UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

mmWave Sensing for Vital Sign Monitoring and User Identification: Enhancing Accuracy and Robustness via Deep Learning

by

Yingqi Wang

A THESIS SUBMITTED
IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Sydney, Australia

2025

Certificate of Original Authorship

I, Yingqi Wang declare that this thesis, is submitted in fulfilment of the requirements for the award of the degree of Doctor of Philosophy, in the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research was supported by an Australian Government Research Training Program (RTP) Scholarship doi.org/10.82133/C42F-K220.

Signature: Production Note:
Signature removed prior to publication.

Date: 22 September 2025

mmWave Sensing for Vital Sign Monitoring and User Identification: Enhancing Accuracy and Robustness via Deep Learning

by
Yingqi Wang

Abstract

Millimeter-wave (mmWave) sensing has emerged as a powerful sensing modality for non-contact, privacy-preserving human monitoring applications. However, such technology could still lose the efficiency in real-world sensing scenarios. This thesis explores the integration of mmWave sensing for vital sign monitoring and user identification, leveraging deep learning and advanced signal processing to enhance accuracy and robustness in dynamic environments. The first focus of this research is on vital sign monitoring, where mmWave radar is employed to detect respiratory rate, heart rate, arterial pulse, and blood pressure (BP). A multi-modal sensing technology and a novel multi-channel variational mode decomposition (VMD) method are investigated to separate vital sign components from interference caused by body movements and environmental noise. Further, a robust heartbeat and wrist localization algorithm and a physics-driven deep learning approach are developed, incorporating Neural Ordinary Differential Equations (Neural ODEs) and Temporal Convolutional Networks (TCNs) to reconstruct high-level physiological signals from radar Doppler shifts. The system achieves high-accuracy estimation of physiological parameters, demonstrating its feasibility for continuous, non-invasive health monitoring. The second focus is user identification, aiming to associate the user's ID with the previous vital signs sensing. A novel Inverse Synthetic Aperture Radar (ISAR) sensing is investigated to reconstruct body profiles of individuals walking past the radar. A ResNet-based deep learning model with Additive Angular Margin Loss is developed to enhance feature discrimination, enabling high-precision user identification in long-term. Unlike traditional biometric methods, this approach identifies individuals based on their radio imaging profiles, making it robust to occlusions and appearance variations. The contributions of this work include: (1) a novel multi-modal mmWave localization and signal decomposition is proposed and addressed the challenges of low accuracy in practical vital sign monitoring, (2) an adaptive mmWave heart and wrist detection, and the physics-embedded learning scheme is proposed to improve flexibility, robustness and explainability in ECG (Electrocardiogram), pulse and BP monitoring, and (3) an ISAR-based user identification method optimized with deep learning for real-world scenarios, significant improved the performance in long-term mmWave user Re-ID. This research paves the way for future applications in smart healthcare, security, and human-computer interaction.

Supervisor: Prof. Min Xu
School of Electrical and Data Engineering

Acknowledgements

The three-and-half-year Ph.D study at UTS has been a wonderful experience. I would like to acknowledge several people who not only made this thesis well finished, but also brought a lot of support, joy, and happiness into this amazing academic journey.

First and foremost, I would like to sincerely thank my supervisor, Prof. Min Xu, for her continually supervision and encouragement during my Ph.D. study. We have collaborated on a number of awesome projects since I started the Ph.D program three and half years ago. Her sharp intuition and passion for knowledge have influenced me to identify the problems we are having and to discover the novel research perspectives. This would be much helpful for my future career. I feel I am very fortunate to have been supervised by and working with her in the past years.

I would like to express my great thanks and appreciation to my co-supervisor Prof. Andrew Zhang for insightful guidance and firm support on my theoretical study and experiment conduction.

Many thanks to Prof. Qiang Wu and A/Prof. Peiyuan Qing for their valuable advices and suggestions for my candidature assessment one and two, which are much helpful for improving the quality of this thesis. I would like to give thanks to Dr. Zhongqin Wang, who has provided useful insight for methodology and my paperwork.

I would like to acknowledge my labmates in the Aural and Visual Intelligence Lab: Dr. Haimin Zhang, Dr Jiahao Xia, Yongze Wang, Minghui Wei, and Xiaoxu Li.

I would like to acknowledge with gratitude the love of my parents: Mr Yaming Wang and Mrs Pingping Liu, and my partner Dr Xutong Zhang. Their support has always been unconditional. This thesis could not have been well finished without their support.

Finally, I would like to thank the anonymous reviewers for reviewing this thesis.

List of Publications

The contents of this thesis are based on the following papers that will be submitted or have been published or accepted, or preprints that have been under submission or submitted to peer-reviewed journals.

Publications:

1. Wang, Y., Wang, Z., Zhang, J.A., Zhang, H. and Xu, M., 2023. Vital sign monitoring in dynamic environment via mmwave radar and camera fusion. IEEE Transactions on Mobile Computing, 23(5), pp.4163-4180.
2. Wang, Y., Xu, M., Wang, Z., Wang, Y. and Zhang, J.A., 2025. User Re-Identification Through mmWave Radio Imaging. IEEE Internet of Things Journal (Accepted and to be published).
3. Y. Wang, Z. Wang, J. A. Zhang and M. Xu, ”Sensing Single User's ECG, Pulse and Blood Pressure Simultaneously Through mmWave Radar via Physics-Informed Learning,” to be submitted to IEEE Transactions on Mobile Computing.

Contents

List of Figures

List of Tables

Abbreviation

- mmWave: millimetre wave
- FMCW: frequency-modulated continuous-wave
- MIMO: multiple-input multiple-output
- ULA: uniform linear array
- AoA: angle of arrival
- DoA: direction of arrival
- MVDR: minimum variance distortionless response
- Rx-BF: receive beamforming
- Tx-BF: transmit beamforming
- PSD: power spectral density
- PSE: power spectral entropy
- STFT: short-time Fourier transform
- DFT: discrete Fourier transform
- IDFT: inverse discrete Fourier transform
- EMD: empirical mode decomposition
- VMD: variational mode decomposition
- WMC-VMD: weighted multi-channel variational mode decomposition
- IMF: intrinsic mode function
- ADMM: alternating direction method of multipliers
- SSA: singular spectrum analysis

- SVD: singular value decomposition
- SAR: synthetic aperture radar
- ISAR: inverse synthetic aperture radar
- ECG: electrocardiogram
- HR: heart rate
- RR: respiratory rate
- BP: blood pressure
- PTT: pulse transit time
- TCN: temporal convolutional network
- CNN: convolutional neural network
- FCN: fully connected network
- ResNet: residual network
- AAM: additive angular margin (loss)
- PCC: Pearson correlation coefficient
- RMSE: root mean square error
- ME: mean error
- STD: standard deviation
- CDF: cumulative distribution function
- AFOV: angular field of view
- FoV: field of view
- COTS: commercial off-the-shelf
- MWK: Modified Windkessel model
- ODE: ordinary differential equation
- NODE: neural ordinary differential equation

- DP: Doppler profile
- YOLO: You Only Look Once

Chapter 1

Introduction

1.1 Background and Problem Statement

Recent years have witnessed the rapid development of vital sign monitoring. Under the COVID-19 crisis, the demand for vital sign monitoring systems for tracking patients' health conditions becomes more urgent. Traditional vital sign measurements based on PPG and ECG sensors require wearing electrodes or chest bands, which are uncomfortable and inconvenient for some users like infants or burned patients. In this case, the technology of contact-free vital sign monitoring has been indispensable for today's healthcare systems. Further, the personalized healthcare monitoring has been recognized as the trend in the future, the conventional healthcare system always requires to manually registered the user. Hence, there is an urgent requirement that not only the vital signs could be monitored in a contact-free way but also the user identification could be achieved wirelessly.

