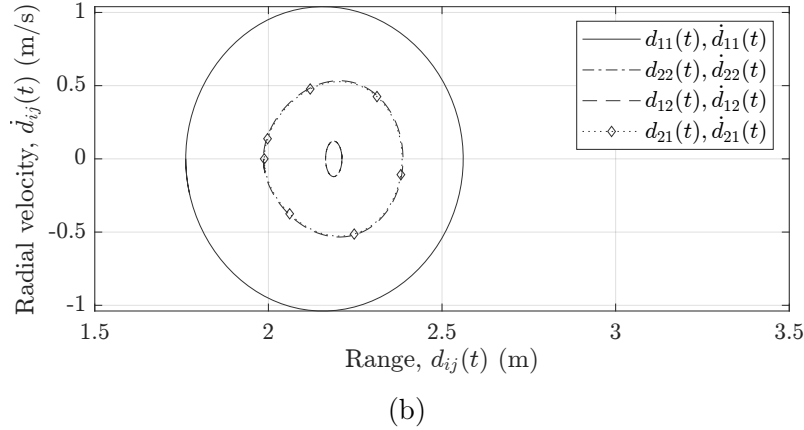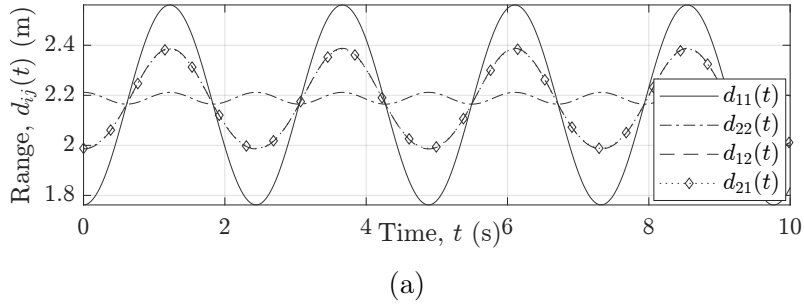
Figure A.8: (a) The analytical radial range components $d_{ij}(t)$ and (b) the analytical radial velocity components $\dot{d}_{ij}(t)$ for $(L_1^{T_x}, L_1^{R_x}) = (L_2^{T_x}, L_2^{R_x})$.

of longer cable lengths $L_2^{T_x}$ and $L_2^{R_x}$ on the radial ranges $d_{ij}(t)$.

The relation in (A.27) is utilized to obtain the measured mean Doppler shift $B_{ij}^{(1)}(t)$ for all channel links $A_i^{T_x}$–$A_j^{R_x}$ in a 2×2 MIMO system. Analogous to the computation of the mean Doppler shift, the mean radial range is obtained from the range profile. The analytical and measured mean Doppler shifts $B_{ij}^{(1)}(t)$ are illustrated in Figure A.10a. Figure A.10b shows the analytical and measured mean Doppler shifts plotted against the range of the moving scatterer S^M. Clearly, a considerable mismatch exists between the analytical and measured mean Doppler shifts due to the interferences.

On the other hand, using the proposed approach, we obtain the segregated non-
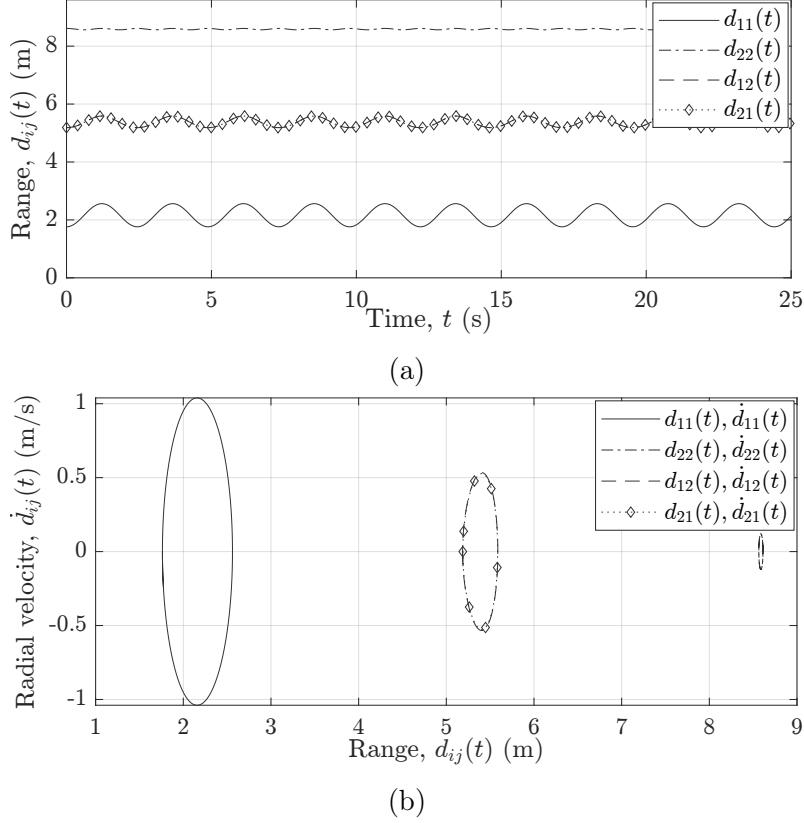
Figure A.9: (a) The analytical radial range components $d_{ij}(t)$ and (b) the analytical radial velocity components $\dot{d}_{ij}(t)$ for $(L_1^{T_x}, L_1^{R_x}) \neq (L_2^{T_x}, L_2^{R_x})$.

linear trajectories of the pendulum as shown in Figure A.11. Figure A.11a illustrates the mean Doppler shift of the pendulum swinging in the xz-plane over a period of 10 seconds. A good match between the measured and the analytical mean Doppler shifts is observed for all channel links $A_i^{T_x} - A_j^{R_x}$. Figure A.11b shows the mean Doppler shift plotted against the mean radial range. Due to the fine Doppler resolution of the FMCW radar, the measured Doppler information matches very well with the analytical results in Figure A.11, whereas an adequate match exists between the analytical and measured range due to an adequate range resolution of the system.

A.7 Conclusions

In this paper, we proposed a unique approach to the problem of cross-channel interferences encountered by the Ancortek SDR-KIT 2400T2R4 MIMO radar system due to its poor interchannel RF isolation. For all subchannels of the MIMO radar system, we observed a significant mismatch between the measured and analytical TV mean Doppler shift due to the problem of cross-channel interference. However, after the application of the proposed interference mitigation method, we found an excellent fit between the measured and analytical TV mean Doppler shift. The proposed approach is optimal and robust in a way that it completely eliminates the cross-channel interferences. The proposed solution works for the Ancortek MIMO

(a)

(b)
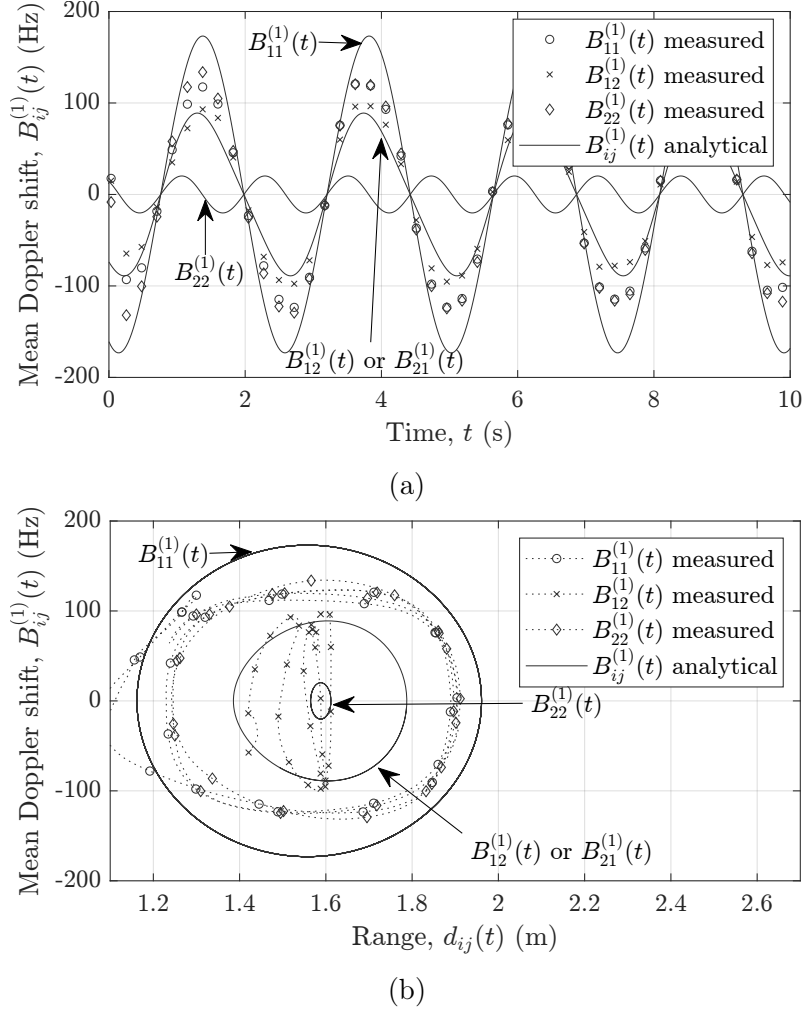
Figure A.10: (a) The measured mean Doppler shift $B_{ij}^{(1)}(t)$ vs. time and (b) the measured mean Doppler shift $B_{ij}^{(1)}(t)$ vs. range $d_{ij}(t)$, where the MIMO radar undergoes cross-channel interferences.

radar system without the need to alter its firmware or hardware. We also presented a channel model to investigate the target's motion in a MIMO system under different target–antenna configurations. A good agreement was found between the geometrical 3D indoor channel model and the measured data. In the proposed solution, the segregation and utilization of the cross-channel component generally lead to an added diversity and improved system capability. Although the proposed approach may find its utility in numerous application areas, we plan to extend this work to orientation-independent human activity recognition. For human activity recognition, we plan to fuse the data from different subchannels of the MIMO radar system to increase the overall classification performance of the system.
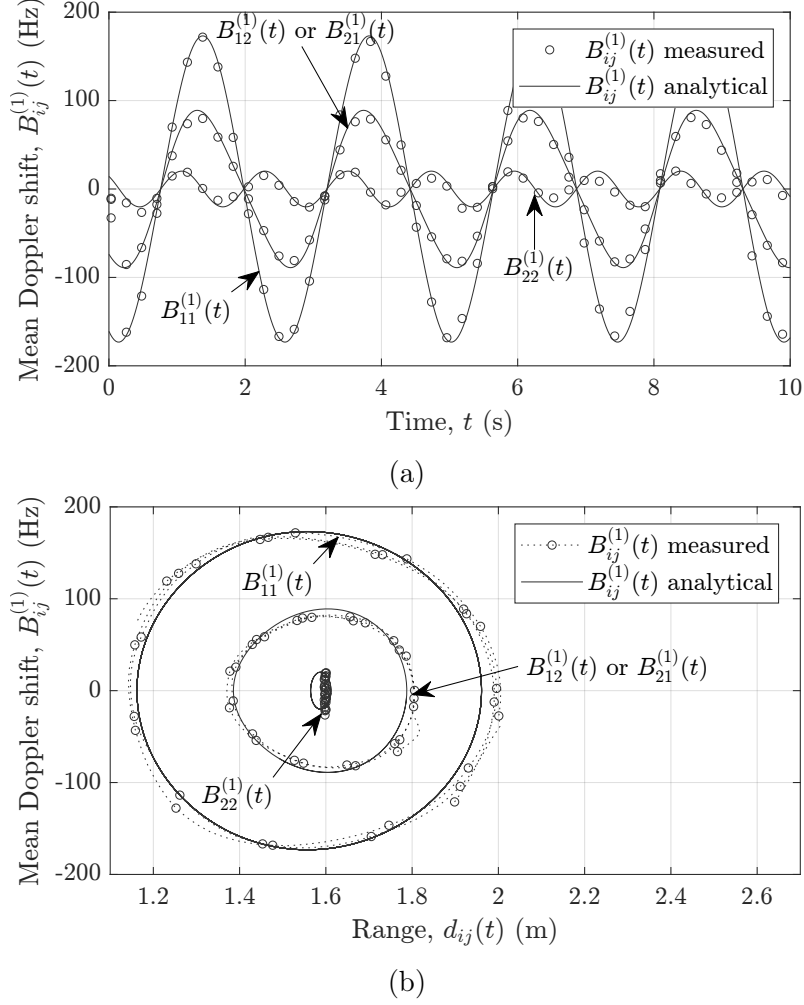
Figure A.11: (a) The measured mean Doppler shift $B_{ij}^{(1)}(t)$ vs. time and (b) the measured mean Doppler shift $B_{ij}^{(1)}(t)$ vs. range $d_{ij}(t)$, where the MIMO radar adopts the proposed interference mitigation scheme.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors may provide the presented measurement data upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

Bibliography

[1] Xiaodong Lu and T. Koga, "DAPS based adaptive tracking system for high-assurance air traffic surveillance," in *2014 Integr. Commun. Navig. Surveill. Conf. Conf. Proc.*, pp. O4–1–O4–8, 2014.

[2] P. J. Gómez-del Hoyo, N. Del-Rey-Maestre, D. Mata-Moya, M. P. Jarabo-Amores, and M. C. Benito-Ortiz, "Coherent detection and 3D tracking stages of a DVB-T based passive radar for terrestrial traffic monitoring," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 524, no. 1, p. 012002, 2019.

[3] D.-A. Nguyen, B. Cho, C. Seo, J. Park, and D.-H. Lee, "Analysis of the optimal frequency band for a ballistic missile defense radar system," *J. Electromagn. Eng. Sci.*, vol. 18, no. 4, pp. 231–241, 2018.

[4] F. Bandini, T. P. Sunding, J. Linde, O. Smith, I. K. Jensen, C. J. Köppl, M. Butts, and P. Bauer-Gottwein, "Unmanned Aerial System (UAS) observations of water surface elevation in a small stream: Comparison of radar altimetry, LIDAR and photogrammetry techniques," *Remote Sens. Environ.*, vol. 237, p. 111487, 2020.

[5] S. Watts, "The ASV 21 maritime surveillance radar," in *2017 IEEE Radar Conf.*, pp. 0027–0032, 2017.

[6] V. Louf, A. Protat, R. A. Warren, S. M. Collis, D. B. Wolff, S. Raunyiar, C. Jakob, and W. A. Petersen, "An integrated approach to weather radar calibration and monitoring using ground clutter and satellite comparisons," *J. Atmos. Ocean. Technol.*, vol. 36, no. 1, pp. 17–39, 2019.

[7] J.-L. Margot, "A Data-Taking System for Planetary Radar Applications," *J. Astron. Instrum.*, vol. 10, no. 01, 2021.

[8] L. Kranold, M. Taherzadeh, F. Nabki, M. Coates, and M. Popovic, "Microwave breast screening prototype: System miniaturization with IC pulse radio," *IEEE J. Electromagn. RF Microwaves Med. Biol.*, vol. 5, no. 2, pp. 168–178, 2021.

[9] Z. Liu, Y. Cai, H. Wang, L. Chen, H. Gao, Y. Jia, and Y. Li, "Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–14, 2021.

[10] C. Waldschmidt, J. Hasch, and W. Menzel, "Automotive Radar — From First Efforts to Future Systems," *IEEE J. Microwaves*, vol. 1, no. 1, pp. 135–148, 2021.

[11] F. Uysal, "Phase-coded FMCW automotive radar: System design and interference mitigation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 270–281, 2020.

[12] A. Frigeri and M. Ercoli, "The ScanMars subsurface radar sounding experiment on AMADEE-18," *Astrobiology*, vol. 20, no. 11, pp. 1338–1352, 2020.

[13] H. Du, T. Jin, Y. He, Y. Song, and Y. Dai, "Segmented convolutional gated recurrent neural networks for human activity recognition in ultra-wideband radar," *Neurocomputing*, vol. 396, pp. 451–464, 2020.

[14] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, 2019.

[15] F. Luo, S. Poslad, and E. Bodanese, "Human activity detection and coarse localization outdoors using micro-Doppler signatures," *IEEE Sens. J.*, vol. 19, no. 18, pp. 8079–8094, 2019.

[16] B. Jokanovic and M. Amin, "Fall detection using deep learning in range-Doppler radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, 2018.

[17] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli : Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.

[18] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A Review," *Sensors*, vol. 21, no. 6, p. 2140, 2021.

[19] N. T. P. Van, L. Tang, A. Singh, N. D. Minh, S. C. Mukhopadhyay, and S. F. Hasan, "Self-identification respiratory disorder based on continuous wave radar sensor system," *IEEE Access*, vol. 7, pp. 40019–40026, 2019.

[20] L. Piotrowsky, T. Jaeschke, S. Kueppers, J. Siska, and N. Pohl, "Enabling high accuracy distance measurements with FMCW radar sensors," *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 12, pp. 5360–5371, 2019.

[21] D. Lee, G. Shaker, and W. Melek, "Imaging of human walking behind the obstacle utilizing pulsed radar technique in the C-band for military surveillance applications," *J. Electr. Eng. Technol.*, vol. 15, no. 3, pp. 1431–1439, 2020.

[22] A. Abuduaini, N. Shiraki, N. Honma, T. Nakayama, and S. Iizuka, "Performance evaluation of multiple human-body localization using bistatic MIMO radar," in *2019 IEEE Asia-Pacific Microw. Conf.*, pp. 575–577, 2019.

[23] S. Waqar, H. Yusaf, S. Sana, M. Waqas, and F. A. Siddiqui, "Reconfigurable monopulse radar tracking processor," in *2018 15th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 805–809, 2018.

[24] F. Bovenga, "Special issue "synthetic aperture radar (SAR) techniques and applications"," *Sensors*, vol. 20, no. 7, 2020.

[25] C. Feng, X. Jiang, M.-G. Jeong, H. Hong, C.-H. Fu, X. Yang, E. Wang, X. Zhu, and X. Liu, "Multitarget vital signs measurement with chest motion imaging based on MIMO radar," *IEEE Trans. Microw. Theory Tech.*, pp. 1–1, 2021.

[26] S. Sana, S. Waqar, H. Yusaf, M. Waqas, and F. A. Siddiqui, "Software defined digital beam forming processor," in *2016 13th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 671–676, 2016.

[27] M. Sadeghi, F. Behnia, R. Amiri, and A. Farina, "Target localization geometry gain in distributed MIMO radar," *IEEE Trans. Signal Process.*, vol. 69, pp. 1642–1652, 2021.

[28] H. Zhang, W. Liu, Z. Zhang, W. Lu, and J. Xie, "Joint target assignment and power allocation in multiple distributed MIMO Radar networks," *IEEE Syst. J.*, vol. 15, no. 1, pp. 694–704, 2021.

[29] P. Wang and H. Li, "Target detection with imperfect waveform separation in distributed MIMO radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 793–807, 2020.

[30] H. Lee, B.-H. Kim, J.-K. Park, and J.-G. Yook, "A novel vital-sign sensing algorithm for multiple subjects based on 24-GHz FMCW Doppler radar," *Remote Sens.*, vol. 11, no. 10, 2019.

[31] N. Malešević, V. Petrović, M. Belić, C. Antfolk, V. Mihajlović, and M. Janković, "Contactless real-time heartbeat detection via 24 GHz continuous-wave Doppler radar using artificial neural networks," *Sensors*, vol. 20, no. 8, 2020.

[32] M. M. Rahman, R. Mdrafi, A. C. Gurbuz, E. Malaia, C. Crawford, D. Griffin, and S. Z. Gurbuz, "Word-level sign language recognition using linguistic adaptation of 77 GHz FMCW radar data," in *2021 IEEE Radar Conf.*, pp. 1–6, 2021.

[33] F. Jin, A. Sengupta, S. Cao, and Y.-J. Wu, "MmWave radar point cloud segmentation using GMM in multimodal traffic monitoring," in *2020 IEEE Int. Radar Conf.*, pp. 732–737, 2020.

[34] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "Mm-pose: Real-time human skeletal posture estimation using mmWave radars and CNNs," *IEEE Sens. J.*, vol. 20, no. 17, pp. 10032–10044, 2020.

[35] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 669–673, 2018.

[36] A. Ledergerber and R. D'Andrea, "A multi-static radar network with ultra-wideband radio-equipped devices," *Sensors*, vol. 20, no. 6, 2020.

[37] N. Avazov, R. Hicheri, M. Muaaz, F. Sanfilippo, and M. Pätzold, "A trajectory-driven 3D non-stationary mm-wave MIMO channel model for a single moving point scatterer," *IEEE Access*, vol. 9, pp. 115990–116001, 2021.

[38] A. Ahmad, J. C. Roh, D. Wang, and A. Dubey, "Vital signs monitoring of multiple people using a FMCW millimeter-wave sensor," in *2018 IEEE Radar Conf.*, pp. 1450–1455, 2018.

[39] R. Hicheri, N. Avazov, M. Muaaz, and M. Patzold, "The transfer function of non-stationary indoor channels and its relationship to system functions of LFMCW radars," in *2021 IEEE 22nd Int. Work. Signal Process. Adv. Wirel. Commun.*, pp. 151–155, 2021.

[40] J. W. Ting, D. Oloumi, and K. Rambabu, "FMCW SAR system for near-distance imaging applications - practical considerations and calibrations," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 450–461, 2018.

[41] A. Meta, P. Hoogeboom, and L. P. Ligthart, "Signal processing for FMCW SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3519–3532, 2007.

[42] N. Avazov, R. Hicheri, and M. Pätzold, "A trajectory-driven SIMO mm-Wave channel model for a moving point scatterer," in *2021 15th Eur. Conf. Antennas Propag.*, pp. 1–5, 2021.

[43] A. Abdelgawwad, A. Borhani, and M. Pätzold, "Modelling, analysis, and simulation of the micro-Doppler effect in wideband indoor channels with confirmation through pendulum experiments," *Sensors*, vol. 20, no. 4, 2020.

Appendix B

Paper B

Title:	Human Activity Signatures Captured Under Different Directions Using SISO and MIMO Radar Systems
Authors:	Sahil Waqar, Muhammad Muaaz, and Matthias Pätzold
Affiliation:	Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway
Journal:	Applied Sciences
DOI:	10.3390/app12041825.

Human Activity Signatures Captured Under Different Directions Using SISO and MIMO Radar Systems

Sahil Waqar, Muhammad Muaaz, and Matthias Pätzold

Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway

E-mails: {sahil.waqar, muhammad.muaaz, matthias.paetzold}@uia.no

Abstract— In this paper, we highlight and resolve the shortcomings of single-input single-output (SISO) millimeter wave (mm-Wave) radar systems for human activity recognition (HAR). A 2×2 distributed multiple-input multiple-output (MIMO) radar framework is presented to capture human activity signatures under realistic conditions in indoor environments. We propose to distribute the two pairs of collocated transmitter–receiver antennas in order to illuminate the indoor environment from different perspectives. For the proposed MIMO system, we measure the time-variant (TV) radial velocity distribution and TV mean radial velocity to observe the signatures of human activities. We deploy the Ancortek SDR-KIT 2400T2R4 mm-Wave radar in a SISO as well as a 2×2 distributed MIMO configuration. We corroborate the limitations of SISO configurations by recording real human activities in different directions. It is shown that, unlike the SISO radar configuration, the proposed MIMO configuration has the ability to obtain superior human activity signatures for all directions. To signify the importance of the proposed 2×2 MIMO radar system, we compared the performance of a SISO radar-based passive step counter with a distributed MIMO radar-based passive step counter. As the proposed 2×2 MIMO radar system is able to detect human activity in all directions, it fills a research gap of radio frequency (RF)-based HAR systems.

Keywords— direction-independent human activity recognition; fall detection; distributed MIMO; FMCW radar; micro-Doppler signatures; aspect angle; multistatic radar systems; passive step counter; DTW; velocity estimation.

B.1 Introduction

B.1.1 General Background

In recent years, the number of application areas of human activity recognition (HAR) has greatly increased, such as remote health assessment [1], smart home [2], smart surveillance [3], human–computer interaction [4], sports [5], autopilots [6], and social robotics [7]. Radio frequency (RF)-based in-home sensing is still considered a developing technology facing some key challenges regarding HAR. However, the attractive features of RF-based HAR systems have brought them to the forefront of indoor HAR systems. Alternatively, vision-based and wearable sensor-based HAR systems have already produced numerous and adequate results. However, unlike RF sensors, visual sensors such as cameras or light detection and ranging (LiDAR) may suffer from issues such as privacy invasion, sensitivity to lighting, and obstructive illumination. On the other hand, wearable sensors such as accelerometers, magnetometers, gyroscopes, and emergency push buttons are radically intrusive, fragile, must be carried by the user, and are prone to user negligence.

RF sensors such as Wi-Fi and radar systems must have robustness to environmental variations, lighting conditions, user's privacy, and nonobstructive illumination. Over the years, two established technologies, Wi-Fi and radar, have been explored in RF sensing for HAR [8, 9, 10, 11]. Unfortunately, commercial Wi-Fi devices suffer from carrier frequency offsets due to hardware limitations and environmental variations [12]. As a consequence of carrier frequency offsets, the phases of the channel frequency response are particularly noisy in commercial Wi-Fi devices and thus hard to utilize [8, 10, 11, 13]. In contrast, with the application of radar signal processing [11, 14, 15, 16], data fusion techniques [17, 18], machine learning, and deep learning algorithms, it will be possible in the coming years to track and classify multiple human activities by means of radar systems in unprecedented complex settings. The work presented in this paper is a step forward in that direction, where we have addressed the problem of the direction-independent recognition of human activities and proposed an effective solution in the context of RF sensing. It should be mentioned that radars have traditionally been deployed by official or governmental entities in application areas such as weather [19], naval [20] and aerial surveillance [21], air defense [22], ground traffic control [23], altimeters [24], geology [25], and astronomical research [26]. However, due to their miniaturization and cost effectiveness, the current radar systems have found utilization in self-driving cars [27, 28, 29], emerging medical solutions [30], and HAR systems [31, 32, 33, 34].

B.1.2 Related Work

One crucial challenge for radar-based HAR systems is the direction of motion of certain activity in relation to the illuminating radar. For instance, a person may fall in a direction either parallel or perpendicular to the radar's boresight. Conceivably, a monostatic single-input single-output (SISO) radar will not be able to detect the fall if the fall direction is perpendicular to the boresight of the radar. This

is due to the fact that the radar systems are merely sensitive to the changes in the scatterer's radial distance with respect to the radar itself. A scatterer moving perpendicular to the radar's boresight has zero Doppler frequency and thus appears as a stationary object to the radar and can therefore not be distinguished from other stationary objects in an indoor environment by the radar system. Generally, a strong degradation of the classification performance is expected for activities with a greater angle of motion to the radar's boresight. The prior state-of-the-art approaches to alleviate this problem are delineated here along with their limitations.

The authors in [35, 36] suggest that a SISO radar positioned on the ceiling can detect fall activities. However, this is not a general solution for direction-independent HAR. For direction-independent HAR, it is understandably compelling to use mono-static beamforming multiple-input multiple-output (MIMO) radars that are capable of measuring the angular information [37]. However, these systems are often limited by their angular resolution, which in turn limits their cross-range resolution significantly. Such monostatic MIMO radars do not perform adequately, especially not for persons relatively far away from the radar, which degrades the overall classification accuracy. The authors in [38] use the phase information of frequency-modulated continuous wave (FMCW) monopulse radars to measure the angle of arrival, but the study is limited to hand gesture sensing. The authors in [39] combine the FMCW mode of radars with interferometry to track vital signs and detect position and life activities, but they do not generally address the challenges caused by the direction of human activities. To improve the angular resolution, direction of arrival algorithms such as the "*estimation of signal parameters via rotational invariance techniques*" (ESPRIT) method and "*multiple signal classification*" (MUSIC) method can also be used [40], but these direction of arrival algorithms usually require a high signal to noise ratio [41]. Although the authors of [18] discuss a multistatic Doppler radar, the study is limited to the detection of armed persons. A bi-static radar configuration is used in [42] to improve the activity classification performance. However, the quality of data can be improved by using a MIMO radar system instead. The performance of another radar configuration is explored in [43] using a MIMO radar system in bi-static configuration, but the study is limited to personnel localization only.

B.1.3 Contributions

The problems faced by the aforementioned SISO and MIMO radar systems motivate us to develop a solution for a direction-independent HAR system. To illuminate the indoor environment from different perspectives, we propose to distribute multiple pairs of collocated transmitter–receiver radar antennas in an indoor environment for a direction-independent human activity detection system (see Section B.4). This multi-perspective view of a distributed MIMO radar system will allow us to render direction-independence for HAR. The micro-Doppler signatures or radial velocity distribution (see Section B.2) obtained using the proposed approach can help us to design HAR systems capable of classifying complex activities. Thus, the proposed

framework is a step forward towards a more pragmatic and sophisticated radar-based HAR system.

The principal contributions of this work are listed as follows:

1. A basic multi-perspective 2×2 MIMO radar system is presented which can be easily scaled to a higher number of transmitter and receiver antennas for better performance.

2. We analyze the time-variant (TV) radial velocity distribution and TV mean radial velocity for the proposed distributed antenna configuration.

3. For the 2×2 MIMO radar configuration, we investigate the impact of a human falling incident and a walking activity on the measured radial velocity distribution and measured mean radial velocity.

4. We analyze the impact of two different activities performed in three different directions on the measured channel characteristics for a SISO and a 2×2 MIMO radar system. We corroborate the limitations of the SISO radar system by analyzing the real radar data.

5. We demonstrate the robustness of the proposed 2×2 MIMO radar system against different directions of the actual human walking and falling activity. We show that the proposed solution is able to detect the human gross motor activity in the horizontal xy-plane.

6. We analyze the performance of a radar-based passive step counter by integrating it with a SISO radar system. It is shown that the radar-based passive step counter, when used with a SISO radar system, may miss some human walking steps or detect false steps depending on the walking direction.

7. It is shown that by deploying the radar-based passive step counter with the proposed 2×2 MIMO radar framework, the step counter would accurately detect the number of steps for all considered human walking directions.

8. Finally, we quantify, compare, and numerically assess the performance of the SISO and 2×2 MIMO radar systems by using the dynamic time warping (DTW) [44] distance metric.

B.1.4   Paper Organization

The paper is organized as follows. The system model and preprocessing techniques are delineated in Section B.2. Section B.3 describes the problems encountered with SISO radar systems for detecting human activities performed in different directions. Section B.4 shows how the proposed 2×2 MIMO radar system overcomes the shortcomings of the SISO radar system. The experimental results for actual human activities are detailed in Sections B.3 and B.4 for the SISO and 2×2 MIMO radar systems, respectively. Finally, Section B.6 draws the conclusions.

B.2   Radar Signal Preprocessing

In this paper, we have adopted an FMCW 2×2 MIMO radar to capture the micro-Doppler signatures of a moving person from different perspectives. The fundamental waveform transmitted by the ith transmitter antenna $A_i^{T_x}$ $(i = 1, 2)$ of the FMCW radar is the chirp waveform [45]

$$c_i(t') = \exp\left[j2\pi\left(f_0 t' + \frac{\gamma}{2}t'^2\right) + j\phi_i\right], \qquad 0 \leq t' < T_{sw} \tag{B.1}$$

where t' denotes the fast time, f_0 is the initial RF frequency, γ is the slope of the linear chirp in the time-frequency domain, ϕ_i is the initial phase, and T_{sw} is the chirp interval. The two transmitters of the 2×2 MIMO radar operate in a time division multiple access (TDMA) mode. For $n = 0, 1, \ldots$, the time windows occupied by the ith transmitter are $(2n + i - 1)T_{sw} \leq t' < (2n + i)T_{sw}$. The chirp waveform is transmitted periodically by each transmitter in their respective time slot. The overall transmit signal $s_i(t, t')$ can be represented as a sum of shifted versions of the fundamental waveform $c_i(t')$ according to

$$s_i(t, t') = \sum_{n=0}^{\infty} c_i(t')\delta(t - t_{n,i}) \tag{B.2}$$

for $i = 1, 2$, where $\delta(\cdot)$ is a Dirac delta function, t is the slow time, and $t_{n,i}$ is the discrete slow time. For TDMA mode, the discrete slow time $t_{n,i}$ is related to chirp interval T_{sw} by $t_{n,i} = (2n + i - 1)T_{sw}$. The expression (B.2) allows us to represent the transmit signal $s_i(t, t')$ as function of two separate time variables.

The wireless channel link between the ith transmitter antenna $A_i^{T_x}$ and jth receiver antenna $A_j^{R_x}$ is denoted by $A_i^{T_x}-A_j^{R_x}$, where $i, j \in \{1, 2\}$. When modeling the human body as a cluster of $\mathcal{L}$ scatterers [46], the beat signal $s_{b,ij}^{(l)}(t, t')$ corresponding to the lth scatterer and the channel link $A_i^{T_x}-A_j^{R_x}$ can be represented by

$$s_{b,ij}^{(l)}(t, t') = \sum_{n=0}^{\infty} a_{ij}^{(l)} \exp\left[j\left(2\pi f_{b,ij}^{(l)} t' + \phi_{ij}^{(l)}\right)\right]\delta(t - t_{n,i}) \tag{B.3}$$

for $l = 1, 2, \ldots, \mathcal{L}$. The symbol $a_{ij}^{(l)}$ stands for the path gain, which is primarily determined by the path loss and the radar cross section. For simplicity, we assume that the path gain $a_{ij}^{(l)}$ is constant within the observation interval. For the lth scatterer, the beat frequency $f_{b,ij}^{(l)}$ and phase $\phi_{ij}^{(l)}$ in (B.3) are given as [47]

$$f_{b,ij}^{(l)} = \frac{2d_{ij}^{(l)}\gamma}{c_0} \tag{B.4}$$

and

$$\phi_{ij}^{(l)} = \frac{4\pi d_{ij}^{(l)}}{\lambda} \tag{B.5}$$

respectively, where $d_{ij}^{(l)}$ is the total propagation distance, which is given by

$$d_{ij}^{(l)} = \frac{1}{2}\left[d_{l,i}^{T_x} + d_{l,j}^{R_x} + L_i^{T_x} + L_j^{R_x}\right]. \tag{B.6}$$

The symbol c_0 represents the speed of light in vacuum and λ is the wavelength. The quantity $d_{l,i}^{T_x}$ in (B.6) is the distance between the transmitter antenna $A_i^{T_x}$ and the lth scatterer. Similarly, the distance between the receiver antenna $A_j^{R_x}$ and the lth scatterer is represented by $d_{l,j}^{R_x}$. The lengths of the RF cables are denoted as $L_i^{T_x}$ and $L_j^{R_x}$ in (B.6) for the transmitter antenna $A_i^{T_x}$ and receiver antenna $A_j^{R_x}$, respectively. In this paper, the transmitter antenna $A_i^{T_x}$ and receiver antenna $A_j^{R_x}$ are collocated for $i = j$. As a consequence, the distances from the transmitter antenna $A_i^{T_x}$ and receiver antenna $A_j^{R_x}$ to the lth scatterer become identical for $i = j$; i.e., $d_{l,i}^{T_x} = d_{l,j}^{R_x}$ ($i = 1, 2$). Moreover, the RF cable lengths of the transmitter antenna $A_i^{T_x}$ and receiver antenna $A_j^{R_x}$ are the same for $i = j$, i.e., $L_i^{T_x} = L_i^{R_x}$ ($i = 1, 2$). For the lth scatterer and the aforementioned antenna placement constraints, the total propagation distance $d_{ij}^{(l)}$ in (B.6) reduces to the sum of the radar radial range $r_{ij}^{(l)}$ and the RF cable length L_i; i.e.,

$$d_{ij}^{(l)} = r_{ij}^{(l)} + L_i \tag{B.7}$$

where the radar radial range $r_{ij}^{(l)}$ is the Euclidean distance between the lth scatterer and the transmitter antenna $A_i^{T_x}$ or the receiver antenna $A_j^{R_x}$, and $L_i = L_i^{T_x} = L_j^{R_x}$ for $i = j$. Note that we distribute multiple pairs of collocated transmitter–receiver antennas; therefore, the transmitter antenna $A_i^{T_x}$ and receiver antenna $A_j^{R_x}$ are not collocated for $i \neq j$. Consequently, to compute the total propagation distance $d_{ij}^{(l)}$, the expression in (B.6) must be used instead of (B.7) for $i \neq j$. With reference to [48], the composite beat signal $s_{b,ij}(t, t')$ is the sum of all beat signals $s_{b,ij}^{(l)}(t, t')$ associated with the cluster of $\mathcal{L}$ scatterers, which can be expressed as

$$s_{b,ij}(t, t') = \sum_{l=1}^{\mathcal{L}} s_{b,ij}^{(l)}(t, t'). \tag{B.8}$$

In an FMCW radar, the composite beat signal $s_{b,ij}(t, t')$ is produced by the quadrature mixer component of the radar. The analog to digital converter (ADC) digitizes the composite beat signal $s_{b,ij}(t, t')$ with a sampling rate of F_s with respect to fast time t'. For each chirp interval T_{sw}, the digitized data are stacked in a raw data matrix $\mathcal{D}_{ij}$. The rows and columns of the raw data matrix $\mathcal{D}_{ij}$ contain samples of the composite beat signal $s_{b,ij}(t, t')$ in the fast-time and slow-time domain, respectively. The slow-time sampling interval is actually equal to the chirp interval T_{sw}.

The fast Fourier transform (FFT) performed on the raw data matrix $\mathcal{D}_{ij}$ with respect to fast time t' and slow time t is known as the range FFT and Doppler FFT, respectively. For the channel link $A_i^{T_x}$–$A_j^{R_x}$, the expression for the range FFT, also known as the beat frequency profile $S_{b,ij}(f_b, t)$, is given as

$$S_{b,ij}(f_b, t) = \int\limits_0^{T_{sw}} s_{b,ij}(t, t') e^{-j2\pi f_b t'} dt' \tag{B.9}$$

where f_b represents the beat frequency. The short-time Fourier transform (STFT) is related to the slow-time domain t and applies on the function resulting from

the range FFT in (B.9). In other words, the data from the beat frequency profile $S_{b,ij}(f_b, t)$ are multiplied by a rectangular window function $W_r(\cdot)$ sliding in slow time t to provide overlapping segments for the FFT operation; i.e.,

$$X_{ij}(f_b, f, t) = \int_{-\infty}^{\infty} S_{b,ij}(f_b, t'') W_r(t'' - t) e^{-j2\pi f t''} dt'' \tag{B.10}$$

where f is the Doppler frequency and t'' denotes the running time.

Finally, the TV micro-Doppler signature $S_{ij}(f, t)$ is obtained by integrating $X_{ij}(f_b, f, t)$ over the beat frequencies f_b from zero to the maximum beat frequency $f_{b,\max}$ and computing the absolute value to the power of 2; i.e., [49]

$$S_{ij}(f, t) = \left| \int_{0}^{f_{b,\max}} X_{ij}(f_b, f, t) df_b \right|^2. \tag{B.11}$$

Note that, according to the Nyquist sampling theorem, the maximum beat frequency $f_{b,\max}$ is equal to $1/2$ of the ADC sampling rate F_s; i.e., $f_{b,\max} = F_s/2$.

The TV mean Doppler shift $B_{ij}^{(1)}(t)$ can be obtained from the micro-Doppler signature $S_{ij}(f, t)$ according to the relation [50]

$$B_{ij}^{(1)}(t) = \frac{\int_{-\infty}^{\infty} f S_{ij}(f, t) df}{\int_{-\infty}^{\infty} S_{ij}(f, t) df}. \tag{B.12}$$

As the Doppler frequency f can be mapped on the radial velocity v according to $v = c_0 f / (2 f_0)$, we can obtain the TV radial velocity profile $V_{ij}(v, t)$ from the TV micro-Doppler signature $S_{ij}(f, t)$ as

$$V_{ij}(v, t) = S_{ij}\left(\frac{2 f_0}{c_0} v, t\right). \tag{B.13}$$

The TV micro-Doppler signature $S_{ij}(f, t)$ in (B.11) is computed from the composite beat signal $s_{b,ij}(t, t')$, and thus it contains the micro-Doppler information of all $\mathcal{L}$ scatterers with their respective strengths. Analogously, the radial velocity profile $V_{ij}(v, t)$ in (B.13) contains the radial velocity information of each scatterer with respect to the slow time t. The strength or power of the lth scatterer in the radial velocity profile $V_{ij}(v, t)$ depends on the path gain $a_{ij}^{(l)}$.

From the TV radial velocity profile $V_{ij}(v, t)$, we can compute the TV radial velocity distribution $p_{ij}(v, t)$ by

$$p_{ij}(v, t) = \frac{V_{ij}(v, t)}{\int_{-\infty}^{\infty} V_{ij}(v, t) dv} \tag{B.14}$$

from which we can obtain the mean radial velocity $\bar{v}_{ij}(t)$ using the relation

$$\bar{v}_{ij}(t) = \int_{-\infty}^{\infty} v p_{ij}(v, t) dv. \tag{B.15}$$

As good descriptive statistics that quantitatively summarize all the main features of the TV micro-Doppler signature $S_{ij}(f,t)$ or radial velocity distribution $p_{ij}(v,t)$, we can utilize the mean Doppler shift $B_{ij}^{(1)}(t)$ or the mean radial velocity $\bar{v}_{ij}(t)$, respectively. The TV mean Doppler shift $B_{ij}^{(1)}(t)$ in (B.12) is basically a weighted arithmetic mean of the TV micro-Doppler signature $S_{ij}(f,t)$. Similarly, the mean radial velocity $\bar{v}_{ij}(t)$ in (B.15) is the weighted average of the velocity components of all $\mathcal{L}$ human body segments computed for each time instance t. The TV radial velocity distribution $p_{ij}(v,t)$ and TV mean radial velocity $\bar{v}_{ij}(t)$ will play an important role in analyzing real-world measurement data in Sections B.3 and B.4.

B.3 Human Activity Signatures Measured by Using a SISO FMCW Radar System

An FMCW radar modulates its transmit signal frequency to detect the radial range of a target. The electromagnetic signal transmitted by the FMCW radar interacts with stationary and non-stationary objects present in the radar's range, thus altering the amplitude, phase, and frequency of the transmitted signal. After applying suitable radar signal preprocessing techniques, the phase and frequency variations of the backscattered signal provide the range and micro-Doppler information of the target (see Section B.2). As part of the preprocessing, the multipath components originating from objects relatively stationary to the radar system are filtered out, thereby making the radar system merely sensitive to the radial component of the object's motion. Therefore, detecting and processing a scatterer's motion perpendicular to the radar's boresight becomes a major challenge. For a human body under observation, the radar signals are reflected off the human body segments. Generally, human body segments can be modeled as a cluster of point scatterers [46]. The TV radial range and micro-Doppler signature caused by such moving body segments can be measured from the backscattered radar signals.

In this section, we primarily focus on the impact of the direction of human activity on the TV radial velocity distribution $p_{ij}(v,t)$ of a SISO radar system, where $i,j = 1$. To highlight the limitations of the SISO radar system deployed in an indoor environment, we first describe the measurement setup. Secondly, we illustrate three different experimental scenarios in which the human activities are performed and analyzed. Thirdly, we show the radial velocity distribution $p_{11}(v,t)$ and mean radial velocity $\bar{v}_{11}(t)$ derived from measurements of a SISO radar system. Finally, the implications of different directions of human activities on the performance of an RF-based step counting algorithm are discussed for the considered SISO radar system.

B.3.1 Measurement Setup

In this paper, we used a commercially available MIMO radar system called Ancortek SDR-KIT 2400T2R4, which is an FMCW millimeter wave (mm-wave) radar operat-

ing at 24 GHz. We configured the radar in SISO mode of operation and used a single transmitter antenna $A_1^{T_x}$ and a single receiver antenna $A_1^{R_x}$. We used a monostatic configuration of the antennas to capture the micro-Doppler information of a human body moving in the xy-plane. In a monostatic configuration, the transmitter and receiver antennas of the SISO radar system are collocated. The SISO FMCW radar system was configured according to the parameters listed in Table B.1. The pulse repetition frequency (PRF) of the SISO FMCW radar system is equal to the inverse of the chirp interval T_{sw}, i.e., PRF $= 1/T_{sw}$. This is due to the fact that we do not require the TDMA mode of operation for the SISO radar system. In other words, the transmitter antenna $A_1^{T_x}$ transmits the chirp waveform $c_i(t')$ in (B.1) continuously in time.

B.3.2   Scenarios of Human Activities

The indoor environment is a laboratory cluttered with fixed items such as wooden furniture, computers, routers, and other miscellaneous laboratory tools and electronics. The direction of activities of a person relative to the SISO radar system are depicted in Figure B.1. To elucidate the direction of human motion in an experiment, we refer throughout this section to a 3×3 grid illustrated in Figure B.1, in which three different scenarios of human movement are represented by different markings. The human activities have been carried out in three different directions in order to demonstrate the limitations of the SISO radar system in the context of HAR. Scenario 1 is a trivial scenario, where a person moves towards the SISO radar system. In this case, the human motion is parallel to the radar boresight, implying that the SISO radar system does not encounter any problem capturing the micro-Doppler signatures with high accuracy. Most of the research on HAR is limited to merely Scenario 1 with the SISO radar systems in a monostatic configuration. However, the shortcomings of SISO radar configuration come to the surface if we consider a human activity that is perpendicular to the boresight of the SISO radar system. Thus, in Scenario 2, when a person moves perpendicular to the radar boresight or moves from the position (x_2, y_3) to (x_2, y_1), then the SISO radar system captures a completely different micro-Doppler signature that is suboptimal for HAR. In Scenario 3, the person moves diagonally in the 3×3 grid of Figure B.1, either from (x_3, y_3) to (x_1, y_1) or from (x_1, y_1) to (x_3, y_3). We expect to acquire adequate human activity signatures in Scenario 3, but not as good as the human activity

Table B.1: System parameters of the SISO experimental setup.

Description	Symbols	Values
RF cable lengths	$(L_1^{T_x}, L_1^{R_x})$	$(0.3, 0.3)$ m
Carrier frequency	f_c	24.125 GHz
Radar's bandwidth	BW	250 MHz
Sweep time	T_{sw}	500 µs
Pulse repetition freq.	PRF	2 kHz
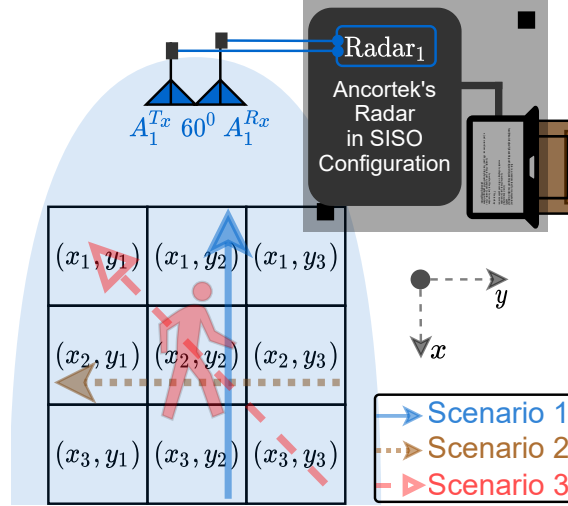
Figure B.1: A SISO radar system in the presence of a moving person in a cluttered indoor environment.

signatures of Scenario 1. For each of Scenarios 1, 2, and 3, we consider a walking and falling activity. The results of the recognized human activities by employing the monostatic SISO radar system are shown in the next subsection for Scenarios 1, 2, and 3.

B.3.3 Results for the Monostatic SISO Configuration

To show the limitations of the aforementioned monostatic SISO radar system, we recorded the human motion in three different directions, which are depicted in Figure B.1. We process the recorded raw radar data by means of radar signal pre-processing, as described in Section B.2. The radial velocity distribution $p_{11}(v, t)$ has been computed by using (B.14), and the mean radial velocity $\bar{v}_{11}(t)$ has been obtained from (B.15) for the human walking and falling activities. Figures B.2 and B.3 show the measured radial velocity distribution $p_{11}(v, t)$ over time t corresponding to a falling and walking activity, respectively. The black dashed lines in Figures B.2 and B.3 represent the measured mean radial velocity $\bar{v}_{11}(t)$. We can see from Figures B.2 and B.3 that the measured mean radial velocity $\bar{v}_{11}(t)$ provides a descriptive statistic, which quantitatively summarizes all the main features of the TV radial velocity distribution $p_{11}(v, t)$.

Recall from Figure B.1 that in Scenario 1, a person moves parallel to the SISO radar boresight—i.e., from position (x_3, y_2) to (x_1, y_2)—whereas in Scenario 2, the person moves perpendicular to the SISO radar boresight—i.e., from position (x_2, y_3) to (x_2, y_1). Figure B.2a,b shows a person falling parallel (Scenario 1) and perpendicular (Scenario 2) to the radar boresight direction, respectively. Note that in Scenario 1, the person suddenly moves closer to the SISO radar system upon falling. Thus, the falling activity produces high and abrupt positive changes of the radial velocity distribution $p_{11}(v, t)$, which leads us to suppose that the fall of a person parallel to the radar boresight will be captured perfectly by the SISO radar system,
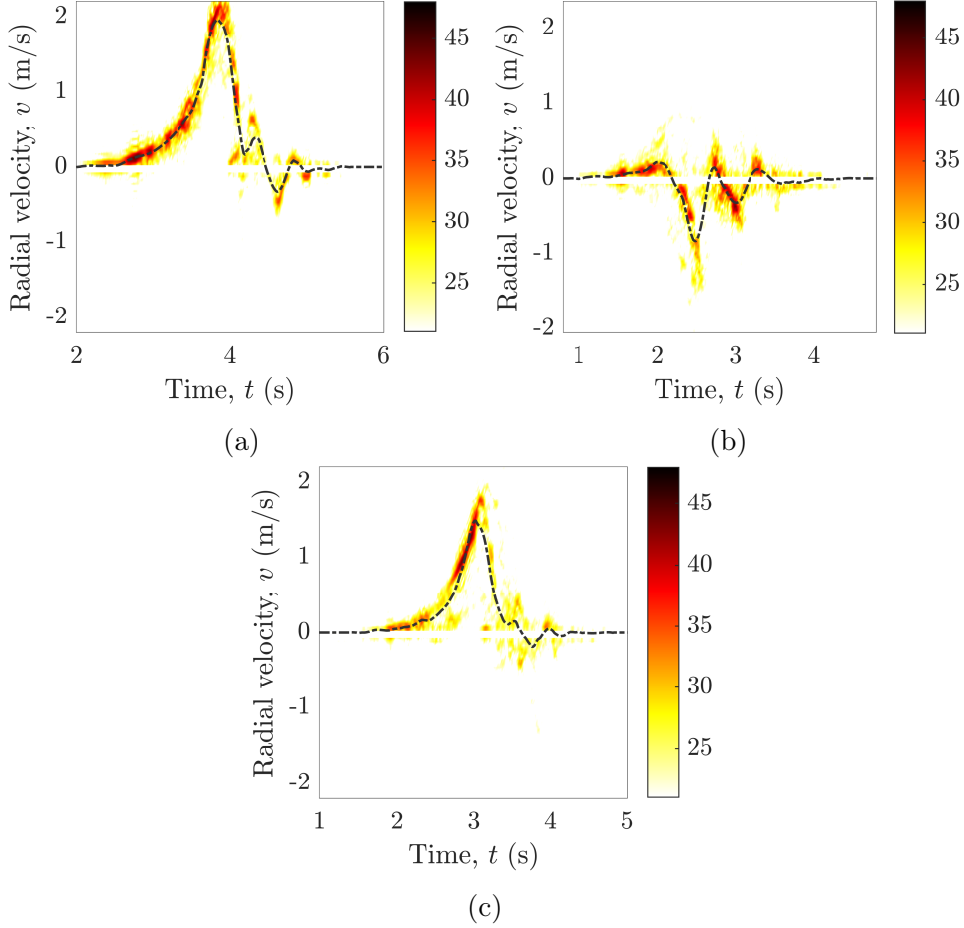
Figure B.2: For a SISO radar system, the measured radial velocity distribution $p_{11}(v,t)$ and mean radial velocity $\bar{v}_{11}(t)$ of a human falling activity for (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

as illustrated in Figure B.2a. On the other hand, a fall perpendicular to the radar boresight does not cause a high and abrupt change in the radial velocity distribution $p_{11}(v,t)$, as shown in Figure B.2b.

The same observations can be made for a person walking in the direction parallel and perpendicular to the radar boresight presented in Figure B.3a,b, respectively. Thus, the activity fingerprints captured by the SISO radar system as shown in Figures B.2 and B.3 are not sufficient to effectively classify different human activities, especially when the direction of the activity is perpendicular to the radar's boresight. Furthermore, in Scenario 3, the person moves diagonally in the 3×3 grid of Figure B.1; i.e., from (x_3, y_3) to (x_1, y_1). The SISO radar system is able to generate adequate human activity signatures in Scenario 3 as shown in Figures B.2c and B.3c for the falling and walking activities, respectively. It should be mentioned that unlike the falling activity related to Scenario 3, the initial and final positions of the walking activity are (x_1, y_1) and (x_3, y_3), respectively. As the person is walking away from the SISO radar system in Scenario 3, the radial velocity components of the TV radial velocity distribution $p_{11}(v,t)$ are negative, as can be seen in Figure B.3c.
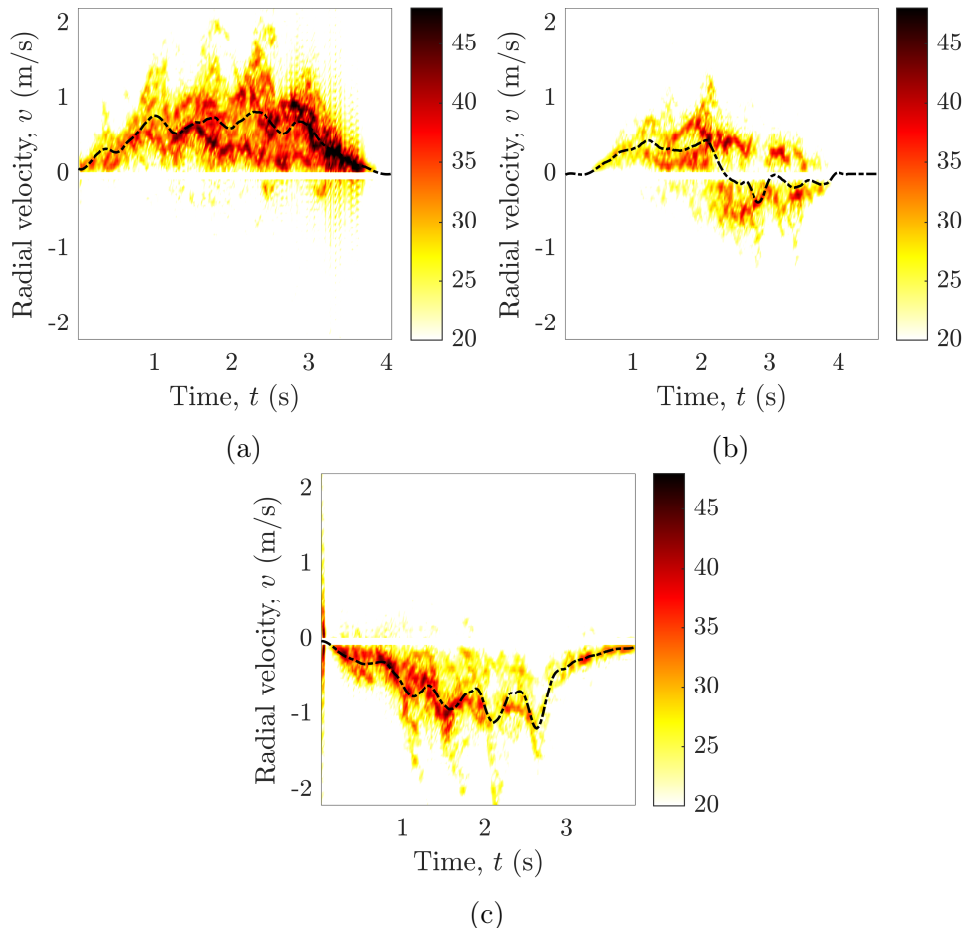
Figure B.3: For a SISO radar system, the measured radial velocity distribution $p_{11}(v, t)$ and mean radial velocity $\bar{v}_{11}(t)$ of a human walking activity for (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

B.3.4 Implications on the Performance of an RF-Based Step Counter

In this subsection, we investigate how the SISO radar system affects the performance of an RF-based step counter under three different experimental scenarios. The radar-based passive step counter [49] was developed in order to count the number of steps in an unobtrusive manner for an SISO FMCW radar system. The authors of [49] compared the performance of their radar-based passive step counter with the wearable Garmin Forerunner 935 step counter. The reported accuracy of the radar-based passive step counter was more than 98%, which was similar to the accuracy of the Garmin Forerunner 935 step counter system. However, the walking activity of a person was restricted to merely Scenario 1, where a person would walk either towards or away from the SISO radar system. This restriction was naturally enforced by the limitations of the SISO radar system in the context of activity direction. In this section, the performance of the radar-based passive step counter is shown for Scenarios 1, 2, and 3.

To detect the number of steps in a particular walk activity, the SISO radar's

raw data are processed according to the block diagram shown in Figure B.4. The Ancortek radar in a SISO configuration produces raw in-phase and quadrature (IQ) data for the wireless channel link $A_i^{T_x}$–$A_j^{R_x}$. The raw IQ data are processed by the radar signal preprocessing block (see Section B.2) to generate the radial velocity distribution $p_{11}(v, t)$ and mean radial velocity $\bar{v}_{11}(t)$. The Savitzky–Golay smoothing filter [51] has been adopted to smooth the mean radial velocity $\bar{v}_{11}(t)$ using the MATLAB command "`smooth(y,span,'sgolay',degree)`", where "y" is the input vector or mean radial velocity vector $\bar{v}_{11}(t)$, "span" is the number of data points used for smoothing, "sgolay" is the Savitzky–Golay smoothing filter, and "degree" is the polynomial degree of the Savitzky–Golay filter. In our experiments, the "span" and "degree" are chosen to be 60 and 4, respectively.

The solid black curves in Figure B.5a–c show the smoothed mean radial velocity $\bar{v}_{11}(t)$ for Scenarios 1, 2, and 3, respectively. The smoothed mean radial velocity $\bar{v}_{11}(t)$ is processed by the RF step detector to detect the number of steps and their corresponding timestamps. The detected steps of the RF step detector are distinctly marked and labeled in Figure B.5. In Scenario 1, the human subject started walking from the position (x_3, y_2) and took four steps towards the position (x_1, y_2). We can see from Figure B.5a that the RF step detection algorithm has successfully detected the four steps. However, in Scenario 2, when the human subject walked with four steps from the position (x_2, y_3) towards the position (x_2, y_1), the SISO radar system is unable to produce an intelligible walking activity signature. Consequently, the RF step detection algorithm is unable to detect all the steps that were taken by the human subject. Apparently, the RF step detector missed one of the four steps in Scenario 2 as shown in Figure B.5b. Moreover, in Scenario 3, where a person walked from the position (x_1, y_1) towards the position (x_3, y_3) taking only four steps, the SISO radar system is able to produce a fair walking activity signature. Thus, the RF step detector is able to detect the four steps successfully as depicted in Figure B.5c. Thus, we can conclude that the monostatic SISO radar system is not sufficient to capture the human micro-Doppler signatures with high precision in all directions.

It is now evident that the aforementioned limitations of the SISO radar system restrict the detection of the scatterer's motion in the horizontal xy-plane. We must observe the environment from different perspectives to effectively detect the scatterer's motion. Thus, we need a distributed RF sensing system to effectively determine the TV trajectories of the object under observation. For this reason, we propose to distribute at least two collocated transmitter–receiver antenna pairs in
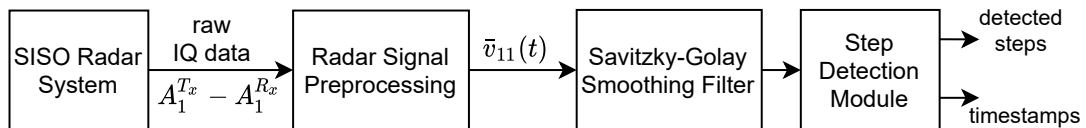
Figure B.4: The block diagram of the RF-based step counter for a SISO radar system.

an indoor environment and preprocess the data for each wireless channel link (see Sections B.2 and B.4). It should be mentioned that for SISO radar-based HAR systems, the classification accuracy will drop as the direction of the human motion relative to the radar's boresight changes from parallel to perpendicular. Furthermore, as we increase the overall system complexity in the context of human activity classification, the performance of a machine learning or deep learning classifier is expected to further degrade for a SISO radar system. For instance, the system complexity is increased by classifying more than two kinds of activities such as falling, walking, sitting on a chair, and standing from a chair. We can now safely assert that a monostatic SISO radar system is not an apposite choice for direction-independent human activity detection.
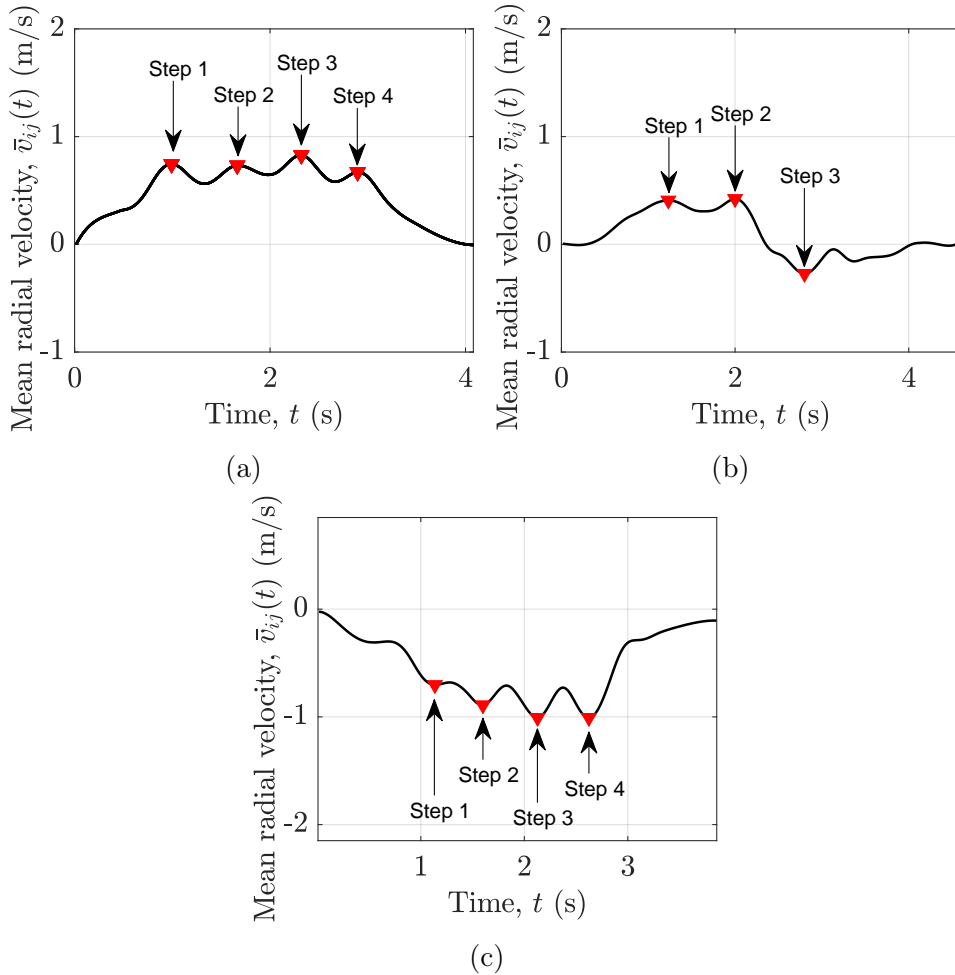
Figure B.5: For a SISO radar, the number of steps detected from the smoothed mean radial velocity $\bar{v}_{11}(t)$ of a human walking activity according to (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

B.4 Human Activity Signatures by Using a Distributed MIMO FMCW Radar System

The lack of the multi-perspective illumination of SISO radar systems is their major limitation, preventing the realization of a direction-independent HAR system (as described in Section B.3). To overcome this limitation, we propose to utilize a MIMO radar system and distribute its antennas in an indoor environment to realize a direction-independent HAR system. We have distributed two pairs of collocated transmitter–receiver antennas to illuminate the indoor environment from different perspectives. For the sake of simplicity, we have limited the number of antennas to four, but the proposed approach can be straightforwardly scaled for a larger number of antennas. A separate radar signal preprocessing chain (as described in Section B.2) has been adopted for each pair of collocated transmitter–receiver antennas. Thus, we have deployed a 2×2 MIMO radar system, which consists of two radar subsystems denoted as Radar_1 and Radar_2. Radar_1 comprises a transmitter antenna $A_1^{T_x}$ and a receiver antenna $A_1^{R_x}$. Analogously, Radar_2 consists of a transmitter antenna $A_2^{T_x}$ and a receiver antenna $A_2^{R_x}$. We propose to position the two radar subsystems such that their boresight axes are orthogonal to each other, which enables the 2×2 MIMO radar system to effectively capture the scatterer motion in the horizontal xy-plane.

In this section, we mainly emphasize the impact of the direction of human activities on the TV radial velocity distribution $p_{ii}(v, t)$ of the link from $A_i^{T_x}$ to $A_i^{R_x}$ for $i \in \{1, 2\}$. First, we discuss the measurement setup for the proposed 2×2 MIMO radar system. Second, three different experimental scenarios are illustrated in which the human activities are performed and analyzed. Third, we discuss the radial velocity distribution $p_{ii}(v, t)$ and mean radial velocity $\bar{v}_{ii}(t)$ derived from measurements of the 2×2 MIMO radar system. Finally, the implications of different directions of human activities on the performance of an RF-based step counter are discussed.

B.4.1 Measurement Setup

To realize the proposed distributed 2×2 MIMO radar configuration, we have used an FMCW mm-wave radar operating at 24 GHz to capture the micro-Doppler information of a human body moving in the xy-plane. The MIMO FMCW radar system operating in the TDMA mode was configured according to the parameters listed in Table B.2. For the Ancortek SDR-KIT 2400T2R4 radar system, we deployed RF cables with different lengths to avoid interchannel interference [47]. Owing to the TDMA mode of operation, the PRF of the 2×2 MIMO FMCW radar system is equal to $1/2$ of the inverse of the chirp interval T_{sw}; i.e., $\text{PRF} = 1/(2T_{sw})$. In TDMA mode, the transmitter antennas $A_1^{T_x}$ and $A_2^{T_x}$ of the radar subsystems Radar_1 and Radar_2, respectively, transmit the chirp waveform $c_i(t')$ alternately and periodically in their respective time slot according to (B.2).

For the proposed distributed 2×2 MIMO configuration, the indoor environment remains exactly the same as described in Section B.3. The activities of a person

Table B.2: System parameters of the 2×2 MIMO experimental setup.

Description	Symbols	Values
RF cable lengths	$(L_1^{T_x}, L_1^{R_x}, L_2^{T_x}, L_2^{R_x})$	$(0.3, 0.3, 7, 7)$ m
Carrier frequency	f_c	24.125 GHz
Radar's bandwidth	BW	250 MHz
Sweep time	T_{sw}	500 µs
Pulse repetition freq.	PRF	1 kHz

are observed in a laboratory cluttered with fixed objects such as electronics, chairs, tables, and other miscellaneous items, as shown in Figure B.6. This Figure B.6 also illustrates the actual experimental setup, where the placement of antennas shows a close resemblance to the proposed 2×2 distributed MIMO radar configuration depicted in Figure B.7. The boresights of the two radar subsystems are orthogonal to each other, which enables the 2×2 MIMO radar system to capture the radial velocity distribution $p_{ii}(v, t)$ using (B.14) regardless the direction of activity. For practical reasons, we have marked the 2×2 MIMO radar's operation region according to the field of view (FOV) of the two radar subsystems (Radar$_1$ and Radar$_2$) as shown in Figure B.6.

B.4.2 Scenarios of Human Activities

To overcome the limitations of the SISO radar system in terms of activity directions, we illuminate the environment from different perspectives via the proposed distributed 2×2 MIMO radar framework as shown in Figure B.7. Recall that Radar$_i$ comprises a transmitter antenna $A_i^{T_x}$ and a receiver antenna $A_i^{R_x}$ ($i = 1, 2$) as illustrated in Figure B.7. The two radar subsystems have a common illumination region as depicted by a 3×3 grid in Figure B.7, which depends on the FOV and the maximum unambiguous range of the radar. This common area essentially limits the region of operation, within which all activities have to be performed. A moving human body under observation is illuminated from two different angles, as shown in Figure B.7. This multi-perspective view helps us to overcome the limitations of the SISO radar system. For instance, if the subject moves in the direction parallel to the boresight of Radar$_2$, then the Doppler frequencies measured with Radar$_2$ change considerably more over a larger range than the Doppler frequencies measured with Radar$_1$. In this case, Radar$_2$ will detect the motion of the subject more effectively than Radar$_1$. However, if the subject moves parallel to the Radar$_1$ boresight, then Radar$_1$ will obtain a much better micro-Doppler signature. The two radar subsystems in Figure B.7 complement each other in the way that when the direction of motion changes from the x-axis to y-axis, the movement signature of the subject gradually disappears from the Radar$_1$ radial velocity distribution $p_{11}(v, t)$ and appears in the Radar$_2$ radial velocity distribution $p_{22}(v, t)$.

To show the effectiveness of the proposed 2×2 MIMO scheme, we recorded the human motion in three different directions. In order to illustrate the direction of
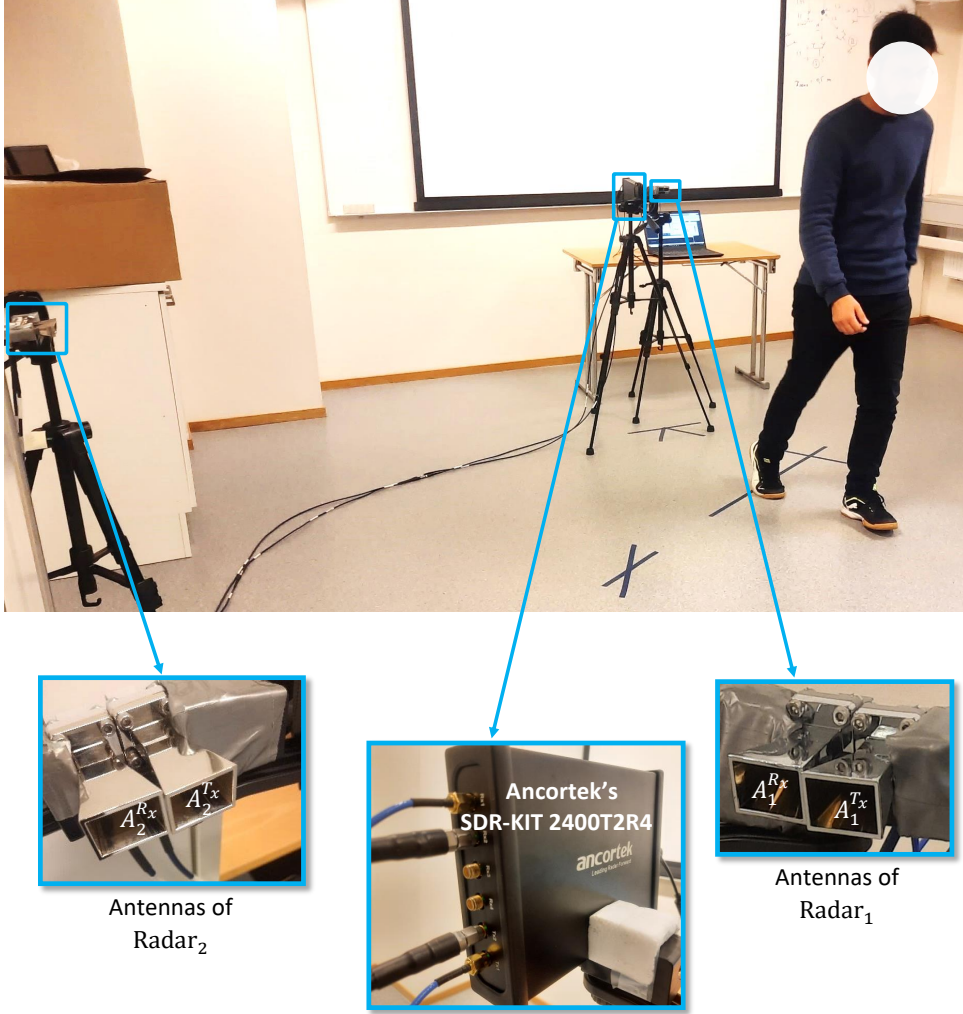
Figure B.6: The distributed 2×2 MIMO radar setup in the presence of a moving person walking in an indoor environment.

human motion in a particular experiment, we refer throughout this section to the 3×3 grid shown in Figure B.7, in which three different scenarios of human movement are represented by different markings. The radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ of the two radar subsystems have been computed, and the results are delineated in the next subsection for a walking and falling activity of a person.

B.4.3 Results for the Proposed 2×2 MIMO Configuration

We have connected the Ancortek MIMO radar system with a signal processor, which processes the MIMO radar's raw IQ data according to the method discussed in Section B.2. For this research, we have recorded the radar's raw IQ data for a 2×2 distributed MIMO radar system, and then we have processed the data offline using MATLAB. Identical but independent radar signal preprocessing chains are implemented for the two subchannels. For each human activity, we have computed the radial velocity distribution $p_{ii}(v, t)$ using (B.14) and mean radial velocity $\bar{v}_{ii}(t)$ using (B.15) for $i \in \{1, 2\}$. Figures B.8 and B.9 correspond to Radar$_2$, which show
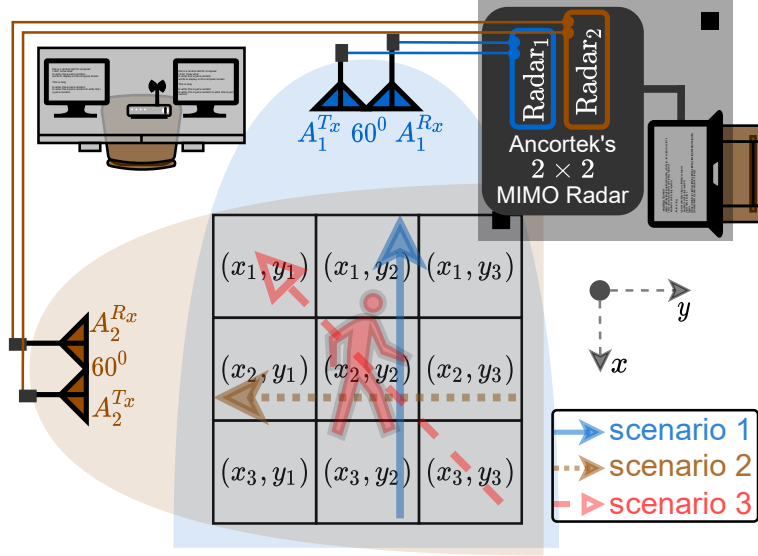
Figure B.7: Two radar subsystems forming a 2×2 distributed MIMO radar system in the presence of a moving person in a cluttered indoor environment.

the TV radial velocity distribution $p_{22}(v, t)$ for the falling and walking activities, respectively. Note that, for the monostatic SISO configuration and 2×2 MIMO configuration in Sections B.3 and B.4, respectively, the positions of the transmitter antenna A_1^{Tx} and receiver antenna A_1^{Rx} are identical. Therefore, Radar_1 of the 2×2 MIMO radar system is identical to the monostatic SISO radar system. As a consequence, the TV radial velocity distributions $p_{11}(v, t)$ corresponding to Radar_1 and the SISO radar system are the same. Recall that the TV radial velocity distributions $p_{11}(v, t)$ corresponding to Radar_1 are shown in Figures B.2 and B.3 for the falling and walking activities, respectively. For all observed activities, we have also computed the mean radial velocity $\bar{v}_{ii}(t)$, which is represented by the black dashed line in Figures B.2, B.3, B.8 and B.9. The falling and walking activities were performed in three different scenarios to see the effect of the human activity direction on the measured TV radial velocity distribution $p_{ii}(v, t)$.

In Figure B.7, we can see that the direction of movement is parallel to the boresight of Radar_1 and orthogonal to the boresight of Radar_2 for Scenario 1. To perform the walking activity in Scenario 1, the person walks from the position (x_3, y_2) towards the position (x_1, y_2) in a straight line. Similarly, to perform a falling activity, the person first stands still on the position (x_3, y_2) and then falls onto a mattress facing towards the antennas of Radar_1. Figures B.2a and B.8a show the impact of the falling activity on the measured TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$, respectively. For the walking activity, Figures B.3a and B.9a show the measured TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$, respectively. Evidently from Figures B.2a, B.8a, B.3a and B.9a, the 2×2 MIMO radar system is able to acquire good multi-perspective human activity signatures for Scenario 1. As the direction of the activity is towards Radar_1 in Scenario 1, it is obvious that Figures B.2a and B.3a give better activity signatures than Figures B.8a and B.9a,
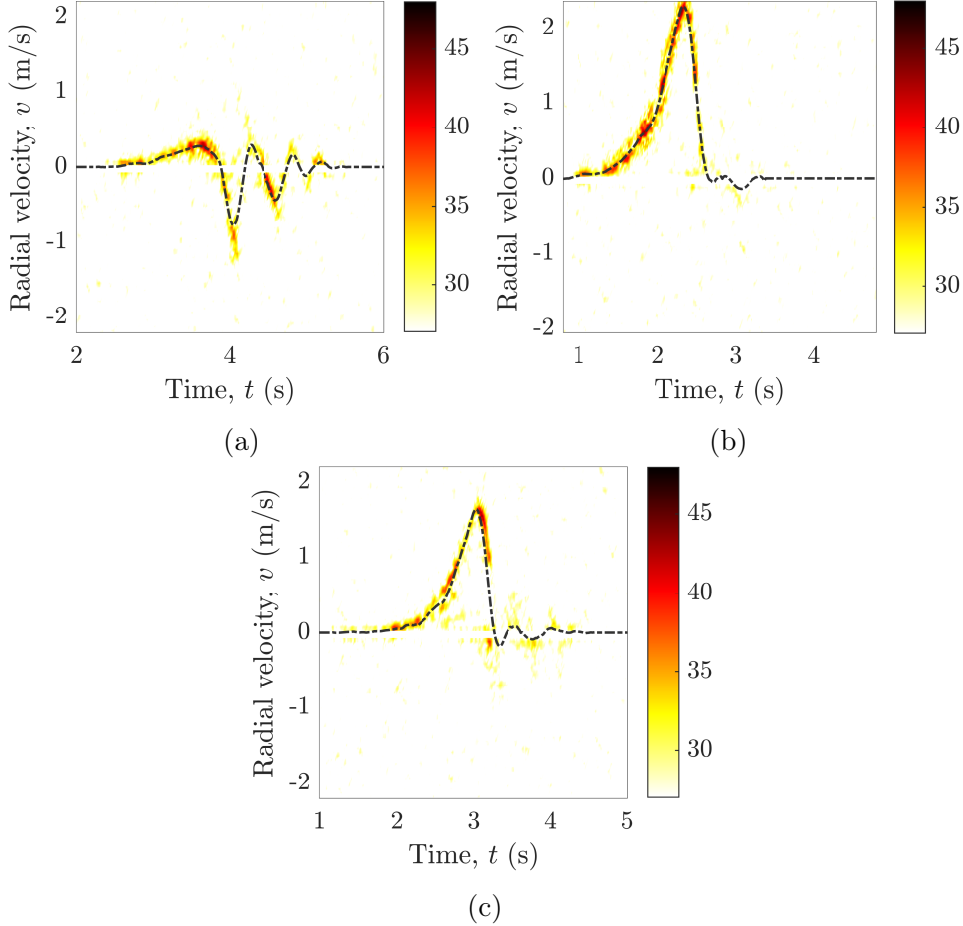
Figure B.8: For Radar$_2$, the measured radial velocity distribution $p_{22}(v, t)$ and mean radial velocity $\bar{v}_{22}(t)$ of a human falling activity in (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

respectively.

The walking and falling activities are repeated for Scenario 2, where the initial and final positions are at (x_2, y_3) and (x_2, y_1), respectively, as shown in Figure B.7. Thus, the direction of motion in Scenario 2 is orthogonal to the boresight of Radar$_1$ and parallel to the boresight of Radar$_2$. In Figures B.2b and B.8b, the TV radial velocity distribution $p_{ii}(v, t)$ of the falling activity is shown for Radar$_1$ and Radar$_2$, respectively. Analogously, Figures B.3b and B.9b show the TV radial velocity distribution $p_{ii}(v, t)$ of the walking activity corresponding to Radar$_1$ and Radar$_2$, respectively. From the measurement results, we can see that the distributed MIMO radar system captures good human activity signatures for Scenario 2 as well. For Scenario 2, as the direction of the activity is towards Radar$_2$, we observe that Figures B.8b and B.9b give better activity signatures than Figures B.2b and B.3b, respectively. In other words, Radar$_2$ produces better human activity signatures than Radar$_1$ for Scenario 2.

In Scenario 3, the direction of movement is roughly diagonal to the boresights of Radar$_1$ and Radar$_2$, as shown in Figure B.7. For the human falling activity, the initial and final positions are (x_3, y_3) and (x_1, y_1), respectively. This is in contrast

(a)

(b)

(c)
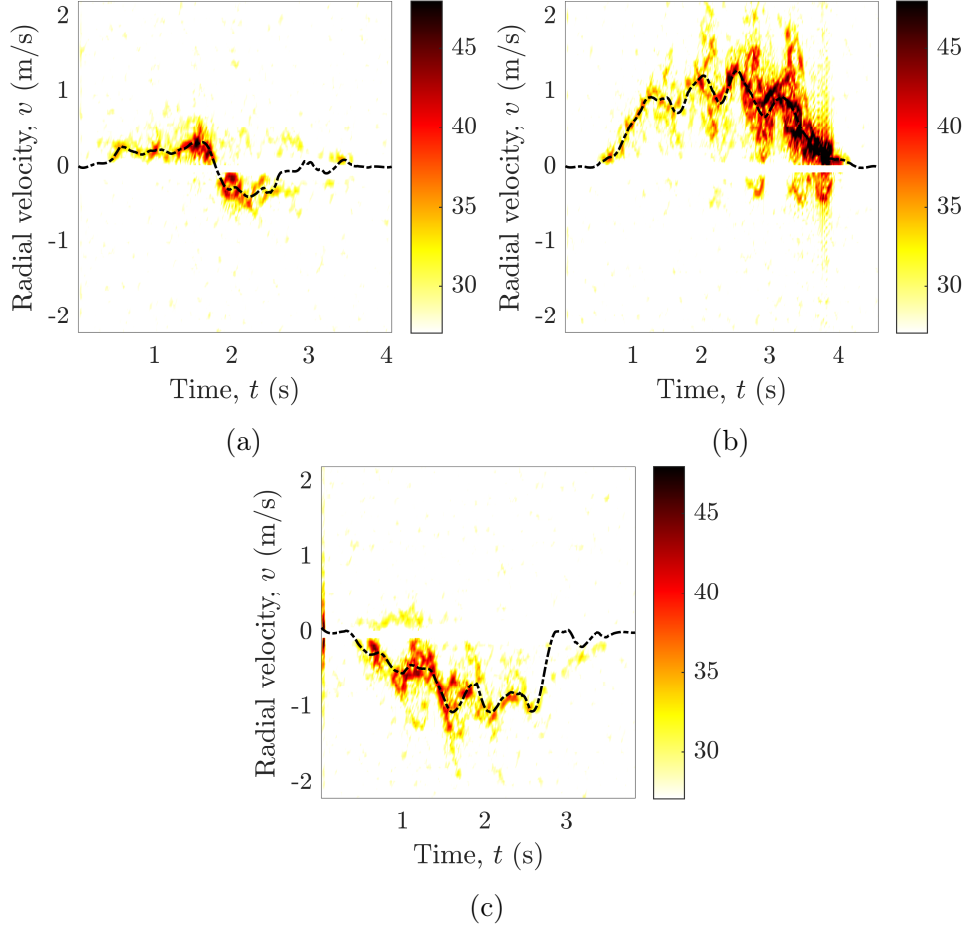
Figure B.9: For Radar$_2$, the measured radial velocity distribution $p_{22}(v, t)$ and mean radial velocity $\bar{v}_{22}(t)$ of a human walking activity in (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

to the walking activity, where the initial and final positions are (x_1, y_1) and (x_3, y_3), respectively. Figures B.2c and B.8c show the TV radial velocity distribution $p_{ii}(v, t)$ of the falling activity with respect to Radar$_1$ and Radar$_2$. Analogously, Figures B.3c and B.9c show the TV radial velocity distribution $p_{ii}(v, t)$ of the walking activity corresponding to Radar$_1$ and Radar$_2$, respectively. Note that both Radar$_1$ and Radar$_2$ capture adequate human activity signatures for Scenario 3. It is evident from the measurement results that unlike the SISO radar system, the distributed MIMO radar system has the innate capability to provide better multi-perspective human activity signatures for all three scenarios.

B.4.4   Implications on the Performance of an RF-Based Step Counter

To illustrate the utility of the proposed distributed 2×2 MIMO radar system, we now investigate the performance of a radar-based passive step counter [49] for the MIMO configuration. As mentioned in Section B.3, the radar-based passive step counter was designed only for human walking activities in accordance with Scenario 1. We have

already shown the performance limitations of the SISO radar system with the radar-based passive step counter module in Section B.3. In this section, we show how the integration of the radar-based passive step counter module with the proposed 2×2 MIMO radar system will mitigate the shortcomings that we encountered with the SISO radar system. Recall that illuminating the indoor environment from different perspectives, as shown in Figure B.7, will enable the distributed 2×2 MIMO radar to capture the walking activity signature regardless of its direction.

A basic block diagram of the 2×2 MIMO radar-based RF step counter is shown in Figure B.10, where it can be seen that the step counter module has been implemented separately for Radar$_1$ and Radar$_2$. To detect the number of steps in a particular walking activity, the 2×2 MIMO radar's raw IQ data are processed by the radar signal preprocessor module (see Section B.2). The radar signal preprocessor module generates the TV mean radial velocities $\bar{v}_{11}(t)$ and $\bar{v}_{22}(t)$ for Radar$_1$ and Radar$_2$, respectively. The TV mean radial velocity $\bar{v}_{ii}(t)$ is smoothed by a Savitzky–Golay filter to be further processed by the RF step detection module.

In Figures B.5 and B.11, the solid black curves show the smoothed TV mean radial velocities $\bar{v}_{11}(t)$ and $\bar{v}_{22}(t)$ for Radar$_1$ and Radar$_2$, respectively. By processing the smoothed mean radial velocity $\bar{v}_{ii}(t)$, the RF step detection module detects the number of steps and their corresponding timestamps for the walking activity. The detected steps of the MIMO radar-based RF step detector are distinctly marked and labeled in Figures B.5 and B.11 for Radar$_1$ and Radar$_2$, respectively. For Scenario 1, the human subject started walking from the position (x_3, y_2) and took four steps towards the position (x_1, y_2). We can see from Figure B.5a that the RF step detection algorithm has successfully detected the four steps for the radar subsystem Radar$_1$. However, the RF step detection algorithm has detected five steps in Figure B.11a due to the poor quality of the data from Radar$_2$. Thus, we can easily discard the data from Radar$_2$ and select the number of steps counted by the RF step counter associated with Radar$_1$.

On the other hand, for Scenario 2, when the human subject walked with four steps from the position (x_2, y_3) towards the position (x_2, y_1), Radar$_2$ is able to produce an intelligible walking activity signature. For Scenario 2, we can choose the
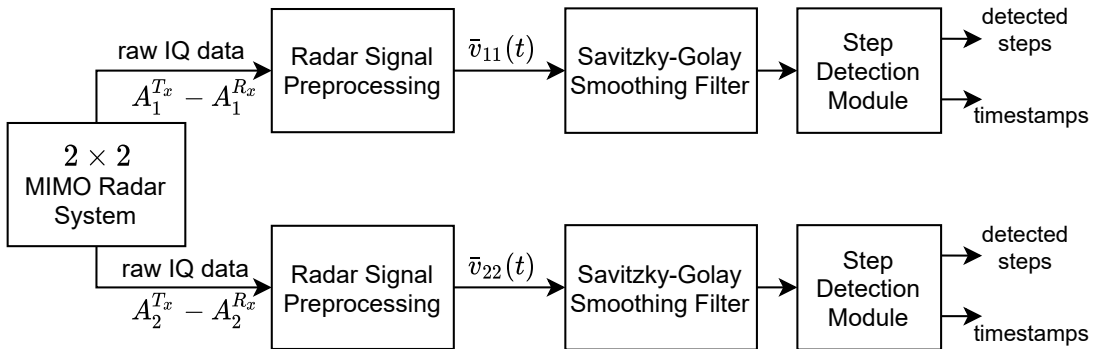
Figure B.10: The block diagram of the RF-based step counter for a 2×2 MIMO radar system.

results from the Radar$_2$-based RF step counter, which accurately counts the number of steps, as shown in Figure B.11b. Therefore, unlike the SISO-based RF step counter, the 2×2 MIMO-based RF step counter is able to detect all the steps that were taken by the human subject in Scenario 1 and 2. Moreover, for Scenario 3, when a person walked from the position (x_1, y_1) towards the position (x_3, y_3) taking only four steps, both radar systems Radar$_1$ and Radar$_2$ are able to produce good walking activity signatures. Thus, the RF step counter detects the four steps successfully as depicted in Figures B.5c and B.11c for Scenario 3. It is now clear that by observing the environment from different perspectives, the proposed 2×2 MIMO radar system is able to detect the scatterer's motion in the horizontal xy-plane. We can conclude that a distributed 2×2 MIMO radar system is sufficient to capture the human activity signatures in all directions.

(a)

(b)

(c)
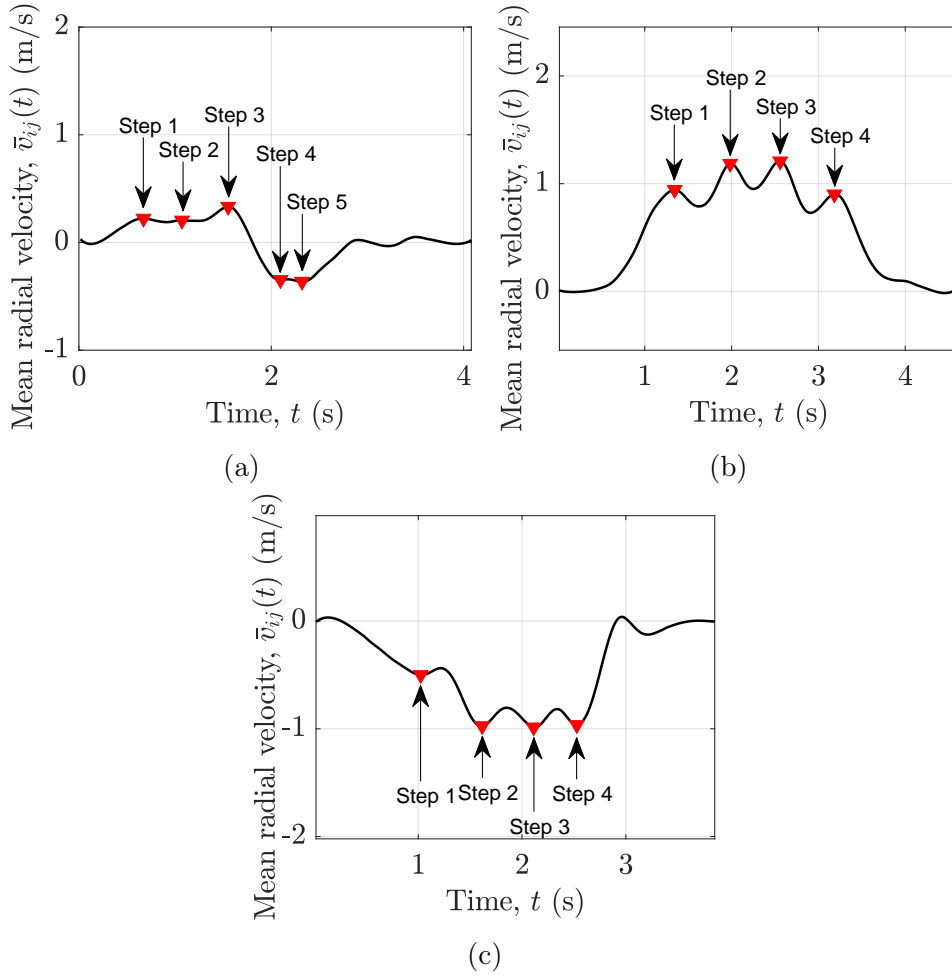
Figure B.11: For Radar$_2$, the number of steps detected from the smoothed mean radial velocity $\bar{v}_{22}(t)$ of a human walking activity according to (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

B.5 Discussion

From the experimental results of Sections B.3 and B.4, we can see that the proposed fundamental 2×2 MIMO radar system ameliorates the limitations posed by the SISO radar configuration. In Figures B.2, B.3, B.8, and B.9, adequate falling and walking activity signatures were observed irrespective of the activity direction. Furthermore, to quantify and numerically assess the performance of the SISO and 2×2 MIMO radar systems for each scenario, we have computed the DTW [44] distance for the walking activity. A numerical analysis of the falling activity signatures yields similar results. Therefore, to be concise, we have only shown the performance of the SISO and 2×2 MIMO radar systems for the walking activity.