Currently, millimeter-wave (mmWave) based sensing methods have been widely investigated in contact-free vital sign monitoring and user identification [1, 2]. They are typically developed based on the FMCW technique and can achieve higher range and velocity resolutions. Furthermore, they enable us to locate and distinguish different targets of interest. In this case, it is becoming the mainstream choice in the field of wireless identification and vital sign detection. However, most of the existing works [3–5] ignore the multipath and dynamic interference caused by the environment and moving persons in actual deployments and setup in an ideal environment. In fact, daily life sensing scenarios could be more complicated. For example, for indoor healthcare, the users may stay at a spot such as sofa, desk or table for a long time and perform activities such as computer work, exercise and entertainment, while other family members may walk past them in the room from time to time. In summary, the problems of accurate localization of static person of interest, suppressing the interference of other moving people, and eliminating the body part motion of the user themselves are still waiting to be addressed.

Meanwhile, the vital signs like breath rate or heartbeat investigated in previous research are preliminary, which is hard for further application such as cardiac abnormality and blood pressure (BP) estimation. Even though there have been some works that using mmWave signal to achieve the higher-level vital signs of ECG, pulse and BP, they still face the problem of strict testing conditions, for example, fixed distance between the user and the device. Beside, the mainstream approaches that recover these complicated vital signs are usually based on deep neural networks that directly fitting the mmWave signals to the ground truth waveforms. They lack of transparency and explainability, leading to poor credibility.

With regard to the mmWave based user identification, recent studies have reveal that using the feature of user's gait is capable of identifying the users at a relatively high accuracy. These features serve as the foundation for applying deep neural networks (DNNs) to identify individuals. However, the Doppler-based gait recognition is vulnerable to the disruptions caused by the movements of arms and torso [6, 7]. Particular, when the users are in a limited space or at a close distance, it could be hard to separate the users via Angle of Arrival (AoA) if the antenna number is small. Therefore, the resolution of point clouds obtained from mmWave radar is relatively low due to the restricted performance of AoA and Doppler estimation. Other approaches that use pointclouds (pcls) created by the users are restricted by the the number of antennas, leading to sparse feature representation. Further, The intra-class change results from the variations of persons themselves, which is a common issue for both computer vision based and wireless signals based long-term Re-ID. Since the appearance or gait of a person can change after a period such as several days and generate distinct sensing features, a trained model may fail to identify the current user, thereby degrading the Re-ID accuracy.

1.2 Thesis Objectives and Contributions

The objectives and main contributions of this thesis can be summarized as follows:

1) Our first research work address the challenges of monitoring vital signs (respiration and heartbeat) in the complex environment. we propose a hybrid multi-objective vital sign monitoring scheme, called Vision-assisted radar Vital Sign Monitor (VaR-VSM). This scheme is capable of monitoring vital signs for the location-fixed human target stably in the presence of interferences from both body movements of the target and the movement of other people. In other words, our scheme can achieve reliable vital sign detection for location-fixed target in a multi-people environment, where the target may have body parts movement and people not being monitored can move around. It firstly fuses a camera and COTS mmWave radar to localize the person of interest. Once the positions of targets are identified, the beamforming techniques are simultaneously performed on Tx and Rx antennas to improve the signal-to-noise ratio (SNR) of received echoes from the targets. To suppress the effect of the subject's body motion, we propose weighted-multi-channel Variational Mode Decomposition (WMC-VMD) which is an enhanced version of VMD. Although our scheme is designed and optimized primarily for sensing location-fixed targets, it is also capable of sensing moving targets and can still work in partially blocked transmission (with light obstacles between radar and targets).

2) We address the challenges of monitoring higher-level vital signs (ECG, arterial pulse and BP), and making the process more robust, flexible and explainable. To this end, we propose to leverage the natural frequency band of the cardiac events and combine with the range and angle estimation, which could adaptively localize the position of heart and wrist, overcoming the problem of fixed distance. Then we propose to embed the physics of the cardiac event and arterial pulse into the deep neural networks, which improve the explainability and accuracy.

3) To address the challenges of user identification, we propose the scheme that employs an improved inverse synthetic aperture radar (ISAR) technique for mmWave radars to achieve a high-resolution human imaging and a multi-dimension deep training strategy for long-term Re-ID. As the person moves relative to the radar, their motion enables the construction of an equivalent enlarged synthetic aperture in the horizontal direction. Via emitting a series of pulses and coherently processing the reflected signals over time, the mmWave radar can construct relatively high-resolution images of the person, by exploring the synthetic aperture. Then we propose to integrate the cosine margin as a training strategy, enhancing the network's ability to handle intra-class changes and thus improving the robustness of Re-ID.

1.3 Thesis outline

In chapter 2, we briefly review related work on mmWave based vital sign monitoring and user identification.

Chapter 3 presents the proposed multimodal method that leverages computer vision and mmWave radar to localize the users of interest, suppress the environmental interference and eliminate the body motion noise in vital sign monitoring.

Chapter 4 introduces the proposed sensing scheme for estimating higher-level vital signs: ECG, pulse and blood pressure. It focuses on localization of heart and wrist, then followed by the physics-driven learning architecture.

In Chapter 5, we introduces the proposed user identification method. A novel ISAR method, combined with the compensation, focusing, and imaging is presented first. Then the neural networks based on cosine margin is introduced to improve the long-term user ID.

Chapter 6 summarizes this thesis and discusses potential directions in our future work.

Chapter 2

Related Work

In this chapter, we briefly review some related work that is relevant to the topic.

2.1 mmWave-based Vital Signs Sensing

The principle of mmWave based vital sign monitoring can be summarized following. Firstly, FFT is performed on the demodulated I/Q pulse signal to estimate the range profile. The peaks are always chosen to be the position of the target. Since the chest motion will cause the phase shift of signal especially on the target point. Hence, the phase demodulation is applies to estimate the phase of current frame on the peaks. The next step is to splice these phases together over each frame since mmWave devices transmits pulses over the time. Based on this operation, some works [8–11] have been conducted. In these works, the authors design the Self-Injection-Locked radar system which works in frequency modulation. Their work yields 5 percent of error rate. [12, 13] proposes another types of mmWave radar based on doppler. This system is capable to extract tiny chest motions while it cannot localize the target. Another work [14] modifies the conventional radar devices and introduce double voltage controlled oscillators and flexible phase-locked loop(PLL) which can is switchable to monitor the vital in the condition of vehicles and obtains over 90 percent of accuracy. In spite that these work all makes progress in contactless vital monitoring, their usage is limited. The reason is that the devices used in these works are customized, which means it is not available in daily life. Except building specific system, commercial mmWave system is more popular in today's research work. The mmWave devices can be generally divided into three types based on the types of transmitted signal: continuous wave(CW), ultra bandwidth(UWB) and frequency modulation continuous wave(FMCW). For CW radar, since the motion of a target will cause doppler effect which will be reflected on the frequency shift, CW radar captures this small frequency shift to detect the vital sign. An example work [15] is carried by Yu et al. The method mainly focuses on the phase extraction and optimization to reconstruct the heartbeat and breath waveform. Commercial UWB radar is more reliable compared with CW radar. Because of the wide bandwidth, UWB signal is easier to separate multiple targets and extract the vital sign based on the methods mentioned above. [16] fully takes advantages of UWB radar. The author applies FFT to estimate range heatmap. In their work, if people are located in different positions, there are some light areas on the heatmap corresponding to these tested targets. After that, the author treats the heatmap as a picture and utilizes the methods of image processing and finally count the number of target area as the number of breath. Their system and strategy is capable of measuring respiration with less than 10 error when there are three people in space. FMCW is famous for its higher range and velocity resolution. Some FMCW radars are with MIMO antenna array which enable them estimate the DoA of the targets which performs better compared with UWB radar. The mainstream FMCW radar usually works in frequency of 24GHz, 77GHz and 120GHz. The validity of these FMCW radars has been proved by these works [17–19] and the results based on these hardware platforms all acquire over 90 percent of accuracy.

Some existing works based on COTS mmWave radar have been conducted to tackle the above-mentioned challenges. The work [20] employs different radar layouts to detect breath during sleeping scenarios when body motion happens. Nonetheless, it requires to manually adjust the sensor position according to different gestures. Another work [21] based on the rotation of Rx and Tx antennas can detect posture change whereas the extra mechanical rotator is required. Vital-Radio [22] considers that the limb motion pattern is aperiodic and uses this property to filter the vital sign signal. WiSpiro [23] trains a neural network to recognize the body motion to navigate the radar to move to the front of the target. But it requires a large rail to place the system.