The DTW algorithm performs a time series analysis by evaluating the similarity between two temporal sequences. In this algorithm, the two temporal sequences are expanded or stretched such that the overall Euclidean distance between the two sequences is minimized, which makes the algorithm robust to any speed variations, accelerations, or decelerations in the data. This algorithm has been widely used in speech [52], gesture [53], and gait [54, 55] recognition. To quantify whether a radar system has captured an adequate human activity signature or not, we first need a reference human activity signature for comparison. The reference activity signature is obtained in favorable conditions, where a person walks in the direction parallel to the boresight of the SISO radar. For each radar subsystem, the DTW distance is computed between the reference human activity signature and the captured human activity signature of a particular scenario. For the SISO radar system, equivalent to Radar_1, we obtain a single DTW distance metric for each walking scenario as shown in the second and fourth columns of Table B.3. However, for the proposed 2×2 MIMO radar system, we obtain two distinct DTW distance metrics, one for each radar subsystem (Radar_1 and Radar_2). For the 2×2 MIMO radar system, the measured DTW distance metrics of Radar_1 and Radar_2 are compared, and the activity signature with the minimum DTW distance metric is chosen as shown in the last column of Table B.3. Note that the radar subsystem with the minimum DTW distance is chosen because its signature would resemble the higher similarity to the reference activity signature.

A performance summary of the SISO and 2×2 MIMO radar system is based on the results shown in Table B.3. It is clear that the SISO radar system is unable to obtain an adequate activity signature in Scenario 2 as its DTW distance is very

Table B.3: The DTW distance metric for the SISO and 2×2 MIMO radar systems.

Scenario #	Dist. of Radar_1	Dist. of Radar_2	Dist. of SISO Radar	Dist. of 2×2 MIMO Radar
1	19.6	81.2	19.6	19.6
2	73.8	12.0	<u>73.8</u>	12.0
3	7.3	9.1	7.3	7.3

large (see the underlined DTW distance metric in Table B.3). Unlike the SISO radar system, the 2×2 MIMO radar system is able to perform well in all three scenarios, as evident from the measured DTW distances and the results from Sections B.3 and B.4.

Although we have presented the basic 2×2 distributed MIMO radar system to successfully capture the human activity signature in all directions, this study does not include a machine learning or deep learning-based classification network to recognize the type of human activity. However, an RF-based direction-independent HAR system using the proposed MIMO radar configuration would be considered as an extension of this research. To develop such a direction-independent HAR system, the data observed by Radar_1 and Radar_2 need to be merged or fused together. As previous studies have shown the performance improvements due to the fusion of the data obtained from either homogeneous sensors [17, 56] or heterogeneous sensors [9], we can expect to achieve similar performance gains by fusing the data of Radar_1 and Radar_2. In all likelihood, the TV radial velocity distribution $p_{ii}(v, t)$ generated by the proposed 2×2 MIMO radar system would ameliorate the classification performance upon appropriate data fusion. Therefore, a learning network based on the proposed 2×2 MIMO radar data would be able to recognize different activities in different directions.

For a computationally efficient machine learning algorithm, we can extract multiple features from the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$. For instance, we have computed a key parameter known as TV mean radial velocity $\bar{v}_{ii}(t)$, which is shown by the black dashed line in Figures B.2, B.3, B.8, and B.9. Higher-order parameters and features can be readily computed for the data obtained from the proposed 2×2 MIMO radar system. Alternatively, a separate convolutional neural network (CNN) [57] may be adopted to extract features from the data of Radar_1 and Radar_2. Then, the obtained features corresponding to Radar_1 and Radar_2 can be merged using a deep neural network for the classification of human activity. Conceivably, the proposed 2×2 MIMO solution combined with a machine learning or deep learning-based classifier would mitigate the concerns regarding a direction-independent HAR system.

B.6   Conclusions

We have proposed a fundamental 2×2 MIMO approach to analyze the radial velocity distribution and mean radial velocity for falling and walking activities. We confirmed the limitations of SISO RF sensing and emphasized the importance of a distributed MIMO RF system in the context of different directions of human activities. Unlike the state-of-the-art monostatic SISO or MIMO radar systems, our proposed 2×2 distributed MIMO radar system enables the realization of a direction-independent HAR system using TV channel characteristics of human activities obtained from different aspect angles. A comparison with the performance of a radar-based passive step counter has been demonstrated for a SISO and a proposed 2×2 MIMO radar

system. It is shown that the 2×2 MIMO radar-based step counter is able to accurately detect the number of walking steps in all scenarios, while the SISO radar-based step counter fails to perform. Moreover, the DTW distance metric is used to numerically assess the performance of the SISO and 2×2 MIMO radar systems.

Although the fundamental approach presented in this paper may be adopted for various applications, we plan to extend this work to a direction-independent HAR system, where we intend to increase the overall performance of the RF-based HAR system by increasing the number of antennas. The proposed 2×2 MIMO radar system can straightforwardly be scaled to an $N \times N$ MIMO radar system. Moreover, from the obtained multi-perspective channel characteristics, multiple features can be extracted for a classical machine learning-based HAR system. For a more complex HAR problem, a deep CNN can be adopted based on the data from the proposed distributed MIMO radar framework.

Bibliography

[1] M. Al-khafajiy, T. Baker, C. Chalmers, M. Asim, H. Kolivand, M. Fahim, and A. Waraich, "Remote health monitoring of elderly through wearable sensors," *Multimed. Tools Appl.*, vol. 78, no. 17, pp. 24681–24706, 2019.

[2] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Electr. Eng. Technol.*, vol. 14, no. 6, pp. 2567–2573, 2019.

[3] M. Rashmi, T. S. Ashwin, and R. M. R. Guddeti, "Surveillance video analysis for student action recognition and localization inside computer laboratories of a smart campus," *Multimed. Tools Appl.*, vol. 80, no. 2, pp. 2907–2929, 2021.

[4] F. Gurcan, N. E. Cagiltay, and K. Cagiltay, "Mapping human–computer interaction research themes and trends from its existence to today: A topic modeling-based review of past 60 years," *Int. J. Hum. Comput. Interact.*, vol. 37, no. 3, pp. 267–280, 2021.

[5] T. Tuncer, F. Ertam, S. Dogan, and A. Subasi, "An automated daily sports activities and gender recognition method based on novel multikernel local diamond pattern using sensor signals," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9441–9448, 2020.

[6] M. Tammvee and G. Anbarjafari, "Human activity recognition-based path planning for autonomous vehicles," *Signal, Image Video Process.*, vol. 15, no. 4, pp. 809–816, 2021.

[7] F. J. Rodriguez Lera, F. Martín Rico, A. M. Guerrero Higueras, and V. M. Olivera, "A context-awareness model for activity recognition in robot-assisted scenarios," *Expert Syst.*, vol. 37, no. 2, 2020.

[8] M. Muaaz, A. Chelli, and M. Patzold, "WiHAR: From Wi-Fi channel state information to unobtrusive human activity recognition," in *2020 IEEE 91st Veh. Technol. Conf.*, pp. 1–7, 2020.

[9] M. Muaaz, A. Chelli, A. A. Abdelgawwad, A. C. Mallofre, and M. Pätzold, "WiWeHAR: Multimodal human activity recognition using Wi-Fi and wearable sensing modalities," *IEEE Access*, vol. 8, pp. 164453–164470, 2020.

[10] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, 2017.

[11] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3617–3628, 2019.

[12] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mob. Comput. Netw.*, pp. 65–76, 2015.

[13] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 363–373, 2016.

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

[15] S. Waqar, H. Yusaf, S. Sana, M. Waqas, and F. A. Siddiqui, "Reconfigurable monopulse radar tracking processor," in *2018 15th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 805–809, 2018.

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

[17] B. Erol, M. G. Amin, and B. Boashash, "Range-Doppler radar sensor fusion for fall detection," in *2017 IEEE Radar Conf.*, pp. 819–824, 2017.

[18] F. Fioranelli, M. Ritchie, and H. Griffiths, "Aspect angle dependence and multistatic data fusion for micro-Doppler classification of armed/unarmed personnel," *IET Radar, Sonar Navig.*, vol. 9, no. 9, pp. 1231–1239, 2015.

[19] V. Louf, A. Protat, R. A. Warren, S. M. Collis, D. B. Wolff, S. Raunyiar, C. Jakob, and W. A. Petersen, "An integrated approach to weather radar calibration and monitoring using ground clutter and satellite comparisons," *J. Atmos. Ocean. Technol.*, vol. 36, no. 1, pp. 17–39, 2019.

[20] S. Watts, "The ASV 21 maritime surveillance radar," in *2017 IEEE Radar Conf.*, pp. 0027–0032, 2017.

[21] Xiaodong Lu and T. Koga, "DAPS based adaptive tracking system for high-assurance air traffic surveillance," in *2014 Integr. Commun. Navig. Surveill. Conf. Conf. Proc.*, pp. O4–1–O4–8, 2014.

[22] D.-A. Nguyen, B. Cho, C. Seo, J. Park, and D.-H. Lee, "Analysis of the optimal frequency band for a ballistic missile defense radar system," *J. Electromagn. Eng. Sci.*, vol. 18, no. 4, pp. 231–241, 2018.

[23] P. J. Gómez-del Hoyo, N. Del-Rey-Maestre, D. Mata-Moya, M. P. Jarabo-Amores, and M. C. Benito-Ortiz, "Coherent detection and 3D tracking stages of a DVB-T based passive radar for terrestrial traffic monitoring," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 524, no. 1, p. 012002, 2019.

[24] F. Bandini, T. P. Sunding, J. Linde, O. Smith, I. K. Jensen, C. J. Köppl, M. Butts, and P. Bauer-Gottwein, "Unmanned Aerial System (UAS) observations of water surface elevation in a small stream: Comparison of radar altimetry, LIDAR and photogrammetry techniques," *Remote Sens. Environ.*, vol. 237, p. 111487, 2020.

[25] A. Frigeri and M. Ercoli, "The ScanMars subsurface radar sounding experiment on AMADEE-18," *Astrobiology*, vol. 20, no. 11, pp. 1338–1352, 2020.

[26] J.-L. Margot, "A Data-Taking System for Planetary Radar Applications," *J. Astron. Instrum.*, vol. 10, no. 01, 2021.

[27] C. Waldschmidt, J. Hasch, and W. Menzel, "Automotive Radar — From First Efforts to Future Systems," *IEEE J. Microwaves*, vol. 1, no. 1, pp. 135–148, 2021.

[28] Z. Liu, Y. Cai, H. Wang, L. Chen, H. Gao, Y. Jia, and Y. Li, "Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–14, 2021.

[29] F. Uysal, "Phase-coded FMCW automotive radar: System design and interference mitigation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 270–281, 2020.

[30] L. Kranold, M. Taherzadeh, F. Nabki, M. Coates, and M. Popovic, "Microwave breast screening prototype: System miniaturization with IC pulse radio," *IEEE J. Electromagn. RF Microwaves Med. Biol.*, vol. 5, no. 2, pp. 168–178, 2021.

[31] H. Du, T. Jin, Y. He, Y. Song, and Y. Dai, "Segmented convolutional gated recurrent neural networks for human activity recognition in ultra-wideband radar," *Neurocomputing*, vol. 396, pp. 451–464, 2020.

[32] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, 2019.

[33] F. Luo, S. Poslad, and E. Bodanese, "Human activity detection and coarse localization outdoors using micro-Doppler signatures," *IEEE Sens. J.*, vol. 19, no. 18, pp. 8079–8094, 2019.

[34] B. Jokanovic and M. Amin, "Fall detection using deep learning in range-Doppler radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, 2018.

[35] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on Doppler radar motion signature," in *2011 5th Int. Conf. Pervasive Comput. Technol. Healthc. Work.*, pp. 222–225, 2011.

[36] L. Liu, M. Popescu, K. C. Ho, M. Skubic, and M. Rantz, "Doppler radar sensor positioning in a fall detection system," in *2012 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 256–259, 2012.

[37] S. Sana, S. Waqar, H. Yusaf, M. Waqas, and F. A. Siddiqui, "Software defined digital beam forming processor," in *2016 13th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 671–676, 2016.

[38] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *2015 IEEE Radar Conf.*, pp. 1491–1496, 2015.

[39] G. Wang, C. Gu, T. Inoue, and C. Li, "A hybrid FMCW-interferometry radar for indoor precise positioning and versatile life activity monitoring," *IEEE Trans. Microw. Theory Tech.*, vol. 62, no. 11, pp. 2812–2822, 2014.

[40] M. Jian, Z. Lu, and V. C. Chen, "Drone detection and tracking based on phase-interferometric Doppler radar," in *2018 IEEE Radar Conf.*, pp. 1146–1149, 2018.

[41] S. Kim, B.-S. Kim, Y. Jin, and J. Lee, "Extrapolation-RELAX estimator based on spectrum partitioning for DOA estimation of FMCW radar," *IEEE Access*, vol. 7, pp. 98771–98780, 2019.

[42] F. Fioranelli, M. Ritchie, and H. Griffiths, "Bistatic human micro-Doppler signatures for classification of indoor activities," in *2017 IEEE Radar Conf.*, pp. 610–615, 2017.

[43] A. Abuduaini, N. Shiraki, N. Honma, T. Nakayama, and S. Iizuka, "Performance evaluation of multiple human-body localization using bistatic MIMO radar," in *2019 IEEE Asia-Pacific Microw. Conf.*, pp. 575–577, 2019.

[44] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series.," in *Proc. 3rd Int. Conf. Knowl. Discov. Data Min.*, pp. 359–370, 1994.

[45] M. Bekar, C. J. Baker, E. G. Hoare, and M. Gashinova, "Joint MIMO radar and communication system using a PSK-LFM waveform with TDM and CDM approaches," *IEEE Sens. J.*, vol. 21, no. 5, pp. 6115–6124, 2021.

[46] V. C. Chen, *The micro-Doppler effect in radar.* Artech House, 2019.

[47] S. Waqar and M. Pätzold, "Interchannel interference and mitigation in distributed MIMO RF sensing," *Sensors*, vol. 21, no. 22, 2021.

[48] A. Lulu and B. G. Mobasseri, "High-resolution range-Doppler maps by coherent extension of narrowband pulses," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 4, pp. 3099–3112, 2020.

[49] M. Muaaz, S. Waqar, and M. Pätzold, "Radar-based passive step counter and its comparison with a wrist-worn physical activity tracker," in *4th Int. Conf. Intell. Technol. Appl. INTAP*, pp. 259–272, Springer International Publishing, 2022.

[50] A. Abdelgawwad, A. Borhani, and M. Pätzold, "Modelling, analysis, and simulation of the micro-Doppler effect in wideband indoor channels with confirmation through pendulum experiments," *Sensors*, vol. 20, no. 4, 2020.

[51] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.

[52] A. Ismail, S. Abdlerazek, and I. M. El-Henawy, "Development of smart healthcare system based on speech recognition using support vector machine and dynamic time warping," *Sustainability*, vol. 12, no. 6, 2020.

[53] M. Fugini, J. Finocchi, and G. Trasa, "Gesture Recognition using Dynamic Time Warping," in *2020 IEEE 29th Int. Conf. Enabling Technol. Infrastruct. Collab. Enterp.*, pp. 279–282, 2020.

[54] M. Muaaz and R. Mayrhofer, "Smartphone-based gait recognition: from authentication to imitation," *IEEE Trans. Mob. Comput.*, vol. 16, no. 11, pp. 3209–3221, 2017.

[55] M. Muaaz and R. Mayrhofer, "Orientation independent cell phone based gait authentication," in *12th Int. Conf. Adv. Mob. Comput. Multimedia, MoMM 2014*, pp. 161–164, 2014.

[56] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 669–673, 2018.

[57] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-Sense: a passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems," *Ann. Telecommun.*, vol. 77, no. 3-4, pp. 163–175, 2022.

Appendix C

Paper C

Direction-Independent Human Activity Recognition Using a Distributed MIMO Radar System and Deep Learning

Sahil Waqar, Muhammad Muaaz, and Matthias Pätzold

Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway

E-mails: {sahil.waqar, muhammad.muaaz, matthias.paetzold}@uia.no

Abstract— Modern monostatic radar-based human activity recognition (HAR) systems perform very well as long as the direction of human activities is either towards or away from the radar. The monostatic single-input single-output (SISO) and monostatic multiple-input multiple-output (MIMO) radar systems cannot detect motion of an object that moves perpendicularly to the radar's boresight axis. Due to this physical layer limitation, today's radar-based HAR systems fail to classify multi-directional human activities. In this article, we resolve this typical but critical physical layer problem of contemporary HAR systems. We propose a HAR system underlying a distributed MIMO radar configuration, where multiple antennas of a millimeter wave MIMO radar system (Ancortek SDR-KIT 2400T2R4) are distributed in an indoor environment. In our proposed HAR system, we have two independent and identical monostatic radar subsystems that irradiate and capture the multi-directional human movement from two perspectives, which allows to compute two distinct time-variant radial velocity distributions. A feature extraction network extracts numerous features from the measured time-variant radial velocity distributions, which are then fused by a multiclass classifier to detect five types of human activities. The proposed multi-perspective MIMO-radar-based HAR system achieves a classification accuracy of 98.52%, which surpasses the accuracy of SISO radar-based HAR system by more than 9%. Our approach resolves the physical layer limitations of modern HAR systems that are based on either monostatic SISO or monostatic MIMO radar systems.

Index Terms— Deep learning, direction-independent human activity recognition (HAR), fall detection, feature fusion, multistatic radar, multiview radar sensing, orientation-independent HAR.

C.1   Introduction

C.1.1   General Background

Studies have shown a considerable amount of progress in the area of human activity recognition (HAR) over the past few years [1, 2, 3, 4]. The steady interest in HAR is due to its extensive range of applications. Over the years, HAR systems have proven their usefulness in application areas such as social robotics [5], autonomous driving [6], sports [7, 8], home automation [9], healthcare [10], automated video analysis [11], and human–computer interaction [12].

To date, numerous diverse sensing modalities have been adopted to effectuate the HAR task. However, each modality may exhibit its distinct advantages and disadvantages [13]. For instance, due to the ongoing advancements in computer vision techniques, HAR systems based on vision sensors have produced remarkable results [14, 15]. However, vision sensors are often criticized because they are very susceptible to lighting conditions, occlusion, and can violate user privacy. Wearable sensors [16, 17, 18, 19] on the other hand, despite being very effective HAR sensors, are generally criticized for being fragile, obtrusive, and vulnerable to user negligence. Also, the need to be worn indefinitely renders the wearable sensors impractical and inconvenient, especially for elderlies or infirmed persons. By taking into consideration the aforementioned shortcomings, recently HAR systems based on radio frequency (RF) sensing techniques have been preferred more and more despite new challenges and hurdles.

Lately, many researchers have studied and eventually leaned towards Wi-Fi and radar systems for the HAR purpose [20, 21, 22, 23]. Unlike radar systems, commercial grade Wi-Fi routers have the channel frequency response with notably noisy phases [24, 25, 26, 27]. In contrast, commercial coherent radar systems conserve the phase information within their coherent processing interval (CPI) [28]. Thus, small phase variations corresponding to non-stationary scatterers in an environment can be easily processed by coherent signal processing techniques [29, 30]. This is one of the reasons why coherent radar systems have been preferred over Wi-Fi devices to capture the propagation phenomena caused by complex human activities. In the context of RF sensing, the recognition of human activity often relies on exploiting the micro-Doppler phenomenon [31, 32, 33, 34] to discern the specific type of activity being performed. Thanks to recent advancements in the areas of radar techniques and machine/deep learning, the classification and tracking of a wide range of human activities in complex environments will be within reach in a few years.

C.1.2   Problem Description

A major problem of radar-based HAR systems is their inability to generate an adequate micro-Doppler signature in a situation where a person moves perpendicularly to the radar boresight axis. Our article is a step in this direction, in which we propose a pragmatic solution to the problem of direction-independent HAR. Thus, we will look into the classification of five different types of human activities performed

in different directions. Single monostatic radar-based HAR systems do not consider the direction of human motion and thus tend to fail in classifying human activities performed in different directions.

Erol and Amin [35] reported the average classification performance at different aspect angles for a human falling activity. For a human fall parallel to the radar boresight axis or at 0^0 aspect angle, the classification accuracy was 96%; at 60^0 aspect angle, the classification accuracy dropped to 85%; and at 90^0 aspect angle (falling perpendicularly to the radar boresight axis), the classification accuracy plummeted to 45% rendering the HAR system futile. Similarly, for six human activities, Ding et al. [36] reported a decrease in classification performance from 95.8% to 86.7% by changing the radar's viewing angle from 15^0 to 30^0.

C.1.3   Related Work

Some of the approaches to mitigate the problem of the direction of human motion are discussed here along with their shortcomings. In [37, 38] it was shown that by positioning a radar on the ceiling, a human falling in different directions can be detected, but the solution cannot be generalized to classify more complex human activities. To realize a direction-independent HAR, it is tempting to employ a monostatic beamforming multiple-input multiple-output (MIMO) radar system with the capability of measuring the target's angle [39, 40]. But in practice, commercial beamforming radar systems have poor angular and cross-range resolutions due to their limited hardware resources. Thus, for applications such as short-range hand gesture sensing, where the cross-range resolution is not a concern, Molchanov et al. [41] rightly utilized the angular information of a single-input multiple-output (SIMO) frequency-modulated continuous wave (FMCW) monopulse radar. Unfortunately, the approach cannot be extended to direction-independent HAR systems because of the radar's poor cross-range resolution. Recently, HAR systems are realized by using three-dimensional (3D) point cloud data generated by millimeter wave (mm-wave) monostatic MIMO radar systems [42, 43]. But 3D point cloud data also suffer from the problem of poor cross-range resolution. For better angle estimation or, equivalently, cross-range resolution, more advanced signal processing techniques such as the *"estimation of signal parameters via rotational invariance techniques (ESPRIT)"* [44] and *"multiple signal classification (MUSIC)"* algorithms [45, 46] are usually employed, but these estimation techniques demand a high signal-to-noise ratio [47]. Alternatively, a single-input single-output (SISO) bistatic radar system [48] is a good choice for HAR. However, an even better choice for the direction-independent HAR are multi-perspective multistatic MIMO radar systems. They can provide the best multi-view signatures of human activities, as we will see in this article.

C.1.4 Proposed Approach for HAR

To overcome the aforementioned issues and drawbacks of monostatic SISO, SIMO, and beamforming MIMO radar-based HAR systems, we develop a multi-perspective 2×2 distributed MIMO radar system to realize a direction-independent HAR system. In our approach, two radar subsystems, each consisting of one transmit and one receive antenna and their own independent signal preprocessing units, are spatially distributed to irradiate the environment from different perspectives (see Section C.3). This multistatic MIMO radar framework enables us to detect and classify different types of human activities independent of their respective directions.

Human body segments can be modelled by N moving scatterers, which reflect back the radar signals to the radar receiver. The scatterers' distinct time-variant (TV) radial velocity components can be described by the so-called TV radial velocity distribution (see Section C.2). The TV radial velocity distributions at the output of the radar's signal preprocessor are in fact the input feature maps to our classifier, which is based on a deep convolutional neural network (DCNN). We use deep learning methods to automatically extract the features from the TV radial velocity distributions of the MIMO radar system to finally classify the type of human activity regardless of its direction of motion.

Conventionally, it was not uncommon to manually extract features in single-variable and joint-variable domains to classify human activities using machine learning techniques, such as support vector machine (SVM), with a well-documented classification accuracy of 90% [49]. Widely adopted conventional machine learning algorithms in conjunction with domain-based feature engineering usually have theoretical foundations and are computationally less expensive when compared to deep learning algorithms. However, manual feature engineering is quite cumbersome and requires specific expertise. Determining the relevance and significance of features for identifying specific motion artifacts is also a complicated task. Large differences in manually measured features were found in different individuals monitored for health status, body height, and habits [50]. Therefore, to account for the intricate attributes of human motion, and to overcome the aforementioned challenges associated with the manual feature engineering, deep learning algorithms are preferred [51].

To train and test our SISO and MIMO radar-based direction-independent HAR classifiers (see Section C.5), we recorded a novel HAR dataset, where the human activities were performed in several directions in the two-dimensional (2D) horizontal xy-plane. In this regard, we denote the recorded HAR dataset with the superscript "(2D)" as HAR$^{(2D)}$ (see Section C.4). For a conventional monostatic SISO radar-based HAR classifier that contains the human movement merely along the one-dimensional (1D) x-axis or the monostatic radar's boresight, we denote the recorded HAR dataset accordingly by HAR$^{(1D)}$.

C.1.5 Contributions

The MIMO radar-based HAR system presented in this article is a stride forward towards actualizing more advanced RF-based HAR systems. The main contributions of the research are as follows:

1. For our direction-independent HAR system, we have addressed a critical physical layer problem of monostatic radar systems related to the target's aspect angle.

2. For a monostatic SISO and the multistatic MIMO radar configurations, we have analyzed the variations in measured channel characteristics for five types of human activities (falling, walking, standing, sitting, picking). We also studied the effects of different directions of human activities by analyzing the TV radial velocity distributions of the MIMO radar system (see Section C.3).

3. We composed a completely novel HAR dataset, denoted as HAR$^{(2D)}$, by using the multi-perspective 2×2 MIMO radar configuration (see Section C.4). We recorded real human activities by using a commercial mm-wave radar system known as Ancortek SDR-KIT 2400T2R4. The HAR$^{(2D)}$ dataset consists of five types of human activities performed by six different persons in several directions.

4. By using the HAR$^{(2D)}$ dataset and its derivative or subset dataset denoted as HAR$^{(1D)}$, we have developed and analyzed three different HAR systems (see Section C.5): (a) a SISO radar-based conventional HAR system, (b) a SISO radar-based direction-independent HAR system, and (c) a MIMO radar-based direction-independent HAR system. The proposed 2×2 MIMO radar-based HAR system is capable of recognizing human gross motor activities regardless of the aspect angle or direction of motion, and it is straightforwardly scalable to a higher number of antennas for a more complex human activity classification task.

5. For the three HAR systems, we accordingly designed three different DCNN-based multiclass classifiers. The DCNN classifier extract features automatically from the radar's TV radial velocity distribution before classifying an activity. For the distributed MIMO radar-based classifier, feature level fusion has been adopted, which virtually combines the target's information from different aspect angles, and thereby eradicates the limitations that emerge due to the direction of motion.

6. The classification performances of the three HAR systems have been assessed and compared quantitatively. It is shown that the proposed HAR system, based on the multi-perspective 2×2 MIMO radar framework, improves the classification accuracy of the monostatic SISO radar-based HAR system from 88.98% to 98.52%.

C.1.6 Article Organization

The article organization is as follows. Section C.2 describes the MIMO radar system model and the deep learning methods that are utilized in this research. A critical problem of modern SISO and monostatic MIMO radar-based HAR systems and its solution is discussed in Section C.3. The data acquisition campaign is described in Section C.4. In Section C.5, a conventional and a direction-independent SISO radar-based HAR system, as well as a direction-independent MIMO radar-based HAR system are presented. Lastly, Section C.6 draws the conclusions.

C.2 System Overview

C.2.1 MIMO Radar Signal Preprocessing

An FMCW 2×2 MIMO radar system periodically transmits a chirp waveform $c_i(t')$, which can be expressed as [52]

$$c_i(t') = \exp\left[j\left(\phi_i + 2\pi f_0 t' + \gamma \pi t'^2\right)\right], \qquad 0 \le t' < T_{sw} \tag{C.1}$$

where $i = 1, 2$. The symbol ϕ_i is the initial phase term, f_0 is the initial frequency, and γ is the slope of the chirp waveform. The symbols t' and T_{sw} in (C.1) are the fast time and duration of the chirp, respectively. We adopted a time division multiple access (TDMA) scheme, where the transmitter antenna $A_i^{T_x}$ periodically transmits the chirp waveform $c_i(t')$ in separate time windows, which are defined as $(2n + i - 1)T_{sw} \le t' < (2n + i)T_{sw}$ for $n = 0, 1, \ldots$ and $i = 1, 2$. With the help of the Dirac delta function $\delta(\cdot)$, we can express the transmit signal $s_i(t', t)$ in terms of fast time t' and slow time t as [53]

$$s_i(t', t) = \sum_{n=0}^{\infty} c_i(t')\delta(t - T_{n,i}). \tag{C.2}$$

The symbol $T_{n,i}$ in (C.2) is the discrete slow time that depends on the chirp duration T_{sw} according to $T_{n,i} = (2n + i - 1)T_{sw}$.

For a 2×2 MIMO radar, the notation $A_i^{T_x}$–$A_k^{R_x}$ describes the wireless link between the transmitter antenna $A_i^{T_x}$ and the receiver antenna $A_k^{R_x}$. The transmit signal $s_i(t', t)$ interacts with $\mathcal{L}$ stationary and non-stationary scatterers present in the wireless link $A_i^{T_x}$–$A_k^{R_x}$, where $i, k \in \{1, 2\}$. Let the symbols $d_{ik}^{(l)}$, c_0, and λ denote the propagation distance of the lth scatterer, speed of light, and radar's wavelength, respectively. Then, the beat frequency $f_{b,ik}^{(l)}$ and the phase $\phi_{ik}^{(l)}$ of the lth scatterer are given by $f_{b,ik}^{(l)} = 2d_{ik}^{(l)}\gamma/c_0$ and $\phi_{ik}^{(l)} = 4\pi d_{ik}^{(l)}/\lambda$, respectively, where $l = 1, 2, \ldots, \mathcal{L}$. For the wireless link $A_i^{T_x}$–$A_k^{R_x}$ and the lth scatterer, the received beat signal $s_{b,ik}^{(l)}(t', t)$ can be expressed as [53]

$$s_{b,ik}^{(l)}(t', t) = \sum_{n=0}^{\infty} a_{ik}^{(l)} \exp\left[j\left(2\pi f_{b,ik}^{(l)} t' + \phi_{ik}^{(l)}\right)\right] \delta(t - T_{n,i} - \tau_{ik}^{(l)}) \tag{C.3}$$

where $a_{ik}^{(l)}$ is the gain, which is assumed to be constant within the radar's CPI. The propagation delay $\tau_{ik}^{(l)}$ in (C.3) is related to the beat frequency $f_{b,ik}^{(l)}$ by $\tau_{ik}^{(l)} = f_{b,ik}^{(l)}/\gamma$. At the radar receiver, the composite beat signal $s_{b,ik}(t', t)$ is simply the sum of all $\mathcal{L}$ beat signals, i.e,

$$s_{b,ik}(t', t) = \sum_{l=1}^{\mathcal{L}} s_{b,ik}^{(l)}(t', t) \tag{C.4}$$

We obtain the beat frequency function $S_{b,ik}(f_b, t)$ by computing the Fourier transform of the beat signal $s_{b,ik}(t', t)$ over the fast time t', i.e., [54]

$$S_{b,ik}(f_b, t) = \int_0^{T_{sw}} s_{b,ik}(t', t) e^{-j2\pi f_b t'} dt' \tag{C.5}$$

where f_b is the beat frequency. The beat frequency function $S_{b,ik}(f_b, t)$ in (C.5) further undergoes a short-time Fourier transform (STFT) over the slow time t. Subsequently, the square of the STFT results in the TV micro-Doppler signature $S_{ik}(f, t)$, which is given as

$$S_{ik}(f, t) = \left| \int_0^{f_{b,\max}} \int_{-\infty}^{\infty} S_{b,ik}(f_b, t'') W_r(t'' - t) e^{-j2\pi f t''} dt'' df_b \right|^2 \tag{C.6}$$

where f represents the Doppler frequency, $f_{b,\max}$ is the maximum beat frequency, t'' is the running time, and $W_r(\cdot)$ is a window function, which is in our case a rectangular function with a width of $64T_{sw}$.

Finally, the TV radial velocity distribution $p_{ik}(v, t)$ is obtained from the TV micro-Doppler signature $S_{ik}(f, t)$ according to [53]

$$p_{ik}(v, t) = \frac{S_{ik}\left(\frac{2f_0}{c_0}v, t\right)}{\int_{-\infty}^{\infty} S_{ik}\left(\frac{2f_0}{c_0}v, t\right) dv} \tag{C.7}$$

where v represents the radial velocity. Note that the human body is composed of body segments and each body segment contains several scatterers that reflect back the RF signals to the radar. Each scatterer on a human body segment has a unique TV radial velocity component due to its spatially distinct motion. The TV radial velocity distribution $p_{ik}(v, t)$ contains the radial velocity components from all the scatterers on the human body. We use the expression in (C.7) to obtain the TV radial velocity distribution $p_{ik}(v, t)$ of the recorded human activities. The TV radial velocity distribution $p_{ik}(v, t)$ is converted into an image in the time-velocity domain, which is basically an input feature map to the DCNN, as described in Section C.2.2.

C.2.2 Deep Learning

In this section, a supervised learning-based multiclass classification method is delineated. Assume a d-dimensional mth feature vector $\boldsymbol{x}_m$ that belongs to a feature

space $\mathcal{X}$. This feature space $\mathcal{X}$ is a proper subset of the real coordinate space $\mathbb{R}^d$, meaning that $\boldsymbol{x}_m \in \mathcal{X} \subset \mathbb{R}^d$. For the entire number of classes C, the mth label y_m is an element of a label space $\mathcal{Y} = \{1, 2, \ldots, C\}$, i.e., $y_m \in \mathcal{Y}$. A dataset D is defined as $\{(\boldsymbol{x}_m, y_m)\}_{m=0}^{M-1}$, where M is the total number of labeled training samples.

We aim to design a DCNN-based classifier function $\mathcal{C}_f$ that maps the input feature space $\mathcal{X}$ into the label space $\mathcal{Y}$, i.e., $\mathcal{C}_f : \mathcal{X} \mapsto \mathcal{Y}$. An empirical risk $R_J(\mathcal{C}_f)$ corresponding to the categorical cross-entropy loss function J_{CCE} and the classifier function $\mathcal{C}_f$ is given as [55, 56]

$$
\begin{aligned}
R_J(\mathcal{C}_f) &= \mathbb{E}_D \left\{ J_{CCE} \left(\mathcal{C}_f(\boldsymbol{x}; \boldsymbol{\theta}), y_{\boldsymbol{x}} \right) \right\} \\
&= -\frac{1}{M} \sum_{m=0}^{M-1} \sum_{c=0}^{C-1} y_m^c \log \mathcal{C}_f^c(\boldsymbol{x}_m; \boldsymbol{\theta})
\end{aligned}
\tag{C.8}
$$

where $\mathbb{E}_D\{\cdot\}$ denotes the expectation operator that is performed over the empirical distribution, which can either be the dataset D or a mini-batch from the dataset D. In (C.8), the symbol $\boldsymbol{\theta}$ is a vector of trainable parameters defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_L)$, where L depends on the complexity of the classifier. The symbol y_m^c in (C.8) is the cth entity of the mth one-hot encoded label vector $\boldsymbol{y}_m$, which means $y_m^c \in \{0, 1\}$ such that $(\mathbf{1})^\top \boldsymbol{y}_m = 1 \ \forall \ m$, where $\mathbf{1}$ is a C-dimensional vector of ones, and $(\cdot)^\top$ is the transpose operator. The symbol $\mathcal{C}_f^c$ represents the cth element of the classifier function $\mathcal{C}_f$. We have used the softmax layer as an output layer of the deep neural network (DNN), thus $\sum_{c=0}^{C-1} \mathcal{C}_f^c(\boldsymbol{x}_m; \boldsymbol{\theta}) = 1$, and $\mathcal{C}_f^c(\boldsymbol{x}_m; \boldsymbol{\theta}) \geq 0, \ \forall \ m, c, \boldsymbol{\theta}$. The trainable parameters of the vector $\boldsymbol{\theta}$ corresponding to the classifier function $\mathcal{C}_f$ can be obtained by minimizing the empirical risk $R_J(\mathcal{C}_f)$.

The learning process of the DCNN and DNN is the same, but in case of DCNN, the number of trainable parameters is drastically reduced. In a DCNN, convolutional layers are employed to generate the feature maps from their inputs by means of multiple learnable filters. Assume a total number of Q filters in a convolutional layer, then the mth input feature map $\boldsymbol{x}_m$ is convolved with the qth filter. The qth filter is characterized by its trainable weight vector $\mathbf{w}_q$ and bias b_q. Then, the qth output $\mathbf{y}_q$ of the convolutional layer is given by

$$
\mathbf{y}_q = \sum_{m=1}^{M} \sigma(\boldsymbol{x}_m * \mathbf{w}_q + b_q \mathbf{1}), \quad q = 1, 2, \ldots, Q
\tag{C.9}
$$

where the symbol $*$ denotes the convolutional operator. The function $\sigma(\cdot)$ in (C.9) is a rectified linear unit (ReLU) activation function [57] formulated as $\sigma(x) = \max(0, x)$, which mitigates the problems of slow convergence and gradient vanishing [58].

Pooling layers are generally utilized as an abstraction and downsampling tool to progressively reduce the spatial size and redundancies of the extracted feature maps to increase the network's computational efficiency. Moreover, dropout layers are added to the network to improve the network generalizability and to avoid the overfitting problem [59]. After several convolutional layers, the feature maps

are flattened before feeding them to the fully connected dense layers or multilayer perceptron (MLP) layers.

In this research, we use a stochastic optimization technique known as adaptive moment estimation (Adam) [60] to optimize or train the parameters of the vector $\boldsymbol{\theta}$. The Adam algorithm applies adaptive learning rates that are based on the estimates of the first-order moment $\boldsymbol{m}_\kappa$ and second-order moment $\boldsymbol{v}_\kappa$ of the gradient $\boldsymbol{g}_\kappa$ according to

$$\boldsymbol{m}_\kappa = \beta_1 \boldsymbol{m}_{\kappa-1} + (1 - \beta_1)\boldsymbol{g}_\kappa \tag{C.10}$$

and

$$\boldsymbol{v}_\kappa = \beta_2 \boldsymbol{v}_{\kappa-1} + (1 - \beta_2)\boldsymbol{g}_\kappa^2 \tag{C.11}$$

where the symbol κ denotes the iteration number, and the decay factors are denoted by β_1 and β_2. The gradient $\boldsymbol{g}_\kappa$ in (C.10) and (C.11) is the gradient of the stochastic objective function $f(\boldsymbol{\theta}_\kappa) = \min_{\boldsymbol{\theta}_\kappa} R_J(\mathcal{C}_f)$. Note that in the Adam algorithm, element-wise operations are adopted for all the vectors $\boldsymbol{m}_\kappa$, $\boldsymbol{v}_\kappa$, $\boldsymbol{g}_\kappa$, and $\boldsymbol{\theta}_\kappa$. Additionally, to counteract the initialization bias of the moments or to avoid the moments' biasedness towards zero, Kingma and Ba [60] suggested that the first- and second-order moments can be rectified as $\hat{\boldsymbol{m}}_\kappa = \boldsymbol{m}_\kappa/(1 - \beta_1^\kappa)$ and $\hat{\boldsymbol{v}}_\kappa = \boldsymbol{v}_\kappa/(1 - \beta_2^\kappa)$, respectively. Then, for α_κ being the learning rate and ϵ a small constant, the ℓth parameter of the vector $\boldsymbol{\theta}_\kappa$ at the κth iteration can be updated as [60]

$$\theta_{\ell,\kappa} = \theta_{\ell,\kappa-1} - \frac{\alpha_\kappa}{\sqrt{\hat{v}_{\ell,\kappa}} + \epsilon}\hat{m}_{\ell,\kappa} \tag{C.12}$$

where $\ell = 1, 2, \ldots, L$.

By using the Adam optimizer delineated in this section, we perform the parameter optimization of our DCNN-based classifiers (see Section C.5), where our objective function is the minimization of the empirical risk $R_J(\mathcal{C}_f)$ as defined in (C.8).

C.3 Experimental Setup and the Proposed Solution

In the following, we develop a more pragmatic and complex HAR system suitable for detecting human activities with motion in multiple directions. To this end, we utilize the multi-perspective 2×2 distributed MIMO radar configuration [53] (see Fig. C.1) to eventually realize a direction-independent HAR system. The human activities were monitored by using the 2×2 MIMO radar configuration shown in Fig. C.1. This configuration is also used for comparison with conventional SISO radar-based HAR systems, and to find out whether the multi-perspective MIMO radar configuration can mitigate their limitations. We deployed a software-defined radar system known as Ancortek SDR-KIT 2400T2R4, which is an FMCW mm-wave MIMO radar system, and used its transmitter-receiver antennas in a 2×2 configuration. The operating parameters of the Ancortek radar system are delineated in Table C.1.

For the proposed 2×2 MIMO radar-based HAR system, we arrange two radar subsystems, denoted by Radar$_1$ and Radar$_2$, where each radar subsystem has a collocated transmitter and a receiver antenna in a monostatic configuration. Radar$_1$
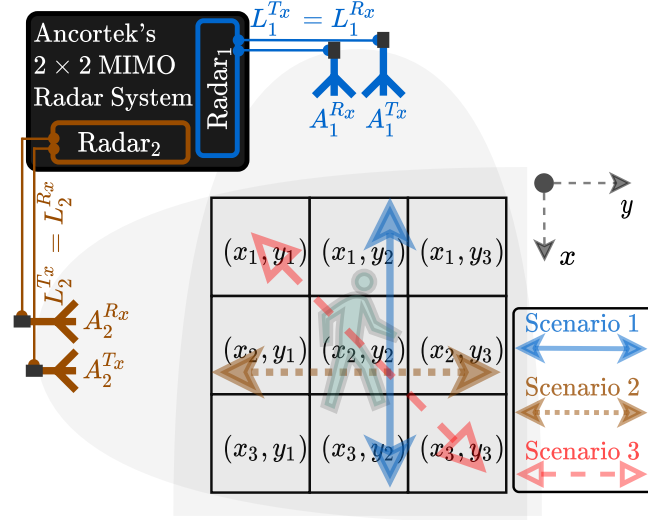
Figure C.1: Measurement setup of the proposed 2×2 MIMO radar-based HAR system consisting of Radar$_1$ and Radar$_2$, where Scenario 1–3 characterize human activities in different directions.

Table C.1: 2×2 MIMO radar system parameters.

Description	Symbols	Values
Carrier frequency	f_c	24.125 GHz
Radar bandwidth	BW	250 MHz
Chirp duration	T_{sw}	500 µs
Pulse repetition frequency	PRF	1 kHz
RF cable lengths	$(L_1^{T_x}, L_1^{R_x}, L_2^{T_x}, L_2^{R_x})$	$(0.3, 0.3, 7, 7)$ m

and Radar$_2$ are distributed in an indoor setting such that the 2×2 MIMO radar system renders a multi-perspective illumination of a target as shown in Fig. C.1, thereby having the potential to overcome the limitations that are posed by the monostatic SISO or monostatic MIMO radar systems in the context of HAR. We operate Radar$_1$ and Radar$_2$ in different time slots according to the TDMA scheme, where both radar subsystems have identical but independent radar signal preprocessing chains (see Section C.2). The radar signal preprocessing chains process the raw in-phase and quadrature (IQ) data recorded by the Ancortek MIMO radar system. For a human activity, the radar signal preprocessing block of Radar$_i$ generates the TV radial velocity distribution $p_{ii}(v, t)$ by using (C.7) for $i \in \{1, 2\}$.

We consider five different types of human activities, which are as follows: falling on a mattress on the floor, walking, standing up from a chair, sitting down on a chair, and picking up an object from the floor. For these activities, the measured TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ are shown in Figs. C.2 and C.3, where the Scenarios 1, 2, and 3 denote the directions of human activities according to Fig. C.1. In Scenario 1 (Scenario 2), the human motion is parallel to the boresight of Radar$_1$ (Radar$_2$), whereas in Scenario 3, the human movement is roughly at 45^0 to the boresights of both radar subsystems, as depicted in Fig. C.1.

Fall (Scenario 1) Fall (Scenario 2) Fall (Scenario 3)
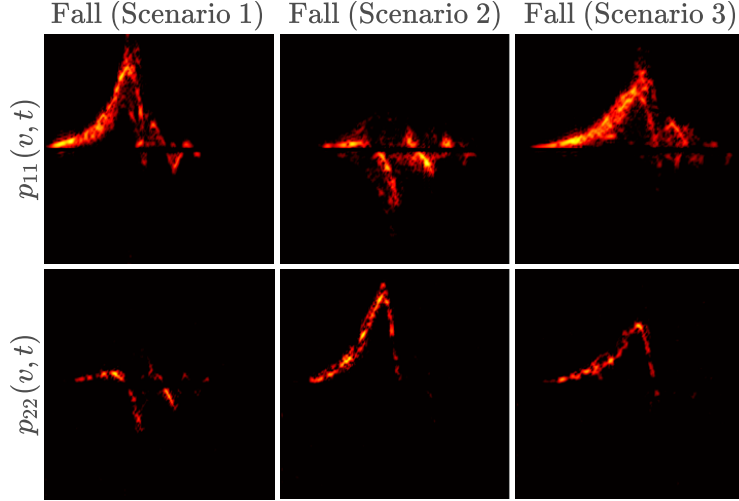
Figure C.2: Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of the "Fall" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over $2-4$ s.

Radar$_1$ and Radar$_2$ complement each other such that when the activity direction changes from the x-axis to the y-axis of Fig. C.1, the activity signature slowly vanishes from the radial velocity distribution $p_{11}(v,t)$ of Radar$_1$ and starts appearing in the radial velocity distribution $p_{22}(v,t)$ of Radar$_2$. For "Fall" activities performed in different directions, the measured radial velocity distributions $p_{11}(v,t)$ and $p_{22}(v,t)$ in the three scenarios vary significantly, as shown in Fig. C.2. We can see from Figs. C.2 and C.3 that Radar$_1$ and Radar$_2$ are unable to acquire optimal human activity signatures in Scenarios 2 and 1, respectively. The suboptimal human activity signatures contribute towards the poor classification performance of a SISO radar-based direction-independent HAR system (see Section C.5.2). Therefore, analogous to a monostatic SISO or monostatic MIMO radar case, a single radial velocity distribution either from Radar$_1$ or Radar$_2$ cannot completely portray a human activity and would not be sufficient for the realization of a direction-independent HAR system. Additionally, Fig. C.3 shows how the TV radial velocity distribution $p_{ii}(v,t)$ changes with the type of human activity. This figure demonstrates that the human activity signature or TV radial velocity distribution $p_{ii}(v,t)$ depends on the type as well as the direction of the human activity. To see the radar signatures corresponding to different multi-directional human activities, please refer to Figs. C.12–C.15 in Appendix C.7.

As our 2×2 distributed MIMO radar-based HAR system generates two distinct activity signatures from two different aspect angles for a human activity, we must fuse or merge the information from the two activity signatures in order to accurately classify the human activity regardless of the direction of motion. In this research, we have implemented a fusion technique at the feature level. For this purpose, for each radar subsystem, Radar$_1$ and Radar$_2$, the features are extracted independently and automatically by several convolutional layers from the radial velocity distri-
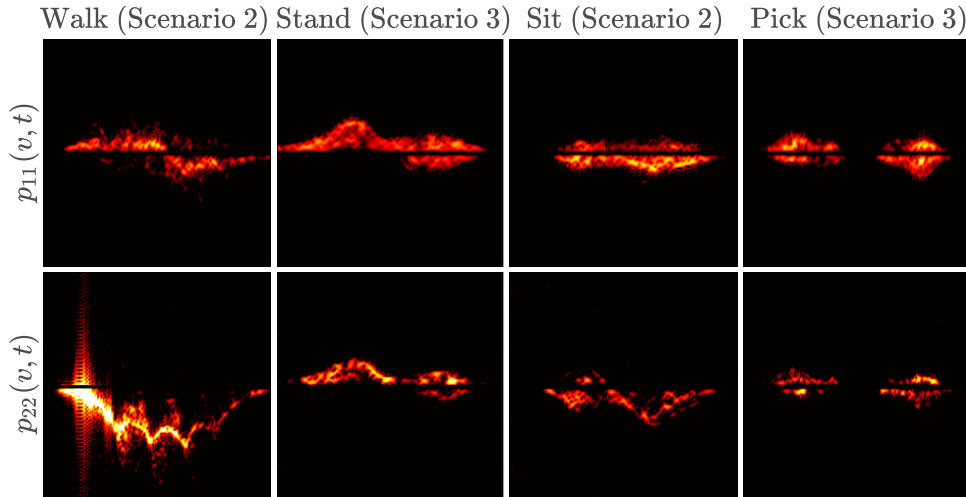
Figure C.3: Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of different human activities, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over $3-5$ s.

butions $p_{11}(v,t)$ and $p_{22}(v,t)$, respectively. The extracted features from the radar subsystems are then merged by the concatenation layer (see Section C.5.3).

In Section C.5, we show how the classification performance of the monostatic radar-based HAR system deteriorates if the human activities take place in the 2D xy-plane, which is depicted by the 3×3 grid in Fig. C.1. We also explain the design of the proposed 2×2 distributed MIMO radar-based HAR system and show how it overcomes the above constraints on the direction of human activity motion. Compared to the SISO radar-based direction-independent HAR system, we see that the proposed 2×2 distributed MIMO radar-based direction-independent HAR system significantly ameliorates the classification accuracy.

C.4 Data Collection

A comprehensive measurement campaign was carried out in an indoor environment consisting of fixed objects, such as chairs, tables, cabinets, computers, and other electronic items. The five types of activities were performed by six different persons, one of them was a female candidate. The human activities were carried out in several directions, with different speeds, and in different locations. For instance, the falling activities were performed in six different directions as depicted by the scenario markers in Fig. C.1. Specifically, the falling activities were executed in the following directions: from (x_3, y_2) to (x_1, y_2), from (x_1, y_2) to (x_3, y_2), from (x_2, y_3) to (x_2, y_1), from (x_2, y_1) to (x_2, y_3), from (x_3, y_3) to (x_1, y_1), and from (x_1, y_1) to (x_3, y_3). The walking activities were performed and recorded in a similar fashion. The other human activities—standing up, sitting down, and picking up an object— were performed accordingly.

In this article, the term HAR$^{(2D)}$ is coined to represent the dataset recorded by the 2×2 MIMO radar system, where the superscript "(2D)" refers to the human

movement in the 2D horizontal xy-plane in Fig. C.1. Therefore, for the direction-independent HAR task, we define FEN$^{(2D)}$, SISO$^{(2D)}$, and MIMO$^{(2D)}$ as a feature extraction network, a SISO radar-based HAR classifier, and a MIMO radar-based HAR classifier, respectively. On the other hand, to denote the human movement along the 1D x-axis of the 3×3 grid in Fig. C.1, we use the superscript "(1D)". Thus, for the conventional 1D HAR task, where the human movement is restricted to Scenario 1 in Fig. C.1, we define HAR$^{(1D)}$, FEN$^{(1D)}$, and SISO$^{(1D)}$ as a dataset recorded by Radar$_1$, a feature extraction network, and a conventional SISO radar-based HAR classifier, respectively.

We need the HAR$^{(2D)}$ dataset to realize the SISO$^{(2D)}$ and MIMO$^{(2D)}$ HAR systems, whereas the HAR$^{(1D)}$ dataset is required for the SISO$^{(1D)}$ HAR system. The details of the HAR$^{(2D)}$ dataset related to the measurement campaign based on the proposed 2×2 MIMO radar framework are shown in Table C.2. As entered in Table C.2, we recorded a total of 1364 activities. For each activity, we generated the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ corresponding to the radar subsystems Radar$_1$ and Radar$_2$, respectively. On the other hand, Table C.3 shows the HAR$^{(1D)}$ dataset, which is a proper subset of the HAR$^{(2D)}$ dataset, i.e., HAR$^{(1D)} \subset$ HAR$^{(2D)}$. Note that the HAR$^{(1D)}$ dataset only contains those activities of the HAR$^{(2D)}$ dataset that were performed parallel to the boresight of Radar$_1$. To implement a conventional monostatic radar-based HAR system (see Section C.5.1), we use only the TV radial velocity distributions $p_{11}(v, t)$ corresponding to the recorded human activities of the HAR$^{(1D)}$ dataset.

Each human activity trial was recorded for 10 seconds. The persons were told to maintain the initial and the final poses before and after performing the activity. Though each activity trial was recorded for 10 seconds, the actual duration of the activity was only 2–5 seconds, depending on the type of the activity and the speed at which the activity was carried out. We applied the active segment detection (ASD) [61] approach to the high-pass filtered in-phase component of the raw activity data to automatically extract an active segment, i.e., the section of the raw activity data corresponding to the actual duration of the activity. The ASD marks the start and end points of the activity by monitoring the variance of the filtered in-phase component of the raw activity data. The identified markers are used to extract active segments from the raw IQ activity data of Radar$_1$ and Radar$_2$. Thereafter,

Table C.2: HAR$^{(2D)}$ dataset recorded by a 2×2 MIMO radar system.

Type of human activity	Person #						Total trials
	1	2	3	4	5	6	
Falling	60	59	18	18	—	—	155
Walking	80	80	39	39	23	24	285
Standing up	102	100	27	27	27	27	310
Sitting down	105	102	27	23	22	25	304
Picking an object	101	103	25	27	27	27	310
Total trials	448	444	136	134	99	103	**1364**

Table C.3: HAR$^{(1D)}$ data subset recorded by Radar$_1$, where the direction of motion of the human activities is restricted to merely Scenario 1.

Type of human activity	Person #						Total trials
	1	2	3	4	5	6	
Falling	20	19	6	6	–	–	51
Walking	20	20	10	9	6	6	71
Standing up	35	32	9	9	9	9	103
Sitting down	35	32	9	7	8	8	99
Picking an object	34	33	9	9	9	9	103
Total trials	144	136	43	40	32	32	**427**

we applied radar signal processing techniques (see Section C.5.3) to compute the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ as given in (C.7).

To demonstrate the utility and effectiveness of our proposed multi-perspective distributed MIMO radar approach, we develop three different types of classifiers or HAR systems. First, we develop a SISO$^{(1D)}$ HAR system underlying a mono-static SISO radar configuration (see Section C.5.1). As conventional monostatic radar-based HAR systems only consider human activities performed along the radar boresight, SISO$^{(1D)}$ uses the HAR$^{(1D)}$ dataset for training and testing purposes. Second, to highlight how the classification performance of a HAR system deteriorates by the introduction of different movement directions, we developed a SISO radar-based direction-independent HAR system denoted as SISO$^{(2D)}$ (see Section C.5.2). Unlike SISO$^{(1D)}$, the SISO$^{(2D)}$ HAR system makes use of the HAR$^{(2D)}$ dataset for training and testing purposes because SISO$^{(2D)}$ is designed to classify human activities in multiple directions of motion. Lastly, to significantly improve the classification performance of the SISO$^{(2D)}$ HAR system, we also developed a 2×2 distributed MIMO radar-based direction-independent HAR system denoted as MIMO$^{(2D)}$ (see Section C.5.3). Analogous to the SISO$^{(2D)}$ HAR system, the proposed MIMO$^{(2D)}$ HAR system uses the HAR$^{(2D)}$ dataset for training and testing purposes, because MIMO$^{(2D)}$ also considers the classification of human activities in multiple directions.

In this work, the recorded data from Person 1 and 2 were divided into training and validation datasets and used for the training phase of the DCNN-based SISO$^{(1D)}$, SISO$^{(2D)}$, and MIMO$^{(2D)}$ classifiers. Of this data, 80% was used to train the classifiers, and 20% was used for validation. The recorded data from the rest of the participants—Person 3, 4, 5, and 6—were reserved to test the trained classifiers or HAR systems. In Section C.5.1, C.5.2, and C.5.3, we elucidate the design and development of the SISO$^{(1D)}$, SISO$^{(2D)}$, and MIMO$^{(2D)}$ HAR systems, respectively, along with their results and discussions.

C.5   SISO and Distributed MIMO Radar-Based HAR Systems

C.5.1   Conventional SISO Radar-Based HAR System

In this section, we describe the design of the $\text{SISO}^{(1D)}$ HAR system, which is analogous to a conventional SISO radar-based HAR system. We show the classification performance of the $\text{SISO}^{(1D)}$ HAR system while restricting the human motion parallel to the boresight of Radar_1. Thus, we consider the $\text{HAR}^{(1D)}$ dataset in Table C.3 for the $\text{SISO}^{(1D)}$ HAR system. Recall that the $\text{HAR}^{(1D)}$ dataset contains only the human activities that were carried out in front of Radar_1 in Scenario 1. For all recorded human activities listed in Table C.3, we generated the TV radial velocity distributions $p_{11}(v, t)$ using the data of Radar_1 and converted the pre-processed data to images of size $224 \times 224 \times 3$. Each image representing a human activity is a color image (see Figs. C.2 and C.3) with 224 pixels in the horizontal and vertical dimensions, and the number 3 refers to the red, green and blue (RGB) color channels.

The images of the radial velocity distributions $p_{11}(v, t)$ are used as input feature maps for the feature extraction network $\text{FEN}^{(1D)}$ as depicted in Fig. C.4. We can see from Fig. C.4, that the first, second, and third convolutional layers of $\text{FEN}^{(1D)}$ contain 32, 48, and 64 filter channels, respectively. The dimension of each 2D learnable filter or kernel, also commonly known as kernel dimension k_d, is 6×6 pixels. For each convolutional layer of the $\text{SISO}^{(1D)}$ network, we set the stride parameter to 1 so that the kernels are moved or strode by one pixel at a time. To avoid the problem of overfitting, we used L2 regularization [62] to penalize and eventually eliminate the spike-like weight vectors. The problems of slow convergence and vanishing gradients were mitigated by using the ReLU activation function on the convolutional layers [58]. Furthermore, each convolutional layer in Fig. C.4 is followed by a max-pool layer and a dropout layer. The max-pool layer is of the order 2×2, which downsamples the output of the convolutional layer by a factor of
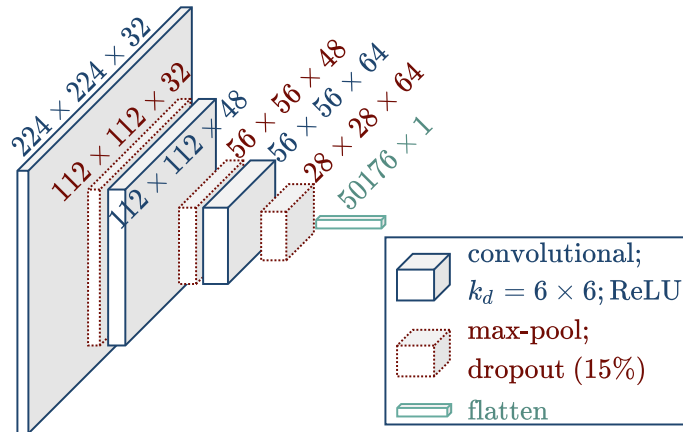
Figure C.4: Feature extraction network $\text{FEN}^{(1D)}$ designed for the $\text{SISO}^{(1D)}$ HAR system.

2. Each max-pool layer is followed by a dropout layer with a dropout rate of 15%. Finally, all the features that are generated by $\text{FEN}^{(1D)}$ are flattened before feeding them to the fully connected layers.

A DCNN-based $\text{SISO}^{(1D)}$ classifier is depicted in Fig. C.5, where $\text{FEN}^{(1D)}$ generates features from the input feature maps or, equivalently, the TV radial velocity distribution $p_{11}(v, t)$. Then, the extracted features undergo two fully connected layers of the order 256×1 and 128×1. As we are classifying five different types of human activities, the second-to-last fully connected layer is followed by an output layer of order 5×1 with the softmax activation function that converts the logits computed by the network into probabilities. To train the $\text{SISO}^{(1D)}$ classifier, the $\text{HAR}^{(1D)}$ dataset (see Table C.3) is divided into training, validation and testing data sets. The training and validation data account for 65.6% of the total data and belong to Persons 1 and 2, while the test data account for 34.4% of the $\text{HAR}^{(1D)}$ dataset belonging to Persons 3, 4, 5, and 6.

In the training phase of the $\text{SISO}^{(1D)}$ HAR system, we used the Adam optimizer to minimize the empirical risk $R_J(\mathcal{C}_f)$ in (C.8) corresponding to the categorical cross-entropy loss function J_{CCE}. Thus, the weights and biases of the DCNN-based $\text{SISO}^{(1D)}$ classifier were optimized by using the Adam optimizer and the examples from the $\text{HAR}^{(1D)}$ dataset. The default values of the decay factors or forgetting factors in (C.10) and (C.11) are equal to $\beta_1 = 0.9$ and $\beta_2 = 0.999$, respectively. In order to prevent division by 0 in (C.12), the value of ϵ was set to be 10^{-8}. A batch size of 32 was adopted in the training phase of the $\text{SISO}^{(1D)}$ classifier. Note that the parameter optimization or training of the three classifiers—$\text{SISO}^{(1D)}$, $\text{SISO}^{(2D)}$, and $\text{MIMO}^{(2D)}$—was performed in the same way with the same values for the network hyperparameters. For all three classifiers, the training history is summarized by the training loss, training accuracy, validation loss, and validation accuracy curves in Fig. C.10. During the training phase, which spans 100 epochs, there is no evidence of overfitting of the $\text{SISO}^{(1D)}$ classifier (see Fig. C.10).
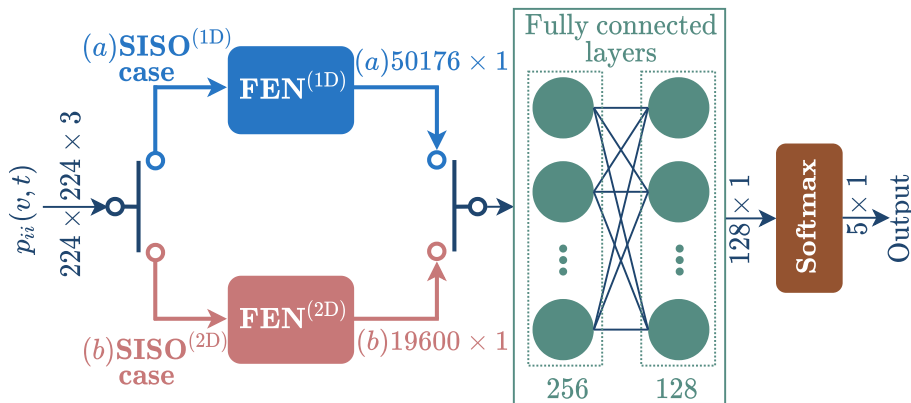
Figure C.5: Architecture of the DCNN classifier for the monostatic SISO radar-based HAR systems, where (a) $\text{SISO}^{(1D)}$ uses $\text{FEN}^{(1D)}$ that outputs a feature vector of dimension 50176×1, and (b) $\text{SISO}^{(2D)}$ uses $\text{FEN}^{(2D)}$ that outputs a feature vector of dimension 19600×1.

We use a confusion matrix shown in Fig. C.6 to summarize and quantitatively assess the overall performance of the trained DCNN-based SISO$^{(1D)}$ classifier. The human activity classification performance of SISO$^{(1D)}$ was evaluated using test-examples from the HAR$^{(1D)}$ dataset. On the y-axis of the confusion matrix, we have the true class of an activity, and the x-axis shows the predicted class of an activity. Thus, for the first five rows and columns of the confusion matrix in Fig. C.6, the diagonal entries show the number of correctly classified human activities, while the non-diagonal entries show the number of misclassified human activities. For example, the first column of the third row shows that a "Stand" activity has been incorrectly predicted or misclassified as a "Fall" activity. Moreover, the first five entries of the last row and last column of the confusion matrix show the precision and recall [63], respectively. Thus, we can see from Fig. C.6 that the walking activity has a 100% recall, and a precision of 96.88%. Most importantly, the overall accuracy of the SISO$^{(1D)}$ classifier is 97.28%, which is indicated by the white color of the sixth entry in the last row and last column of the confusion matrix. It should be noted that using a complex network architecture (FEN$^{(2D)}$) for a smaller dataset (HAR$^{(1D)}$) can lead to overfitting and reduced generalizability. When we conducted experiments by changing the structure of FEN$^{(1D)}$ to FEN$^{(2D)}$ for the SISO$^{(1D)}$ HAR system, as expected, we observed a small decline in the accuracy of the SISO$^{(1D)}$ classifier, which dropped to 96.60% from 97.28%.

In this section, we looked into a conventional SISO radar-based HAR system denoted as SISO$^{(1D)}$ that demonstrated a good classification performance (see Fig. C.6).

	Fall	Walk	Stand	Sit	Pick	
Fall	11	1	0	0	0	91.67% / 8.33%
Walk	0	31	0	0	0	100% / 0%
Stand	1	0	35	0	0	97.22% / 2.78%
Sit	0	0	0	31	1	96.88% / 3.12%
Pick	0	0	1	0	35	97.22% / 2.78%
	91.67% / 8.33%	96.88% / 3.12%	97.22% / 2.78%	100% / 0%	97.22% / 2.78%	97.28% / 2.72%
	Fall	Walk	Stand	Sit	Pick	

True Class (y-axis) / Predicted Class (x-axis)

Figure C.6: Confusion matrix of the results obtained by the SISO$^{(1D)}$ HAR system. The first five entries of the last row and last column show the precision and recall, respectively, whereas the last entry highlighted in dark grey shows the overall accuracy.

The classification performance of the SISO$^{(1D)}$ classifier is comparable to state-of-the-art HAR systems. Analogous to the SISO$^{(1D)}$ classifier or HAR system, most modern HAR systems that are based on either radar or Wi-Fi data are able to classify basic human activities with classification accuracies above 90% [43, 24, 64]. However, in these conventional monostatic radar-based HAR systems, the human subjects' movements are limited to Scenario 1. In Section C.5.2, we extend the human activity recognition problem by considering human motion in the horizontal xy-plane, and investigate how this affects the classification performance of a conventional SISO radar-based direction-independent HAR system.

C.5.2 Direction-Independent SISO Radar-Based HAR System

To provide a comprehensive analysis and ensure a fair comparison, we include the SISO$^{(2D)}$ approach in this section, which is a direction-independent monostatic SISO radar-based HAR system. This inclusion allows us to highlight the limitations of the SISO$^{(2D)}$ HAR system and emphasize the effectiveness of the proposed MIMO$^{(2D)}$ HAR system in addressing diverse directions of human activities. By comparing their performance using the HAR$^{(2D)}$ dataset, we aim to demonstrate the significance of the proposed direction-independent HAR framework. Hence, we use the HAR$^{(2D)}$ dataset as shown in Table C.2 to realize the SISO$^{(2D)}$ HAR system. For all the recorded human activities listed in Table C.2, we generated the TV radial velocity distributions $p_{ii}(v, t)$ by using Radar$_i$ data, where i may be chosen as either 1 or 2. For brevity, we report only the results of the SISO$^{(2D)}$ HAR system trained and tested with the data of Radar$_1$. The TV radial velocity distributions $p_{11}(v, t)$ representing human activity fingerprints were converted into images of the order $224 \times 224 \times 3$ (see Figs. C.2 and C.3), which were used as input feature maps to the feature extraction network FEN$^{(2D)}$ as depicted in Fig. C.7.

The neural network architecture of the SISO$^{(2D)}$ classifier is similar to the SISO$^{(1D)}$ classifier except for a few modifications. For instance, the DCNN-based SISO$^{(2D)}$ classifier uses FEN$^{(2D)}$ instead of FEN$^{(1D)}$ to extract features from the input feature maps or the TV radial velocity distribution $p_{ii}(v, t)$ as shown in Fig. C.5. Compared with FEN$^{(1D)}$ in Fig. C.4, we see that FEN$^{(2D)}$ in Fig. C.7 has an additional convolutional layer, and each convolutional layer has a larger number of filters, i.e., 40, 60, 80, and 100. Consequently, the SISO$^{(2D)}$ HAR system has a greater network complexity and capacity compared to the SISO$^{(1D)}$ HAR system. Note that we needed a more complex DCNN classifier with higher network capacity because (a) SISO$^{(2D)}$ uses a larger HAR$^{(2D)}$ dataset containing 1364 human activity fingerprints instead of 427, and (b) because SISO$^{(2D)}$ aims to classify human activities in different directions, taking into account more diverse, complex, and sometimes suboptimal human activity signatures.

Moreover, the kernel dimension k_d of each 2D learnable filter in FEN$^{(2D)}$ is 5×5 as shown in Fig. C.7. The rest of the specifications of the SISO$^{(2D)}$ and SISO$^{(1D)}$ classifiers are similar in terms of the max-pool layers, dropout layers, stride, batch
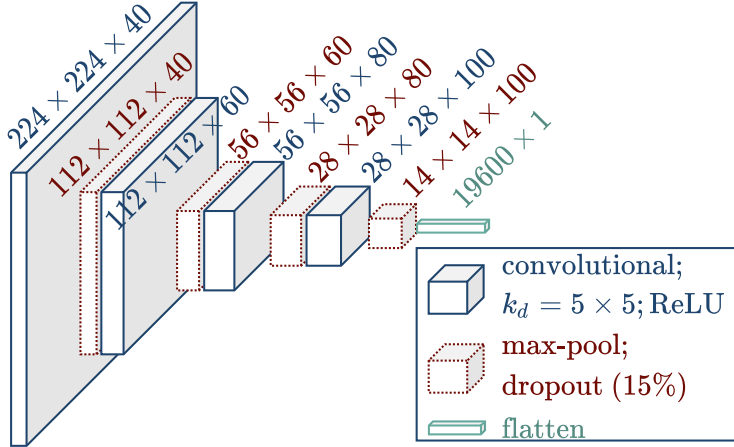
Figure C.7: Feature extraction network FEN$^{(2D)}$ designed for SISO$^{(2D)}$ and MIMO$^{(2D)}$ HAR systems.

size, and activation function. Analogous to SISO$^{(1D)}$, SISO$^{(2D)}$ uses L2 regularization to penalize and eliminate the peaky weight vectors to avoid the overfitting problem. Like SISO$^{(1D)}$, SISO$^{(2D)}$ uses the Adam optimizer to minimize the empirical risk $R_J(\mathcal{C}_f)$ in (C.8) corresponding to the categorical cross-entropy loss function J_{CCE}. In order to train the SISO$^{(2D)}$ classifier, the HAR$^{(2D)}$ dataset is split into training, validation, and testing data sets. The training and validation data is 65.4% of the total data and belongs to Person 1 and 2, whereas the testing data is 34.6% of the HAR$^{(2D)}$ dataset belonging to Person 3, 4, 5, and 6. As mentioned in Section C.5.1, the training history is summarized by the training loss, training accuracy, validation loss, and validation accuracy curves shown in Fig. C.10 for all three classifiers. Note that for the SISO$^{(2D)}$ classifier, there is no evidence of overfitting during the training phase that spans over 100 epochs, as shown in Fig. C.10.

The classification performance of the SISO$^{(2D)}$ direction-independent HAR system was evaluated using the test-examples from the HAR$^{(2D)}$ dataset. Recall that the SISO$^{(2D)}$ HAR system is realized by using the data of Radar$_1$. To summarize and quantitatively assess the classification performance of the SISO$^{(2D)}$ HAR system, we present a confusion matrix in Fig. C.8. The predicted and actual class of a human activity is shown on the x-axis and y-axis of the confusion matrix, respectively. The confusion matrix in Fig. C.8 shows that the overall classification performance of the SISO$^{(2D)}$ HAR system has dropped significantly to only 88.98%. On a partially unrelated note and without going into too much details, we would also like to mention that Radar$_2$ provides relatively poor data quality due to the cross-channel interference problem [65]. Solving the cross-channel interference problem requires the deployment of longer RF cables (see Table C.1), which cause a higher attenuation of the received signal. For this reason, a SISO$^{(2D)}$ direction-independent HAR system realized by using only the data of Radar$_2$ provided an overall classification accuracy of just 83.05%.

Looking at the non-diagonal entries of the confusion matrix in Fig. C.8, we see numerous misclassified human activities, e.g., the "Pick" activity was misclassified

	Fall	Walk	Stand	Sit	Pick	
Fall	36	0	0	0	0	100% 0%
Walk	0	116	9	0	0	92.8% 7.2%
Stand	0	3	100	3	2	92.59% 7.41%
Sit	0	4	7	83	3	85.57% 14.43%
Pick	0	0	15	6	85	80.19% 19.81%
	100% 0%	94.31% 5.69%	76.34% 23.66%	90.22% 9.78%	94.44% 5.56%	88.98% 11.02%
	Fall	Walk	Stand	Sit	Pick	

True Class (rows) — Predicted Class (columns)

Figure C.8: Confusion matrix of the results obtained by the SISO$^{(2D)}$ HAR system, where SISO$^{(2D)}$ was trained and tested by using Radar$_1$ data. The first five entries of the last row and last column show the precision and recall, respectively, whereas the last entry shows the overall accuracy.

15 times as the "Stand" activity by the SISO$^{(2D)}$ HAR system. Therefore, the worst precision of the system is 76.34% corresponding to the "Stand" activity, and the worst recall is observed as 80.19% for the "Pick" activity. Interestingly, the precision and recall are 100% for the "Fall" activity, which implies that the SISO$^{(2D)}$ HAR system learned to classify the human falling activity in all directions. Unfortunately, this is not true for the other four types of human activity, which have diverse and relatively complex radial velocity distributions that vary in different directions (see Figs. C.2 and C.3).

In this section, a direction-independent SISO radar-based HAR system (SISO$^{(2D)}$) was investigated, which showed significant degradation in its classification performance for human motion in different directions. For the simpler case of human motion, or when the human motion was restricted to Scenario 1 in Fig. C.1, the overall classification accuracy of the SISO$^{(1D)}$ HAR system was 97.28%. However, when we complicated the human motion by considering the different directions of motion, the classification accuracy dropped to 88.98% for the SISO$^{(2D)}$ HAR system. The deterioration of the classification performance manifested by the SISO$^{(2D)}$ HAR system comes from the physical limitations of monostatic SISO radar systems. These physical limitations of monostatic radar systems can be overcome by the 2×2 distributed MIMO radar configuration of Fig. C.1 to eventually realize a direction-independent MIMO$^{(2D)}$ HAR system. In Section C.5.3, we will see how the MIMO$^{(2D)}$ HAR system ameliorates the shortcomings of the SISO$^{(1D)}$ and SISO$^{(2D)}$ HAR systems altogether.

C.5.3   2×2 MIMO Radar-Based Direction-Independent HAR System

We now elucidate the design of our proposed 2×2 distributed MIMO radar-based direction-independent HAR system denoted as $\text{MIMO}^{(2D)}$. Considering the different directions of human activities in the horizontal xy-plane in Fig.C.1, we use the $\text{HAR}^{(2D)}$ dataset (see Table C.2) to eventually realize the $\text{MIMO}^{(2D)}$ HAR system. In this section, we demonstrate that unlike the $\text{SISO}^{(2D)}$ HAR system, our proposed $\text{MIMO}^{(2D)}$ HAR system is able to recognize the human activities with a very good classification performance for the $\text{HAR}^{(2D)}$ dataset. For all the recorded human activities listed in Table C.2, we computed the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ by using Radar_1 and Radar_2 data, respectively. The TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ were converted separately into images of the order $224 \times 224 \times 3$ (see Figs. C.2 and C.3), which served as input feature maps to the feature extraction network $\text{FEN}^{(2D)}$ as depicted in Fig. C.7.

Although the neural network architecture of the $\text{MIMO}^{(2D)}$ and $\text{SISO}^{(2D)}$ HAR systems are quite different in Fig. C.9 and Fig. C.5, respectively, the building blocks, hyperparameter values, and training processes of the two networks are very similar. For instance, the $\text{MIMO}^{(2D)}$ and $\text{SISO}^{(2D)}$ HAR systems use the same specifications related to kernel dimension k_d, max-pool layers, dropout layers, stride, batch size, activation function, regularizer, and Adam optimizer (refer to Section C.5.2 for more details). Moreover, the same feature extraction network $\text{FEN}^{(2D)}$ in Fig. C.7 has been adopted for the $\text{MIMO}^{(2D)}$ and $\text{SISO}^{(2D)}$ HAR systems. However, unlike the $\text{SISO}^{(2D)}$ HAR system, the $\text{MIMO}^{(2D)}$ HAR system uses two identical feature extraction blocks as depicted in Fig. C.9 for the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$. The two $\text{FEN}^{(2D)}$ blocks of $\text{MIMO}^{(2D)}$ HAR system extract unique features automatically and independently from the two radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$. In Fig. C.9, we can see that these features are then merged using a concatenation layer, which is followed by MLP and softmax layers
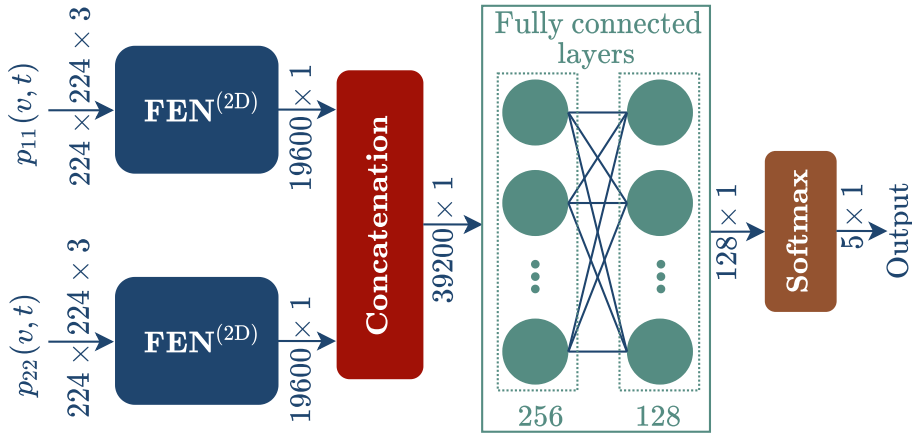
Figure C.9: Architecture of the proposed $\text{MIMO}^{(2D)}$ HAR system with two independent $\text{FEN}^{(2D)}$ blocks to generate feature vectors that are fused by the concatenation layer for subsequent classification.

to eventually classify the human activities.

Analogous to the SISO$^{(2D)}$ HAR system, the MIMO$^{(2D)}$ HAR system also uses the HAR$^{(2D)}$ dataset for the training and testing purposes. However, for the MIMO$^{(2D)}$ HAR system, the main difference is that the activity fingerprints from both radar subsystems shown in Fig. C.1 are simultaneously utilized to classify the human activities. In other words, for the classification of a human activity, two distinct multi-perspective radial velocity distributions $p_{11}(v,t)$ and $p_{22}(v,t)$ produced by Radar$_1$ and Radar$_2$, respectively, are processed at once by the MIMO$^{(2D)}$ HAR system. Therefore, in the MIMO$^{(2D)}$ HAR system, we utilized 2728 images or equivalently 1364 pairs of images corresponding to 1364 human activities of the HAR$^{(2D)}$ dataset. The HAR$^{(2D)}$ dataset was split into training, validation and testing data sets, where the training and validation data was 65.4% of the total data belonging to Person 1 and 2, and the testing data was 34.6% of the total data belonging to Person 3, 4, 5, and 6. Recall that the training history is summarized by the training loss, training
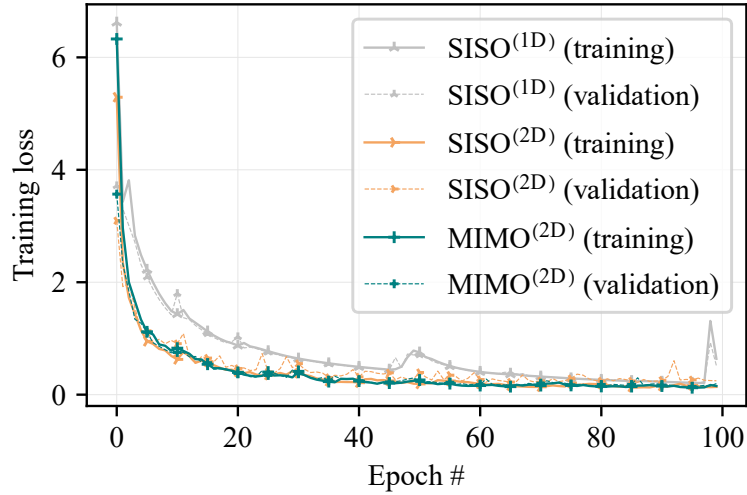
(a)
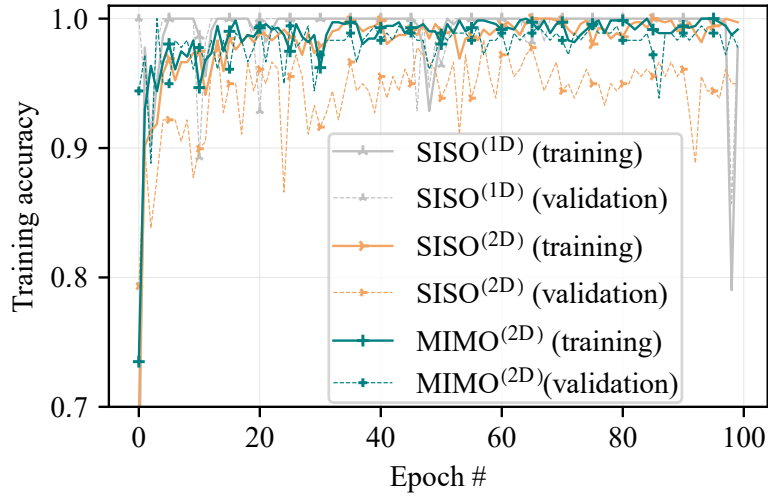
(b)

Figure C.10: Training history for the SISO$^{(1D)}$, SISO$^{(2D)}$, and MIMO$^{(2D)}$ HAR systems: (a) training losses, and (b) training accuracies over 100 epochs.

accuracy, validation loss, and validation accuracy curves as depicted in Fig. C.10 for all three classifiers or HAR systems. Note that this figure does not reveal any signs of overfitting during the training phase of the MIMO$^{(2D)}$ HAR system.

In Fig. C.11, we present a confusion matrix to quantitatively assess the overall classification performance of the MIMO$^{(2D)}$ direction-independent HAR system. The human activity classification performance of the MIMO$^{(2D)}$ HAR system was evaluated over the test examples from the HAR$^{(2D)}$ dataset. In the test examples, the number of falling activities is comparatively low because it is difficult to carry out a real-life "Fall" activity. Nevertheless, the train–test split ratio is roughly 77 : 23 for the "Fall" activity. In the confusion matrix in Fig. C.11, the overall classification performance of the MIMO$^{(2D)}$ direction-independent HAR system comes out to be 98.52%, which is a significant improvement over the classification accuracy of 88.98% achieved by the SISO$^{(2D)}$ direction-independent HAR system. Looking at the non-diagonal entries of the confusion matrix in Fig. C.11, we see only seven misclassified human activities. We can observe that the worst precision of the MIMO$^{(2D)}$ HAR system is 95.54% corresponding to the "Stand" activity, and the worst recall is observed as 94.34% for the "Pick" activity. Note that the increase in the classification performance is basically due to the multi-perspective illumination of the environment by the proposed 2×2 distributed MIMO radar-based HAR system.

We addressed a human activity recognition task in a complex situation, where we considered the human motion in the horizontal xy-plane in Fig. C.1. To mitigate the shortcomings of the SISO$^{(2D)}$ HAR system in relation to the human activity direction, we illuminated the subject from different aspect angles by using the proposed 2×2 MIMO radar-based direction-independent HAR system denoted as MIMO$^{(2D)}$,

Figure C.11: Confusion matrix of the results obtained by the proposed MIMO$^{(2D)}$ HAR system with an overall accuracy of 98.52%.

which demonstrated a remarkably good classification performance as summarized by the confusion matrix in Fig. C.11. As evident from the classification performance of the MIMO$^{(2D)}$ HAR system, the physical limitations of the monostatic radar systems were successfully mitigated by the multi-perspective 2×2 distributed MIMO radar configuration. Therefore, by addressing and rectifying the fundamental radar problem at the physical layer, we were able to design a radar-based HAR system that was capable of recognizing human activities independent of their directions with a classification accuracy close to 100%.

C.6   Conclusion

In this article, we analyzed and resolved a crucial physical layer problem of state-of-the-art monostatic SISO, SIMO and MIMO radar-based HAR systems, which primarily arises due to the target's aspect angle. Thus, a more pragmatic and more complex HAR problem has been elucidated in this research in the context of RF sensing, where we improve the activity recognition task by considering multiple directions of human activities. A novel HAR dataset (HAR$^{(2D)}$) was recorded by using the proposed multi-perspective 2×2 MIMO radar framework. We developed and analyzed three different HAR systems, denoted as SISO$^{(1D)}$, SISO$^{(2D)}$, and MIMO$^{(2D)}$, by using our HAR$^{(2D)}$ dataset and its sub-dataset HAR$^{(1D)}$.

Analogous to most modern radar-based HAR systems, the SISO$^{(1D)}$ HAR system was able to classify human activities with a classification accuracy of 97.28%. However, in this conventional monostatic radar-based HAR approach, the movement of the human subjects were restricted along the radar's boresight axis. By developing and analyzing the monostatic SISO$^{(2D)}$ HAR system and considering the human activities taking place in the 2D xy-plane, we substantiated a significant deterioration in the classification performance from 97.28% to 88.98%. The deterioration of the classification performance manifested by the SISO$^{(2D)}$ HAR system came from the inherent physical layer limitations of the monostatic SISO radar systems. To overcome these physical layer issues and drawbacks experienced by today's radar-based HAR systems, we utilized a multi-perspective 2×2 distributed MIMO radar system to realize a direction-independent HAR system that was capable of recognizing human gross motor activities regardless of the aspect angle or direction of motion. To eradicate the limitations that emerge due to the direction of motion, feature level fusion was adopted in the DCNN-based MIMO$^{(2D)}$ classifier, which virtually combines the target's information from different aspect angles.

For the HAR$^{(2D)}$ dataset, it was shown that the proposed multi-perspective MIMO$^{(2D)}$ HAR system significantly outperforms the monostatic SISO$^{(2D)}$ HAR system. Compared with the SISO$^{(2D)}$ HAR system, the proposed MIMO$^{(2D)}$ HAR system significantly improved the classification accuracy from 88.98% to 98.52%. Therefore, the physical layer limitations of the monostatic SISO radar-based HAR systems were successfully mitigated by the proposed MIMO$^{(2D)}$ HAR system.

The MIMO$^{(2D)}$ HAR system presented in this article paves a way forward towards

actualizing a more realistic and more advanced radar-based HAR system. To further enhance the classification performance, we plan to use the bistatic components of the 2×2 MIMO radar system, which are the TV radial velocity distributions $p_{12}(v,t)$ and $p_{21}(v,t)$. For more aspect angle coverage and a more complex HAR problem, we plan to extend the fundamental distributed 2×2 MIMO radar system to a larger MIMO antenna configuration.

C.7   Appendix

Multi-Directional Human Activity Signatures

In this appendix, we provide the measured radial velocity distributions $p_{ii}(v,t)$ of the four types of human activities: walking, standing up from a chair, sitting down on a chair, and picking up an object from the floor.
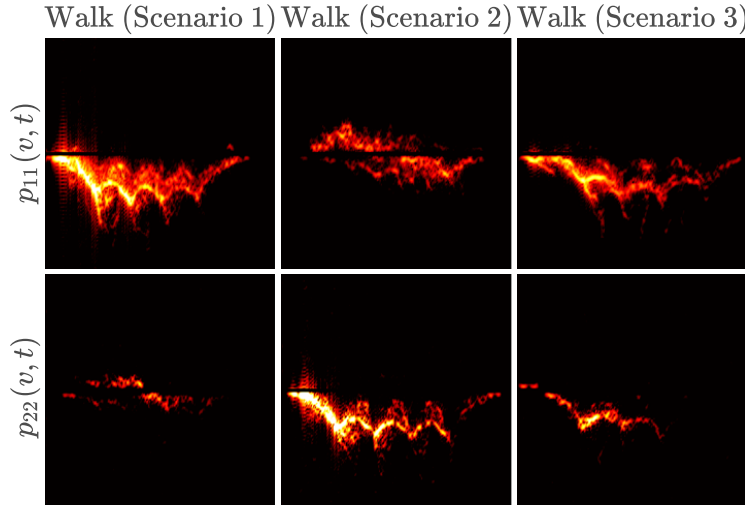
Figure C.12: Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of the "Walk" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over $3 - 5$ s.
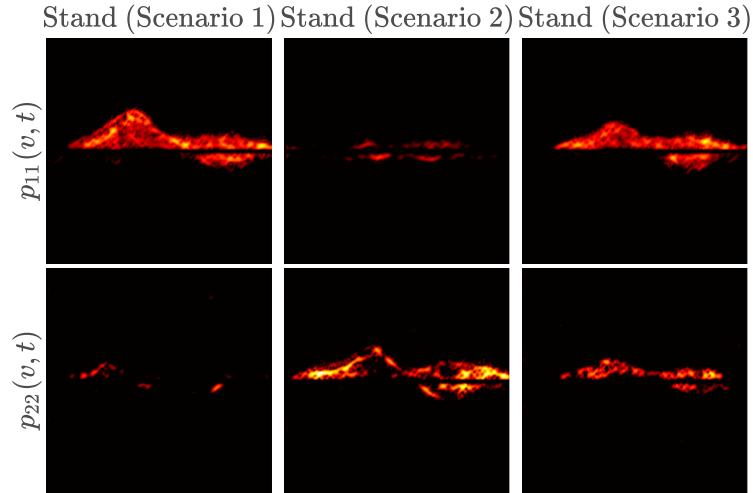
Figure C.13: Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v, t)$ of the "Stand" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over $2 - 3$ s.
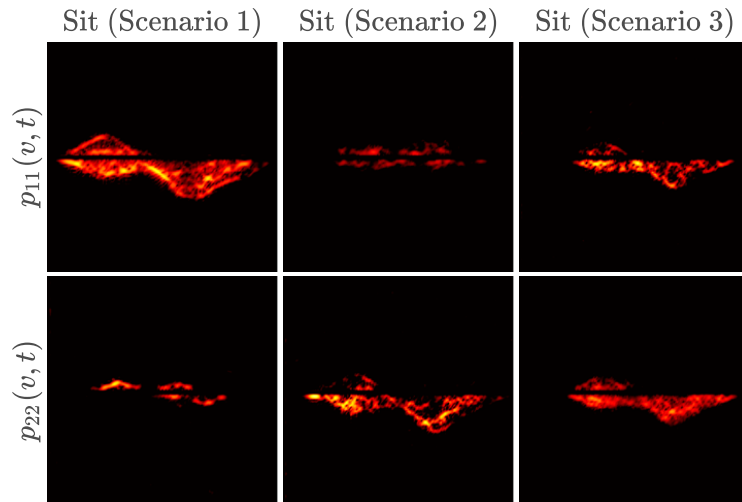
Figure C.14: Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v, t)$ of the "Sit" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over $2 - 3$ s.
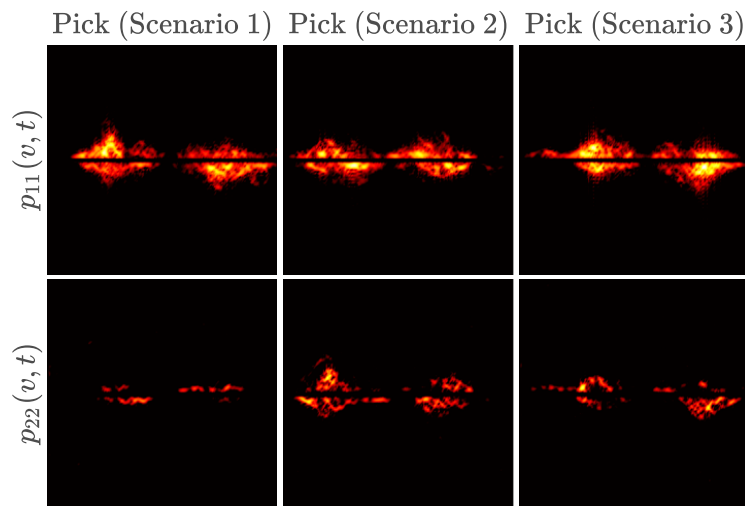
Figure C.15: Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of the "Pick" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over $2-4$ s.