Despite that no work has fully resolved the second challenge, there are still some inspiring works that have demonstrated the potential. Independent component analysis (ICA) based algorithms have been investigated in the works [24, 25], which are capable of separating different vital signs from the targets. However, ICA requires the dimension of observation to be equal to or higher than the number of sources, which is not always available for radars. Beam sensing-based methods are investigated in mmVital [26]. In this work, reflection loss is calculated to localize and direct the antennas to point toward the target. Similarly, Vimo [27] leverages 2-D antenna scanning to localize the object and applies smoothing spline and dynamic programming (DP) to get RR and HR. Its disadvantage is that scanning the beam over the whole space will cost extra time, which will adversely affect real-time performance. Beamforming of MIMO radar has also been found to be effective to enhance the signal from the targets [28–30], but these works do not provide experiments to validate the efficiency in the circumstances of moving interference sources.

Recently, many data-driven deep learning models [31, 32] are proposed to filter out the original heartbeat and breath waveform. RF-SCG [33] applies a convolutional neural network (CNN) to reconstruct accurate seismocardiography and effectively overcome the impact of the dynamic environment. MoVi-Fi [34] and MoRe-Fi [35] respectively exploit multiple layer Multilayer perceptron (MLP) and Variational Encoder-decoder network (VAE) to eliminate the body part motion interference and refine heart and breath waveform. However, these methods always require a large amount of well-labeled training data collected in the RF fields, which is a very time-consuming process. Regarding to the arterial and blood pressure (BP), some works also prove the feasibility of using mmWave radio to detect them from a period of [36–38]. However, the requirement of good mmWave radio signal quality and dedicate hardware cannot be neglected [39]. The work [40] releases the fixed hand requirement by placing a UWB radar towards a sitting subject while this limits the use case and restrict the users.

2.2 mmWave-based Human Identification/Re-identification

Radio-based methods mainly depends on the moving patterns of different human. However, since mmWave has higher range and doppler resolution, it is possible reconstruct more fine-grained moving features such as the signatures of leg and arms. One typical research [41] combines mmWave and RFID to recognize the human ID. In their experiments, RFID labels are attached on the wall in different height. The operation of this research is that when people walk past the RFID labels, their body part in different height level will reflect the signal. For different people, the property of the reflect signal will differ with each other, which can be taken advantages to identify people. In their methods, firstly, FFT is performed over the echo signal and the peak is chosen to be the location of the movement. After that, the wavelet transform are utilized as feature extractor. Finally, DTW is used to classify people. Although this method achieves high performance(over 85 percent of accuracy), there still some limitations. according to their results, it indicates that with the increasing number of RFID tags, the performance will be better. Placing more RFID tags will cost extra space to place and cost extra payment, which might be unpractical in real cases.

Another kind of approaches is based on recognize the point cloud generated by humans [42, 43]. The procedures of this method can be summarized as: Performing range FFT over all samples; Performing 2D FFT over all channels over all range bins to obtain azimuth elevation heatmap; Using CFAR to detect the point on each heatmap and then splice them. In this way, the point could will be obtained when people are moving in space. Typical works are carried out by Peijun et.al. [44, 45]. They propose a LSTM based deep neural network to classify the people. In their work, the system firstly outputs the 3D point cloud to as the potential position of each target. After that, a Kalman filter is applied to track the target. The Hungarian algorithm is leveraged to associate those points to compose the human object in a 3D contour. Finally, the points in each contour are introduced into the DNN for classification. However, this methods can only work in the condition of near field. This is caused the coarse point cloud. With the range increasing, the point cloud be will less and insufficient to represent the human features. Also according to their result, the accuracy will boost when the recording duration goes up, that means this will have adverse impact on real time performance. Another work [46] designs the transformer-based learning method to improve the generalization ability, which could achieve 93 percent accuracy on different scenes. The adversarial strategy is also induced to improve the performance and overcome the sparsity for learning mmWave pointclouds, and it has been proved to be valid and obtain higher accuracy than previous pointnet based method [47].

Another work [48] propose an idea that the daily behavior can be used to identify human ID. This assumption is based on that different people should have different habit. For example, in a family, some people will spend more time to make up themselves while another tend to study in another place. During these process, their track is also different which can be utilized as the essential information to recognize human. In their experiments, they place two devices to capture the signal. Both of them perform the 2D FFT over range and angle axes to respectively estimate the range-azimuth and range-elevation heatmaps. After that, they calculate the track vertically and horizontally. To leverage the information as much as possible, the heatmaps are also be considered since it can represent the reflection property of different individuals. In their algorithm, they propose the deep neural network with two branches. In each branch, the track and heatmap are multiplied and finally they are concatenated together then placd into a fully connected network. Even though their work achieves high accuracy(over 95 percent), it is still limited by the long-term data collection. That means the training data should contain long period(over 24 hour per set). In this case, the training dataset will be too large to train and converge in a common device. Another work [49] also focuses on human ID based on long term information. In this work, the heatmap generated by respectively perform 2D FFT to estimate the two heatmaps as abovementions. The different point with last work is that only 90 frames of heatmaps are used for feature extraction. This rapidly improve the computational efficiency while the representative features are still weak. To extract more refined features for human ID, gait information starts to draw attention. Generally, gait pattern are obtained by micro-doppler signatures which is yielded by performing STFT or wavelet transform to get time-frequency spectrum. Since the gait pattern of people tends to differ with each other, it is possible to use a deep learning based model to learn to classify human targets. [50] leverages the CNN model to recognize gait pattern. Before that, the author analyses the range variance when people walk past the radar since the range variation will adversely affect the gait tracking. To remove the attenuation caused by range, they make the raw signal to subtract the mean value over each frame to enhance the sigature when the range is far. After that, they set a threshold to compress the 2D signatures into 1D series. This action not only boost the computational efficiency but also remove the redundancy of the data thereby it obtains high accuracy(over 90 percent). However, this work cannot address the issue that when there are two or more persons walking past the device since the working pattern will overlap in micro-doppler signature and it is hard to separate them only depending 2D spectrum. This problem is tackled by [51] where the MIMO is fully leveraged to separate gait pattern. In this work, before extract micro-doppler signatures, the spatial FFT is performed to estimate the AoA of each target, after that, by estimate the range and velocity, it is easy to separate each target in a 3D coordinate(AoA, range and velocity). Finally, an encoder and a decoder based on CNN are utilized for human ID recognition. This work is capable to extract the gait information for at most 4 people. Even in condition of 4 people, the accuracy is still over 95 percent. However, the result of this work is limited by the distance between each subject. From the result, when the distance between people is too closed, the accuracy will drop. Another work [52] is also based on analysing the gait pattern. The essential point is that the author performs 2D FFT in fast time and slow time axes to obtain range-dopper first. Then the author sum up over each frame to compress the 2D heatmap into 1D. Similar with previous work, the author performs spatial FFT to estimate the direction of each person. After removing the noise via classic filter and threshold, a K-means is conducted to separate each target from the point clutter.

Chapter 3

Vision-assisted mmWave Vital Signs monitoring

To estimate the vital signs in a complex practical scenario, which usually contains those interference that could be unpredictable but inevitable, such as moving passengers, body part motions and static objects, the computer vision based modal is introduced and fused with mmWave signals, which could further improve the sensing robustness and accuracy performance.