Bibliography

[1] F. Gu, M. H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu, "A survey on deep learning for human activity recognition," *ACM Comput. Surv.*, vol. 54, no. 8, 2022.

[2] M. Straczkiewicz, P. James, and J.-P. Onnela, "A systematic review of smartphone-based human activity recognition methods for health research," *npj Digit. Med.*, vol. 4, no. 1, p. 148, 2021.

[3] Y. Guo, Y. Chu, B. Jiao, J. Cheng, Z. Yu, N. Cui, and L. Ma, "Evolutionary Dual-Ensemble Class Imbalance Learning for Human Activity Recognition," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 4, pp. 728–739, 2022.

[4] Z. Ma, L. T. Yang, M. Lin, Q. Zhang, and C. Dai, "Weighted Support Tensor Machines for Human Activity Recognition with Smartphone Sensors," *IEEE Trans. Ind. Informatics*, 2021.

[5] F. J. Rodriguez Lera, F. Martín Rico, A. M. Guerrero Higueras, and V. M. Olivera, "A context-awareness model for activity recognition in robot-assisted scenarios," *Expert Syst.*, vol. 37, no. 2, 2020.

[6] M. Tammvee and G. Anbarjafari, "Human activity recognition-based path planning for autonomous vehicles," *Signal, Image Video Process.*, vol. 15, no. 4, pp. 809–816, 2021.

[7] K. Host and M. Ivašić-Kos, "An overview of Human Action Recognition in sports based on Computer Vision," *Heliyon*, vol. 8, no. 6, p. e09633, 2022.

[8] T. Tuncer, F. Ertam, S. Dogan, and A. Subasi, "An automated daily sports activities and gender recognition method based on novel multikernel local diamond pattern using sensor signals," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9441–9448, 2020.

[9] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Electr. Eng. Technol.*, vol. 14, no. 6, pp. 2567–2573, 2019.

[10] M. Al-khafajiy, T. Baker, C. Chalmers, M. Asim, H. Kolivand, M. Fahim, and A. Waraich, "Remote health monitoring of elderly through wearable sensors," *Multimed. Tools Appl.*, vol. 78, no. 17, pp. 24681–24706, 2019.

[11] M. Rashmi, T. S. Ashwin, and R. M. R. Guddeti, "Surveillance video analysis for student action recognition and localization inside computer laboratories of a smart campus," *Multimed. Tools Appl.*, vol. 80, no. 2, pp. 2907–2929, 2021.

[12] F. Gurcan, N. E. Cagiltay, and K. Cagiltay, "Mapping human–computer interaction research themes and trends from its existence to today: A topic modeling-based review of past 60 years," *Int. J. Hum. Comput. Interact.*, vol. 37, no. 3, pp. 267–280, 2021.

[13] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "A review of multi-modal human activity recognition with special emphasis on classification, applications, challenges and future directions," *Knowledge-Based Syst.*, vol. 223, p. 106970, 2021.

[14] J. Roche, V. De-Silva, J. Hook, M. Moencks, and A. Kondoz, "A multimodal data processing system for LiDAR-based human activity recognition," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10027–10040, 2022.

[15] W. Qi, N. Wang, H. Su, and A. Aliverti, "DCNN based human activity recognition framework with depth vision guiding," *Neurocomputing*, vol. 486, pp. 261–271, 2022.

[16] M. A. A. Al-qaness, A. Dahou, M. A. Elaziz, and A. M. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," *IEEE Trans. Ind. Informatics*, vol. 19, no. 1, pp. 144–152, 2023.

[17] Y. Tang, L. Zhang, F. Min, and J. He, "Multiscale deep feature learning for human activity recognition using wearable sensors," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 2106–2116, 2023.

[18] Y. Tang, L. Zhang, Q. Teng, F. Min, and A. Song, "Triple cross-domain attention on human activity recognition using wearable sensors," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 5, pp. 1167–1176, 2022.

[19] W. Huang, L. Zhang, W. Gao, F. Min, and J. He, "Shallow convolutional neural networks for human activity recognition using wearable sensors," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021.

[20] L. Zhang, W. Cui, B. Li, Z. Chen, S. Member, M. Wu, S. Member, and T. S. Gee, "Privacy-preserving cross-environment human activity recognition," *IEEE Trans. Cybern.*, pp. 1–11, 2021.

[21] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, "RF-based human activity recognition using signal adapted convolutional neural network," *IEEE Trans. Mob. Comput.*, vol. 22, no. 1, pp. 487–499, 2023.

[22] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022.

[23] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, and M. Wu, "Two-stream convolution augmented transformer for human activity recognition," *35th AAAI Conf. Artif. Intell. AAAI 2021*, vol. 35, no. 1, pp. 286–293, 2021.

[24] M. Muaaz, A. Chelli, and M. Patzold, "WiHAR: From Wi-Fi channel state information to unobtrusive human activity recognition," in *2020 IEEE 91st Veh. Technol. Conf.*, pp. 1–7, 2020.

[25] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3617–3628, 2019.

[26] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, 2017.

[27] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 363–373, 2016.

[28] Z. Peng, J. M. Munoz-Ferreras, Y. Tang, C. Liu, R. Gomez-Garcia, L. Ran, and C. Li, "A portable FMCW interferometry radar with programmable low-IF architecture for localization, ISAR imaging, and vital sign tracking," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 4, pp. 1334–1344, 2017.

[29] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, "SimHumalator: An open-source end-to-end radar simulator for human activity recognition," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 3, pp. 6–22, 2022.

[30] S. Waqar, H. Yusaf, S. Sana, M. Waqas, and F. A. Siddiqui, "Reconfigurable monopulse radar tracking processor," in *2018 15th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 805–809, 2018.

[31] L. Pallotta, M. Cauli, C. Clemente, F. Fioranelli, G. Giunta, and A. Farina, "Classification of micro-Doppler radar hand-gesture signatures by means of Chebyshev moments," in *2021 IEEE 8th Int. Work. Metrol. Aerosp.*, pp. 182–187, 2021.

[32] C. Clemente, L. Pallotta, A. De Maio, J. J. Soraghan, and A. Farina, "A novel algorithm for radar classification based on doppler characteristics exploiting orthogonal Pseudo-Zernike polynomials," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 417–430, 2015.

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

[34] L. Pallotta, C. Clemente, A. De Maio, J. J. Soraghan, and A. Farina, "Pseudo-Zernike moments based radar micro-Doppler classification," in *2014 IEEE Radar Conf.*, pp. 0850–0854, may 2014.

[35] B. Erol and M. Amin, "Effects of range spread and aspect angle on radar fall detection," in *2016 IEEE Sens. Array Multichannel Signal Process. Work.*, pp. 1–5, 2016.

[36] C. Ding, H. Hong, Y. Zou, H. Chu, X. Zhu, F. Fioranelli, J. Le Kernec, and C. Li, "Continuous human motion recognition with a dynamic range-Doppler

trajectory method based on FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6821–6831, 2019.

[37] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on Doppler radar motion signature," in *2011 5th Int. Conf. Pervasive Comput. Technol. Healthc. Work.*, pp. 222–225, 2011.

[38] L. Liu, M. Popescu, K. C. Ho, M. Skubic, and M. Rantz, "Doppler radar sensor positioning in a fall detection system," in *2012 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 256–259, 2012.

[39] P. Nallabolu, L. Zhang, H. Hong, and C. Li, "Human presence sensing and gesture recognition for smart home applications with moving and stationary clutter suppression using a 60-GHz digital beamforming FMCW radar," *IEEE Access*, vol. 9, pp. 72857–72866, 2021.

[40] S. Sana, S. Waqar, H. Yusaf, M. Waqas, and F. A. Siddiqui, "Software defined digital beam forming processor," in *2016 13th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 671–676, 2016.

[41] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *2015 IEEE Radar Conf.*, pp. 1491–1496, 2015.

[42] C. Yu, Z. Xu, K. Yan, Y.-R. Chien, S.-H. Fang, and H.-C. Wu, "Noninvasive human activity recognition using millimeter-wave radar," *IEEE Syst. J.*, vol. 16, no. 2, pp. 3036–3047, 2022.

[43] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proc. 3rd ACM Work. Millimeter-wave Networks Sens. Syst.*, pp. 51–56, 2019.

[44] J. Pan, M. Sun, Y. Wang, and X. Zhang, "An enhanced spatial smoothing technique with ESPRIT algorithm for direction of arrival estimation in coherent scenarios," *IEEE Trans. Signal Process.*, vol. 68, pp. 3635–3643, 2020.

[45] Z. Wang, Z. Yang, S. Wu, H. Li, S. Tian, and X. Chen, "An improved multiple signal classification for nonuniform sampling in blade tip timing," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7941–7952, 2020.

[46] M. Jian, Z. Lu, and V. C. Chen, "Drone detection and tracking based on phase-interferometric Doppler radar," in *2018 IEEE Radar Conf.*, pp. 1146–1149, 2018.

[47] S. Kim, B.-S. Kim, Y. Jin, and J. Lee, "Extrapolation-RELAX estimator based on spectrum partitioning for DOA estimation of FMCW radar," *IEEE Access*, vol. 7, pp. 98771–98780, 2019.

[48] F. Fioranelli, M. Ritchie, and H. Griffiths, "Bistatic human micro-Doppler signatures for classification of indoor activities," in *2017 IEEE Radar Conf.*, pp. 610–615, 2017.

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

[50] B. Jokanovic and M. Amin, "Fall detection using deep learning in range-Doppler radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, 2018.

[51] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, 2019.

[52] M. Bekar, C. J. Baker, E. G. Hoare, and M. Gashinova, "Joint MIMO radar and communication system using a PSK-LFM waveform with TDM and CDM approaches," *IEEE Sens. J.*, vol. 21, no. 5, pp. 6115–6124, 2021.

[53] S. Waqar, M. Muaaz, and M. Pätzold, "Human activity signatures captured under different directions using SISO and MIMO radar systems," *Appl. Sci.*, vol. 12, no. 4, 2022.

[54] M. Muaaz, S. Waqar, and M. Pätzold, "Radar-based passive step counter and its comparison with a wrist-worn physical activity tracker," in *4th Int. Conf. Intell. Technol. Appl. INTAP*, pp. 259–272, Springer International Publishing, 2022.

[55] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.

[56] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, pp. 8792–8802, 2018.

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[58] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," *Proc. Int. Jt. Conf. Neural Networks*, pp. 2684–2691, 2017.

[59] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[60] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv Prepr.*, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980.

Bibliography

[61] M. Muaaz and R. Mayrhofer, "Smartphone-based gait recognition: from authentication to imitation," *IEEE Trans. Mob. Comput.*, vol. 16, no. 11, pp. 3209–3221, 2017.

[62] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *7th Int. Conf. Learn. Represent. ICLR*, 2019.

[63] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. ICML*, pp. 233–240, 2006.

[64] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-Sense: a passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems," *Ann. Telecommun.*, vol. 77, no. 3-4, pp. 163–175, 2022.

[65] S. Waqar and M. Pätzold, "Interchannel interference and mitigation in distributed MIMO RF sensing," *Sensors*, vol. 21, no. 22, 2021.

Appendix D

Paper D

A Simulation-Based Framework for the Design of Human Activity Recognition Systems Using Radar Sensors

Sahil Waqar and Matthias Pätzold

Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway

E-mails: {sahil.waqar, matthias.paetzold}@uia.no

Abstract— Modern human activity recognition (HAR) systems are designed using large amounts of experimental data. So far, real-data-driven or experimental-based HAR systems using Wi-Fi or radar systems have shown considerable results. However, the acquisition of large, clean, and labeled training datasets remains a crucial impediment to the progress of experimental-based HAR systems. Therefore, in this paper, a paradigm shift from the experimental to a fully simulation-based design of HAR systems is proposed in the context of radar sensing. An end-to-end simulation framework is proposed as a proof-of-concept that can simulate realistic millimeter-wave radar signatures for synthesized human motion. We designed a human motion synthesis tool that emulates different types of human activities and generates the spatial trajectories accordingly. These trajectories are processed by a geometric model with respect to user-defined antenna configurations. Considering the long- and short-time stationarity of wireless channels, we synthesize the raw in-phase and quadrature data and process the data to simulate the radar signatures for emulated human activities. Finally, a simulated and a real HAR dataset were used to train and test a simulation-based HAR system, respectively, which gave an average (maximum) classification accuracy of 94% (98.4%). The main advantage of the proposed simulation framework is that the training effort for radar-based classifiers, e.g., gesture recognition systems, can be minimized drastically.

Index Terms— Data augmentation, data generation, deep learning, human activity recognition (HAR), micro-Doppler analysis, Mixamo animation, motion synthesis, multiclass classification, radar simulation.

D.1 Introduction

D.1.1 Background

Within the domain of radio frequency (RF) sensing, an important and continuously evolving research area is human activity recognition (HAR), where the classification performance greatly depends on the quality, impartiality, and comprehensiveness of experimental data. Such merits of empirical data are hard to come by, especially when dealing with real humans as subjects. Over the years, researchers have endeavoured to classify different types of human activities using several sensing modalities, such as vision [1, 2], wearable [3, 4, 5], and RF sensors [6, 7, 8, 9, 10].

Various sensor types have been employed in HAR systems, each with distinct advantages and limitations [11]. Vision sensors, driven by advanced computer vision methods, have shown significant success in HAR [1]. However, they are vulnerable to lighting conditions and privacy concerns, unlike RF sensors. Wearable sensors [4], though effective, face criticism due to their fragility, intrusiveness, and reliance on user care. The need for continuous wear renders them impractical, particularly for elderly and ill individuals. Hence, RF sensors, particularly millimeter wave (mm-wave) radars, have garnered growing interest despite the challenges and complexities they entail [12].

In this research, we primarily focus on developing a human activity classification system using mm-wave radar technology. The collection of radar micro-Doppler signatures corresponding to real human subjects is a time-consuming, expensive, and laborious task. The recorded radar dataset usually has a narrow scope because of its validity for a particular scenario and fixed radar parameters. To create diverse training datasets for radar-based HAR systems, a simulation-based approach becomes a compelling and viable alternative.

We design a fully simulation-based HAR system that exclusively relies on simulated radar data for training and validation. Unlike conventional methods, we avoid the use of real radar micro-Doppler signatures during these stages. Instead, experimental measurements from a real mm-wave radar system are solely employed for testing, showcasing our simulation-based HAR system's real-world performance. To ensure accurate radar system modeling and realistic radar micro-Doppler signature simulation, we adopt scatterer-level signal modeling (see Sect. D.6). This proof-of-concept approach facilitates the generation of diverse simulated radar micro-Doppler signatures, thereby providing essential training data for simulation-based HAR systems.

D.1.2 Our Approach

In this paper, we present an end-to-end simulation framework for HAR using frequency-modulated continuous wave (FMCW) radar systems. First, we design a human motion synthesis tool using the Unity software [13] from Unity Software Inc. that emulates different types of human activities and accordingly generates the three-dimensional (3D) trajectories for the virtual markers placed on a humanoid charac-

ter. The 3D marker trajectories are processed by a geometric model (see Sect. D.5) with respect to a user-defined antenna configuration. Taking into account the long- and short-time stationarity properties of wireless channels and using our radar signal synthesizer, we simulate the raw in-phase and quadrature (IQ) components. Finally, the radar micro-Doppler signatures or, equivalently, the time-variant (TV) radial velocity distributions are generated for several types of emulated human activities.

Our proposed simulation-based framework offers several advantages over experimental-based designs, such as flexibility to simulate radar datasets with specific distributions or target motion characteristics, ability to augment training data, cost-effectiveness, and mitigation of legal and privacy issues. With the proposed simulation framework, we can augment human motion data at a motion-synthesis layer, e.g., by varying an avatar's size and speed. The proposed simulation framework gives control over several radar parameters as well, thus it enables us to generate different types of training datasets corresponding to different radar-operating conditions and different applications. Above all, the proposed simulation framework drastically minimize the overall training effort of radar-based HAR systems.

Note that our simulation-based framework, basically designed for HAR, has versatile applications across various domains, including gesture recognition [14], sports [15], autonomous vehicles [16], social robotics [17], and smart homes [18]. In this research, our validation process involves real experiments covering five human activities, highlighting the effectiveness of our proof-of-concept. The core strength of this simulation-based framework, however, lies in its innovative capability to translate motion capture (MoCap) data into radar data (see Sect. D.4 and Sect. D.6), making it adaptable to a wide array of real-world scenarios. The availability of extensive online MoCap data repositories like Mixamo [19], covering domains such as sports, multimedia, healthcare, and more, further enhances the framework's applicability. With our proposed framework, these repositories can be used to simulate radar signatures for a multitude of real-world scenarios. For instance, in healthcare, we demonstrate the framework's capability for fall detection, providing a tangible example of its real-world utility. In sports, our solution can be extended to simulate radar signatures for activities such as running, swimming, and various exercises, thereby enhancing its practicality.

Changes in radar configurations in practical applications, driven by shifts in operational requirements, technological advancements, and emerging applications, necessitate the generation of new datasets. For emerging radar-based classifiers, the need to simulate new datasets is inevitable as it aligns with the dynamic nature of radar sensors. This constraint is common to all radar-based classifiers, whether realized using simulation or experimental data. Our simulation-based framework stands out for its efficient and rapid generation of diverse datasets for new or modified radar configurations, presenting a more resource-efficient alternative compared to the classifiers based on experimental data. The ability to swiftly and easily adapt to varied radar configurations stands as a distinctive strength of our proposed framework.

D.1.3 Contributions

The multiple contributions of this research can be summarized as follows:

1. We propose a novel end-to-end simulation framework to avoid the need of real radar data for training. By using the proposed simulation framework, large quantities of realistic synthetic radar data are generated for human gross motor activities. It is worth noting that the proposed simulation framework is also useful for many other radar-based classifiers, for instance, gesture recognition systems.

2. We leverage a geometrical 3D indoor channel model (see Sect. D.5) to simulate TV radial distances from the spatial trajectories of an avatar with 21 non-stationary virtual markers. By employing the proposed approach, we emulate and diversify various human activities by varying parameters such as location, speed, acceleration, deceleration, and avatar's height. Our unique simulation framework offers the flexibility to augment data at the motion-synthesis layer, enabling the generation of diverse and customizable datasets for training HAR systems.

3. We simulate high-fidelity radar signatures, namely TV range distribution, TV radial velocity distribution, and TV mean velocity for the emulated human activities. By computing the dynamic time warping (DTW) distance metric [20], it is shown that the simulated radar signatures closely resemble the radar signatures measured in reality. This shows the effectiveness of our simulation framework, which can simulate realistic radar signatures for adults and children alike, and can even be extended to simulate realistic radar data for animals, vehicles and airplanes.

4. For the radar-signal synthesis, we expound the long- and short-time stationarity properties of the indoor wireless channel (see Sect. D.6). The short-time stationarity assumption is quite advantageous because it significantly simplifies the synthesis of the radar signal.

5. Through our proposed approach, we establish a novel simulated HAR dataset to train our simulation-based HAR system, which was developed by using a deep convolutional neural network (DCNN). This dataset comprises simulated radar signatures computed from emulated human activities. To demonstrate the practical relevance of our simulation-based HAR system, we tested its performance on unseen data acquired with a real mm-wave FMCW radar system involving real persons. The mean (maximum) classification accuracy of the fully simulation-based HAR system was 94% (98.4%). The classification performance of the proposed simulation-based HAR system over the experimental dataset demonstrates the utility and efficacy of our proposed end-to-end simulation framework.

D.1.4 Paper Organization

The paper is structured as follows. Section D.2 presents the related work, and Section D.3 gives an overview of conventional and the simulation-based HAR systems. Our human motion synthesis module is elucidated in Section D.4. Section D.5 details the 3D geometrical model. The synthesis of realistic radar data is explained in Section D.6. Section D.7 describes the processing of the radar data. The design, training, and testing of the simulation-based HAR system is detailed in Section D.8. Finally, Section D.9 draws the conclusions.

D.2 Related Work

Recently, the availability of commercial mm-wave sensors has led to the development of numerous human-centric research areas. For instance, many studies have been conducted on radar-based HAR systems [21, 22, 23], sign language [24] and gesture [25, 26] recognition systems. So far, most of the studies have focused on HAR systems that are realized by utilizing the scarcely available recorded radar data [27]. In [27], for instance, the HAR classifier was based on a long short-term memory (LSTM) neural network and was trained on manually labeled 3D point cloud data. The authors of [28] addressed the problem of HAR in multi-angle scenarios by exploiting measured characteristics of a mm-wave radar, such as received power, range, Doppler frequencies, azimuth, and elevation. Another problem with experimental data collected with radar systems is the reusability of the data. Generally, the recorded data of the radar system is not reusable due to its fixed operating parameters and antenna configurations. When the operating conditions of the radar system are fixed, the few-shot learning scheme [29] is useful to enhance the capability of the already trained HAR system.

To address the lack of real radar data, some studies have suggested to use data augmentation techniques. For instance, the authors of [30] proposed a data augmentation technique based on a generative adversarial network (GAN) to create diverse micro-Doppler signatures for human activities. Apart from GANs, a self-supervised HAR approach has been recently proposed to tackle the issue of limited labeled data [31]. The authors of [32] presented a technique called supervised few-shot adversarial domain adaptation for HAR. This approach addresses the challenge of having a limited amount of radar training data available for a particular scenario. Moreover, the authors of [33] proposed a rotational shift method to augment radar point cloud data. Recently, a two-stage domain adaptation scheme was presented in [34] to address the lack of training data for radar-based HAR systems. For data augmentation, they used a GAN-based domain-translation network to translate the simulated spectrograms into measurement-like spectrograms with the help of small measurement datasets. Even with this data augmentation technique, it is not possible to completely get rid of the tedious data collection process that requires real radar datasets and real human subjects.

The lack of publicly available real radar datasets, the limited reusability of radar

data, and the resource-intensive data collection are the main factors driving us to pursue fully simulation-based HAR system development. So far, only a handful of studies have been carried out in this direction. To model the intricate details of human motion, high-fidelity MoCap systems are preferred to eventually reanimate more realistic and complex human motion [35]. In [36], the authors formulated Doppler modulations and established equations for micro-Doppler effects caused by various micro motions such as vibration, rotation, tumbling, and coning. They validated these formulations through simulation studies. A simulation tool has been developed recently called SimHumalator, which simulates the target echoes for passive Wi-Fi radar (PWR) scenarios [37]. The authors of [38] developed a simulation tool that characterizes the near-field radar cross-section of a walking person in the K-band, but the approach is not suitable to model the finer details of human motion.

D.3 System Overview

In the following, let us first describe the basic building blocks of a conventional HAR system, which is employed solely to evaluate the proposed simulation-based HAR system.

D.3.1 A Conventional HAR System

The building blocks of a conventional (experimental-based) HAR system are depicted in Fig. D.1(a). For each human activity, the mm-wave radar system produces real raw IQ data. The IQ data is subsequently processed by the radar signal processing module to generate the real micro-Doppler signature or, equivalently, the TV radial velocity distribution capturing the characteristics of a human activity (see Sect. D.7). The TV radial velocity distributions of the recorded human activities are stored in files and represent the real radar dataset, as shown in Fig. D.1(a). Generally, the real radar data samples are used to train the experimental-based HAR classifier. Subsequently, a portion of the real radar data is used to test the performance of experimental-based HAR classifier, as shown in Fig. D.2. In this research, we are mainly interested in devising a simulation-based HAR system that matches the performance of state-of-the-art HAR systems. Thus, we will only use the entire recorded dataset from our conventional HAR system to test our proposed simulation-based HAR system, as shown in Fig. D.2. An overview of the proposed simulation-based HAR system is presented in the following subsection.

D.3.2 Proposed Simulation-Based HAR System

Conventionally recorded training datasets may not be reusable as they are only valid for specific radar parameters and a specific antenna configurations. A revision or redesign of even a single radar parameter may render the training dataset useless, e.g., the redesign of the radar system using a different pulse repetition interval (PRI). Therefore, a pragmatic alternative is proposed in this paper to overcome the

aforementioned issues associated with the acquisition of large training datasets. We propose a fully simulation-based approach, as shown in Fig. D.1(b), to develop a real-world HAR system. For training the HAR classifier, our simulation-based approach enables easy generation of a large amount of training data without involving real human subjects and a real radar system, which makes the simulation-based approach very feasible and pragmatic.

The overall view of the proposed simulation-based approach is shown in Fig. D.1(b). We start with six types of basic animations: standing still, falling, walking in two steps, standing up, sitting down, and picking up an object. Based on these six basic animations, we synthesize five different types of human activities: Falling on the floor, walking forward with more than two steps, standing up from a chair, sitting down on a chair, and picking up an object from the ground. The first 3D simulations block in Fig. D.1(b), implemented in the Unity software [13], synthesizes the human motion in the 3D space and generates the corresponding TV 3D trajectories of moving body segments, such as head, arms, legs, hands, and chest (see Sect. D.4.3). Subsequently, the geometrical 3D indoor channel model converts the TV 3D trajectories into the TV radial distances with respect to the positions of the
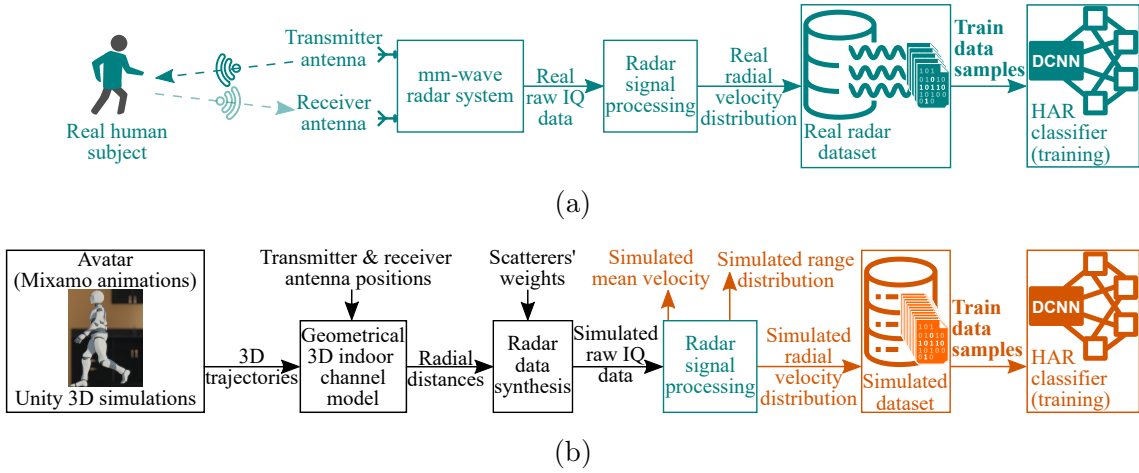
Figure D.1: (a) Design of conventional HAR systems that require real human subjects and a real radar system for training. (b) Design of the proposed simulation-based HAR system that only needs the simulated radar dataset for training.
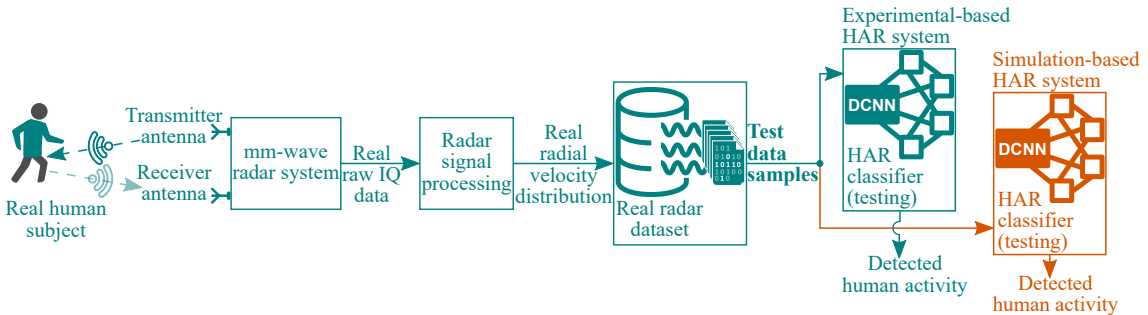
Figure D.2: Testing of the conventional (experimental-based) and the proposed simulation-based HAR systems on unseen real radar data samples.

transmitter and receiver antennas of the radar system (see Sect. D.5). For the TV radial distances and a set of scatterers' weights (see Sect. D.6), our radar data synthesizer in Fig. D.1(b) simulates the raw IQ data in the slow- and fast-time domains. Note that the virtual markers in our simulation framework are analogous to the real scatterers on the human body segments, which scatter the electromagnetic energy to the receive antenna of the radar system (see Sect. D.4.3). Finally, the radar signal processing block of our simulation framework generates the simulated range distribution, the simulated radial velocity distribution, and the simulated mean velocity for a synthesized human activity.

In this study, we refrain from using simulated range distributions for HAR due to their limited intelligibility resulting from the radar systems' restricted range resolution. Additionally, we solely use the mean velocity for comparison, not for HAR, as it contains less information about scatterers' velocity compared to the radial velocity distribution. This is detailed in Sect. D.7 and evident from Figs. D.6–D.8 as well. The simulated radial velocity distributions corresponding to the synthesized human activities are stored in a simulated radar dataset. We have developed our simulation-based HAR system by training it using only the examples from our simulated radar dataset as depicted in Fig. D.1(b). The simulation-based HAR system was designed by using a DCNN approach. In order to demonstrate the practical significance of our simulation-based HAR system, we need to test its performance on unseen data collected by a real radar system and real human subjects. Therefore, we recorded real human activities (falling, walking, picking, standing and sitting) in front of a mm-wave radar system to create a real radar dataset, which is used to test our simulation-based HAR system, as shown in Fig. D.2. It is noteworthy that the raw IQ data of the real and simulated radar are similar in structure. Therefore, we used the same radar signal processing block to process the real and simulated raw IQ data.

The proposed simulation framework emulates five distinct human activities and generates corresponding simulated radial velocity distributions for moving body segments. The simulated radar signatures (radial velocity distributions) are used to train the DCNN-based HAR classifier (see Fig. D.1(b)). Real mm-wave radar signatures are used to test the simulation-based HAR system as shown in Fig. D.2. The details of the individual components of the proposed simulation-based HAR system are explained in the following sections.

D.4 Human Motion Synthesis

In this section, we elucidate the first component of our simulation framework, which is the synthesis of the human activities for our simulation-based HAR system.

D.4.1 Basic Humanoid Animations

To synthesize realistic human activities, we use a pre-rigged 3D humanoid character and six types of basic humanoid animations from a well-known source called Mix-

amo [19]. It is a royalty-free library from Adobe Inc. offering countless realistic humanoid animations, which have been created with the help of professional actors and real-world MoCap systems [39]. We used the following animations from the Mixamo online library: idle, walking, falling, standing, picking, and sitting. In the idle animation, the avatar stands still in a natural upright posture, which causes a negligible in-place motion of all body segments. The walking animation consists of two steps in a forward direction on a flat floor. The falling animation portrays the avatar imitating a heart attack and collapsing abruptly to the ground. In the standing animation, the avatar gradually rises from a sitting position, while in the picking animation, it retrieves an object from the ground. In the sitting animation, the avatar is first in the idle upright position and then sits down on a chair.

We imported the six basic animations into the Unity software using the Filmbox (FBX) file format with a frame rate of 60 frames per second (fps). While importing an animation from the Mixamo's online library, a keyframe reduction parameter must be configured to optimize the animation data. We have refrained from applying keyframe reduction to the animation data, as this could in some cases alter or degrade the animation itself. In fact, we used linear interpolation in the Unity software to upscale the frame rate of the animations from 60 fps to 2000 fps to emulate and match the radar's pulse repetition frequency (PRF).

D.4.2 Unity Animation System

Among other things, the Unity animation system estimates the spatial positions of the avatar's body segments between frames by performing an interpolation operation. While shape-preserving interpolation methods, such as spline interpolation, can offer more accurate representations of the motion data, they often come with higher computational costs. In the context of our framework, where we aim to synthesize motion data at a high frame rate of 2000 fps, computational efficiency is an important consideration. Linear interpolation provides a computationally efficient solution while still preserving the general shape and trajectory of the motion. Moreover, it is important to keep the animation frame rate f_r equal to the real radar's PRF, because the PRF samples the motion of an object and thereby dictates the maximum measurable radial velocity $v_{\max}$ according to $v_{\max} = \text{PRF} \cdot \lambda/4$, where λ is the wavelength of the radar transmit signal. Analogously, in our simulation framework, the frame rate f_r dictates the maximum synthesizable radial velocity $v'_{\max}$ according to $v'_{\max} = f_r \lambda/4$. Any motion of the avatar with a radial velocity component greater than the maximum synthesizable radial velocity $v'_{\max}$ reverts to a lower velocity, just as in a real radar system. The Unity animation system is, by and large, quite versatile in supporting a wide range of animation techniques, e.g., procedural, MoCap and keyframe animations.

We use the Unity's animation state controller to enable the transition of the avatar between the six basic animation states. To synthesize a realistic human walking activity, we first need to switch between the basic idle and walking animations to merely start and end the overall walking activity. In addition, we need to

gradually increase and decrease the walking speed during the transition periods of the emulated activity. These natural and smooth transitions with gradual acceleration and deceleration are provided by a special type of state in the Unity's animation state machine called the blend tree.

Note that we do not have any animation data for the transition periods. This gap is filled by blend trees. When emulating a human walking activity and transitioning between idle and walking animation, the blend tree state dynamically creates new animation data in the 3D space in real time by aptly varying the avatar limbs to different degrees. With the help of blend trees, we can thus seamlessly transition from (to) idling animation to (from) walking animation with varying speeds while blending the two animations during the transition period. The human falling, standing, sitting, and picking activities are synthesized straightforwardly by combining the idle animation with the respective falling, standing, sitting, and picking animations.

D.4.3 3D Trajectories and Data Augmentation

We have synthesized five realistic human activities in the Unity software. In this subsection, we will explain how to capture the 3D trajectories of the synthesized motion for the five types of human activities. First, we need to place several virtual markers on different body segments of the avatar, as shown in Fig. D.3. These virtual markers are simulated point scatterers that resemble real scatterers on a human body.

In order to thoroughly capture the movements of the avatar, we placed a total of 21 virtual markers on different segments of the avatar body, which are represented by numbered stars in Fig. D.3. The body segments associated with the virtual markers in Fig. D.3 are listed in the ascending order: upper head, lower head, neck, right shoulder, left shoulder, right arm, left arm, upper spine, spine, lower spine, right
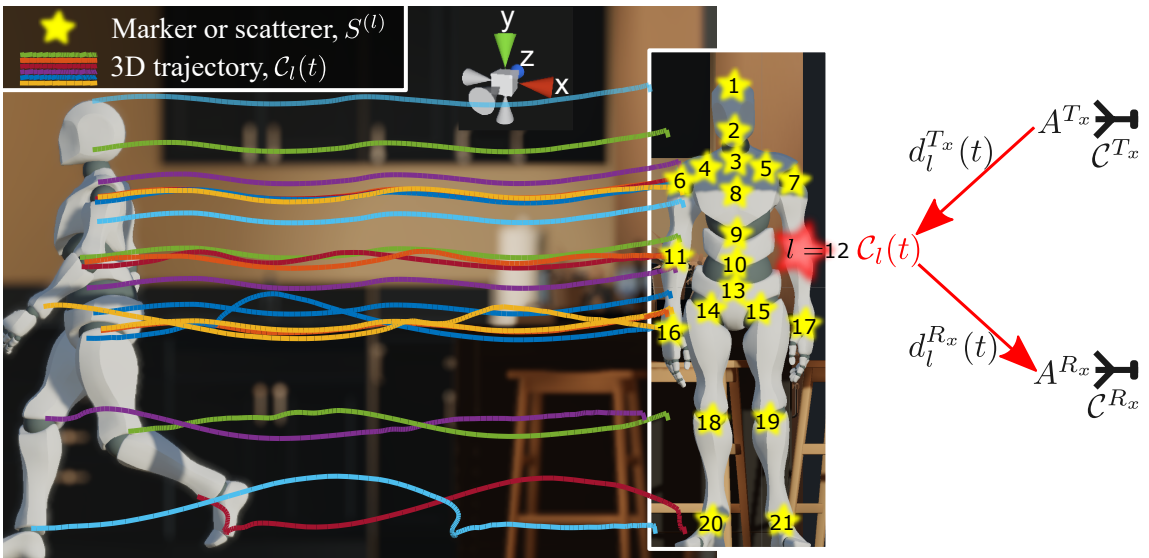
Figure D.3: Emulated propagation scenario composed of a radar system and a moving avatar with 21 non-stationary virtual markers.

forearm, left forearm, hips, right upper leg, left upper leg, right hand, left hand, right leg, left leg, right foot, and left foot. We need to spatially track the virtual markers and record the corresponding TV 3D trajectories of the virtual markers for the synthesized human activities. For instance, for a walking activity consisting of three steps in a forward direction, we can visualize the progression of the TV 3D trajectories associated with the 21 virtual markers, as represented by the colored curves in Fig. D.3. With the ability to synthesize the human activities and the corresponding 3D trajectories, we have created a dataset of diverse human activities that is used to train the simulation-based HAR classifier.

The synthesized human activities can be augmented and diversified in the Unity software by varying the emulation parameters such as the avatar's location, speed, acceleration, and deceleration. Thus, for each type of human activity, ten additional activity samples were generated by varying the above emulation parameters. For example, for the synthesized walking activities, random accelerations (decelerations) were assumed during the transition from the idling (walking) state to the walking (idling) state. For the five types of human activities, a total of fifty activity samples were generated in the Unity software. The TV 3D trajectories were recorded for the synthesized human activity samples. Subsequently, the TV 3D trajectories were exported to MATLAB for further data augmentation and processing. Using the geometrical 3D indoor channel model, which is detailed in the following section, we simulated eight slightly different radar antenna positions by moving virtually the transmitter and receiver antennas laterally for data augmentation. We also scaled the weights of the scatterers (see Fig. D.1(b) and Sect. D.6.1) to vary the power levels of the simulated radar signatures for further data augmentation.

D.5   Geometrical 3D Indoor Channel Model

In this section, we formulate a geometrical 3D indoor channel model corresponding to an indoor propagation scenario equipped with a radar system. Analogous to the real 3D indoor propagation scenario, the emulated indoor environment is shown in Fig. D.3. The emulated (real) propagation scenario is composed of a moving avatar (human) with $\mathcal{L}$ non-stationary virtual markers (scatterers), where the lth virtual marker (scatterer) is denoted by $S^{(l)}$ and $l = 1, 2, \ldots, \mathcal{L}$. In our simulation framework, the total number of virtual markers is $\mathcal{L} = 21$. The geometrical channel model is used to compute the TV radial distances between the $\mathcal{L}$ virtual markers (scatterers) and the radar transmit and receive antennas. In the simulation framework, the radar antennas can be placed freely as per the designer's requirements. Note that the antenna configuration greatly affects the simulated radar signatures. So, we can easily optimize the transmit and receive antenna positions with the help of the proposed simulation framework. For this research, the transmit and receive antennas of the radar system are placed in a monostatic configuration.

The radar transmit antenna A^{T_x} and receive antenna A^{R_x} are respectively placed at fixed positions $\mathcal{C}^{T_x} = [x^{T_x}, y^{T_x}, z^{T_x}]^\top$ and $\mathcal{C}^{R_x} = [x^{R_x}, y^{R_x}, z^{R_x}]^\top$, where $[\cdot]^\top$ repre-

sents the vector transpose operation. In Fig. D.3, $\mathcal{C}_l(t) = [x_l(t), y_l(t), z_l(t)]^\top$ is the TV 3D trajectory of the lth scatterer, $d_l^{T_x}(t)$ denotes the TV distance between the lth scatterer and the transmitter antenna A^{T_x}, and $d_l^{R_x}(t)$ denotes the TV distance between the lth scatterer and the receiver antenna A^{R_x}. Let $\|\cdot\|$ represents the Euclidean norm, then the TV distances $d_l^{T_x}(t)$ and $d_l^{R_x}(t)$ can be expressed as [40]

$$d_l^{T_x}(t) = \left\| \mathcal{C}_l(t) - \mathcal{C}^{T_x} \right\| \tag{D.1}$$

and

$$d_l^{R_x}(t) = \left\| \mathcal{C}_l(t) - \mathcal{C}^{R_x} \right\| \tag{D.2}$$

respectively. For the lth non-stationary virtual marker (scatterer), the TV radial distance $d_l(t)$ can be obtained as

$$d_l(t) = \frac{1}{2} \left[d_l^{T_x}(t) + d_l^{R_x}(t) \right]. \tag{D.3}$$

It is evident from (D.1)–(D.3) that the geometrical channel model maps the 3D trajectory $\mathcal{C}_l(t)$ to the TV radial distance $d_l(t)$ for a particular antenna configuration $\{\mathcal{C}^{T_x}, \mathcal{C}^{R_x}\}$. For the monostatic configuration, we have $\mathcal{C}^{T_x} = \mathcal{C}^{R_x}$, and thus $d_l(t) = d_l^{T_x}(t) = d_l^{R_x}(t)$. In the context of radar sensing, the TV radial distance $d_l(t)$ characterizes the synthesized motion of the lth virtual marker (scatterer), which is used to simulate the radar raw IQ data in the fast- and slow-time domain as explained in the next section.

For all $\mathcal{L}$ virtual markers of the synthesized falling, walking, picking, standing and sitting activities, we used the geometrical channel model to simulate the TV radial distances $d_l(t)$, as shown in Fig. D.4. As the $\mathcal{L}$ virtual markers are spatially distributed on the avatar's body segments, they have distinct TV radial distances, which are represented by the colored curves in Fig. D.4. In the simulated falling activity of Fig. D.4, some virtual markers exhibit more variations than others because the virtual markers on the lower body segments are less mobile than the virtual markers on the upper body segments. The simulated walking activity in Fig. D.4 exhibits its periodic nature and it consists of three walking steps. Compared to the falling and walking activities, the radial distances of the virtual markers do not vary as much in the simulated standing, sitting, and picking activities.
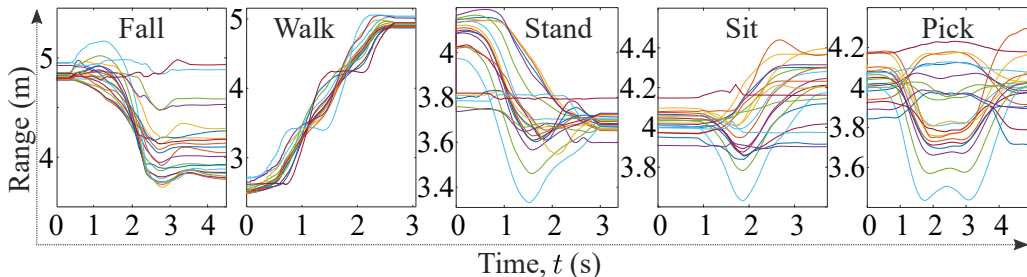
Figure D.4: Simulated TV radial distances $d_l(t)$ of 21 virtual markers for the five activities.

D.6   Radar Data Synthesis

One of the main modules of our simulation framework is the radar data synthesis module, which simulates realistic raw IQ data by emulating an FMCW radar system. The simulated raw IQ data depends entirely on the TV radial distances $d_l(t)$ of all $\mathcal{L}$ virtual markers (scatterers) and their respective weights which are modeled in Sect. D.6.1. In the following subsections, we elucidate the synthesis of the radar baseband signal called beat signal and explore the relevant stationary and non-stationary aspects of the indoor wireless channel.

D.6.1   Beat Signal Synthesis

The FMCW radar system periodically emits RF pulses, where the intra-pulse modulation is a linear chirp [41] waveform $c(t')$, where t' denotes the fast-time domain [42]. These RF pulses, also called the transmitted chirp signals $c(t')$, are reflected to the radar receiver by several scatterers in the environment. In this paper, we only consider and model the non-stationary scatterers because the stationary scatterers do not cause any Doppler shift and can therefore easily be filtered out in the radar signal preprocessing unit [43]. Furthermore, we assume that the $\mathcal{L}$ scatterers are stationary in the fast time t' and non-stationary in the slow time t as explained in the following subsection. From the lth non-stationary scatterer, a copy of the transmitted chirp waveform $c(t')$ is received with the TV propagation delay $\tau^{(l)}(t)$, which is proportional to the TV range (radial distance) of the lth scatterer $d_l(t)$ according to $\tau^{(l)}(t) = 2d_l(t)/c_0$, where c_0 denotes the speed of light.