3.1 Ideal Signal Models in COTS MmWave Radar Systems

The operation of a mmWave radar has been investigated in detail in previous works [30,53]. An FMCW radar periodically transmits a series of linearly-increasing frequency signals, called chirps. The period of transmission is T, called pulse repetition interval (PRI). Each chirp lasts the time duration T_c. During each chirp, the frequency starts from f_c and increases by the bandwidth B. Then the transmitted signal is given by

$$x_T(t) = A_T e^{j2\pi(f_c t + \frac{B}{2T_c}t^2 + \varphi_0)}. \tag{3.1}$$

where A_T is the transmitted amplitude and φ_0 is the initial phase which is a constant. For a static target located at radial range R, after the round-trip delay $\tau = \frac{2R}{c}$ (c is speed of light), the received signal due to the reflected factor β is described as

$$y(t) = \beta A_T e^{j2\pi(f_c(t-\tau) + \frac{B}{2T_c}(t-\tau)^2 + \varphi_0)}. \tag{3.2}$$

After I/Q demodulation and range correlation, we can omit the constant phase value φ_0 and $\frac{B}{2T_c}\tau^2$ (small in practice) and then obtain

$$y(t) = A_R e^{j2\pi(\alpha R t + \frac{2R}{\lambda})}, \tag{3.3}$$

where $\alpha = \frac{2B}{cT_c}$, λ is the wavelength of the transmitted signal, A_k is the amplitude of response of the received signal.

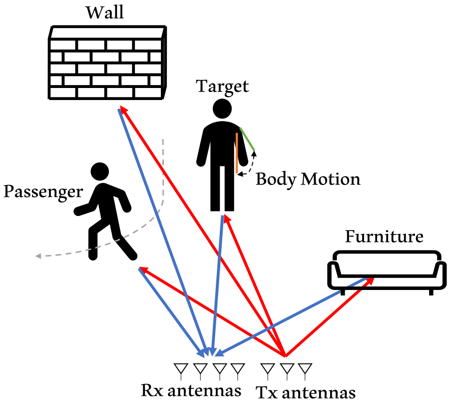


Figure 3.1 : Signal propagation in practical environment. It contains the static objects (wall and furniture), other passenger(s) and the user.

In the case of MIMO mmWave radar with a uniform linear array (ULA), the interval between neighboring antennas is d. If the target is located at the angle of θ with respect to the radar, the signal received by the k^{th} antenna is expressed as:

$$y_k(t) = A_k e^{j2\pi(\alpha Rt + \frac{2R}{\lambda} + \frac{kdsin\theta}{\lambda})}. \tag{3.4}$$

3.2 Our Model for Vital Sign Monitoring in Dynamic Scenario

As shown in Fig. 3.1, the above-mentioned model cannot work well in practice due to the static multipath, the body motion of the target and the dynamic environment (when another person passes by). In this case, we rebuild the signal model by adding the following three interference factors: static objects, target of monitoring and other moving people.

1) Static Objects. Assuming that there are L static objects in space and the l-th is located at the range R_l and angle θ_l. The total received signal received by the k^{th} antenna is expressed as:

$$y_{st}(t,k) = \sum_{l=0}^{L-1} A_{l,k} e^{j2\pi(\alpha R_l t + \frac{2R_l}{\lambda} + \frac{kdsin\theta}{\lambda})}, \tag{3.5}$$

where $A_{l,k}$ is the amplitude of the signal reflected by the l^{th} object and received by the k^{th} antenna.

2) Monitored Target. The signal model that includes the target is more complicated.

Since the chest movement and body motion will cause small range variation, even though the target is at the same position, the signal propagation distance R will vary over time. Therefore, we re-write the range as

$$R(t) = R_{tar}(t) + \triangle R_{tar}(t), \tag{3.6}$$

where $\triangle R_{tar}(t)$ represents the small time-varying shift caused by the body movement. Here $R_{tar}(t)$ denotes the nominal range at t. For a monitored target who stays in the same spot, the nominal range is constant. Thus, the received signal of the target person at the k^{th} antenna can be represented as:

$$y_{tar}(t, k) = A_{tar,k} e^{j2\pi(\alpha R_{tar} t + \frac{2R_{tar}+2\triangle R_{tar}(t)}{\lambda} + \frac{kdsin\theta_{tar}}{\lambda})}. \tag{3.7}$$

3) Other Moving People. For other people walking in space, the signal models not only contain the body motion $\triangle R_p(t)$, with a similar expression to that for the target, but also includes the changes on nominal range $R_p(t)$, angle $\theta_p(t)$ and reflected factor. Assuming that there are P people (in addition to the target person) moving in space, the received signal is expressed by:

$$y_m(t, k) = \sum_{p=0}^{P-1} A_{p,k}(t) e^{j2\pi(\alpha R_p(t) t + \frac{2R_p(t)+2\triangle R_p(t)}{\lambda} + \frac{kdsin\theta_p(t)}{\lambda})}. \tag{3.8}$$

4) Signal Model after Sampling. The signal is sampled and discretized by an Analog/Digital Converter(ADC) after I/Q demodulation. Hence, let

$$t = nT_f + mT_c \tag{3.9}$$

be the time of the n^{th} sample in the m^{th} chirp and T_f denotes the sample interval during a chirp. Then the signal can be represented by the fast time n, slow time m and antenna k in three dimensions. After simplification, the signal in the dynamic scenario is finally expressed by:

$$y_{dy}(n, m, k) = y_{st}(n, m, k) + y_{tar}(n, m, k) + y_m(n, m, k), \tag{3.10}$$

where

$$\begin{cases} y_{st}(n,m,k) = \sum_{l=0}^{L-1} A_{l,k} e^{j2\pi(\alpha R_l n + \frac{2R_l}{\lambda} + \frac{kdsin\theta_l}{\lambda})}, \\ y_{tar}(n,m,k) \\ = A_{tar,k} e^{j2\pi(\alpha R_{tar} n + \frac{2R_{tar}+2\triangle R_{tar}(m)}{\lambda} + \frac{kdsin\theta_{tar}}{\lambda})}, \\ \\ y_m(n,m,k) \\ = \sum_{p=0}^{P-1} A_{p,k}(m) e^{j2\pi(\alpha R_p(m) n + \frac{2R_p(m)+2\triangle R_p(m)}{\lambda} + \frac{kdsin\theta_p(m)}{\lambda})}. \end{cases} \tag{3.11}$$

Our model reveals that in order to obtain vital signs, we need to first estimate the target's location (range R_{tar} and angle θ_{tar}) of the target and then extract the phase ($\triangle R_{tar}$). Specifically, moving people and static objects are major hurdles for locating the target. Their signal amplitudes could be larger than the target's and be time-varying (moving people). Also, the moving people will cause noise superimposed on the phase of the target when they walk close to the target.

Furthermore, $\triangle R_{tar}$ contains four major components: (a) heartbeat, (b) respiration, (c) vibration caused by body part movement, (d) noise caused by moving people nearby. It is noticed that the four components of the received signal could overlap in the frequency domain, which reveals that a classical filter cannot separate vital signs from other components. In the following, we aim to propose a novel scheme to achieve accurate localization and adaptive refinement of vital signs.

3.3 Overview

This work presents a fusion system *VaR-VSM* that enables dynamic vital sign monitoring using a camera and a MIMO mmWave radar. Our scheme is developed by referring to the COTS mmWave device TI AWR1843. It contains 3 transmitting antennas and 4 receiving antennas. A camera is mounted closely with the radar, which works with the radar synchronously. Fig. 3.2 illustrates the workflow of our system. The function of each major processing module is summarized below and will be discussed in detail later.