In FMCW radar systems, the quadrature mixture module downconverts the received passband signal and produces the complex baseband signal, also known as the composite beat signal $s_b(t', t)$ [42]. The raw IQ data from the FMCW radar system is the digitized version of the composite beat signal $s_b(t', t)$. In FMCW radar systems, the analog to digital converter (ADC) samples the composite beat signal $s_b(t', t)$ in fast time t' with the sampling interval T_s. For the radar's coherent processing interval (CPI), in which the phase of the scatterers is preserved [44], the discrete fast-time samples are arranged in the fast- and slow-time domain to form the raw IQ data matrix $\mathcal{D}$, i.e.,

$$
\mathcal{D} = \begin{bmatrix} s_b(0,0) & s_b(T_s,0) & \cdots \\ s_b(0,T_{sw}) & s_b(T_s,T_{sw}) & \cdots \\ \vdots & \vdots & \vdots \\ s_b(0,(N_c-1)T_{sw}) & s_b(T_s,(N_c-1)T_{sw}) & \cdots \end{bmatrix}
$$

(D.4)

$$
\begin{bmatrix} s_b(T_{sw}-T_s,0) \\ s_b(T_{sw}-T_s,T_{sw}) \\ \vdots \\ s_b(T_{sw}-T_s,(N_c-1)T_{sw}) \end{bmatrix}
$$

where T_{sw} is the chirp duration and N_c is the number of chirps in the CPI of the radar system.

Now we want to model the composite beat signal $s_b(t', t)$, so that we can synthesize the raw IQ data of the FMCW radar system. Let $s_b^{(l)}(t', t)$ be the beat signal corresponding to the lth virtual marker, then the received composite beat signal $s_b(t', t)$ can be expressed as [42]

$$s_b(t', t) = \sum_{l=1}^{\mathcal{L}} s_b^{(l)}(t', t). \tag{D.5}$$

Note that in (D.5), the composite beat signal $s_b(t', t)$ is composed of $\mathcal{L}$ distinct beat signals $s_b^{(l)}(t', t)$ corresponding to $\mathcal{L}$ virtual markers. In particular, for the lth virtual marker, the beat signal $s_b^{(l)}(t', t)$ is fully characterized by its TV path gain $a^{(l)}(t)$, beat frequency $f_b^{(l)}(t)$, phase $\phi^{(l)}(t)$, and propagation delay $\tau^{(l)}(t)$ according to

$$s_b^{(l)}(t', t) = \sum_{n=0}^{\infty} a^{(l)}(t) \exp\left[j\left(2\pi f_b^{(l)}(t)t' + \phi^{(l)}(t)\right)\right] \times$$
$$\delta(t - \tau^{(l)}(t) - T_n) \tag{D.6}$$

where T_n is the discrete slow time that relates to the chirp duration T_{sw} by $T_n = nT_{sw}$ for $n = 0, 1, \ldots$. The function $\delta(\cdot)$ in (D.6) represents the Dirac delta function.

For the lth virtual marker, the syntheses of the TV beat frequency $f_b^{(l)}(t)$, phase $\phi^{(l)}(t)$, and propagation delay $\tau^{(l)}(t)$ in (D.6) are solely determined by the TV radial distance $d_l(t)$. Also, the lth TV path gain $a^{(l)}(t)$ is inversely proportional to the lth TV radial distance $d_l(t)$. The beat frequency $f_b^{(l)}(t)$ associated with the lth virtual marker can be modeled according to

$$f_b^{(l)}(t) = \frac{2d_l(t)\gamma}{c_0} \tag{D.7}$$

where γ is the slope of the chirp waveform $c(t')$. The phase $\phi^{(l)}(t)$ of the lth virtual marker is related to the radial distance $d_l(t)$ according to

$$\phi^{(l)}(t) = \frac{4\pi d_l(t)}{\lambda}. \tag{D.8}$$

Recall that the lth propagation delay component $\tau^{(l)}(t)$ in (D.6) can be obtained as $\tau^{(l)}(t) = 2d_l(t)/c_0$. Thus, the synthesis of the lth beat signal $s_b^{(l)}(t', t)$ in (D.6) is mainly determined by the lth TV radial distance $d_l(t)$.

We use the TV path gain $a^{(l)}(t)$ in (D.6) to model and simulate the amount of energy reflected back to the radar receiver from the lth scatterer (virtual marker). Thus, in the synthesis of the lth beat signal $s_b^{(l)}(t', t)$, the TV path gain $a^{(l)}(t)$ simulates the power or strength of the lth virtual marker. In this research, for the sake of simplicity, we have used $\mathcal{L} = 21$ time-invariant path gains, i.e., $a^{(l)}(t) = a^{(l)}$. For the five types of simulated human activities, we have accordingly used five sets of path gains to synthesize the composite beat signal $s_b(t', t)$ in (D.5). The body surface area [45] and the real TV radar signatures (see Sect. D.7) helped us adjust the path gains of the $\mathcal{L}$ virtual markers corresponding to the five types of simulated human activities. Note that multiple sets of path gains can be used to simulate multiple radar signatures for a single simulated human activity.

D.6.2 Long- and Short-Time Stationarity of the Channel

In this subsection, we explain the long- and short-time stationarity of the indoor wireless channel. Since the transmitter and receiver antennas are spatially fixed, the non-stationarity of the wireless channel is due to the motion of a human subject. For fixed antennas, the wireless channel is non-stationary due to the motion of the scatterers. We assume that the channel is long-time non-stationary or, equivalently, non-stationary over the slow time t. But in the fast time t', we assume that the channel is stationary over the limited duration of a chirp waveform T_{sw}. This assumption simplifies the synthesis of the radar beat signal. In the following, we will see that the short-time stationarity of the channel basically comes down to the radar's range resolution denoted as d_{res}, which is related to the radar's bandwidth B according to $d_{\text{res}} = c_0/2B$.

In the radar signal processing module (see Sect. D.7), we first apply the fast Fourier transform (FFT) to each row of the raw data matrix $\mathcal{D}$, called the range FFT. The frequency resolution f_{res} of the range FFT is equal to the inverse of the observation interval T_{sw}, i.e., $f_{\text{res}} = 1/T_{sw}$ [46]. For a row of the raw data matrix $\mathcal{D}$ and a slow-time instant t_0, the range FFT computes the spectrum containing the beat frequencies $f_b^{(l)}(t_0)$ corresponding to $d_l(t_0)$ for $l = 1, 2, \ldots, \mathcal{L}$. To resolve the spectral components corresponding to the $\mathcal{L}$ scatterers (virtual markers), the scatterers (virtual markers) must be at least f_{res} apart in the spectrum or, equivalently, d_{res} apart in the range (see (D.7)).

Let Δd_l and $\Delta f_b^{(l)}$ denote the overall change in the lth radial distance $d_l(t)$ and beat frequency $f_b^{(l)}(t)$, respectively, over one chirp duration T_{sw}. Then, a small change in the lth radial distance Δd_l results in a small change in the lth beat frequency $\Delta f_b^{(l)}$ according to (D.7). In practice, these changes are insignificant over the chirp duration T_{sw} and are not discernible in the spectrum of (D.5), such that $\Delta d_l \ll d_{\text{res}}$ and $\Delta f_b^{(l)} \ll f_{\text{res}}$, especially for indoor channels. Thus, the lth beat frequency $f_b^{(l)}(t_0)$ is assumed to be constant at the slow-time instant t_0 and over the fast-time duration $t_0 < t' < t_0 + T_{sw}$. Therefore, the wireless channel is assumed to be short-time stationary, which makes the synthesis of the discrete beat signals $s_b^{(l)}(t', t_0 + T_n)$ fairly simple for $n = 0, 1, \ldots$ and $l = 1, 2, \ldots, \mathcal{L}$. For instance, for the lth radial distance $d_l(t_0)$, the real and imaginary components of the lth beat signal $s_b^{(l)}(t', t_0)$ in (D.6) can be digitally synthesized as simple tone signals with fixed frequency $f_b^{(l)}(t_0)$ and phase $\phi^{(l)}(t_0)$.

D.7 Radar Signal Processing

This section describes the radar signal processing module of Fig. D.1, which can be used to process either the simulated or the real raw IQ data. First, the FFT operation is performed on the rows of the raw data matrix $\mathcal{D}$ (see (D.4)) to obtain

the beat frequency function $S_b(f_b, t)$, which can be expressed as [47]

$$S_b(f_b, t) = \int_0^{T_{sw}} s_b(t', t)e^{-j2\pi f_b t'} dt' \tag{D.9}$$

where f_b denotes the beat frequency. Subsequently, the short-time Fourier transform (STFT) of the beat frequency function $S_b(f_b, t)$ is carried out over the slow-time domain t to acquire the beat- and Doppler-frequency function $X(f_b, f, t)$ [42], which is given as

$$X(f_b, f, t) = \int_{-\infty}^{\infty} S_b(f_b, t'')W_r(t'' - t)e^{-j2\pi f t''} dt'' \tag{D.10}$$

where f and t'' represent the Doppler frequency and running time, respectively. The function $W_r(\cdot)$ in (D.10) represents a rectangular window function spanning over the slow-time duration of $64T_{sw}$.

Note that in (D.10), the beat- and Doppler-frequency function $X(f_b, f, t)$ can be integrated with respect to the Doppler frequency f (beat frequency f_b) to acquire the TV beat-frequency (micro-Doppler) signature. Thus, the expressions for the TV beat-frequency signature $S'(f_b, t)$ and the TV micro-Doppler signature $S(f, t)$ are given as

$$S'(f_b, t) = \left| \int_0^{\mathrm{PRF}} X(f_b, f, t)df \right|^2 \tag{D.11}$$

and

$$S(f, t) = \left| \int_0^{f_{b,\max}} X(f_b, f, t)df_b \right|^2 \tag{D.12}$$

respectively, where $f_{b,\max}$ is the maximum beat frequency. By using the TV beat-frequency signature $S'(f_b, t)$ in (D.11), we can express the TV range distribution $p'(r, t)$ as

$$p'(r, t) = \frac{S'\left(\frac{2\gamma}{c_0}r, t\right)}{\int_{-\infty}^{\infty} S'\left(\frac{2\gamma}{c_0}r, t\right)dr}. \tag{D.13}$$

The symbol r in (D.13) denotes the radar range, which is related to the beat frequency f_b by $r = c_0 f_b/(2\gamma)$.

Similarly, the TV radial velocity distribution $p(v, t)$ can be expressed as [42]

$$p(v, t) = \frac{S\left(\frac{2f_0}{c_0}v, t\right)}{\int_{-\infty}^{\infty} S\left(\frac{2f_0}{c_0}v, t\right)dv} \tag{D.14}$$

where v is the radial velocity and f_0 is the carrier frequency, which are related to the Doppler frequency f according to $v = c_0 f/(2f_0)$. Finally, the TV mean radial velocity $\bar{v}(t)$ can be obtained as

$$\bar{v}(t) = \int_{-\infty}^{\infty} vp(v, t)dv. \tag{D.15}$$

The TV mean radial velocity $\bar{v}(t)$ in (D.15) encapsulates the dominant characteristics of the TV radial velocity distribution $p(v,t)$ [42]. It provides a measure of the average radial velocity of all body segments at time t. Recall that the scatterers found on the human body segments reflect the electromagnetic energy back to the radar system. When a human body (avatar) moves, each scatterer (virtual marker) follows a spatially distinct trajectory and thus has a distinct TV radial velocity component with respect to the radar system. The TV radial velocity components corresponding to the scatterers (virtual markers) appear in the TV radial velocity distribution $p(v,t)$. Similarly, the TV range components of all the scatterers (virtual markers) appear in the TV range distribution $p'(r,t)$.

For the real radar data, the TV range distribution $p'(r,t)$ is not very intelligible due to the limited range resolution d_{res} of the radar system. Therefore, the real TV range distribution $p'(r,t)$ is usually not used for HAR. However, as an example, we show in Fig. D.5 the simulated TV range distribution $p'(r,t)$ with a range resolution d_{res} of 75 mm for the simulated fall, walk, stand, sit and pick activities. We also show the simulated (real) TV radial velocity distributions $p(v,t)$ for these five types of simulated (real) activities in Fig. D.6 (Fig. D.7). We can clearly see the striking similarities between the simulated and the real TV radial velocity distributions $p(v,t)$ in Fig. D.6 and Fig. D.7, respectively. It is worth noting that we will use the images of the simulated (real) TV radial velocity distributions $p(v,t)$ to train (test) the proposed simulation-based HAR system (see Sect. D.8). Finally, for the five types of simulated and real activities, the TV mean radial velocities $\bar{v}(t)$ are shown in Fig. D.8(a) and D.8(b), respectively.

The similarities between the simulated and real radar results in Figs. D.6, D.7, D.8(a), and D.8(b) demonstrate the quality of the data generated by the proposed simulation framework. Furthermore, to quantify the similarity between the simulated and real radar signatures, we employ the DTW algorithm [20]. This DTW distance metric measures the resemblance between TV mean radial velocities $\bar{v}(t)$ of simulated and real human activities (see Fig. D.8). The normalized DTW distances, presented in Table D.1, indicate the efficacy of our simulation-based approach in capturing the kinematic characteristics of various human activities. Notably, for all the activities, the DTW distance metric is minimized when comparing a given simulated
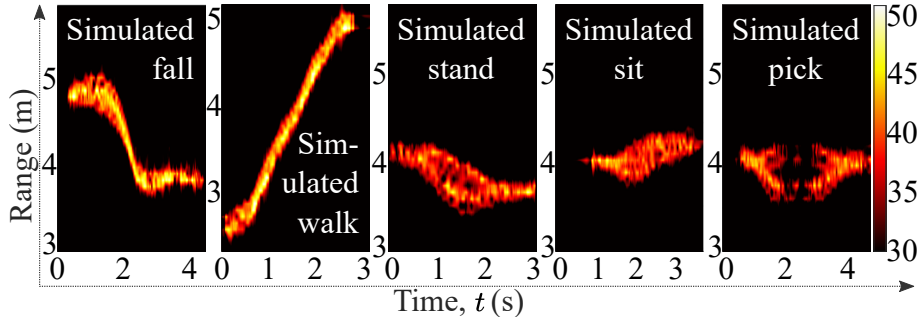
Figure D.5: Simulated TV range distributions $p'(r,t)$ for the emulated human activities.
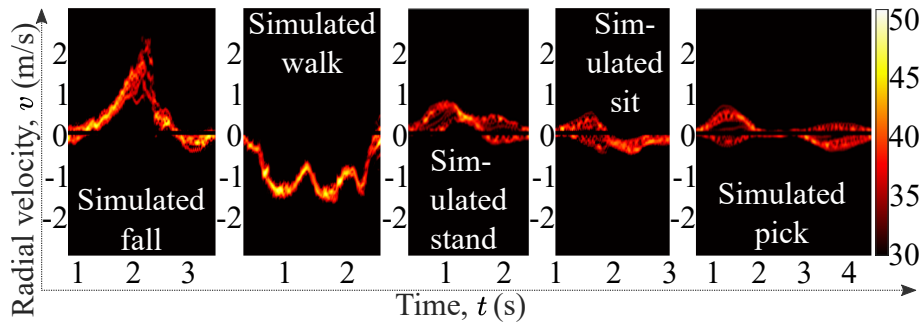
Figure D.6: Simulated TV radial velocity distributions $p(v, t)$ for the emulated human activities.
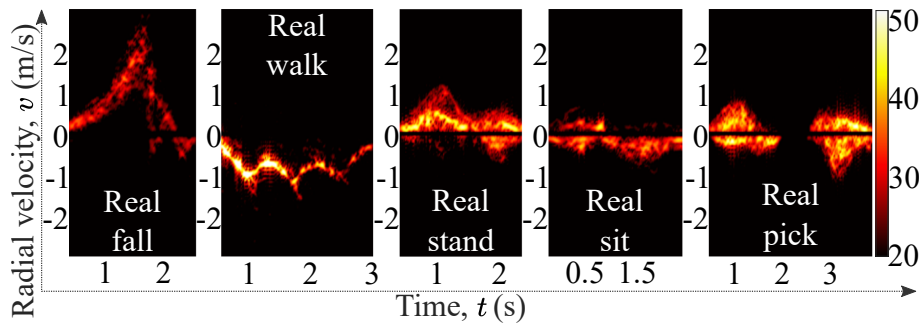
Figure D.7: Real TV radial velocity distributions $p(v, t)$ for the real human activities.
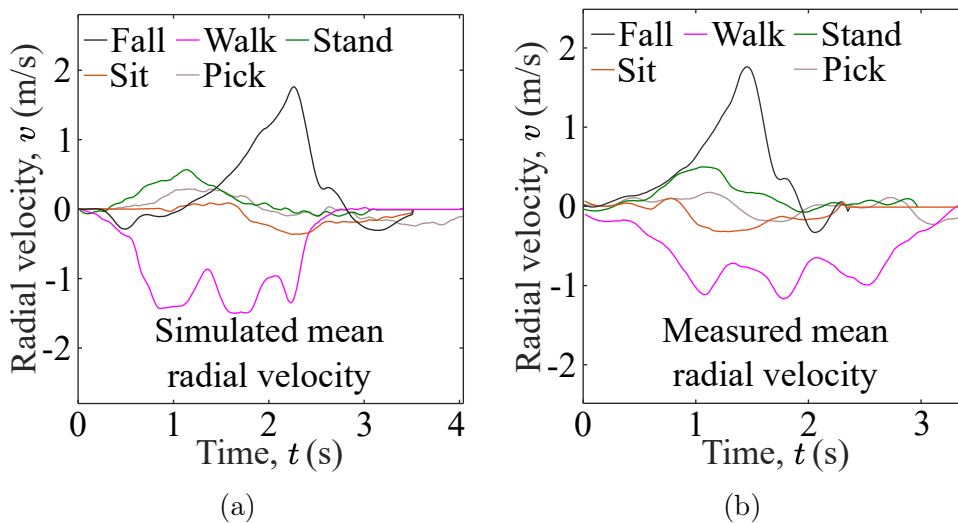
Figure D.8: TV mean radial velocities $\bar{v}(t)$ for (a) the emulated and (b) real human activities.

activity to its corresponding real counterpart. For instance, for the walking activity, the DTW distance of 0.07 between the simulated and real TV mean radial velocities $\bar{v}(t)$ highlights the precise simulation of the walking pattern. This trend persists across all simulated activities, affirming the fidelity of our simulation framework in accurately simulating real-world radar signatures. Note that the DTW distance between certain activities, such as sitting and picking, is smaller. This is due to their closely aligned patterns, thereby making them relatively challenging to classify.

Table D.1: The DTW distances between the simulated and real TV mean radial velocities $\bar{v}(t)$.

| Activity type | | Real | | | | |
		fall	walk	stand	sit	pick
Simulated	fall	**0.03**	0.36	0.18	0.26	0.20
	walk	0.49	**0.07**	0.55	0.36	0.34
	stand	0.07	0.26	**0.01**	0.10	0.05
	sit	0.13	0.08	0.16	**0.05**	0.07
	pick	0.08	0.19	0.04	0.03	**0.01**

D.8 Simulation-Based HAR System: Realization and Testing

In this section, we first explain how we realized the proposed HAR system using a DCNN-based multiclass classifier and how we trained it using only the simulated HAR dataset. A range of variations of the DCNN classifier is systematically analyzed through model ablations, facilitating the process of model analysis and selection. Subsequently, we demonstrate the performance of our trained simulation-based HAR classifier on the unseen real radar data using the best DCNN model.

D.8.1 Supervised Learning Using Simulated HAR Dataset

First and foremost, we need a simulated HAR dataset for training purposes. To create the simulated HAR dataset, we first synthesized the human motion using the Unity software as described in Section D.4. The position, speed, acceleration, and deceleration parameters were randomly varied in the Unity software to synthesize ten unique activity samples for each of the five activity types: falling, walking, standing, sitting, and picking. Subsequently, the spatial trajectories of these fifty activity samples were imported into MATLAB for further data augmentation. For each activity sample, eight slightly different radar positions $\{\mathcal{C}^{T_x}, \mathcal{C}^{R_x}\}$ and three different power levels were simulated in MATLAB. Low, medium, and high power levels were simulated by scaling the $\mathcal{L}$ weights $a^{(l)}(t)$ of the scatterers. In conclusion, the simulated HAR dataset consists of five types of human activities, ten different emulations of each activity type, eight radar positions, and three power levels. Thus, the total number of simulated TV radial velocity distributions $p(v,t)$ was 1200 in our simulated HAR dataset, which was used to train the DCNN-based HAR classifier.

The simulated TV radial velocity distributions $p(v,t)$ (see Fig. D.6) were transformed into images of dimension $224 \times 224 \times 3$. Thus, for each image, the number of pixels in the horizontal and vertical dimensions are 224 and the number of color channels are 3 (red, green and blue). These 1200 images serve as input feature maps to the DCNN-based HAR classifier as shown in Fig. D.9. The four convolutional layers of the DCNN classifier in Fig. D.9 contain 32, 64, 128, and 256 filter channels, respectively, which extract features from the simulated TV radial velocity distribu-
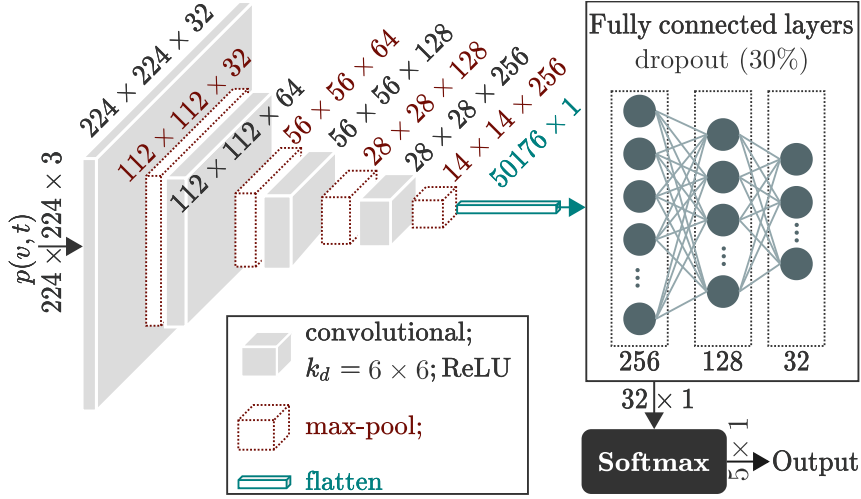
Figure D.9: Design of our DCNN-based HAR classifier that uses the simulated (real) HAR dataset for its training (testing).

tion $p(v, t)$. Each filter in a convolutional layer is a two-dimensional (2D) trainable kernel with the dimension k_d equal to 6×6 pixels. Note that, for the DCNN classifier shown in Fig. D.9, the network complexity (depth of hidden layers), kernel dimension k_d, max-pool layers, learning rate l_r, and other hyperparameters were determined through systematic analysis of a range of model variations. Further details are provided in the subsequent subsections.

To train the weights of the kernels, the L2 regularization technique was adopted to overcome the potential issue of overfitting [48]. The stride parameter was set to 1 in the DCNN, so that the feature-extraction filters stepped by one pixel. We employed the rectified linear unit (ReLU) function to alleviate the problem of vanishing gradients [49]. In Fig. D.9, the convolutional layers are followed by the max-pool layers of the order 2×2. The purpose of the max-pool layers is to reduce the redundancies by downsampling the output of the convolutional layers by a factor of 2. The features extracted by the multiple layers of the convolutional filters are flattened prior to the multilayer perceptron (MLP) layers (see Fig. D.9).

In the DCNN, the feature vector of dimension 50176×1 is obtained from the input TV radial velocity distribution $p(v, t)$. Then, the feature vector undergoes three MLP layers of dimensions 256, 128 and 32 with a dropout rate of 30%, as shown in Fig. D.9. The dropout layers mitigate the problems related to overfitting and generalizability of the network [50]. Finally, the softmax layer of order 5×1 was employed to compute the probabilities corresponding to the five types of human activities. For the training and validation of our simulation-based HAR classifier, we used our simulation dataset with a training–validation split ratio of $80 : 20$. To optimize the weights and biases of our simulation-based HAR system in Fig. D.9, we adopted the adaptive moment estimation (Adam) optimizer [51] and the human activity samples from our simulated dataset. The decay factors β_1, β_2, and the parameter ϵ of the Adam optimizer were set to 0.9, 0.999, and 10^{-8}, respectively, and the batch size was set to 32.

D.8.2 Real Data Collection and Model Variations

To test our proposed simulation-based HAR system, we used real human activities recorded by Ancortek's mm-wave FMCW radar system. During the measurement campaign, the operating parameters such as the carrier frequency f_c, bandwidth BW, chirp duration T_{sw}, and PRF of the mm-wave radar system were set to 24.125 GHz, 250 MHz, 500 μs, and 2 kHz, respectively. Note that the same values were chosen in the radar simulation model. The antennas of the real and simulated radar systems were chosen to be placed in a monostatic configuration.

We conducted in-depth experiments with Ancortek's mm-wave radar system in an indoor propagation scenario to compose the real radar-based HAR dataset consisting of five types of human activities, namely falling, walking, picking, sitting, and picking. Five male adults and one female adult repeatedly performed the human activities in the presence of various indoor objects. The mm-wave radar's IQ data corresponding to the real human activities were processed by the radar signal processing module to generate the TV radial velocity distributions $p(v, t)$. Note that the human activities were recorded for more than 5 seconds, but the actual duration of the activities was mostly 3 seconds (see Fig. D.7). The total number of radar signatures in the real radar dataset is 306.

For our simulation-based HAR system, we systematically explored various DCNN network configurations, detailed in Table D.2. Utilizing simulated and real radar signatures, we respectively trained and tested the DCNN classifiers with varying depths and complexities of the convolutional neural network (CNN) and MLP layers. Models 4, 5, and 6 demonstrated mean accuracies exceeding 86% with standard deviations (SDs) of less than 5% (see Table D.2). It's important to highlight that other DCNN models with lower and higher complexities displayed suboptimal performance, as indicated by the mean test accuracies in Table D.2. Subsequently, we systematically determined optimal hyperparameters, including kernel dimension k_d and learning rate l_r, for Models 4–6. Among these, Model 6 emerged as the most promising classifier, achieving the average (maximum) accuracy of 94% (98.4%) with optimized hyperparameters. The average percentage accuracies of Model 6 across different kernel dimensions k_d and learning rates l_r are depicted through the curves in Fig. D.10(a) and the heatmap in Fig. D.10(b).

D.8.3 Testing of the Simulation-Based HAR System Employing Model 6

The train–test (or simulation–real) data split ratio was 80 : 20. From the real radar-based HAR dataset, the 306 TV radial velocity distributions $p(v, t)$ corresponding to the real human subjects were used to test our trained simulation-based HAR system. The confusion matrix presented in Fig. D.11 shows the performance of our simulation-based HAR system (see Fig. D.9), specifically focusing on the trained model with the maximum performance. The x- and y-axis of the confusion matrix correspond to the predicted and true class of a human activity, respectively. Thus,

Table D.2: Mean classification accuracies of the DCNN models.

Model	CNN layers	MLP layers	Trainable parameters	Mean accuracy $\pm$ SD (%)
0	$[16, 32]$	$[32, 16]$	$3,232,117$	80.5 ± 7.5
1	$[16, 32, 48]$	$[48, 16]$	$1,882,805$	81.9 ± 5.7
2	$[16, 16, 32, 32, 48, 48]$	$[48, 32, 16]$	$658,485$	77.4 ± 16.5
3	$[32, 48, 64, 80]$	$[128, 64, 32]$	$2,371,557$	85.3 ± 4.7
4	$[32, 64, 72, 80]$	$[256, 128, 32]$	$4,502,205$	87 ± 3.3
5	$[32, 64, 96, 128]$	$[256, 128, 32]$	$7,201,029$	86.2 ± 4.7
6	$[32, 64, 128, 256]$	$[256, 128, 32]$	$14,434,725$	86.7 ± 4.8
7	$[48, 128, 256, 512]$	$[256, 128, 32]$	$31,853,109$	47.3 ± 31.1
8	$[48, 128, 512, 512]$	$[256, 128]$	$37,747,957$	66.4 ± 30.8
9	$[48, 128, 256, 256, 512, 512]$	$[256, 128, 64]$	$43,654,645$	19.7 ± 5.5

(a)

(b)

Figure D.10: Model 6 performance analysis: (a) the mean accuracy curves and (b) mean accuracy heatmap for kernel dimensions k_d and learning rates l_r.

the first five diagonal elements of the confusion matrix represent the number of correct classifications. The number of misclassifications is represented by the off-diagonal elements in the first five rows and columns of the confusion matrix. For

Figure D.11: Confusion matrix of the simulation-based HAR classifier with a classification accuracy of 98.4% on real data.

example, a "walking" activity was misclassified as a "falling" activity, as shown in the second row of the first column. Fig. D.11 also shows that four "sitting" activities were misclassified as "picking" activities as indicated by the fourth row of the fifth column. In the confusion matrix, the precision and recall quantities [52] are shown by the green color in the last row and last column, respectively. The worst precision and recall values are 95.9% and 92.3%, respectively. Most importantly, the overall classification accuracy of our simulation-based HAR system is 98.4% as shown by the white entry in Fig. D.11.

Note that the classification accuracy of our simulation-based HAR system is similar to today's real or experimental-based HAR systems [21, 53, 54]. However, the proposed simulation-based approach is quite unique in that it effortlessly generates a large amount of high-quality simulation data for training purposes. In the context of radar-based HAR, it is difficult to claim the superiority of one method or system over another as these systems are designed to address different constraints and resolve distinct problems. Nevertheless, in Table D.3, we have reported the performance of various state-of-the-art HAR systems using classification accuracy as the base metric.

The joint domain and semantic transfer learning (JDS-TL) [8] approach employed semi-supervised transfer learning (TL) and domain adaptation on partially labeled radar data to achieve an accuracy of 87.6%, as shown in Table D.3. Utilizing a hybrid architecture of CNNs and recurrent neural networks (RNNs) for spatial-temporal pattern extraction, the hybrid CNN-RNN [55] approach achieved a classification accuracy of 90.8% in recognizing human activities. Through a combination of convolutional auto encoder (CAE)-based unsupervised feature learning and multi-view data fusion, the CNN-LSTM method in [56] achieved an accuracy of 92%. The few-shot adversarial domain adaptation (FS-ADA) [32] method learned a common feature space from existing and new datasets, yielding a 91.6% accuracy

in radar-based HAR despite limited training data. The aforementioned state-of-the-art HAR systems relied on experimental-based training datasets, as outlined in Table D.3. Now, let's turn our attention to HAR systems trained with either partially (GAN-based) simulated datasets or fully simulated datasets.

To tackle kinematic inconsistencies associated with GAN-based data synthesis, the multibranch generative adversarial network (MBGAN) system in [57] employed physics-aware GAN-based techniques to synthesize micro-Doppler signatures, achieving 89.2% classification accuracy. For dataset augmentation, [58] employed a Wasserstein refined generative adversarial network with gradient penalty (WRGAN-GP) to generate synthetic micro-Doppler spectrograms. Vid2Doppler [59] employed cross-domain translation to generate synthetic Doppler signatures from videos, achieving an accuracy of 81.4% through entirely simulated training data. In contrast, our proposed simulation-based framework translated MoCap data into radar data via channel modeling, achieving a mean (maximum) accuracy of 94% (98.4%) using entirely simulated training data.

In this section, we have explained the design of the proposed simulation-based HAR system. It is worth noting that the proposed simulation framework of Fig. D.1(b) can be easily extended to other mm-wave radar-based application areas, such as gesture classification. The only difference would be to animate different types of gestures in the Unity software, while the rest of the modules of Fig. D.1(b) would remain the same.

D.9 Conclusion

The development of the modern radar-based HAR systems is mostly hindered by the scarce, unbalanced and partial datasets, because the acquisition of real radar data is not an easy task, especially for real human subjects. Therefore, in this paper, we alleviated the problems related to data scarcity for radar-based HAR classifiers. As a proof-of-concept, we presented an end-to-end simulation framework that synthesizes human motion and simulates the realistic mm-wave FMCW radar signatures. By generating large amounts of high-quality synthetic data, the proposed simulation framework significantly decreases the overall training effort of radar-based HAR systems. We used the synthetic and real data to train and test the HAR system, respectively. The proposed simulation-based HAR system demonstrated a classification accuracy of 98.4% on the unseen real radar data. Since the proposed end-to-end simulation framework reduces the involvement of real human subjects, it is crucial to improve the capabilities of future radar-based HAR classifiers.

In addition, the proposed simulation framework provides control over numerous radar and target parameters, such as avatar speed, acceleration, deceleration, height, position, motion type, radar antenna configuration, frequency, PRF, and bandwidth. This allows us to generate different types of radar datasets corresponding to different radar-operating conditions and different applications. Additionally, the proposed framework enables us to augment the data at the motion synthesis layer. Thus, at

Table D.3: State-of-the-art radar-based HAR approaches and their performance.

HAR method	System description	Training dataset	Dataset simulation method	Accuracy
JDS-TL [8]	Semisupervised TL and domain adaptation using partially (10%) labeled radar data for HAR	Experimental	None	87.6%
Hybrid CNN-RNN [55]	Distributed radar network with spatial-temporal pattern extraction for HAR	Experimental	None	90.8%
Multi-view CNN-LSTM [56]	CAE-based unsupervised feature learning and multi-view data fusion via LSTM	Experimental	None	92%
FS-ADA [32]	Learning common feature space of pre-existing and newly acquired training datasets	Experimental	None	91.6%
MBGAN [57]	Physics-aware GAN-based systems that utilize limited real data to synthesize micro-Doppler signatures	Simulated and experimental	GAN-based data synthesis using limited real data	89.2%
WRGAN-GP [58]	Generation and refinement models for realistic spectrogram synthesis	Simulated and experimental	GAN-based data synthesis using limited real data	94.9%
Vid2Doppler [59]	Cross-domain translation: Creating synthetic Doppler signatures from videos of human activities	Simulated	Translation of video dataset to Doppler dataset	81.4%
Proposed simulation-based framework	Simulating RF signals and radar signatures by using simulated 3D trajectories of virtual point scatterers	Simulated	Translating MoCap data to radar data using sensor modeling	Mean: 94% Max: 98.4%

the base motion synthesis layer, the target motion characteristics can be randomized to generate impartial, unbiased or balanced datasets that can be used to train radar-based classifiers.

In the proposed simulation framework, the scatterer-level modelling of the radar signal opens up new avenues of research for the radar-based classifiers. For instance, different optimization techniques can be explored to further improve the quality of the simulated radar signal and ultimately the simulated radar signatures. Furthermore, the work presented in this paper can be extended to classify other types of everyday human activities. The proposed approach can also be used to actualize other mm-wave radar-based classifiers, such as gesture recognition. We anticipate that the proposed end-to-end simulation framework will empower future radar-based classifiers with enhanced capabilities. We plan to extend the proposed simulation framework to multiple-input multiple-output radar systems incorporating multi-directional HAR.

Bibliography

[1] W. Qi, N. Wang, H. Su, and A. Aliverti, "DCNN based human activity recognition framework with depth vision guiding," *Neurocomputing*, vol. 486, pp. 261–271, 2022.

[2] J. Roche, V. De-Silva, J. Hook, M. Moencks, and A. Kondoz, "A multimodal data processing system for LiDAR-based human activity recognition," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10027–10040, 2022.

[3] Y. Tang, L. Zhang, F. Min, and J. He, "Multiscale deep feature learning for human activity recognition using wearable sensors," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 2106–2116, 2023.

[4] M. A. A. Al-qaness, A. Dahou, M. A. Elaziz, and A. M. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," *IEEE Trans. Ind. Informatics*, vol. 19, no. 1, pp. 144–152, 2023.

[5] Y. Tang, L. Zhang, Q. Teng, F. Min, and A. Song, "Triple cross-domain attention on human activity recognition using wearable sensors," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 5, pp. 1167–1176, 2022.

[6] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-Sense: a passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems," *Ann. Telecommun.*, vol. 77, no. 3-4, pp. 163–175, 2022.

[7] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, "RF-based human activity recognition using signal adapted convolutional neural network," *IEEE Trans. Mob. Comput.*, vol. 22, no. 1, pp. 487–499, 2023.

[8] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022.

[9] L. Zhang, W. Cui, B. Li, Z. Chen, S. Member, M. Wu, S. Member, and T. S. Gee, "Privacy-preserving cross-environment human activity recognition," *IEEE Trans. Cybern.*, pp. 1–11, 2021.

[10] M. Muaaz, A. Chelli, and M. Patzold, "WiHAR: From Wi-Fi channel state information to unobtrusive human activity recognition," in *2020 IEEE 91st Veh. Technol. Conf.*, pp. 1–7, 2020.

[11] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions," *Knowledge-Based Syst.*, vol. 223, p. 106970, 2021.

Bibliography

[12] S. Waqar, M. Muaaz, and M. Pätzold, "Direction-independent human activity recognition using a distributed MIMO radar system and deep learning," *IEEE Sens. J.*, vol. 23, no. 20, pp. 24916–24929, 2023.

[13] "Unity Real-Time Development Platform | 3D, 2D VR & AR Engine." Unity, accessed Feb. 20, 2023. [Online]. Available: https://www.unity.com/.

[14] S. Ahmed, W. Kim, J. Park, and S. H. Cho, "Radar-based air-writing gesture recognition using a novel multistream CNN approach," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 23869–23880, 2022.

[15] T. Tuncer, F. Ertam, S. Dogan, and A. Subasi, "An automated daily sports activities and gender recognition method based on novel multikernel local diamond pattern using sensor signals," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9441–9448, 2020.

[16] M. Tammvee and G. Anbarjafari, "Human activity recognition-based path planning for autonomous vehicles," *Signal, Image Video Process.*, vol. 15, no. 4, pp. 809–816, 2021.

[17] F. J. Rodriguez Lera, F. Martín Rico, A. M. Guerrero Higueras, and V. M. Olivera, "A context-awareness model for activity recognition in robot-assisted scenarios," *Expert Syst.*, vol. 37, no. 2, 2020.

[18] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Electr. Eng. Technol.*, vol. 14, no. 6, pp. 2567–2573, 2019.

[19] "Get animated." Mixamo, accessed Feb. 20, 2023. [Online]. Available: https://www.mixamo.com/#/.

[20] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series.," in *Proc. 3rd Int. Conf. Knowl. Discov. Data Min.*, pp. 359–370, 1994.

[21] W.-Y. Kim and D.-H. Seo, "Radar-based human activity recognition combining range–time–Doppler maps and range-distributed-convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[22] G. Lee and J. Kim, "Improving human activity recognition for sparse radar point clouds: A graph neural network model with pre-trained 3D human-joint coordinates," *Appl. Sci.*, vol. 12, no. 4, 2022.

[23] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from FMCW radar with Bi-LSTM networks," *IEEE Sens. J.*, vol. 20, no. 22, pp. 13607–13619, 2020.

[24] M. M. Rahman, E. A. Malaia, A. C. Gurbuz, D. J. Griffin, C. Crawford, and S. Z. Gurbuz, "Effect of kinematics and fluency in adversarial synthetic data generation for ASL recognition with RF sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 2732–2745, 2022.

[25] Z. Xia and F. Xu, "Time-space dimension reduction of millimeter-wave radar point-clouds for smart-home hand-gesture recognition," *IEEE Sens. J.*, vol. 22, no. 5, pp. 4425–4437, 2022.

[26] H. Liu, A. Zhou, Z. Dong, Y. Sun, J. Zhang, L. Liu, H. Ma, J. Liu, and N. Yang, "M-gesture: Person-independent real-time in-air gesture recognition using commodity millimeter wave radar," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3397–3415, 2022.

[27] Z. Yu, A. Taha, W. Taylor, A. Zahid, K. Rajab, H. Heidari, M. A. Imran, and Q. H. Abbasi, "A radar-based human activity recognition using a novel 3-D point cloud classifier," *IEEE Sens. J.*, vol. 22, no. 19, pp. 18218–18227, 2022.

[28] Y. Zhao, A. Yarovoy, and F. Fioranelli, "Angle-insensitive human motion and posture recognition based on 4D imaging radar and deep learning classifiers," *IEEE Sens. J.*, vol. 22, no. 12, pp. 12173–12182, 2022.

[29] Z. Liu, C. Wu, and W. Ye, "Category-extensible human activity recognition based on Doppler radar by few-shot learning," *IEEE Sens. J.*, vol. 22, no. 22, pp. 21952–21960, 2022.

[30] Y. Yang, Y. Zhang, Y. Lang, B. Li, S. Guo, and Q. Tan, "GAN-based radar spectrogram augmentation via diversity injection strategy," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[31] Y. Zhou, C. Xie, S. Sun, X. Zhang, and Y. Wang, "A self-supervised human activity recognition approach via body sensor networks in smart city," *IEEE Sens. J.*, 2023.

[32] X. Li, Y. He, J. A. Zhang, and X. Jing, "Supervised domain adaptation for few-shot radar-based human activity recognition," *IEEE Sens. J.*, vol. 21, no. 22, pp. 25880–25890, 2021.

[33] C. Yu, Z. Xu, K. Yan, Y.-R. Chien, S.-H. Fang, and H.-C. Wu, "Noninvasive human activity recognition using millimeter-wave radar," *IEEE Syst. J.*, vol. 16, no. 2, pp. 3036–3047, 2022.

[34] Y. Yang, Y. Zhang, H. Ji, B. Li, and C. Song, "Radar-based human activity recognition under the limited measurement data support using domain translation," *IEEE Signal Process. Lett.*, vol. 29, pp. 1993–1997, 2022.

[35] M. J. Bocus, W. Li, S. Vishwakarma, R. Kou, C. Tang, K. Woodbridge, I. Craddock, R. McConville, R. Santos-Rodriguez, K. Chetty, and R. Piechocki, "OPERAnet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors," *Sci. Data*, vol. 9, no. 1, 2022.

[36] V. C. Chen, F. Li, S. S. Ho, and H. Wechsler, "Micro-doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, 2006.

[37] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, "SimHumalator: An open-source end-to-end radar simulator for human activity recognition," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 3, pp. 6–22, 2022.

[38] G. Manfredi, P. Russo, A. De Leo, and G. Cerri, "Efficient simulation tool to characterize the radar cross section of a pedestrian in near field," *Prog. Electromagn. Res. C*, vol. 100, pp. 145–159, 2020.

[39] Mixamo, "Mixamo motion capture (behind the scenes)." YouTube, Aug. 24, 2012. [Online]. Available: https://www.youtube.com/watch?v=q7jWMd7ix98.

[40] S. Waqar and M. Pätzold, "Interchannel interference and mitigation in distributed MIMO RF sensing," *Sensors*, vol. 21, no. 22, 2021.

[41] S. Sana, S. Waqar, H. Yusaf, M. Waqas, and F. A. Siddiqui, "Software defined digital beam forming processor," in *2016 13th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 671–676, 2016.

[42] S. Waqar, M. Muaaz, and M. Pätzold, "Human activity signatures captured under different directions using SISO and MIMO radar systems," *Appl. Sci.*, vol. 12, no. 4, 2022.

[43] S. Waqar, H. Yusaf, S. Sana, M. Waqas, and F. A. Siddiqui, "Reconfigurable monopulse radar tracking processor," in *2018 15th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 805–809, 2018.

[44] Z. Peng, J. M. Munoz-Ferreras, Y. Tang, C. Liu, R. Gomez-Garcia, L. Ran, and C. Li, "A portable FMCW interferometry radar with programmable low-IF architecture for localization, ISAR imaging, and vital sign tracking," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 4, pp. 1334–1344, 2017.

[45] M. Mance, M. Prutki, A. Dujmovic, M. Miloševic, V. Vrbanovic-Mijatovic, and D. Mijatovic, "Changes in total body surface area and the distribution of skin surfaces in relation to body mass index," *Burns*, vol. 46, no. 4, pp. 868–875, 2020.

[46] J. W. Ting, D. Oloumi, and K. Rambabu, "FMCW SAR system for near-distance imaging applications - practical considerations and calibrations," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 450–461, 2018.

[47] M. Muaaz, S. Waqar, and M. Pätzold, "Radar-based passive step counter and its comparison with a wrist-worn physical activity tracker," in *4th Int. Conf. Intell. Technol. Appl. INTAP*, pp. 259–272, Springer International Publishing, 2022.

[48] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *7th Int. Conf. Learn. Represent. ICLR*, 2019.

[49] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," *Proc. Int. Jt. Conf. Neural Networks*, pp. 2684–2691, 2017.

[50] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[51] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv Prepr.*, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980.

[52] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. ICML*, pp. 233–240, 2006.

[53] Y. Zhao, H. Zhou, S. Lu, Y. Liu, X. An, and Q. Liu, "Human activity recognition based on non-contact radar data and improved PCA method," *Appl. Sci.*, vol. 12, no. 14, p. 7124, 2022.

[54] M. Muaaz, S. Waqar, and M. Pätzold, "Orientation-independent human activity recognition using complementary radio frequency sensing," *Sensors*, vol. 23, no. 13, p. 5810, 2023.

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

[56] H.-U.-R. Khalid, A. Gorji, A. Bourdoux, S. Pollin, and H. Sahli, "Multi-view CNN-LSTM architecture for radar-based human activity recognition," *IEEE Access*, vol. 10, pp. 24509–24519, 2022.

[57] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, "Physics-aware generative adversarial networks for radar-based human activity recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 3, pp. 2994–3008, 2023.

[58] L. Qu, Y. Wang, T. Yang, and Y. Sun, "Human activity recognition based on WRGAN-GP-synthesized micro-Doppler spectrograms," *IEEE Sens. J.*, vol. 22, no. 9, pp. 8960–8973, 2022.

[59] K. Ahuja, Y. Jiang, M. Goel, and C. Harrison, "Vid2Doppler: Synthesizing Doppler radar data from videos for training privacy-preserving activity recognition," in *Proc. 2021 CHI Conf. Hum. Factors Comput. Syst.*, pp. 1–10, may 2021.