1) Camera and mmWave Radar Fusion Localization. The COTS mmWave radar

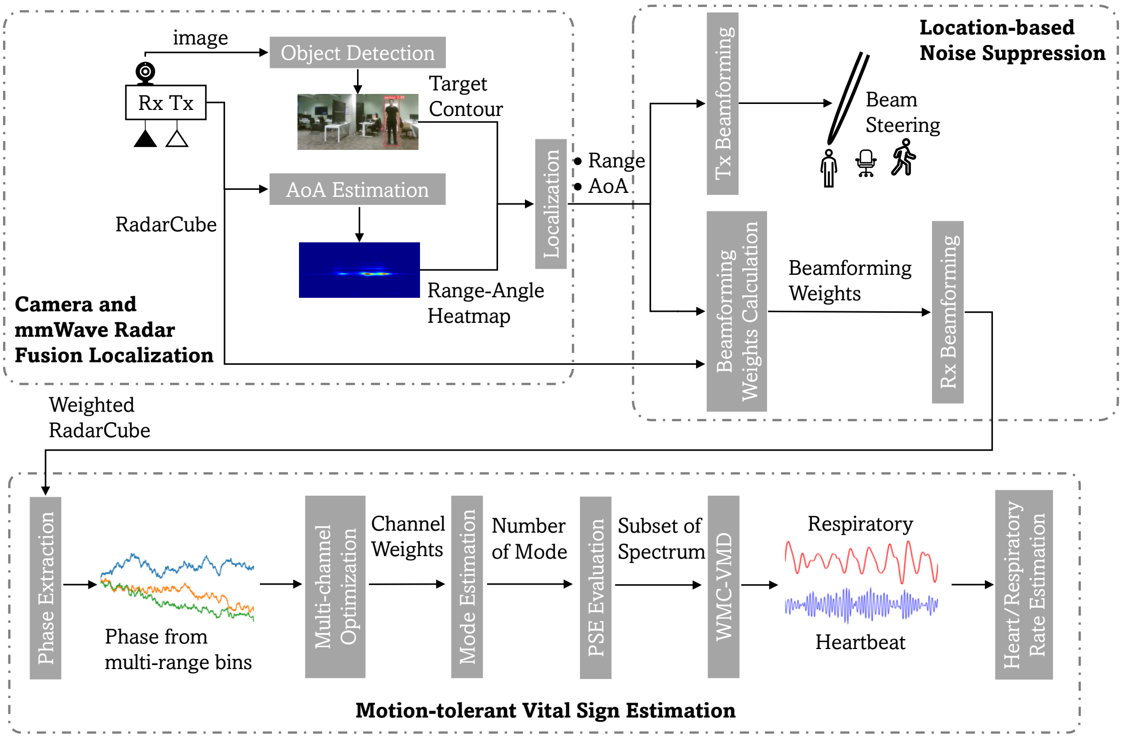


Figure 3.2 : System workflow

performs the 1D range FFT on the raw RadarCube in the fast time dimension. This step is directly completed in the radar hardware layer. We use the 1D FFT data output for AoA estimation and obtain the range-angle heatmap. Meanwhile, the camera captures the image stream and we perform a deep-learning based object detection framework on the images. Thus, we fuse the range-angle heatmap and contours of detected persons for accurate localization. The details will be described in Section 3.4.

2) Location-based Noise Suppression. Once the position of each target is identified, beamforming techniques are adopted to improve the SNR of the signal reflected by the target and mitigate the interference from other people. This consists of Tx-BF (transmitting beamforming) and Rx-BF (receiving beamforming). *Tx-BF* is implemented by shifting the phase of each transmitter. The main beam formed by Tx-BF can point towards the target. *Rx-BF* determines a set of weights that form a receive beam pointing to target too. It enables the received signal to yield maximum power in this direction. This part will be presented in detail in Section 3.5.

Figure 3.3 : Mmwave radar and camera. The center of the antenna array is aligned with the camera.

3) Motion-tolerant Vital Sign Estimation. To reduce the impact of motion interference, we propose a vital signal reconstruction method by optimizing the variance of the signal phase over successive range bins. Further, we propose the approaches of signal mode estimation and PSE (Power Spectral Entropy) evaluation to automatically measure the mode number of the signal and accelerate the computation. Based on this, heart rate and respiration rate are calculated by performing FFT over the reconstructed vital sign waveform. The details will be presented in Section 3.6.

3.4 Camera and mmWave Radar Fusion Localization

This section describes a method to pinpoint the subjects of interest by fusing the target contour from images and the range-angle heatmap from the mmWave data. It focuses on using computer vision framework and range-angle heatmap to obtain the angle and distance of the users.

3.4.1 Target Contour Acquisition from Camera

The image stream is captured by a camera mounted on the radar (shown in Fig. 3.3). The resolution of the image is 1920 × 1080 pixels. The frame rate is set to 30 FPS. We manually calibrate the orientation and height of the camera lens to align the Field of View (FOV) of the camera with that of the mmWave radar. Once an image is collected, we apply an object detector YOLO V5 [54] to detect all targets of interest, each of which is bounded with a rectangle box on the image. Since the YOLO detector enables to classify as many

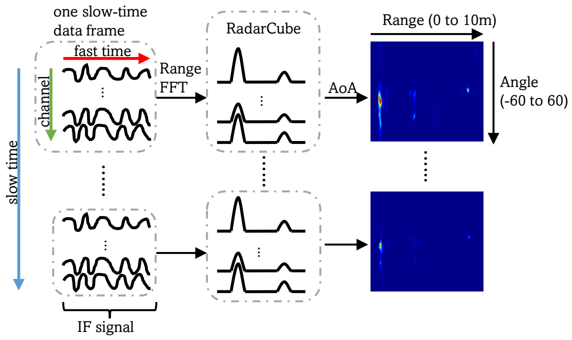


Figure 3.4 : Heatmap generation flow

as 80 classes, there exist many redundant and undesirable bounding boxes containing non-human objects. Hence, those bounding boxes only including the person are automatically selected by reprogramming YOLO. Further, there still exist many similar bounding boxes around each person target. To eliminate the ambiguity of the target of interest, we utilize the non-maximum suppression (NMS) [54] algorithm to remove those redundant boxes and obtain the optimal one.

3.4.2 Range-angle Heatmap from MmWave Radar

As shown in Fig. 3.4, the heatmap is generated by stacking the AoA spectrum over all range bins. Firstly, the radar receives the reflected signal and performs I/Q demodulation. The output of this step is called intermediate-frequency (IF) signal. As described in Section 3.2, the spectrum of range can be estimated by performing FFT on IF signal of all channels over the fast time dimension. In this paper, the RadarCube is defined as the output of this step. The spatial FFT method is commonly used to obtain the AoA spectrum. However, there only exists a small number of receiving elements in the COTS mmWave radar, evenwhen the virtual-array technique is used. For example, the mmWave radar used in our system has a total of 8 virtual receiving antennas in the azimuth direction. The FFT method could not provide a high enough AoA resolution. Although the method of zero-padding is often utilized in FFT, there exists serious spectral leakage (as the zero-padding will induce the high-frequency components and this contributes to the sidelobe besides each frequency component) and is incapable of separating those closely located objects. Instead, we introduce the minimum variance distortionless response (MVDR) [55] algo-

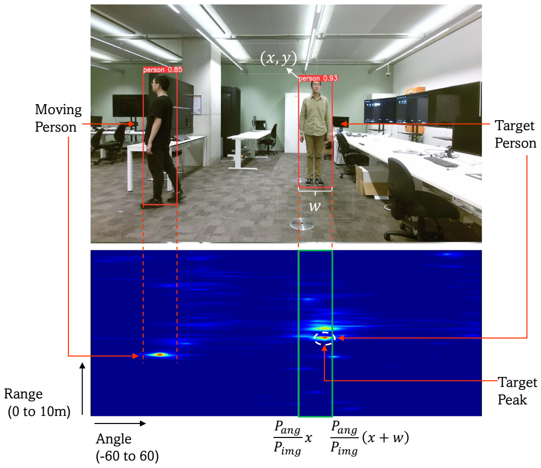


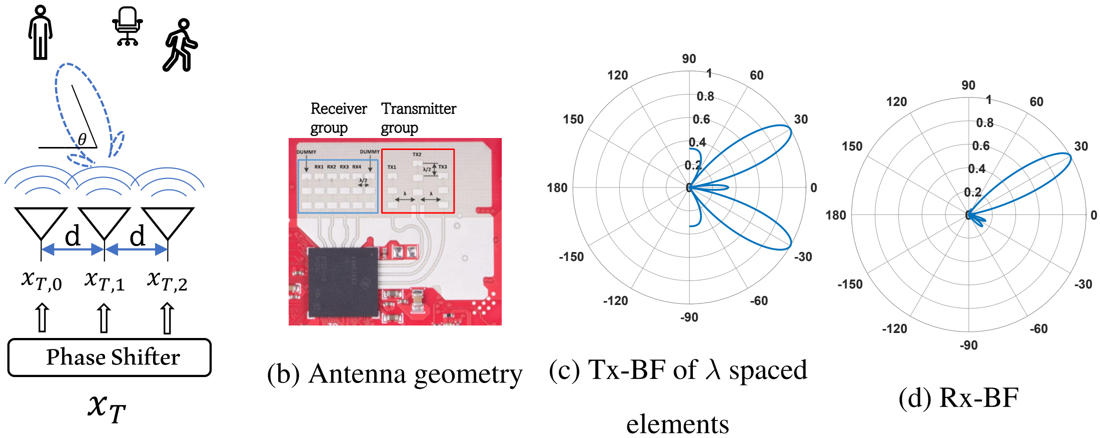
Figure 3.5 : Localization. The output contour of YOLO is combined with the high-power area of the range-angle heatmap to find the user.

rithm to obtain fine-grained AoA spectrum, because MVDR uses covariance matrix to model the noise, thereby it could suppress the power leakage from non-target direction.