Appendix E

Paper E

A Paradigm Shift from an Experimental-Based to a Simulation-Based Framework Using Motion-Capture Driven MIMO Radar Data Synthesis

Sahil Waqar[1], Muhammad Muaaz[1], Stephan Sigg[2], and Matthias Pätzold[1]

[1]Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway
[2]Department of Communications and Networking, Aalto University, 00076 Espoo, Finland

E-mails: [1]{sahil.waqar, muhammad.muaaz, matthias.paetzold}@uia.no, [2]stephan.sigg@aalto.fi

Abstract— The development of radar-based classifiers driven by empirical data can be highly demanding and expensive due to the unavailability of radar data. In this paper, we introduce an innovative simulation-based approach that addresses the data scarcity problem, particularly for our multiple-input multiple-output (MIMO) radar-based direction-independent human activity recognition (HAR) system. To simulate realistic MIMO radar signatures, we first synthesize human motion and generate corresponding spatial trajectories. From these trajectories, a received radio frequency (RF) signal is synthesized using our MIMO channel model, which considers the non-stationary behavior of human motion and the multipath components originating from the scatterers on human body segments. Subsequently, the synthesized RF signals are processed to simulate MIMO radar signatures for various human activities. The proposed simulation-based direction-independent HAR system achieves a classification accuracy of 97.83% when tested with real MIMO radar data. A significant advantage of our simulation-based framework lies in its ability to facilitate multi-stage data augmentation techniques at the motion-layer, physical-layer, and signal-layer syntheses. This capability significantly reduces the training workload for radar-based classifiers. Importantly, our simulation-based proof-of-concept is applicable to single-input single-output and MIMO radars in monostatic, bistatic, and multistatic configurations, making it a versatile solution for realizing other radar-based classifiers, such as gesture classifiers.

Index Terms— Aspect angle, data augmentation, data synthesis, deep learning, distributed MIMO radar simulation, human activity recognition (HAR), micro-Doppler analysis, motion capture, motion synthesis, multiclass classification, virtual reality.

E.1 Introduction

E.1.1 Background

The generation of area-specific synthetic data has been an important topic of interest among researchers [1, 2]. Device-specific or sensor-tailored simulation models help generate realistic sensory data and have been used to realize real-world solutions [3, 4]. Given the increasing prevalence of machine learning and artificial intelligence methodologies and applications today, the importance of the concept of device-specific synthetic data generation, as well as the significance of sensor modeling, cannot be overstated. For many sensing modalities such as magnetometer, infrared, light detection and ranging (LiDAR), sonar, and radar, data scarcity often hinders the realization of machine learning-based solutions [5, 6]. Sensor-tailored simulation models mitigate the data scarcity problem by providing clean and labeled synthetic datasets for various real-world conditions. Such synthetic datasets are important to develop machine learning-based applications, e.g., medical imaging [7].

Human activity recognition (HAR) [8, 9, 10, 11, 12, 13] remains an important and active research area facing the challenge of data scarcity, especially when using radio frequency (RF) sensors such as Wi-Fi [14] and radar [15, 16, 17]. Furthermore, for multiple-input multiple-output (MIMO) radar systems with user-defined (required) operating parameters and antenna configurations, readily-available HAR datasets are almost non-existent. Optimal radar operating conditions and antenna configurations are often not known in advance for different environmental conditions and applications. Synthetic data generation is therefore a pragmatic and promising approach to realizing radar-based classifiers, offering tremendous design control and system flexibility in a cost-effective manner. Realizing HAR systems through a simulation-based approach poses two main challenges: (a) how to synthesize human activities, and (b) how to simulate single-input single-output (SISO) and MIMO radar signatures for the synthesized human activities. Before going into further details of synthetic data generation and our proposed simulation-based approach, we first provide an overview of the relevant research in the following subsection.

E.1.2 Related Work

The ongoing miniaturization and commercialization of radar sensors, as well as many Internet of Things (IoT) sensors, have encouraged the development of human-centric applications, including HAR. Small-scale radar systems are increasingly preferred by researchers for the development of HAR systems [18, 19], gesture [20, 21] and sign language [22] recognition systems. Realizing empirical-data-driven (experimental-based) HAR systems is often very challenging due to the low availability of recorded radar datasets. Among other challenging and monotonous tasks, the development of experimental-based HAR systems requires the involvement of human subjects, an actual SISO or MIMO radar system, and the manual labeling of the recorded data. The authors of [23] used manually labeled point cloud data to train the HAR system, which was built upon a long short-term memory (LSTM) network. By utilizing the

measured features of a millimeter wave (mm-wave) radar, the authors in [24] tackled the issue of HAR in multi-view settings.

Recent studies have shown that, to some extent, data augmentation techniques can reduce the scarcity of empirical data for HAR systems. For instance, a rotation-shift technique was utilized in [25] to expand the three-dimensional (3D) point cloud dataset. A generative adversarial network (GAN)-based data augmentation technique was adopted in [26] to create varied radar signatures of human activities. The use of the few-shot learning method was suggested by the authors of [27], which offers a unique way of augmenting the capabilities of pre-trained and pre-existing HAR systems. According to a recent study [28], a two-stage domain adaptation approach can be used to alleviate the data scarcity issue as well. With this approach, the simulated micro-Doppler signatures can be translated into measurement-like micro-Doppler signatures by using small real datasets. Note that even with such data augmentation methods, time-consuming and tedious data collection cannot be avoided.

Radar-based classifiers may face unique challenges in different situations and application areas, which may necessitate the adaptation of radar antenna configurations and operating conditions. This exacerbates the problem of data scarcity in radar systems because the training dataset recorded from a radar system in one scenario, may not be applicable and useful in another. Therefore, the synthetic data generation is the way to realize radar-based HAR systems. To date, only a few studies have been conducted in the context of RF sensing that deal with synthetic data generation for HAR. In this regard, the utilization of motion capture (MoCap) systems [29] is an effective means of modeling and reanimating complex human motion for further motion synthesis. For passive Wi-Fi radar (PWR), the authors of [30] devised a system, namely SimHumalator, to generate target returns. In [31], a simulation tool was created to evaluate the radar cross-section of a walking individual in close proximity. However, this technique is inadequate for reproducing detailed and complex human movements.

E.1.3 Our Approach

In this paper, we present a proof-of-concept that overcomes the problems related to radar data scarcity, offers significant design control and flexibility of the radar system, and allows the simulation of unbounded, clean and labeled radar datasets. We emulate a 2×2 MIMO radar system with the help of our proposed simulation-based framework to realize a simulation-based direction-independent HAR system. First, we devise an activity simulation module that synthesizes multiple types of human activities in a virtual environment by using the 3D animation tools from the Unity [32] and MotionBuilder [33] software. An appropriate avatar or a humanoid character, equipped with multiple simulated point scatterers on its body segments, is used to reanimate MoCap data in these programs (see Sect. E.3). Subsequently, we generate spatial trajectories corresponding to all simulated point scatterers or body segments of the avatar, which effectively characterize the overall humanoid

motion.

The spatial trajectories of the body segments are processed by our channel model, which simulates the received RF signal from a frequency-modulated continuous wave (FMCW) radar system for software-defined antenna positions. While simulating the raw in-phase and quadrature (IQ) components of a received baseband signal, our channel model takes into account the multipath components originating from the non-stationary simulated (real) point scatterers with distinct time-variant (TV) propagation delays (see Sect. E.4). In the proposed channel model, the long- and short-time stationarity characteristics of the scatterers are considered in an indoor wireless propagation environment. Additionally, to train the 2×2 MIMO radar-based direction-independent HAR system (see Sect. E.7), we simulated five types of multi-directional human activities by rotating the transmitter and receiver antennas of the emulated MIMO radar system (see Sect. E.5).

Unlike conventional or experimental-based designs of HAR systems, the proposed simulation-based approach is highly versatile and offers numerous advantages. Our simulation-based approach is capable of simulating diverse training datasets to meet various radar-based applications and a wide range of operational requirements. For monostatic/bistatic/multistatic SISO/MIMO radar systems, the scatterer-level modeling of moving objects in our simulation-based framework opens up new research opportunities to further fine-tune the simulated radar signatures, such as TV micro-Doppler signatures (TV radial velocity distributions) and TV range distributions (see Sect. E.4 and Sect. E.6). For example, the TV path gains of the scatterers (simulated point scatterers) can be adjusted or optimized to improve and augment the simulated radar signatures. Moreover, the simulation-based framework provides multi-stage data augmentation techniques (see Sect. E.5), which allow us to generate diverse and high-quality SISO/MIMO radar datasets in a flexible and cost-effective manner. For instance, at the motion-layer synthesis data augmentation stage, various animation parameters and avatar characteristics, e.g., speed and height, can be arbitrarily varied to simulate a range of human motions. Most importantly, the proposed simulation-based framework radically reduces the workload and resources for classifier training. As our simulation-based approach is versatile, it can be easily extended to implement many other SISO/MIMO radar-based classifiers, such as air-writing gesture classification [34].

E.1.4 Contributions

The key findings and contributions of this study can be delineated as follows:

1. This research proposes a simulation-based framework to significantly minimize the data collection workload required for devising real-world radar-based HAR systems. The simulation-based framework is capable of synthesizing realistic, diverse, and clean datasets for MIMO radar systems, regardless of their configuration: monostatic, bistatic, or multistatic. Although this study focuses on a 2×2 MIMO radar-based direction-independent HAR system, the utility of the

simulation-based framework extends beyond the HAR application, making it also valuable for other radar-based applications, e.g., sign language detection.

2. We have developed a MoCap-data-driven activity simulation module that enables the synthesis of multiple types of human activities in a virtual environment. For a total of 21 simulated point scatterers placed on body segments of an avatar, the activity simulation module generates 3D trajectories that essentially characterize the overall human motion. Our activity simulation module can integrate motion data from diverse sources, including biomechanical, wearable, and optical MoCap systems (see Sect. E.3). Additionally, the activity simulation module can generate arbitrary software-defined motion data.

3. We formulate a MIMO channel model that simulates realistic RF data or raw IQ data by using the spatial trajectory data from non-stationary simulated point scatterers. In this channel model, we study and simulate the TV propagation delays corresponding to the multipath components emanating from the non-stationary simulated point scatterers on the avatar's body segments. The proposed MIMO channel model helps generate unlimited radar datasets and provides extensive design control and versatility.

4. We present multi-stage data augmentation techniques for motion-layer synthesis, physical-layer synthesis, and signal-layer synthesis. For example, in the proposed simulation-based framework, we first diversified the target motion data in the motion-layer synthesis using the activity simulation module. And subsequently in the physical-layer synthesis, we augmented the radar data by varying physical layer parameters such as radar orientation. Lastly, we further augmented the radar data by using multiple sets of simulated point scatterers' weights (TV path gains) at the signal-layer synthesis of the proposed simulation-based framework. The multi-stage, simulation-based data augmentation techniques allowed us to vary target motion characteristics and antenna configurations, simulate multiple radar sensors, and transform uni-directional motion data to multi-directional motion data.

5. For the simulated human activities, we generated high-quality MIMO radar signatures, such as TV radial velocity distribution and mean radial velocity, which closely resemble the actual MIMO radar signatures of actual human activities. This demonstrates the efficacy of the proposed simulation-based framework, which is highly versatile as it can be effortlessly extended to simulate radar signatures for various other moving objects such as cars, drones, and aircraft.

6. By employing our simulation-based framework, we generated a unique simulated dataset to train/realize classifiers based on (deep) machine learning. The training dataset for HAR incorporates simulated radar patterns, derived from software-defined avatar movements. This approach proves highly advan-

tageous and practical as the training data is developed entirely from scratch, eliminating the need for real individuals and an actual MIMO radar system.

7. For the 2×2 MIMO radar framework, we realized a simulation-based HAR system by employing a deep convolutional neural network (DCNN). The system employed multiperspective simulated radar signatures as input features. To showcase the practical applicability of our simulation-driven HAR system, we evaluated its performance using actual mm-wave radar data collected from actual individuals. Our simulation-based multiperspective HAR system achieved an impressive classification accuracy of 97.83%, providing compelling evidence for its effectiveness.

E.1.5 Paper Organization

The article is divided into eight sections. Section E.2 deals with the system design and the general structures of the conventional and the proposed approaches. Human motion capture and synthesis techniques are presented in Section E.3. Section E.4 details channel modeling and simulation. Multi-stage data augmentation approaches are elucidated in Section E.5. Section E.6 discusses the generation of MIMO radar signatures. Section E.7 presents the design, training, and testing phases of our simulation-based direction-independent HAR system. Finally, we conclude our research in Section E.8.

E.2   System Design

In this section, we discuss a conventional experimental-based design of a HAR system and the proposed simulation-based realization of a HAR system. We also discuss problems of conventional HAR systems and how the proposed end-to-end simulation framework resolves them. Note that SISO radar-based HAR systems struggle to classify multi-directional human activities [35, 36]. To classify different types of multi-directional human activities, we need multiple radar subsystems illuminating the environment from different perspectives. Therefore, in the following subsections, we consider multi-directional human activities recorded by a multiperspective distributed MIMO radar system.

E.2.1   Conventional Experimental-Based Designs of HAR Systems

In radar sensing, state-of-the-art experimental-based HAR systems [19, 18, 17, 36, 15, 14, 9, 13, 23, 24] generally face challenges, such as data scarcity and their adaptability to environmental conditions. As an example of state-of-the-art experimental-based designs, we considered a direction-independent HAR system implemented with a mm-wave 2×2 MIMO radar system, as shown in Fig. E.1(a). In Fig. E.1(a), Radar$_i$ represents the ith radar subsystem of the distributed MIMO radar system, $A_i^{T_x}$ is

the ith transmitter antenna, and $A_i^{R_x}$ is the ith receiver antenna for $i = 1, 2$. Note that the two horn antennas, namely $A_i^{T_x}$ and $A_i^{R_x}$, are arranged in a monostatic configuration for Radar_i. In the conventional experimental-based HAR system of Fig. E.1(a), six human subjects performed the following types of multi-directional activities: falling on a mattress, walking, standing up from a chair, sitting down on a chair, and picking up an object from the floor.

The distributed MIMO radar system simultaneously illuminates the human subject from two aspect angles and generates the corresponding raw IQ data, as shown in Fig. E.1(a). Then, the radar signal processing block (see Sect. E.6) generates the TV micro-Doppler signatures or, equivalently, the TV radial velocity distributions for Radar_1 and Radar_2. These recorded radar signatures (TV radial velocity distributions) are accumulated to create a real radar dataset. In conventional experimental HAR systems, the real radar data set is usually divided into a training subset and a testing subset to train and test these HAR systems, respectively. However, for this research, we only use the experimentally obtained radar dataset to test our proposed simulation-based HAR system (see Fig. E.2).

Similar to any multiclass classifier, radar-based HAR systems require extensive amounts of recorded data for their training. However, unlike other sensing modalities such as cameras, radar systems often suffer from data scarcity. To experimentally design a HAR system, real human subjects must perform various types of activities in front of the MIMO radar system in multiple directions. These requirements make data collection time-consuming and costly. Additionally, the recorded radar training dataset usually cannot be reused for different antenna configurations and operating conditions. For instance, changing the position of a transmitter or a receiver antenna
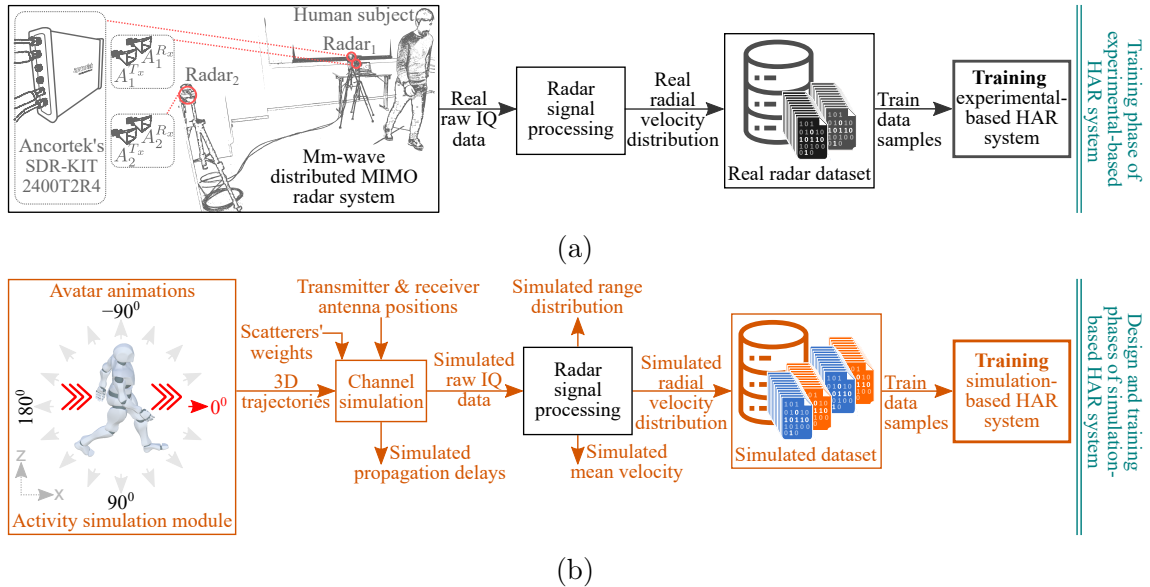
Figure E.1: (a) Design of conventional (experimental-based) direction-independent HAR systems that require human subjects and a MIMO radar system for their training. (b) Design of the proposed simulation-based HAR system that requires the simulated radar signatures for its training.
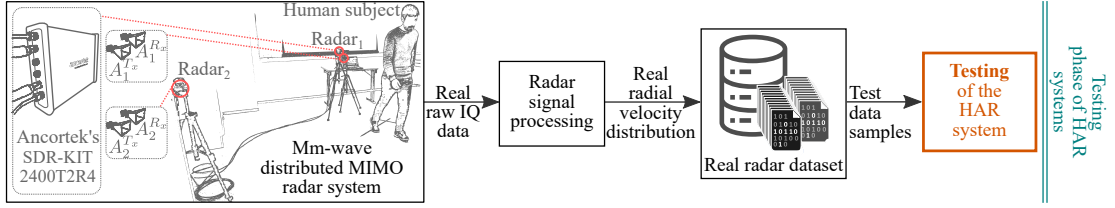
Figure E.2: The testing phase of both experimental and simulation-based direction-independent HAR systems. In the testing phase, the performance of the simulation-based HAR system is evaluated against unseen real radar signatures.

of the MIMO radar system can invalidate the entire recorded training dataset.

E.2.2 A Simulation-Based Design of HAR Systems

In this paper, we propose a feasible alternative to overcome the aforementioned limitations of radar-based classifiers, particularly with regard to the scarcity of radar data. To develop real-world HAR systems, we propose a comprehensive simulation-based framework that utilizes MoCap systems to synthesize realistic MIMO radar data, as depicted in Fig. E.1(b). The objective is to generate a simulated MIMO radar-based training dataset by seamlessly simulating a large number of realistic MIMO radar signatures without real human subjects and a physical radar system.

The block diagram in Fig. E.1(b) provides a general overview of the proposed end-to-end simulation framework for HAR systems. In Fig. E.1(b), the activity simulation module synthesizes the five types of human activities in the 3D space from motion data collected by the MoCap systems (see Sect. E.3). The activity simulation module simulates 3D trajectories corresponding to different body segments of an avatar, e.g., head, neck, torso, and upper and lower extremities. To simulate the human activities in multiple directions as shown in Fig. E.1(b), we rotate the positions of the transmitter antenna $A_i^{T_x}$ and receiver antenna $A_i^{R_x}$ in our simulation-based framework for $i = 1, 2$ (see Sect. E.5.2). For a desired antenna configuration of the MIMO radar system, our channel simulation module first transforms the 3D trajectories into TV propagation delays. Then, the channel simulation module generates realistic RF or raw IQ data for the simulated TV propagation delays and a set of scatterer weights. Eventually, the radar signal processor arranges the simulated raw IQ data in the fast- and slow-time domain and processes it to simulate realistic radar signatures, i.e., range distribution, radial velocity distribution (micro-Doppler signature), and mean velocity (mean Doppler shift).

We synthesize numerous examples of the five types of human activities, simulate the corresponding radial velocity distributions (micro-Doppler signatures), and store them in our simulated radar dataset, as shown in Fig. E.1(b). The proposed simulation-based framework has no limits on the generation of simulation data. The simulated radar dataset is used to train the simulation-based HAR system, which is based on a DCNN architecture. To demonstrate the practical importance and the generalizability of this simulation-based framework, we need to evaluate its perfor-

mance in a real scenario. Therefore, the proposed simulation-based HAR system is evaluated on a previously unseen real radar dataset acquired with a mm-wave distributed MIMO radar system and real human subjects, as shown in the testing phase of Fig. E.2. Note that we used an identical radar signal processing block in Fig. E.1 because the simulated and real RF signals are structurally indistinguishable. More details on each block of the simulation-based HAR system are provided in the following sections.

E.3   Human Motion Capture and Synthesis

This section explores several ways of capturing and synthesizing human motion. First, biomechanical modeling and its limitations will be briefly discussed. Second, wearable sensors as a means of MoCap systems are briefly mentioned. Third, we discuss optical motion capture systems such as Mixamo [37] and Qualisys [38]. It is important to highlight that the proposed simulation-based framework allows incorporating synthesized or recorded motion data from diverse sources such as biomechanical, wearable, and optical MoCap systems. Lastly, we explain the process of generating 3D trajectories of human body segments using software such as Unity [32] and Autodesk's MotionBuilder [33]. These software programs (3D animation tools) help us augment the motion data at the motion-layer synthesis.

E.3.1   Biomechanical Modeling of Human Body Segments

The utility of biomechanical modeling [39] for human body segments is undeniable, yet its complexity is inherently high, primarily due to the intricate nature of the human body. Also, it is difficult to develop generalizable biomechanical models because individuals differ in physiology, anatomy, and motor function. Moreover, the interaction between the human body and the environment can further increase the complexity of a biomechanical model.

Obtaining high-fidelity motion data of human body segments can be more feasible and accessible through MoCap repositories and systems such as Mixamo and Qualisys. In addition, the Unity and MotionBuilder software provide a cost-effective and pragmatic alternative to biomechanical modeling, enabling the seamless and dynamic simulation of new motion data in a virtual environment. Therefore, we use MoCap systems to capture the human motion and employ 3D animation tools from MotionBuilder and Unity to synthesize and subsequently augment human motion.

E.3.2   Wearable Motion Capture Systems

Wearable MoCap systems offer a versatile and cost-effective solution for capturing human movement data. The sensors, typically accelerometers and gyroscopes, are often integrated into garments to capture data on the orientation and acceleration of body segments. In this area, Rokoko Smartsuit Pro [40] is a viable choice with multiple inertial sensors for real-time tracking of an individual's skeletal movements.

It facilitates seamless transfer of motion data to various applications such as sports, biomechanical analysis, and virtual reality. Compared to optical MoCap systems, wearable MoCap systems have limitations in terms of accuracy. Additionally, wearable MoCap systems can suffer from magnetic interference, which can affect the precision of the MoCap data.

E.3.3 Optical Motion Capture Systems

We used Mixamo and Qualisys optical MoCap systems to capture motion data for human activities. Mixamo is an online platform that offers an extensive selection of readily available MoCap data captured from real performers [41]. Our Qualisys MoCap system was based on six Miqus M3 cameras connected in a daisy chain, capable of tracking passive reflective markers placed on a subject at 340 frames per second (fps). The Qualisys MoCap system includes proprietary Qualisys track manager (QTM) software that provides an interface for tasks such as camera configuration and calibration, session setup and organization, marker-set definition, and MoCap measurements. Furthermore, QTM offers a suite of tools for marker labeling, data processing, analysis, and the export of MoCap data, thereby enabling seamless integration with third-party software. The camera system was calibrated according to the QTM guidelines to ensure accurate tracking of the markers and capturing their position and orientation in 3D space. Next, 41 passive reflective markers were attached to a full body suit. The participant wore the suit, and we recorded a MoCap trial to generate an automatic identification of markers (AIMs) model. This model applies computer vision, localization, and motion estimation techniques to detect and track markers, facilitating an automated workflow for identifying and labeling markers. Once the AIM model was created, the skeleton solver function of QTM was used to calibrate the skeleton based on the marker positions. Next, a person's motion data was recorded for four activities: normal walking, standing up from a chair, sitting down onto a chair from a standing position, and picking up a small object from the floor. The recorded skeleton data was then exported in the Filmbox (FBX) file format and further processed in the MotionBuilder software. Note that for the falling activity, the MoCap data was relatively difficult to collect due to markers attached to the body. Therefore, we obtained MoCap data of the falling activity from Mixamo [37], a freely-accessible online platform. In the next step, we import the acquired MoCap data into specialized software such as Unity or MotionBuilder, which are equipped with powerful tools that allow for the creation of comprehensive, meticulous and lifelike 3D animations.

E.3.4 3D Trajectories of Human Body Segments

By using the basic MoCap data and the 3D animation tools, we synthesized, augmented and visualized five human activities: falling on the floor, walking in an indoor environment, standing up from and sitting down on a chair, and picking up an object from the floor. Initially, the human activities were simulated and varied in a single

direction or at an aspect angle of zero degrees with the help of 3D animation tools, as shown in Fig. E.1(b). However, we also needed to synthesize multi-directional human activities to realize a simulated MIMO radar-based direction-independent HAR system. Instead of using 3D animation tools, we simulated multi-directional human activities more conveniently and efficiently by spatially rotating the transmitter and receiver antennas of the radar subsystem, Radar_i (see Sect. E.5.2).

Following the synthesis of the human movement, we extract the spatial trajectories corresponding to each body segment of the avatar. To track the different body segments, 21 simulated point scatterers were placed on the avatar (see Fig. E.3), these model the actual body scatterers that backscatter the transmitted RF signal to the receiver antennas of the 2×2 distributed MIMO radar system. We recorded the TV positions (trajectories) of the simulated point scatterers in the 3D space for the simulated human activities. For example, the 3D trajectories of the simulated point scatterers for a simulated walking activity are shown in Fig. E.3.

At the outset, only 34 MoCap files were recorded, each representing one of the five distinct types of human activities. We visualized these activities using the Unity and MotionBuilder 3D animation tools, and computed the corresponding 3D trajectories. To expand, the total number of synthesized human activities to 84, we applied data augmentation at the motion-layer synthesis using the Unity and MotionBuilder software (see Sect. E.5.1). Subsequently, we processed the 3D trajectories in MATLAB for further data augmentation at the physical- and signal-layer syntheses. Although data augmentation at the motion-layer synthesis may require some attention to motion details, the physical-layer synthesis and signal-layer synthesis data augmentation stages in the proposed simulation-based framework are fairly automated. With the help of such multi-stage data augmentation techniques, we generated 2826 micro-Doppler signatures (TV radial velocity distributions) for each radar subsystem of the MIMO radar system. Section E.5 provides more details on the multi-stage data augmentation techniques furnished by the proposed simulation-based framework.

E.4   Channel Modeling and Simulation

In this section, we first present a geometrical 3D indoor channel that models an indoor propagation scenario using the proposed simulation-based framework (see Fig. E.3). Second, we investigate the multipath components caused by non-stationary simulated (real) point scatterers on avatar (human) body segments and simulate the corresponding TV propagation delays for a human activity. Lastly, we explain how the simulated propagation delays can be used to synthesize a received RF signal, specifically for an FMCW 2×2 MIMO radar system.

E.4.1   Geometrical Channel Model

We model and simulate a 3D channel for an indoor environment, which consists of a 2×2 distributed MIMO radar system, a moving person, and stationary miscellaneous

items such as furniture and electronics, as illustrated in Fig. E.1(a). Recall that Radar_i represents the ith radar subsystem of the distributed MIMO radar system, $A_i^{T_x}$ is the ith transmitter antenna, and $A_i^{R_x}$ is the ith receiver antenna for $i = 1, 2$. Let $[\cdot]^\top$ denote the vector transpose operation. Then, the position of the ith transmit (receive) antenna $A_i^{T_x}$ ($A_i^{R_x}$) of the 2×2 MIMO radar system is represented by $\mathcal{C}_i^{T_x} = [x_i^{T_x}, y_i^{T_x}, z_i^{T_x}]^\top$ ($\mathcal{C}_i^{R_x} = [x_i^{R_x}, y_i^{R_x}, z_i^{R_x}]^\top$), as illustrated in Fig. E.3.

A virtual propagation environment that resembles a real geometrical 3D indoor channel is depicted in Fig. E.3. In a real propagation environment, a moving human subject has countless non-stationary scatterers. For this research, we model these non-stationary bodily scatterers with $\mathcal{L} = 21$ non-stationary simulated point scatterers on a moving avatar, as shown in Fig. E.3. Moreover, in Fig. E.3, $\mathcal{C}_l(t) = [x_l(t), y_l(t), z_l(t)]^\top$ denotes the TV spatial trajectory of the lth marker $S^{(l)}$, $d_{l,i}^{T_x}(t)$ ($d_{l,i}^{R_x}(t)$) represents TV Euclidean distance between the lth marker $S^{(l)}$ and the ith transmit antenna $A_i^{T_x}$ (receive antenna $A_i^{R_x}$), where $i = 1, 2$ and $l = 1, 2, \ldots, \mathcal{L}$.

For the lth marker $S^{(l)}$ and the ith radar subsystem Radar_i, the TV radial distance $d_{l,i}(t)$ is equal to one-half of the overall propagation distance, i.e., $d_{l,i}(t) = (d_{l,i}^{T_x}(t) + d_{l,i}^{R_x}(t))/2$. Fig. E.3 shows that the antenna configuration $\{\mathcal{C}_i^{T_x}, \mathcal{C}_i^{R_x}\}$ of the ith radar subsystem, Radar_i, follows a monostatic configuration, where $\mathcal{C}_i^{T_x} = \mathcal{C}_i^{R_x}$ for $i = 1, 2$. This leads to the following simplification: $d_{l,i}(t) = d_{l,i}^{T_x}(t) = d_{l,i}^{R_x}(t)$. The obtained TV radial distances $d_{l,i}(t)$ of the $\mathcal{L}$ non-stationary simulated point scatterers play an important role in simulating the TV propagation delays $\tau_i^{(l)}(t)$, as explained in the following subsection.
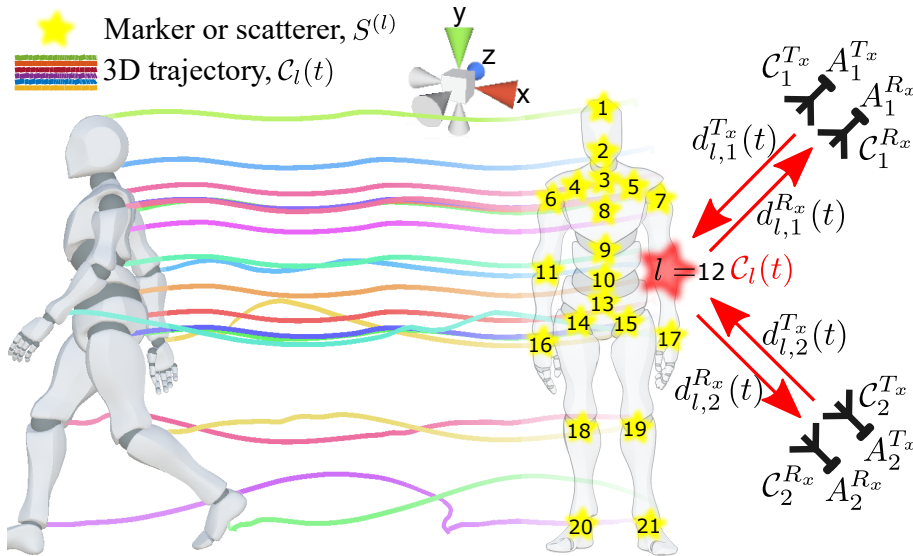

Figure E.3: Virtual 3D propagation environment comprising a non-stationary avatar with 21 simulated point scatterers on its body segments and a simulated 2×2 multiperspective MIMO radar system.

E.4.2 Modeling of Multipath Components Caused by Human Body Segments

RF signals generally experience multipath propagation, particularly in indoor environments with numerous stationary and non-stationary reflective objects. In Fig. E.3, the transmitted RF signal takes on multiple propagation paths, traveling from the transmitter antenna to the receiver antenna via multiple real (simulated) point scatterers on the human (avatar) body segments. Recall that in our simulation-based framework, the 21 simulated point scatterers on the avatar's body segments basically model the actual bodily scatterers that scatter the transmitted RF signal back to the receiver antennas of the 2×2 distributed MIMO radar system. For this study, by virtue of the cross-channel interference mitigation technique [16], we assume that the two radar subsystems, $Radar_1$ and $Radar_2$, of Ancortek's mm-wave radar system do not interfere with each other.

In the proposed simulation-based framework, we only consider multipath components originating from the $\mathcal{L} = 21$ non-stationary dominant and non-dominant scatterers located on various body segments of the avatar, as shown in Fig. E.3. The multipath components originating from stationary dominant scatterers, such as walls, furniture, and floor, are excluded from the analysis because they are easily filtered out through signal preprocessing. Moreover, the bistatic components of the 2×2 distributed MIMO radar systems are not considered for this study. However, if required, the bistatic components of the 2×2 distributed MIMO radar system can be easily simulated in the proposed simulation-based framework.

The receiver antennas receive the multipath components, or multiple copies of the transmitted RF signal, with distinct TV propagation delays $\tau_i^{(l)}(t)$. For $Radar_i$, the lth TV propagation delay $\tau_i^{(l)}(t)$ is related to the lth TV radial distances $d_{l,i}(t)$ according to the relation $\tau_i^{(l)}(t) = 2d_{l,i}(t)/c_0$, where c_0 is the speed of light. Within the framework of radar sensing, the synthesized motion is completely characterized by the simulated TV propagation delays $\tau_i^{(l)}(t)$, as explained in the subsequent section.

For the five distinct types of simulated human activities and $Radar_1$, Fig. E.4 shows the simulated TV propagation delays $\tau_1^{(l)}(t)$ of the $\mathcal{L} = 21$ simulated point scatterers. The lth TV propagation delay $\tau_i^{(l)}(t)$ depends solely on the spatial trajectory of the lth marker. Therefore, when a person suddenly falls, the abrupt change in the spatial positions of the upper-body segment is reflected in the corresponding TV propagation delays $\tau_1^{(l)}(t)$, as illustrated in Fig. E.4. In Fig. E.4, the TV propagation delays $\tau_1^{(l)}(t)$ demonstrate the repetitive nature of the walking activity. By analyzing the TV propagation delays $\tau_1^{(l)}(t)$, it is evident that the simulated walking activity comprised four steps towards $Radar_1$. In contrast, the TV propagation delays $\tau_1^{(l)}(t)$ in Fig. E.4 for the other three types of simulated human activities in place, namely sitting, standing up and picking up an object, show smaller variations corresponding to the mobility of the simulated point scatterers.
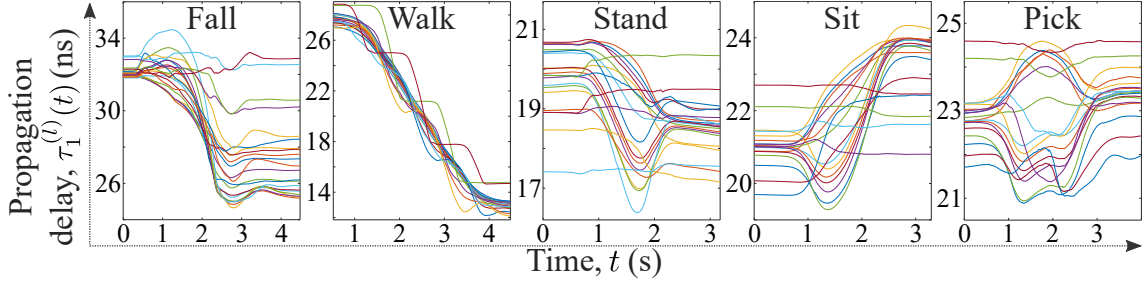
Figure E.4: The simulated TV propagation delays $\tau_1^{(l)}(t)$ of the $\mathcal{L}$ simulated point scatterers for the five distinct human activities and Radar$_1$.

E.4.3 Channel Modelling for Radio-Frequency Sensing

This section elucidates the simulation of a composite RF signal or equivalently, raw IQ data in fast time t' and slow time t, corresponding to a specific motion. To simulate the composite RF signal of Radar$_i$, we need user-defined scatterer weights, a user-defined antenna configuration $\{\mathcal{C}_i^{T_x}, \mathcal{C}_i^{R_x}\}$, and the simulated TV propagation delays $\tau_i^{(l)}(t)$ corresponding to the spatial trajectories of the simulated point scatterers for a specific motion or a human activity (see Fig. E.1(b) and Fig. E.4). For this study, we consider the $\mathcal{L}$ bodily scatterers to be long-time non-stationary over the slow time t, and short-time stationary over a limited chirp duration T_{sw} [42]. In the following, for the FMCW 2×2 distributed MIMO radar system placed in the indoor wireless channel, we synthesize the complex baseband signal called the composite beat signal $s_{b,i}(t', t)$ [43], where $i = 1, 2$. Additionally, we discuss an interpolation procedure that is integral to our channel-simulation module of Fig. E.1(b), as it mitigates the issues of aliasing in the Doppler domain.

FMCW radar systems operate by repetitively emitting a chirp waveform $c(t')$ [44], which is scattered back to the receiver antenna by multiple stationary and non-stationary scatterers present on the human body segments and other objects in the environment. A quadrature mixture element integrated into the receiver chain of the FMCW 2×2 distributed MIMO radar system is responsible for transforming the incoming passband RF signal into complex baseband (composite beat) signal $s_{b,i}(t', t)$. The received complex baseband signal $s_{b,i}(t', t)$ is sampled in the fast-time domain by the analog to digital converter (ADC) module of the receiver with the discrete sampling interval T_s in the fast-time domain. Subsequently, for the coherent processing interval (CPI) of the ith radar subsystem, Radar$_i$, the discrete samples of the received complex baseband signal $s_{b,i}(t', t)$ are organized in fast- and slow-time domains. During the CPI, the phase of Radar$_i$ is preserved. This organization or rearrangement of the discrete fast- and slow-time samples results in the radar's raw

IQ data matrix $\mathcal{D}_i$ [42], which can be expressed as

$$
\mathcal{D}_i =
\begin{bmatrix}
s_{b,i}(0,0) & s_{b,i}(T_s,0) & \cdots \\
s_{b,i}(0,T_{sw}) & s_{b,i}(T_s,T_{sw}) & \cdots \\
\vdots & \vdots & \vdots \\
s_{b,i}(0,(N_c-1)T_{sw}) & s_{b,i}(T_s,(N_c-1)T_{sw}) & \cdots
\end{bmatrix}
$$

(E.1)

$$
\begin{bmatrix}
s_{b,i}(T_{sw}-T_s,0) \\
s_{b,i}(T_{sw}-T_s,T_{sw}) \\
\vdots \\
s_{b,i}(T_{sw}-T_s,(N_c-1)T_{sw})
\end{bmatrix}
$$

where N_c represents the number of chirps present within the CPI of the FMCW radar system.

We want to synthesize the actual received complex baseband signal $s_{b,i}(t',t)$ of the FMCW 2×2 distributed MIMO radar system, so that we can simulate the radar's raw IQ data matrices $\mathcal{D}_i$ for $i=1,2$. The received complex baseband signal $s_{b,i}(t',t)$ of Radar$_i$ can be synthesized by adding up the $\mathcal{L}$ distinct beat signals $s_{b,i}^{(l)}(t',t)$ [42, 43], each corresponding to the lth multipath component originating from the lth simulated point scatterer, i.e.,

$$
s_{b,i}(t',t) = \sum_{l=1}^{\mathcal{L}} s_{b,i}^{(l)}(t',t).
$$

(E.2)

For Radar$_i$, the lth beat signal $s_{b,i}^{(l)}(t',t)$ or the lth multipath component can be simulated by using the expression [42]

$$
s_{b,i}^{(l)}(t',t) = \sum_{n=0}^{\infty} a_i^{(l)}(t)\exp\left[j\left(2\pi f_{b,i}^{(l)}(t)t' + \phi_i^{(l)}(t)\right)\right] \times
$$
$$
\delta(t - \tau_i^{(l)}(t) - T_n)
$$

(E.3)

where $a_i^{(l)}(t)$, $f_{b,i}^{(l)}(t)$, and $\phi_i^{(l)}(t)$ denote the TV path gain, beat frequency, and phase of the lth beat signal $s_{b,i}^{(l)}(t',t)$, respectively, and $\delta(\cdot)$ denotes the Dirac delta function. The symbol T_n in (E.3) represents the nth discrete slow-time instance, which is determined by the chirp duration T_{sw}, such that $T_n = nT_{sw}$, where n is a non-negative integer. Let γ represent the slope of the chirp signal. Then, the lth TV beat frequency $f_{b,i}^{(l)}(t)$ of Radar$_i$ in (E.3) is given by $f_{b,i}^{(l)}(t) = \tau_i^{(l)}(t)\gamma$. The ith and lth TV phase $\phi_i^{(l)}(t)$ component is determined by the TV propagation delay $\tau_i^{(l)}(t)$ according to $\phi_i^{(l)}(t) = 2\pi f_0 \tau_i^{(l)}(t)$, where f_0 is the carrier frequency.

The TV path gain $a_i^{(l)}(t)$ in (E.3) models the strength of the lth marker in the received signal. For Radar$_i$ and $\mathcal{L}$ simulated point scatterers, we use time-invariant path gains $a_i^{(l)}$ in (E.3) to avoid unnecessary complexity. Therefore, we have $a_i^{(l)}(t) = a_i^{(l)}$. In this study, for the five types of synthesized human activities, the values of the time-invariant path gains $a_i^{(l)}$ are adjusted by investigating the

actual TV radial velocity distributions $p_i(v,t)$ (see Sect. E.6) and the body surface area [45]. It is worth noting that by using different sets of time-invariant path gains, we can augment the radar data at the signal-layer synthesis for a synthesized human activity (see Sect. E.5.3).

We consider the $\mathcal{L}$ bodily scatterers to be long-time non-stationary over the slow time t, and short-time stationary over the fast time t' for a limited chirp duration T_{sw} [42]. Thus, the TV propagation delays $\tau_i^{(l)}(t)$, beat frequencies $f_{b,i}^{(l)}(t)$, and phases $\phi_i^{(l)}(t)$ of the $\mathcal{L}$ simulated point scatterers are only a function of the slow time t. For Radar$_i$ and the kth slow-time instant t_k (kth row of the raw IQ data matrix $\mathcal{D}_i$ in (E.1)), the short-time stationarity assumption simplifies the synthesis of the discrete complex baseband signal $s_{b,i}(t', t_k)$ for a synthesized human activity. At the slow-time instant t_k, the IQ components of the complex baseband signal $s_{b,i}(t', t_k)$ can be digitally simulated as a sum of tone signals, i.e., $s_{b,i}(t', t_k) = \sum_{l=1}^{\mathcal{L}} s_{b,i}^{(l)}(t', t_k)$, where the lth tone signal $s_{b,i}^{(l)}(t', t_k)$ has the constant (time-invariant) beat frequency $f_{b,i}^{(l)}(t_k)$ and phase $\phi_i^{(l)}(t_k)$.

Within the framework of radar sensing, the synthesized motion can be completely characterized by the simulated TV propagation delays $\tau_i^{(l)}(t)$ of the $\mathcal{L}$ simulated (real) point scatterers. The $\mathcal{L}$ TV propagation delays $\tau_i^{(l)}(t)$ are computed from the TV spatial trajectories $\mathcal{C}_l(t)$ of the $\mathcal{L}$ simulated point scatterers, which are animated with a fixed frame interval denoted by T_f. Therefore, the frame interval T_f is the slow-time sampling interval of the simulated TV spatial trajectories $\mathcal{C}_l(t)$ and the propagation delays $\tau_i^{(l)}(t)$. In actual radar systems, the slow-time sampling interval is equal to the radar's pulse repetition interval (PRI), which is smaller (better) than the frame interval T_f. Concretely, for the actual (simulated) raw IQ data matrix $\mathcal{D}_i$ in (E.1), the slow-time sampling interval T_{sw} is equal to the radar's PRI (frame interval T_f). Thus, to ensure that the simulated frame interval T_f is equal to the radar's PRI, we interpolate the spatial trajectories or the simulated TV propagation delays $\tau_i^{(l)}(t)$ in our simulation framework. This is necessary because the upper limit of the actual (synthesizable) radial velocity, denoted by $v_{\max}$ ($v'_{\max}$), is determined by the radar's PRI (animation's frame interval T_f). Let λ denote the wavelength, then we have $v_{\max} = \lambda/(4 \cdot \text{PRI})$, and $v'_{\max} = \lambda/(4T_f)$.

E.5 Multi-Stage Data Augmentation

In this section, we explore multi-stage data augmentation techniques (see Fig. E.5) provided by the proposed simulation-based framework that allow us to simulate large quantities of quality radar signatures. First, we discuss a motion-layer data augmentation technique, where various animation parameters and avatar characteristics, e.g., size and speed, can be randomly varied to synthesize a variety of human motions. We then explain data augmentation of the physical layer that allows us to vary numerous physical-layer configurations and the radar's operating parameters, e.g., number of antennas and their setup and PRI. Lastly, we delve into a data augmentation technique at the signal-layer synthesis.

MoCap data sources
- Online repositories
- Vision sensors
- Wearable sensors
- Biomechanical models

MoCap data files = 34

→ 3D trajectories →

Motion-layer synthesis
- Retarget MoCap data using Unity, MotionBuilder, etc.
- Vary avatar parameters, such as size location, speed, and acceleration
- Simulate user-defined number of bodily scatterers

Simulated human activity files = 84

→ 3D trajectories →

Physical-layer synthesis
- Simulate radar sensor nodes
  - SISO or MIMO $\{C_i^{T_x}, C_i^{R_x}\}$
  - Monostatic, bistatic, or multistatic
- Convert uni-directional to multi-directional motion data

Radar$_1$ propagation delays

Radar$_2$ propagation delays

...

Signal-layer synthesis
- User-defined scatterer weights (TV path gains)
- Vary backscattered energy
- Vary radar operating parameters
  - PRI, carrier frequency, etc.