3.4.3 Stationary Person Localization

Once the range-angle heatmap is acquired, the 2D position of the monitored target could be estimated by fusing the contour and peaks in the heatmap. At first, the angular field of view (AFOV) of the camera ranges from -60 degrees to +60 degrees, which is set by the manufacturer. The AFOV of the mmWave radar can cover the range of -90 degrees to +90 degrees. However, the main lobe of the radar transmitter used in the experiment only focuses within the range of about ±60 degrees. The beam gain in other directions is much less than the main lobe's, which makes the radar hard to accurately sense the objects. In this case, the searching range of AoA in the MVDR algorithm is set to ±60 degrees, thereby achieving the AFOV matching between the camera and mmWave radar.

In real environments, there may exist multiple persons. We leverage the camera data to filter out the stationary targets of interest for vital sign recognition. Note that only the locations of these targets are required to be unchanged, while their body parts may still have some movement. The output of the image-based YOLO detector includes the coordinate of the start point (x, y), width w and height h of all people contours. When a person is moving, its start point and width may change over time. For the target person



(a) BF Illustration

(b) Antenna geometry

(c) Tx-BF of λ spaced elements

(d) Rx-BF

Figure 3.6 : Beamforming in AWR1843

these values are almost constant. In this case, we can set two thresholds to remove those bounding boxes of moving targets. If the change in the start points (or widths) of the bounding boxes corresponding to a person is larger than the predefined threshold, we consider him/her as the moving target and thereby ignore the bounding box. Then can map the pixels of the stationary persons in the image into the angle bins on the heatmap. The optimal AoA of each target can be obtained by searching the maximum values within the width of the corresponding bounding box. This can be expressed by:

$$R_{tar}, \theta_{tar} = \underset{\substack{R\in[0,10] \\ \theta\in\frac{P_{ang}}{P_{img}}[x,x+w]}}{\text{argmax}} \quad \boldsymbol{H_m}(R, \theta), \tag{3.12}$$

where $\boldsymbol{H_m}$ denotes the range-angle heatmap estimated by MVDR (0 to 10 represents the corresponding range values on the heatmap and θ is obtained by remapping the image pixels to the heatmap). P_{ang} and P_{img} respectively represent the numbers of angle bins of range-angle heatmap and pixels of the image.

3.5 Location-based Noise Suppression

In this section, we exploit the location-based beamforming technique to suppress the impact of range bin overlapping caused by moving persons. They may have the same distance from a static person to the mmWave radar but in different directions. Tx-BF and Rx-BF are combined to handle this issue.

3.5.1 Beamforming at Transmitter

Tx-BF can physically control the transmitter beam to form the desired beam pattern. The main beam lobe can be changed to steer towards the direction of interest by controlling the phase of the transmitting signal on each antenna (shown in Fig. 3.6a).

The COTS mmWave radar used in our system has three Tx antennas spaced by d. As described above, the position of each stationary target to be tracked is calculated. To steer the beam towards the target direction θ_{tar}, we firstly define the phase-shift beamforming weights vector,

$$\boldsymbol{w}_{tx} = [1, e^{j2\pi \frac{d}{\lambda} sin\theta_{tar}}, e^{j2\pi \frac{2d}{\lambda} sin\theta_{tar}}]. \tag{3.13}$$

Then we multiply the original transmitting signal $x_T(t)$ by the beamforming weight $\boldsymbol{w}_{tx}$ to obtain

$$\begin{cases} x_{T,0}(t) = A_T e^{j2\pi(f_c t + \frac{B}{2T_c}t^2 + \varphi_0)} \\ x_{T,1}(t) = A_T e^{j2\pi(f_c t + \frac{B}{2T_c}t^2 + \frac{d}{\lambda} sin\theta_{tar} + \varphi_0)} \\ x_{T,2}(t) = A_T e^{j2\pi(f_c t + \frac{B}{2T_c}t^2 + \frac{2d}{\lambda} sin\theta_{tar} + \varphi_0)} \end{cases}. \tag{3.14}$$

In this case, the radar can transmit the main beam in a specific direction θ_{tar}. It can enhance the signal strength of the stationary person in a certain direction and attenuate the reflected signals of moving persons in other directions.

Unfortunately, if the spacing between two adjacent antennas is larger than half a wavelength, the modified radiation pattern has multiple main beams. Fig. 3.6b illustrates the antenna layout of the radar in our experiment. The distance between each transmitter is λ, the wavelength of the carrier signal. The radar generates two main lobes, shown in Fig. 3.6c. The desired direction is 30 degrees while there is an extra beam pointing towards -30 degrees. In this case, if there is an undesired object located at -30 degrees, the signal will still be interfered since there is no attenuation on the beam response from that angle. In the following, we will use the receiving beamforming in the digital domain to further suppress the interference from other undesired directions.

3.5.2 Beamforming at Receiver

Rx-BF is performed over the receiving antennas. It aims to use a set of weights to maximize the power from the specific direction in digital domain.

To maximize the received signal power for each target, we can use the estimated AoA θ to compute the beamforming weights $\boldsymbol{w}_{rx}$, i.e.,

$$\boldsymbol{w}_{rx} = [1, e^{j2\pi \frac{d_r}{\lambda} sin\theta}, ..., e^{j2\pi \frac{(L-1)d_r}{\lambda} sin\theta}], \tag{3.15}$$

where L and d_r denote the number and distance of receiving antennas, respectively.

Finally, we use the above RX weights to combine all received signals and then obtain

$$\bar{y}(t) = \sum_{l=0}^{L-1} w_{rx}^{*}(l) x_l(t) = \boldsymbol{w}_{rx}^{H} \boldsymbol{x}(t), \tag{3.16}$$

where the symbols $*$ and H denote the conjugate and conjugate transpose operations, respectively. Fig. 3.6d shows the extra lobe is eliminated by Rx-BF and the overall beam concentrates the direction of 30 degree.

3.6 Motion-tolerant Vital Sign Estimation

This section describes the proposed weighted multi-channel variational mode decomposition (WMC-VMD) and uses it to estimate vital signs. The optimization weights calculation, parameter selection and acceleration methods are also introduced in this section.

3.6.1 Vital Sign Estimation by Weighted Multi-channel Variational Mode Decomposition

To estimate vital signs, we need to extract phase from the target's range bin by phase demodulation and compensation [28, 56]. However, there are at least 4 components contained in extracted phase sequence: heartbeat, respiration, body movement and range overlapping. Although the target has been separated by beamforming, there are still two issues.

Firstly, the short-term body motion of the target may cause interference signals with frequencies close to those of the respiration and heartbeat signals. Therefore, traditional filters, such as Butterworth filter, cannot accurately extract vital signs. The VMD algorithm [57] has the potential to address this issue. It aims to estimate the Intrinsic Mode Functions (IMFs) with the assumption that each mode of the signal is around a central frequency with a narrow bandwidth. Then It leverages Alternating Direction Method of Multipliers (ADMM) to concurrently calculate the central frequency and mode function.

However, another issue is that the slight body motion may result in the change of the predefined range bin calculated in Section 3.4. It is hard to predict which range bin the person will be located at in each sampling time. The only prior knowledge is that the nominal position of the target is roughly unchanged. Therefore, it is necessary to fuse the phase sequence from the adjacent range bins of the target. In this case, the VMD becomes ineffective since it only supports the decomposition of the single-dimension signal. The improved MS-VMD [58] is capable of dealing with the high-dimension signals by equally adding up each dimension in the optimization procedure. But this operation cannot pinpoint the signal from the true range bin the target is located in.

In this chapter, we propose weighted multi-channel VMD (WMC-VMD), which is capable of adaptively combining the phase sequence from multiple range bins to estimate the IMFs. In WMC-VMD, we introduce a set of adaptive weights w_l (the notation l denotes the l^{th} range bin) to sum up the phase sequence from different range bins. These weights can eliminate the offsets among different range bins and remap the phase sequence in the same scale. Since the phase sequence from each range bin contains the vital sign, each sequence is optimized individually.

The proposed algorithm WMC-VMD will be detailed in Algorithm 1, consisting of the following steps: 1) applying Hilbert transform in phase sequence to obtain the analytic signal (depicted by the inner sum and modulated by a complex exponential to shift its frequency spectrum to baseband), 2) for each mode u_k, mixing it with a complex exponential of the center frequency to shift its spectrum to the baseband, 3) different with VMD-based methods, the baseband of u_k is estimated by squared l^2-norm of the gradient with the operation of weighted summation on each phase sequence from multi-range bins. Specifically, WMC-VMD is modeled as:

$$\begin{aligned} &\min_{u_k, w_k} \sum_{k=1}^{K} \sum_{l=1}^{L} \left\| \partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t} \right\|_2^2 \\ &s.t. \sum_{k=1}^{K} u_k(t) = w_l s_l(t), \end{aligned} \tag{3.17}$$

where K and L respectively represent the number of modes and channels, w_l denotes the adaptive weight (the calculation will be introduced in Section 3.6.2), and $s_l(t)$ represents the phase sequence of the l^{th} range bin. The augmented Lagrangian function is formed to solve this problem (This is a classic augmented Lagrangian formulation used to solve constrained optimization problems with equality constraints.). With the constraints of weighted multiple linear equations, it becomes:

$$\begin{aligned}\mathscr{L} = &\sum_{k=1}^{K}\sum_{l=1}^{L}\left\|\partial_t[(\delta(t)+\frac{j}{\pi t}) * u_k(t)]e^{-j\omega_k t}\right\|_2^2 \\ &+\sum_{l=1}^{L}\left\|w_l s_l(t)-\sum_{k=1}^{K}u_k(t)\right\|_2^2+\sum_{l=1}^{L}\langle\lambda_l(t), w_l s_l(t)-\sum_{k=1}^{K}u_k(t)\rangle.\end{aligned} \tag{3.18}$$

This task is divided into updating the IMF and center frequency of each mode. Based on Parseval/Plancherel Fourier isometry under the l^2-norm, we can convert this optimization problem from the time domain to frequency domain and perform iterative sub-optimization by ADMM, the final updating results of IMF and center frequency can be obtained as

$$\hat{u}_k^{n+1}(\omega) = \frac{\sum_{l=1}^{L}[w_l S_l(\omega) - \sum_{i\neq k}\hat{u}_i^{n+1}(\omega) + \frac{\hat{\lambda}_l^n(\omega)}{2}]}{L + 2(\omega - {\omega_k}^n)^2}, \tag{3.19}$$

and

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega\left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}. \tag{3.20}$$

3.6.2 Weights Calculation

In the above section, a set of adaptive weights are introduced to fuse the phase sequences from range bins. Specifically, when a target is staying stationary, the range bins around the target all contain the vital signs but with different amplitudes. If there is a small range shift, other range bins will have stronger vital signs. Since this process could occur at any time, the phase sequence from one range bin could be suppressed for a moment while others are on the contrary. Based on this, we propose a method that aims to identify a set of weights $\boldsymbol{w}$ to minimize the variance of signal from multiple channels. By multiplying these weights in WMC-VMD, the variability of phase sequence from all range bins will be normalized to the same scale to improve the decomposition results. To calculate the weights, we define the following optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{w}}{min}\ \boldsymbol{w}^T\boldsymbol{s}\boldsymbol{s}^T\boldsymbol{w} \\ &s.t. \sum_{l=0}^{L-1} w_l = 1, \end{aligned} \tag{3.21}$$

where $\boldsymbol{s} \in R_{L\times N}$ is N phase samples extracted from L continuous range bins.

To solve this constrained optimization problem, we apply the Lagrange multipliers to obtain the weight $\boldsymbol{w}$,

$$\boldsymbol{w} = \frac{(\boldsymbol{s}\boldsymbol{s}^T)^{-1}\boldsymbol{I}}{\boldsymbol{I}^T(\boldsymbol{s}\boldsymbol{s}^T)^{-1}\boldsymbol{I}}, \tag{3.22}$$

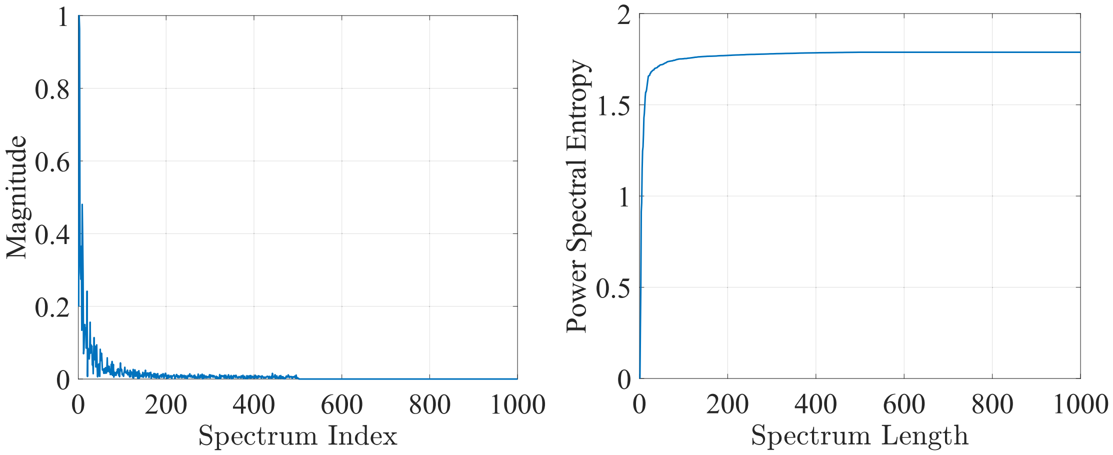
where $\boldsymbol{I}$ is a vector with a size of $1 \times L$ and each element is 1.

3.6.3 Parameter Selection

In the existing VMD-based algorithm, the number of modes K needs to be identified as a prior. The selection of K will significantly impact on the decomposition and optimization. Intuitively, K is supposed to be 4 (heartbeat, respiration, body movement and passenger's range overlapping) based on the empirical assumption for the number of signal sources. However, this cannot work efficiently because in the real environment, when the target moves the body parts, the mode number cannot be guaranteed to remain unchanged. For instance, the target may move the arms together with the legs (even the torso). Since the movements caused by arms and legs are much more salient than heartbeat and breath, K needs to be at least 5 instead of 4. If K were still set to 4, the decomposition results would only contain the IMFs of arms, legs, range overlapping and respiration but miss heartbeat.

In WMC-VMD, we leverage singular spectrum analysis (SSA) to automatically select K. At first, the Hankel matrix is constructed by reforming the weighted phase sequence signal:

$$\boldsymbol{X} = \begin{bmatrix} s_1 & s_2 & s_3 & \cdots & s_n \\ s_2 & s_3 & s_4 & \cdots & s_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_m & s_{m+1} & s_{m+2} & \cdots & s_N \end{bmatrix}, \tag{3.23}$$



(a) Spectrum of extracted phase sequence (b) Visualization of PSE

Figure 3.7 : PSE Analysis

where N is the total number of samples, $n = N - m + 1$ and M is the window length. Next the singular value decomposition (SVD) is applied on $\boldsymbol{H} = \boldsymbol{X}\boldsymbol{X}^T$, and we can obtain

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{3.24}$$

where $\boldsymbol{\Sigma}$ is the singular value matrix which is rectangular diagonal and the elements are sorted in the decreasing order. We choose the mode number K as the number of the largest values having 70+percentile power out of the total singular values.