Simulated RF files = 5652

Radar$_1$ Simulated IQ signal
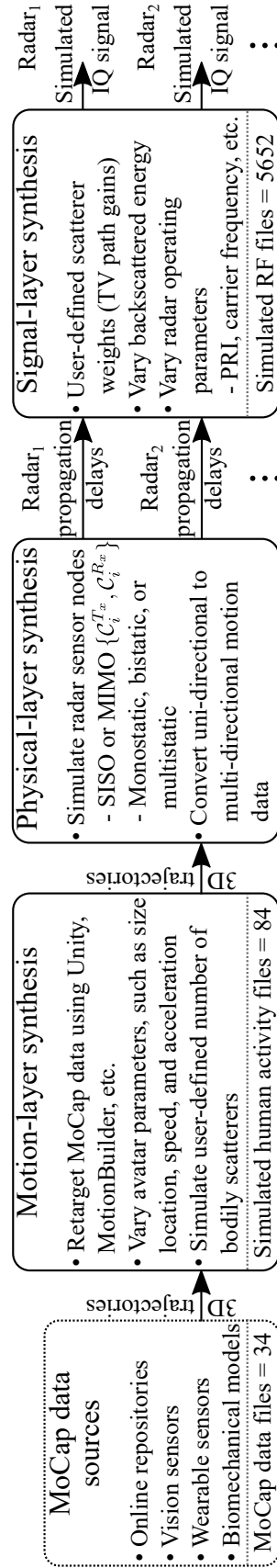
Radar$_2$ Simulated IQ signal

...

Figure E.5: The proposed multi-stage data augmentation techniques offered by our simulation-based framework.

E.5.1 Motion-Layer Synthesis

For the five types of distinct human activities, we acquired a small and basic MoCap dataset from the Mixamo platform and the Qualisys MoCap system. A person with a height of about 1.74 m performed the activities several times in a room equipped with the Qualisys MoCap system. The MoCap dataset we acquired comprised only 34 MoCap files, each representing one of the five types of activities. The 3D animation tools from both Unity and MotionBuilder software were used to visualize the basic MoCap data for the human activities. We complemented the basic MoCap data with the 3D animation tools to render realistic and diverse motion data.

In this study, one of our objectives is to synthesize a large amount of data representing real human motions at the motion-layer synthesis of our simulation-based framework. To this end, we first adjusted the height of the avatar in the MotionBuilder software by reducing it to 1.52 m (5 ft) and increasing it to 1.83 m (6 ft). We then aligned the MoCap data to the avatars with different sizes to account for the effects of avatar dimensions and extended the data on the motion-layer synthesis. Therefore, in the Unity and MotionBuilder software, the total number of synthesized human activities were increased to 84 by applying data augmentation at the motion-layer synthesis stage, as indicated by Fig. E.5. Note that we can synthesize complex, varied, and entirely new sequences of human movements by using the blend tree animation tool in the Unity software that facilitates seamless transitions between multiple humanoid animations. For the augmented human-motion data (synthesized human activities), we computed TV spatial trajectories (see Sect. E.3.4) and imported them into MATLAB for further data augmentation at the physical- and signal-layer syntheses (see Fig. E.5).

E.5.2 Physical-Layer Synthesis

The simulation-based framework allows the adjustment of the radar operating parameters and physical layer configurations, e.g. PRI, carrier frequency f_c, bandwidth B_w, and antenna configuration $\{\mathcal{C}_i^{T_x}, \mathcal{C}_i^{R_x}\}$. Through these adjustments it is possible to both extend the simulated radar data and simulate specific scenarios. At the physical-layer synthesis data augmentation stage, appropriate antenna configurations $\{\mathcal{C}_i^{T_x}, \mathcal{C}_i^{R_x}\}$ were chosen to simulate the two radar subsystems, Radar$_1$ and Radar$_2$, as shown in Fig. E.6. To maintain consistency with the actual 2×2 distributed MIMO radar system depicted in Fig. E.1(a), the emulated radar system's operating parameters, such as PRI, carrier frequency f_c, and bandwidth B_w, were kept the same.

We first simulated different positions of the radar subsystems, Radar$_1$ and Radar$_2$, by using the rotation matrix $R_y(\theta_{Ri})$, which can be expressed as [46]

$$R_y(\theta_{Ri}) = \begin{bmatrix} \cos\theta_{Ri} & 0 & \sin\theta_{Ri} \\ 0 & 1 & 0 \\ -\sin\theta_{Ri} & 0 & cos\theta_{Ri} \end{bmatrix} \tag{E.4}$$

where θ_{Ri} denotes the clockwise angular rotation along the y-axis for Radar$_i$ and
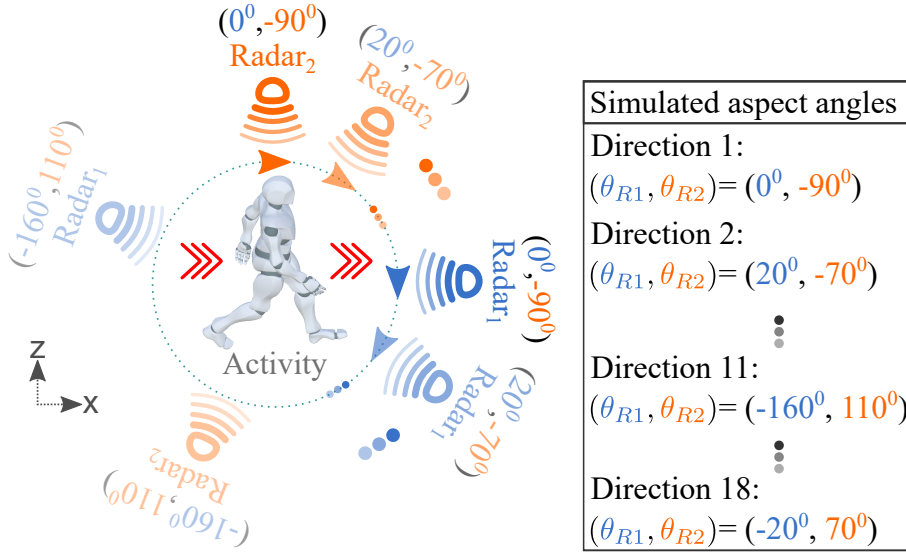
Figure E.6: Simulation of a multiperspective 2×2 MIMO radar system and multidirectional human activities.

$i = 1, 2$. Initially, the simulated radar subsystems, Radar_1 and Radar_2, were placed at $\theta_{R1} = 0°$ ($\mathcal{C}_1^{T_x} \approx \mathcal{C}_1^{R_x} \approx [3, 1.1, 0]^\top$) and $\theta_{R2} = -90°$ ($\mathcal{C}_2^{T_x} \approx \mathcal{C}_2^{R_x} \approx [0, 1.1, 3]^\top$), respectively, i.e., $\mathcal{C}_2^{T_x} = R_y(-90°)\mathcal{C}_1^{T_x}$ and $\mathcal{C}_2^{R_x} = R_y(-90°)\mathcal{C}_1^{R_x}$. In other words, Radar_2 can be simulated by simply rotating Radar_1 counterclockwise by 90 degrees along the y-axis, as illustrated in Fig. E.6. Using this method, we emulated a 2×2 distributed MIMO radar system, similar to the actual radar system in Fig. E.1(a), to simulate the MIMO radar signatures. Note that, with the use of the rotation matrix $R_y(\theta_{Ri})$, any number of radar subsystems, sensors, or nodes can be simulated at the physical-layer synthesis data augmentation stage.

Recall that the human activities were initially simulated with 3D animation tools in a single direction or at an aspect angle of zero degrees. However, to develop a simulated MIMO radar-based direction-independent HAR system, we required multidirectional human activities. Compared to the motion-layer synthesis, the required multi-directional human activities can be simulated more easily and efficiently at the physical-layer synthesis data augmentation stage. The multi-directional human activities are simulated by spatially rotating the transmitter and receiver antennas of the radar subsystem Radar_i, for $i = 1, 2$. The angular difference between the two radar subsystems is always kept at 90 degrees, i.e., $\theta_{R1} - \theta_{R2} = 90°$, as depicted in Fig. E.6. The different rotations of Radar_1 and Radar_2 (θ_{R1}, θ_{R2}) correspond to the different directions of the human activities, where $(\theta_{R1}, \theta_{R2}) \in [-180°, 180°)$. We simulated 18 different directions of the human activities at the physical-layer synthesis data augmentation stage, namely Direction 1 to Direction 18, as illustrated in Fig. E.6. For instance, for a human activity, Direction 11 in Fig. E.6 corresponds to the scenario, where $(\theta_{R1}, \theta_{R2}) = (-160°, 110°)$.

To summarize, at the physical-layer synthesis, we first simulated the two radar subsystems, Radar_1 and Radar_2, to emulate the 2×2 distributed MIMO radar system. Secondly, by using the rotation method, we simulated the multi-directional hu-

man activities by simultaneously rotating the two radar subsystems, as illustrated in Fig. E.6. Thus, our proposed simulation-based framework includes a physical-layer synthesis data augmentation stage, which efficiently and conveniently transforms and augments uni-directional motion data into multi-directional motion data and single radar data into multiple radar data.

E.5.3   Signal-Layer Synthesis

The signal-layer synthesis data augmentation stage of the proposed simulation-based framework allows to simulate realistic and diverse TV radial velocity distributions $p_i(v, t)$ (micro-Doppler signatures) for a single human activity. Using (E.3), we can simulate numerous multipath components corresponding to the stationary and non-stationary scatterers in the received complex baseband signal $s_{b,i}(t', t)$ (see (E.2)). In this research, multipath components originating from stationary scatterers, such as walls and furniture, are not considered, as they can be effectively filtered out during the signal preprocessing stage. However, if necessary, the signal-layer synthesis can simulate numerous complex propagation scenarios, e.g., those with or without radar clutter, by adjusting the path gains $a_i^{(l)}(t)$, beat frequencies $f_{b,i}^{(l)}(t)$, and phases $\phi_i^{(l)}(t)$ of the lth beat signal $s_{b,i}^{(l)}(t', t)$ for Radar$_i$.

For the five types of synthesized human activities, we first adjusted the values of the time-invariant path gains $a_i^{(l)}$ by looking into the actual radar signatures (TV radial velocity distributions $p_i(v, t)$ (see Sect. E.6)) and the body surface area [45]. Subsequently, we augmented the simulated radar signatures by varying the power levels (time-invariant path gains $a_i^{(l)}$) of the individual multipath component. Therefore, at the signal-layer synthesis of the proposed simulation-based framework, we augmented the radar data by using different sets of time-invariant path gains $a_i^{(l)}$ for the five types of synthesized human activities.

In this section, we discussed three data augmentation techniques implemented at multiple layers of the proposed simulation-based framework: the motion-layer synthesis, physical-layer synthesis, and signal-layer synthesis. By applying these multi-stage data augmentation techniques, we simulated 2826 TV radial velocity distributions $p_i(v, t)$ (micro-Doppler signatures) for each radar subsystem of the 2×2 MIMO radar system. In other words, for the two radar subsystems, Radar$_1$ and Radar$_2$, a total of 5652 TV radial velocity distributions $p_i(v, t)$ were simulated. To conclude, the multi-stage data augmentation methods in the proposed simulation-based framework are quite useful and they allowed for increased variability, realism, and diversity in the simulated radar dataset. With these methods, we were able to transform and augment the basic motion data (34 MoCap files) into 5652 radar signatures, which indicates the utility of the proposed simulation-based approach for realizing radar-based classifiers.

E.6   MIMO Radar Signatures

In this section, we delineate the radar signal processing module of Fig. E.1 that generates the MIMO radar signatures: range distribution, TV radial velocity distribution $p_i(v,t)$ (micro-Doppler signature), and mean velocity (mean Doppler shift). For $i = 1, 2$, the radar signal processing module transforms the actual and the simulated complex baseband signals $s_{b,i}(t',t)$ into the TV radial velocity distributions $p_i(v,t)$. The first step is to compute the beat frequency function $S_{b,i}(f_b,t)$ as [47]

$$S_{b,i}(f_b,t) = \int_0^{T_{sw}} s_{b,i}(t',t)e^{-j2\pi f_b t'}dt' \tag{E.5}$$

where f_b refers to the beat frequency.

Let f and $f_{b,\max}$ denote the Doppler frequency and maximum beat frequency, respectively. Then, the micro-Doppler signatures $S_i(f,t)$ are obtained from the beat frequency function $S_{b,i}(f_b,t)$ according to the relation [35]

$$S_i(f,t) = \left| \int_0^{f_{b,\max}} \int_{-\infty}^{\infty} S_{b,i}(f_b,t'')W_r(t''-t)e^{-j2\pi f t''}dt''df_b \right|^2 \tag{E.6}$$

where t'' denotes the running time, and $W_r(\cdot)$ denotes a rectangular window function that spans over 64 chirp intervals. According to [43], the TV radial velocity distribution $p_i(v,t)$ can be obtained as

$$p_i(v,t) = \frac{S_i\left(\frac{2f_0}{c_0}v,t\right)}{\int_{-\infty}^{\infty} S_i\left(\frac{2f_0}{c_0}v,t\right)dv} \tag{E.7}$$

where v denotes the radial velocity. From the TV radial velocity distribution $p_i(v,t)$ in (E.7), we can compute the TV mean radial velocity $\bar{v}_i(t)$ as [43]

$$\bar{v}_i(t) = \int_{-\infty}^{\infty} vp_i(v,t)dv. \tag{E.8}$$

For Radar$_i$, the TV beat-frequency signatures $S_i'(f_b,t)$ can be computed as

$$S_i'(f_b,t) = \left| \int_0^{\text{PRF}} \int_{-\infty}^{\infty} S_{b,i}(f_b,t'')W_r(t''-t)e^{-j2\pi f t''}dt''df \right|^2 \tag{E.9}$$

where PRF is the pulse repetition frequency of the radar system, and $i = 1, 2$. Finally, for the 2×2 MIMO radar system, the TV range distribution $p_i'(r,t)$ can be obtained as [42]

$$p_i'(r,t) = \frac{S_i'\left(\frac{2\gamma}{c_0}r,t\right)}{\int_{-\infty}^{\infty} S_i'\left(\frac{2\gamma}{c_0}r,t\right)dr}. \tag{E.10}$$

Recall that the real (simulated) point scatterers on the human (avatar) body segments, each with unique TV radial velocity components, scatter the transmitted RF signal back to the receiver antennas of the 2×2 distributed MIMO radar system. For Radar$_i$ and the $\mathcal{L}$ distinct non-stationary real (simulated) point scatterers, the TV radial velocity distribution $p_i(v,t)$ in (E.7) indicates the strengths of the radial velocity components over the slow time t (see Fig. E.7). The TV mean radial velocity $\bar{v}_i(t)$ in (E.8), obtained from the TV radial velocity distribution $p_i(v,t)$, shows the weighted average of the velocity components of all $\mathcal{L}$ real (simulated) bodily scatterers over the slow time t (see Fig. E.8). Moreover, the strengths of the TV radial distances of all $\mathcal{L}$ non-stationary real (simulated) point scatterers over the slow time t are provided by the TV range distributions $p_i'(r,t)$. Due to the current practical limitations of radar systems, the TV range distributions $p_i'(r,t)$ are not usually used to realize HAR systems, so their simulation results are omitted for brevity. However, for completeness and possible future applications, we have included the expression in (E.10) to simulate the TV range distribution $p_i'(r,t)$.

In Section E.5, we saw that multi-directional human activities can be simulated by simultaneously rotating the two radar subsystems, Radar$_1$ and Radar$_2$, as shown in Fig. E.6. For some of the 18 directions and all five types of simulated (actual) human activities, the simulated (actual) TV radial velocity distributions, $p_1(v,t)$ and $p_2(v,t)$, are shown in Fig. E.7(a) (Fig. E.7(b)). The images of the simulated (ac-
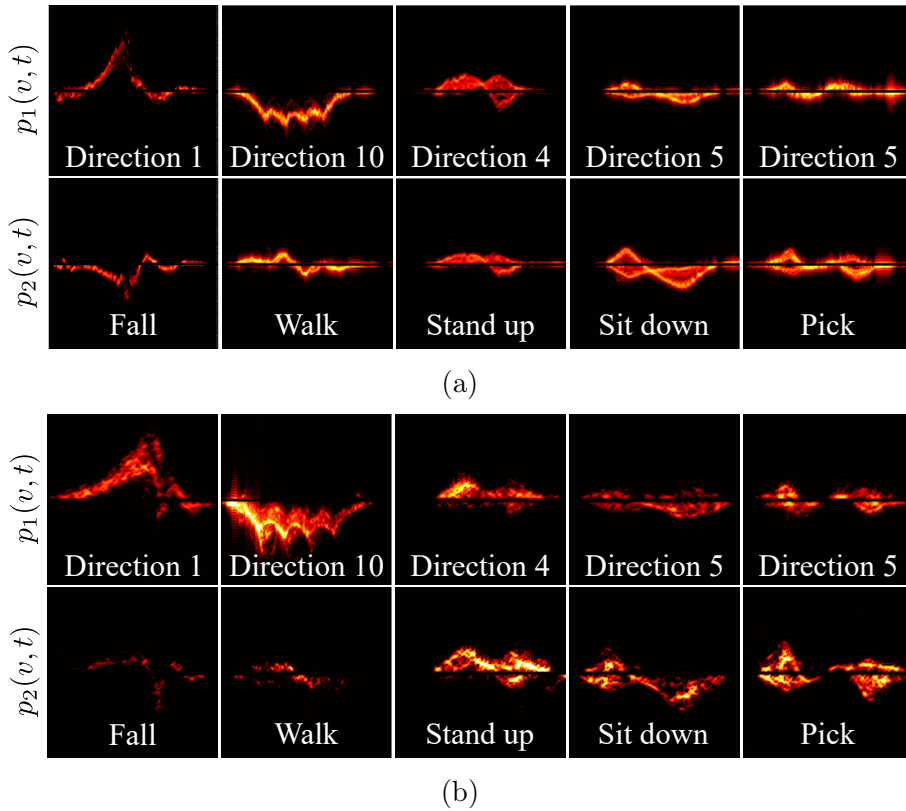

(a)

(b)

Figure E.7: (a) Simulated TV radial velocity distributions, $p_1(v,t)$ and $p_2(v,t)$, for the emulated multi-directional human activities. (b) Real TV radial velocity distributions, $p_1(v,t)$ and $p_2(v,t)$, for the real multi-directional human activities.
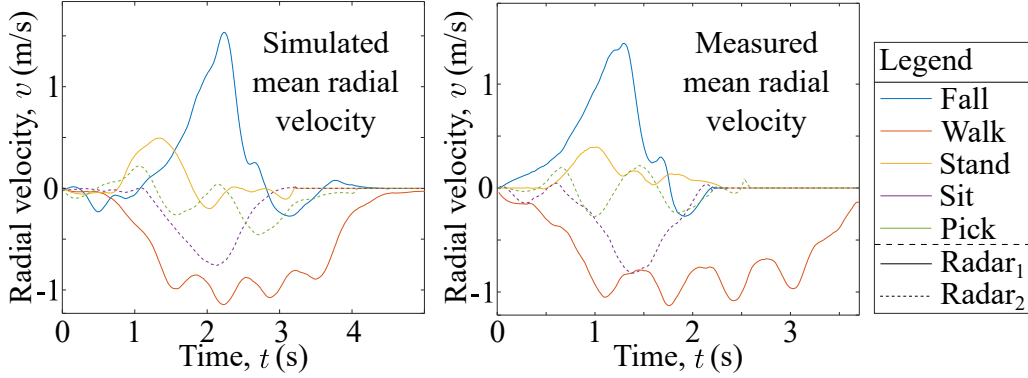
Figure E.8: Simulated and measured TV mean radial velocities $\bar{v}_1(t)$ and $\bar{v}_2(t)$ of Radar$_1$ (—) and Radar$_2$ ($\cdots$), respectively.

tual) TV radial velocity distributions, $p_1(v,t)$ and $p_2(v,t)$, were used to train (test) the proposed 2×2 MIMO radar-based direction-independent HAR system. In the subsequent section, the two colored images of the TV radial velocity distributions, $p_1(v,t)$ and $p_2(v,t)$, will serve as input feature maps to the HAR system. Moreover, for the five types of human activities and the two radar subsystems, Radar$_1$ and Radar$_2$, the simulated and actual TV mean radial velocities $\bar{v}_i(t)$ are depicted in Fig. E.8. The utility and effectiveness of the proposed simulation-based framework is evident from the high-fidelity simulated radar signatures, which are quite similar to the actual radar signatures, as exemplified by Figs. E.7 and E.8.

To quantitatively assess the similarity between simulated and real radar signatures, we employ the dynamic time warping (DTW) algorithm [48]. Table E.1 presents the normalized DTW distances between the real and simulated TV mean radial velocities $\bar{v}_i(t)$ from Fig. E.8 across five human activities. Remarkably, the DTW distance metric is minimized for each activity, indicating close resemblance between the simulated and real radar signatures. For example, for the sitting activity, a DTW distance of 0.01 between the simulated and real TV mean radial velocities $\bar{v}_i(t)$ demonstrates precise replication of this pattern. This consistent trend across all activities confirms the accuracy of our approach in simulating realistic radar data.

Table E.1: The DTW distance metric is calculated for the simulated and real (actual) TV mean radial velocities $\bar{v}_i(t)$ of Fig. E.8.

| Activity type | Real | | | | |
	fall	walk	stand	sit	pick
fall	**0.03**	0.30	0.14	0.21	0.14
walk	0.28	**0.01**	0.36	0.03	0.22
stand	0.06	0.24	**0.03**	0.11	0.05
sit	0.17	0.05	0.17	**0.01**	0.08
pick	0.09	0.13	0.08	0.04	**0.02**

(Simulated)

E.7 Simulation-Based HAR System

This section elucidates the training and testing phases of our simulation-based direction-independent HAR system that was realized by using a DCNN-based multiclass classifier. First, we look into the design of the HAR classifier and its training with the simulated radar dataset. Then, to demonstrate the practical importance and the generalizability of our proposed simulation framework in real-world scenarios, we used a real 2×2 MIMO radar dataset to evaluate the classification performance of the trained simulation-based direction-independent HAR system.

E.7.1 Design of the Simulation-Based HAR System

To develop a simulation-based HAR system, we first created a large labeled dataset of simulated radar signatures. For Radar_i $(i = 1, 2)$ of the 2×2 MIMO radar system and the five types of humanoid activities, we simulated 2826 TV radial velocity distributions $p_i(v, t)$ by using the proposed multi-stage data augmentation techniques of our simulation-based framework (see Sect. E.5). Thus, the simulated radar dataset consisted of a total of 5652 simulated TV radial velocity distributions $p_i(v, t)$, which were used to train the proposed simulation-based direction-independent (multiperspective) HAR classifier.

The simulation-based direction-independent HAR system comprises two feature extraction networks (FENs) and a multilayer perceptron (MLP) network. Fig. E.9(a) illustrates the FEN that computes relevant features from the simulated (actual) TV radial velocity distributions $p_i(v, t)$ during the training (testing) phase for the ith radar subsystem, Radar_i. It consists of four convolutional layers, containing 64, 72, 80, and 96 two-dimensional (2D) trainable kernels with dimension k_d either equal to 4×4 pixels or 3×3 pixels. Each 2D kernel uses the rectified linear unit (ReLU) activation function to avoid the problem of vanishing gradients [49]. The max-pool layers were employed to reduce redundancies in the feature maps. To avoid overfitting the training data, we used dropout layers with the dropout rates of 10% and 15% for the FEN and MLP, respectively. The flatten layer of our FEN rearranges the extracted features into a vector of order 18816×1, as shown in Fig. E.9(a).

The two FENs in the DCNN-based multiperspective HAR system are identical, as shown in Fig. E.9(b). As Radar_1 and Radar_2 illuminate the indoor environment from multiple perspectives, the extracted features from the two TV radial velocity distributions, $p_1(v, t)$ and $p_2(v, t)$, are merged by the multiperspective feature fusion block, as shown in Fig. E.9(b). Subsequently, based on the received multiperspective features, the MLP network is trained to detect the type of the human activity. The multiperspective feature fusion block enables the HAR classifier to recognize the human activities regardless of their directions. Note that the design of this multiperspective deep neural network closely resembles the architecture reported in [35]. To train the parameters of our DCNN-based multiperspective HAR classifier, we used the adaptive moment estimation (Adam) optimizer [50] and the simulated radar signatures of multi-directional human activities. The training dataset, comprising

2826 pairs of simulated TV radial velocity distributions $p_i(v, t)$, was further divided into training and validation subsets in an $80 : 20$ ratio. During the training phase, our DCNN-based multiperspective HAR classifier showed no signs of overfitting, as demonstrated by the training and validation curves in Fig. E.10.

E.7.2   Testing of the Simulation-Based HAR System

To evaluate the performance of the trained 2×2 MIMO radar-based multiperspective HAR classifier in a real-world setting, we used a real radar dataset recorded by Ancortek SDR-KIT 2400T2R4, as shown in Fig. E.2. The operating parameters and antenna configurations of the real and the simulated 2×2 MIMO radar systems were kept similar for consistency. Specifically, we set the PRI, carrier frequency f_c, and bandwidth B_w of the real and simulated MIMO radar systems to 0.5 ms, 24.125 GHz, and 250 MHz, respectively. For Radar$_1$ and Radar$_2$, the antennas were placed at $\mathcal{C}_1^{T_x} \approx \mathcal{C}_1^{R_x} \approx [3, 1.1, 0]^\top$ and $\mathcal{C}_2^{T_x} \approx \mathcal{C}_2^{R_x} \approx [0, 1.1, 3]^\top$, respectively.

A total of 875 multi-directional human activities were recorded with the 2×2
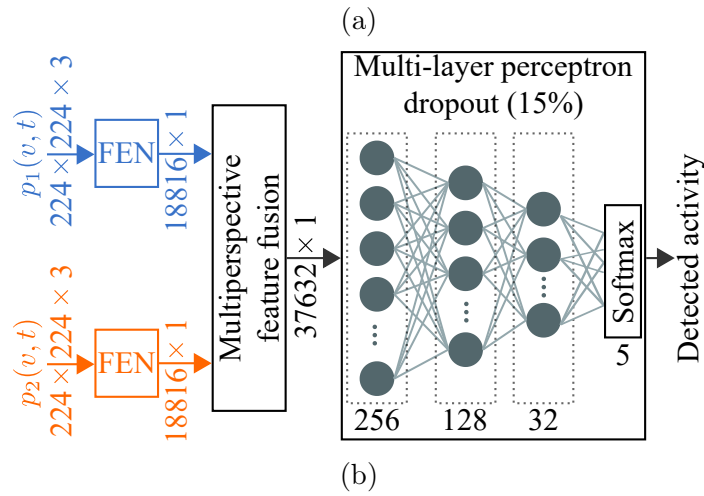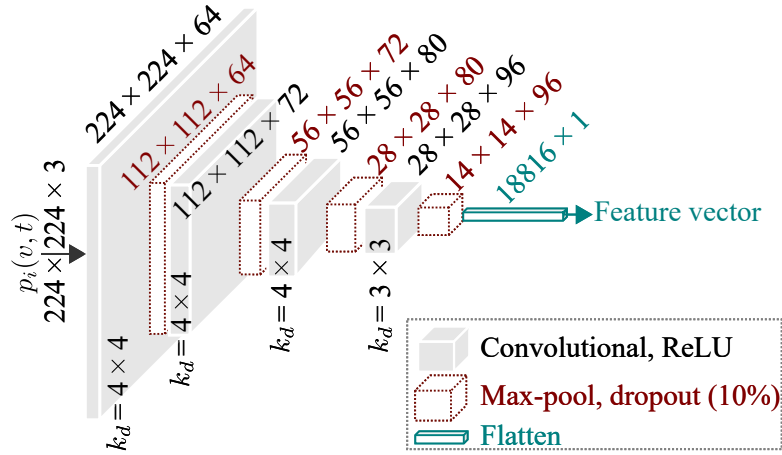

Figure E.9: (a) Feature extraction network, FEN, based on convolutional filters. (b) The proposed DCNN-based multiperspective HAR classifier that is trained (tested) using the simulated (real) HAR dataset.
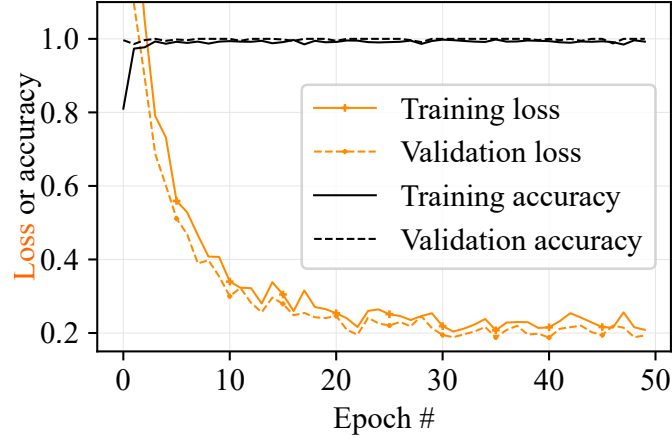
Figure E.10: Training history of our simulation-based direction-independent HAR system.

MIMO radar system from six human subjects, including a female participant. Thus, the real radar dataset consisted of 1750 TV radial velocity distributions $p_i(v, t)$ (micro-Doppler signatures) for the two radar subsystems, Radar_1 and Radar_2. As a direct result of this extensive measurement campaign, the simulation–real (training–testing) data ratio approximately came out to be $76\!:\!24$. Our simulation-based framework enabled the realization of the simulation-based direction-independent HAR system, which exhibited remarkable performance and efficacy in the real world, as demonstrated by the confusion matrix in Fig. E.11. For each of the five types of multi-directional human activities, the number of correct classifications is represented by the first five diagonal entries of the confusion matrix. The green colored entries in the last row and column exhibit the precision and recall [51] in Fig. E.11. Finally, the white colored entry of the confusion matrix shows the overall classification accuracy of our simulation-based direction-independent HAR system, which is 97.83%. As our test dataset was sufficiently balanced, the macro average F1-score [52] came out to be approximately 97.6%, which is close to the overall classification accuracy.

For RF-based HAR systems, asserting the superiority of one method proves challenging, given their tailored design to address diverse research challenges. Nonetheless, Table E.2 presents the performance of various contemporary HAR systems, utilizing classification accuracy for comparison. Notably, the measurement-based HAR methods and those partially utilizing measurement data demonstrate strong classification accuracies. The Vid2Doppler [53] method, which translates video to radar data, achieves an accuracy of 81.4%, while our simulation-based approach, converting MoCap data to radar data, achieves a higher accuracy of 97.8%, both utilizing entirely simulated training data.

This section demonstrated the utility and efficacy of the simulation-based framework in the real world. The classification accuracy of the simulation-based direction-independent HAR system is comparable to the current HAR systems [59, 18], with the additional consideration of the multi-directional human activity recognition problem. Moreover, our simulation-based framework is unique in its abil-

	Fall	Walk	Stand	Sit	Pick	
Fall	94	2	0	1	0	96.91% 3.09%
Walk	2	151	0	0	0	98.69% 1.31%
Stand	2	0	163	0	0	98.79% 1.21%
Sit	0	1	3	223	4	96.54% 3.46%
Pick	4	0	0	0	225	98.25% 1.75%
	92.16% 7.84%	98.05% 1.95%	98.19% 1.81%	99.55% 0.45%	98.25% 1.75%	97.83% 2.17%
	Fall	Walk	Stand	Sit	Pick	

True Class (vertical axis) — Predicted Class (horizontal axis)

Figure E.11: Confusion matrix of our simulation-based multiperspective HAR classifier with a classification accuracy of 97.83%.

ity to generate realistic, diverse, and unlimited labeled MIMO radar datasets with software-defined operating parameters and configurations. Therefore, the proposed simulation-based framework in Fig. E.1(b) can be readily used to develop other SISO and MIMO radar-based classifiers, e.g., for sign language detection.

E.8   Conclusion

The progression of SISO and MIMO radar-based classifiers is primarily impeded by the unavailability of large labeled training datasets. Therefore, as a proof-of-concept, we have presented in this work a simulation-based approach to address the concern of data scarcity for monostatic, bistatic, and multistatic SISO and MIMO radar systems. Although our focus was on realizing a 2×2 MIMO radar-based direction-independent HAR system, the utility of our simulation-based framework extends beyond HAR applications.

The proposed simulation-based framework provides the flexibility to synthesize software-defined human movements using MoCap data-driven activity simulation. We proposed a MIMO channel model to convert simulated 3D trajectories into received RF signals, while considering a user-defined antenna configuration of a distributed MIMO radar system and the multipath components emanating from the non-stationary simulated point scatterers. The synthesized RF signals were further processed to simulate the multiperspective MIMO radar signatures used to implement our simulation-based direction-independent HAR system.

To generate a diverse training dataset for radar-based HAR systems, we introduced multi-stage data augmentation techniques at the motion-layer synthesis, physical-layer synthesis, and signal-layer synthesis within our simulation-based

Table E.2: Comparing the classification performance of state-of-the-art RF-based HAR approaches.

Approach	Training data type & brief description	Accuracy
Wi-Sense [14]	Involves measurement data; Obtaining fingerprints from the Wi-Fi channel state information (CSI)	97.8%
Convolutional neural network (CNN)-LSTM [54]	Involves measurement data; Unsupervised learning and data fusion using LSTM	92%
Few-shot adversarial domain adaptation (FS-ADA) [55]	Involves measurement data; Discovering shared feature spaces in training datasets	91.6%
CNN-recurrent neural network (RNN) [56]	Involves measurement data; Extracting patterns over space and time	90.8%
Joint domain and semantic transfer learning (JDS-TL) [12]	Involves measurement data; Utilizing 10% labeled radar data for domain adaptation	87.6%
Wasserstein refined generative adversarial network with gradient penalty (WRGAN-GP) [57]	Synthesizing data with GANs utilizing limited measurements; Employing a refinement model to synthesize high-fidelity spectrograms	94.9%
Multibranch generative adversarial network (MBGAN) [58]	GAN-centered data generation using limited real data; Physics-informed method improving GANs for accurate micro-Doppler synthesis	89.2%
Vid2Doppler [53]	Simulation-based training; Converting video data into Doppler data	81.4%
Our simulation- based approach	Simulation-based training; Converting motion data into radar signatures	97.8%

framework. The multi-stage data augmentation techniques helped to gain absolute control over various factors, such as avatar size, location, velocity, acceleration, PRI, and radar antenna configuration. By using these techniques, we augmented the basic MoCap data to 5652 micro-Doppler signatures, drastically minimizing the overall training workload and demonstrating the effectiveness of our simulation-based approach for realizing radar-based classifiers. Our MIMO radar-based HAR system trained on the simulated micro-Doppler signatures achieved classification accuracy of 97.83% when tested with actual radar data. As our study eliminates the need for direct involvement of human participants and an actual radar system, we believe that the proposed proof-of-concept will be of great importance for training future SISO/MIMO radar-based classifiers.

Our MIMO channel model opens up new research perspectives for modelling received RF signals at the scatterer level. For example, future studies can explore

the optimization of scatterer-level parameters, such as the simulated TV path gains. A limitation of this research is that the methods discussed are not directly applicable to the moving clutter scenario where the radar antennas are non-stationary. This research gap is beyond the scope of this work and can be addressed in future studies.

Acknowledgments

Bibliography

[1] C. Hu, S. Hudson, M. Ethier, M. Al-Sharman, D. Rayside, and W. Melek, "Sim-to-real domain adaptation for lane detection and classification in autonomous driving," in *2022 IEEE Intell. Veh. Symp.*, pp. 457–463, 2022.

[2] M. Hernandez, G. Epelde, A. Alberdi, R. Cilla, and D. Rankin, "Synthetic data generation for tabular health records: A systematic review," *Neurocomputing*, vol. 493, pp. 28–45, 2022.

[3] B. Kiefer, D. Ott, and A. Zell, "Leveraging synthetic data in object detection on unmanned aerial vehicles," in *2022 26th Int. Conf. Pattern Recognit.*, pp. 3564–3571, 2022.

[4] K. Hagn and O. Grau, "Improved sensor model for realistic synthetic data generation," in *Comput. Sci. Cars Symp.*, pp. 1–9, 2021.

[5] M. A. Bansal, D. R. Sharma, and D. M. Kathuria, "A systematic review on data scarcity problem in deep learning: Solution and applications," *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 1–29, 2022.

[6] H. Tang and K. Jia, "A new benchmark: On the utility of synthetic data with Blender for bare supervised learning and downstream domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 15954–15964, 2023.

[7] V. Thambawita, P. Salehi, S. A. Sheshkal, S. A. Hicks, H. L. Hammer, S. Parasa, T. de Lange, P. Halvorsen, and M. A. Riegler, "SinGAN-Seg: Synthetic training data generation for medical image segmentation," *PLoS One*, vol. 17, no. 5, 2022. doi:10.1371/journal.pone.0267976.

[8] M. A. A. Al-qaness, A. Dahou, M. A. Elaziz, and A. M. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," *IEEE Trans. Ind. Informatics*, vol. 19, no. 1, pp. 144–152, 2023.

[9] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, "RF-based human activity recognition using signal adapted convolutional neural network," *IEEE Trans. Mob. Comput.*, vol. 22, no. 1, pp. 487–499, 2023.

[10] Y. Tang, L. Zhang, F. Min, and J. He, "Multiscale deep feature learning for human activity recognition using wearable sensors," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 2106–2116, 2023.

[11] Y. Tang, L. Zhang, Q. Teng, F. Min, and A. Song, "Triple cross-domain attention on human activity recognition using wearable sensors," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 5, pp. 1167–1176, 2022.

Bibliography

[12] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022.

[13] L. Zhang, W. Cui, B. Li, Z. Chen, S. Member, M. Wu, S. Member, and T. S. Gee, "Privacy-preserving cross-environment human activity recognition," *IEEE Trans. Cybern.*, pp. 1–11, 2021.

[14] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-Sense: a passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems," *Ann. Telecommun.*, vol. 77, no. 3-4, pp. 163–175, 2022.

[15] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from FMCW radar with Bi-LSTM networks," *IEEE Sens. J.*, vol. 20, no. 22, pp. 13607–13619, 2020.

[16] S. Waqar and M. Pätzold, "Interchannel interference and mitigation in distributed MIMO RF sensing," *Sensors*, vol. 21, no. 22, 2021. doi:10.3390/s21227496.

[17] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3617–3628, 2019.

[18] W.-Y. Kim and D.-H. Seo, "Radar-based human activity recognition combining range–time–Doppler maps and range-distributed-convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[19] G. Lee and J. Kim, "Improving human activity recognition for sparse radar point clouds: A graph neural network model with pre-trained 3D human-joint coordinates," *Appl. Sci.*, vol. 12, no. 4, 2022. doi:10.3390/app12042168.

[20] Z. Xia and F. Xu, "Time-space dimension reduction of millimeter-wave radar point-clouds for smart-home hand-gesture recognition," *IEEE Sens. J.*, vol. 22, no. 5, pp. 4425–4437, 2022.

[21] H. Liu, A. Zhou, Z. Dong, Y. Sun, J. Zhang, L. Liu, H. Ma, J. Liu, and N. Yang, "M-gesture: Person-independent real-time in-air gesture recognition using commodity millimeter wave radar," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3397–3415, 2022.

[22] M. M. Rahman, E. A. Malaia, A. C. Gurbuz, D. J. Griffin, C. Crawford, and S. Z. Gurbuz, "Effect of kinematics and fluency in adversarial synthetic data generation for ASL recognition with RF sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 2732–2745, 2022.

[23] Z. Yu, A. Taha, W. Taylor, A. Zahid, K. Rajab, H. Heidari, M. A. Imran, and Q. H. Abbasi, "A radar-based human activity recognition using a novel 3-D point cloud classifier," *IEEE Sens. J.*, vol. 22, no. 19, pp. 18218–18227, 2022.

[24] Y. Zhao, A. Yarovoy, and F. Fioranelli, "Angle-insensitive human motion and posture recognition based on 4D imaging radar and deep learning classifiers," *IEEE Sens. J.*, vol. 22, no. 12, pp. 12173–12182, 2022.

[25] C. Yu, Z. Xu, K. Yan, Y.-R. Chien, S.-H. Fang, and H.-C. Wu, "Noninvasive human activity recognition using millimeter-wave radar," *IEEE Syst. J.*, vol. 16, no. 2, pp. 3036–3047, 2022.

[26] Y. Yang, Y. Zhang, Y. Lang, B. Li, S. Guo, and Q. Tan, "GAN-based radar spectrogram augmentation via diversity injection strategy," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[27] Z. Liu, C. Wu, and W. Ye, "Category-extensible human activity recognition based on Doppler radar by few-shot learning," *IEEE Sens. J.*, vol. 22, no. 22, pp. 21952–21960, 2022.

[28] Y. Yang, Y. Zhang, H. Ji, B. Li, and C. Song, "Radar-based human activity recognition under the limited measurement data support using domain translation," *IEEE Signal Process. Lett.*, vol. 29, pp. 1993–1997, 2022.

[29] M. J. Bocus, W. Li, S. Vishwakarma, R. Kou, C. Tang, K. Woodbridge, I. Craddock, R. McConville, R. Santos-Rodriguez, K. Chetty, and R. Piechocki, "OPERAnet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors," *Sci. Data*, vol. 9, no. 1, 2022. doi:10.1038/s41597-022-01573-2.

[30] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, "SimHumalator: An open-source end-to-end radar simulator for human activity recognition," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 3, pp. 6–22, 2022.

[31] G. Manfredi, P. Russo, A. De Leo, and G. Cerri, "Efficient simulation tool to characterize the radar cross section of a pedestrian in near field," *Prog. Electromagn. Res. C*, vol. 100, pp. 145–159, 2020.

[32] "Unity Real-Time Development Platform | 3D, 2D VR & AR Engine." Unity, accessed Feb. 20, 2023. [Online]. Available: https://www.unity.com/.

[33] "Autodesk MotionBuilder Software." Autodesk, accessed Jun. 21, 2023. [Online]. Available: https://www.autodesk.com/products/motionbuilder/overview?term=1-YEAR&tab=subscription.

[34] S. Ahmed, W. Kim, J. Park, and S. H. Cho, "Radar-based air-writing gesture recognition using a novel multistream CNN approach," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 23869–23880, 2022.

[35] S. Waqar, M. Muaaz, and M. Pätzold, "Direction-independent human activity recognition using a distributed MIMO radar system and deep learning," *IEEE Sens. J.*, vol. 23, no. 20, pp. 24916–24929, 2023.

[36] M. Muaaz, S. Waqar, and M. Pätzold, "Orientation-independent human activity recognition using complementary radio frequency sensing," *Sensors*, vol. 23, no. 13, 2023. doi:10.3390/s23135810.

[37] "Get animated." Mixamo, accessed Feb. 20, 2023. [Online]. Available: https://www.mixamo.com/#/.

[38] "Qualisys | Motion Capture Systems." Qualisys, accessed Jun. 21, 2023. [Online]. Available: https://www.qualisys.com/.

[39] I. Roupa, M. R. da Silva, F. Marques, S. B. Gonçalves, P. Flores, and M. T. da Silva, "On the modeling of biomechanical systems for human movement analysis: A narrative review," *Arch. Comput. Methods Eng.*, vol. 29, no. 7, pp. 4915–4958, 2022.

[40] "Smartsuit Pro II - Quality body motion capture in one simple mobile mocap suit." Rokoko, accessed Jul. 10, 2023. [Online]. Available: https://www.rokoko.com/products/smartsuit-pro.

[41] Mixamo, "Mixamo motion capture (behind the scenes)." YouTube, Aug. 24, 2012. [Online]. Available: https://www.youtube.com/watch?v=q7jWMd7ix98.

[42] S. Waqar and M. Pätzold, "A simulation-based framework for the design of human activity recognition systems using radar sensors," *IEEE Internet Things J.*, 2023. doi:10.1109/JIOT.2023.3344179.

[43] S. Waqar, M. Muaaz, and M. Pätzold, "Human activity signatures captured under different directions using SISO and MIMO radar systems," *Appl. Sci.*, vol. 12, no. 4, 2022.

[44] S. Sana, S. Waqar, H. Yusaf, M. Waqas, and F. A. Siddiqui, "Software defined digital beam forming processor," in *2016 13th Int. Bhurban Conf. Appl. Sci. Technol.*, pp. 671–676, 2016.

[45] M. Mance, M. Prutki, A. Dujmovic, M. Milošević, V. Vrbanovic-Mijatovic, and D. Mijatovic, "Changes in total body surface area and the distribution of skin surfaces in relation to body mass index," *Burns*, vol. 46, no. 4, pp. 868–875, 2020.

[46] A. Vilfan, "Myosin V passing over Arp2/3 junctions: Branching ratio calculated from the elastic lever arm model," *Biophys. J.*, vol. 94, no. 9, pp. 3405–3412, 2008.

[47] M. Muaaz, S. Waqar, and M. Pätzold, "Radar-based passive step counter and its comparison with a wrist-worn physical activity tracker," in *4th Int. Conf. Intell. Technol. Appl. INTAP*, pp. 259–272, Springer International Publishing, 2022.

[48] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series.," in *Proc. 3rd Int. Conf. Knowl. Discov. Data Min.*, pp. 359–370, 1994.

[49] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, pp. 2684–2691, 2017.

[50] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv Prepr.*, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980.

[51] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. ICML*, pp. 233–240, 2006.

[52] H. Schutze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval.* Cambridge University Press, 2008.

[53] K. Ahuja, Y. Jiang, M. Goel, and C. Harrison, "Vid2Doppler: Synthesizing Doppler radar data from videos for training privacy-preserving activity recognition," in *Proc. 2021 CHI Conf. Hum. Factors Comput. Syst.*, pp. 1–10, may 2021.

[54] H.-U.-R. Khalid, A. Gorji, A. Bourdoux, S. Pollin, and H. Sahli, "Multi-view CNN-LSTM architecture for radar-based human activity recognition," *IEEE Access*, vol. 10, pp. 24509–24519, 2022.

[55] X. Li, Y. He, J. A. Zhang, and X. Jing, "Supervised domain adaptation for few-shot radar-based human activity recognition," *IEEE Sens. J.*, vol. 21, no. 22, pp. 25880–25890, 2021.

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

[57] L. Qu, Y. Wang, T. Yang, and Y. Sun, "Human activity recognition based on WRGAN-GP-synthesized micro-Doppler spectrograms," *IEEE Sens. J.*, vol. 22, no. 9, pp. 8960–8973, 2022.

[58] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, "Physics-aware generative adversarial networks for radar-based human activity recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 3, pp. 2994–3008, 2023.

[59] Y. Zhao, H. Zhou, S. Lu, Y. Liu, X. An, and Q. Liu, "Human activity recognition based on non-contact radar data and improved PCA method," *Appl. Sci.*, vol. 12, no. 14, p. 7124, 2022.