3.6.4 Computation Acceleration

In WMC-VMD, the optimization is performed in the frequency domain, which requires the Fourier Transform operation over the whole snapshots of samples. From Eqs. 3.18 and 3.19, we can observe that in each iteration, the entire spectral components are used to calculate and update the new IMF and central frequency. Although this process can obtain precise decomposition results, it is time-consuming and storage-costly. Fortunately, the spectrum of the extracted phase sequence is very similar to the tail distribution (shown in Fig. 3.7a), which implies that the main information is gathered in the low-frequency domain. Therefore, to accelerate the optimization and save computation resources, the power spectral entropy (PSE) is introduced to show that the subset of the original spectrum is capable of reconstructing IMF and estimating central frequency effectively. The PSE of a signal is always utilized to measure the information containing in its spectrum based on Shannon entropy [59]. The steps are shown below:

1) Calculate the power spectral density (PSD) of the whole signal and then normalize it:

$$P(\omega_i) = \frac{1}{N}|X(\omega_i)|^2 \tag{3.25}$$

where N is the total number of frequency bins.

2) Normalize the PSD to get the probability density:

$$p_i = \frac{P(\omega_i)}{\sum_i P(\omega_i)} \tag{3.26}$$

3) Calculate the PSE:

$$PSE = -\sum_{i=1}^{N} p_i ln p_i \tag{3.27}$$

Fig. 3.7b illustrates the PSE of extracted phase sequence. When the size of the subspace is chosen to be 1000, corresponding to the entire spectrum, the PSE reaches the highest. On the other hand, the PSE for size 100 is very close to that for the whole spectrum. This reveals that we can choose the first 100 components, which is only $1/10$ of the whole spectral sequence, to replace the entire spectral components used in (18) and (19), thereby the computation time and resources are dramatically reduced.

3.7 Implementation

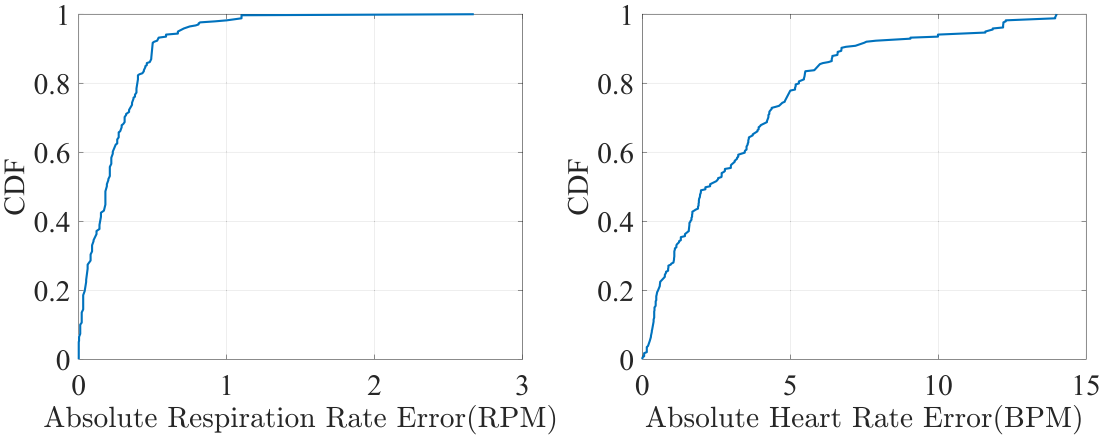
System Setup. The device and experimental setup are shown in Fig. 3.3. The TI 1843 mmWave radar and the webcam are mounted together. The center of Rx and Tx pairs is aligned at the camera lens vertically. Both of them are fixed on a tripod at a height of 1.3 m.

Experiment Environment. The system is deployed in our office shown in Fig. 3.14c, containing multiple chairs, desks, computers, etc. The environment can induce serious multipath interference. Other testing environments include a meeting room and kitchen, shown in Fig. 3.13c.

Evaluation. A target to be tracked sits on a chair or stands. The target is allowed to move the body in a period, which generates motion interference. At the same time,

Algorithm 1 Vital Sign Decomposition based on WMC-VMD

```
$\boldsymbol{w} = \frac{(\boldsymbol{s}\boldsymbol{s}^T)^{-1}\boldsymbol{I}}{\boldsymbol{I}^T(\boldsymbol{s}\boldsymbol{s}^T)^{-1}\boldsymbol{I}}$
$\boldsymbol{s}_c(t) = Hilbert(\boldsymbol{s}(t))$
$\boldsymbol{S}(\omega) = FFT(\boldsymbol{s}_c(t))$
$\hat{\boldsymbol{S}}(\omega) \leftarrow choose\ main\ components\ of\ \boldsymbol{S}(\omega)$
$K \leftarrow Singular\ Spectrum\ Analysis$
Initialize $\{\hat{u}_i^1\}$, $\{\hat{\omega}_i^1\}$, $\hat{\lambda}^1$, $q \leftarrow 0$
while $n \leftarrow n+1$ do
  for $k = 1 : K$ do
    Update $\hat{u}_k$ for all $\omega \geq 0$:
    $\hat{u}_k^{n+1}(\omega) \leftarrow$
    $\frac{\sum_{l=1}^{L}[w_l\hat{S}_l(\omega) - \sum_{i<k}\hat{u}_i^{n+1}(\omega) - \sum_{i>k}\hat{u}_i^{n}(\omega) + \frac{\hat{\lambda}_l^n(\omega)}{2}]}{1+2\alpha(\omega-{\omega_k}^n)^2}$
    Update $\omega_k$:
    $\hat{\omega}_k^{n+1} \leftarrow \frac{\int_0^\infty \omega \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}$
  end for
  for $l = 1 : L$ do
    Dual ascent for all $\omega \geq 0$:
    $\hat{\lambda}_l^{n+1}(\omega) \leftarrow \hat{\lambda}_l^{n}(\omega) + \eta(w_l\hat{S}_l(\omega) - \sum_k \hat{u}_i^{n+1}(\omega))$
  end for
  if Convergence: $\frac{\left\|\hat{u}_k^{n+1} - \hat{u}_k^{n}\right\|_2^{\ 2}}{\left\|\hat{u}_k^{n}\right\|_2^{\ 2}} < \epsilon$ then
    break;
  end if
end while
$\tilde{u}(\omega) \leftarrow zero\ padding\ \hat{u}(\omega)\ to\ original\ length$
$\tilde{u}(t) = IFFT(\tilde{u}(\omega))$
```



(a) Respiration

(b) Heart

Figure 3.8 : Overall Performance

multiple persons walk around the office and induce dynamic interference. The target person is located in a different position in each experiment. Five volunteers are recruited as the target person and moving interference in turn.

Baselines. We compare our scheme with two state-of-the-art works based on SVD+ICA [25] and long-short-term-memory (LSTM) [60]. Specifically, the work [25] applies SVD subspace decomposition to improve the localization and ICA to obtain vital signs. And the work [60] utilizes LSTM-based deep neural network to estimate heart and breath rates in dynamic scenarios.

Ground Truth. The ground truth of the respiration rate and heartbeat rate are captured using a sport watch.

Metric. The error is defined as the absolute value of the difference between the estimated heartbeat rate (HR) and respiration rate (RR) R_e and their corresponding ground truth rate R_g, i.e., $|R_e - R_g|$. The measurement units for breath and heartbeat rate are the respirations per minute (RPM) and beats per minute (BPM), respectively.

Computation Platform. The system is implemented in Python 3.7 and C and runs on a desktop with Intel(R) Core CPU i7-7700 3.6 GHz A4 and 32G memory.

Table 3.1 : Comparison of overall estimation errors

Methods	RR (RPM)	HR (BPM)
SVD + ICA [25]	1.5	3
LSTM [60]	3.22	2.86
This work	0.19	2.29

Table 3.2 : Estimation errors under different distances

Methods	RR (RPM)		HR (BPM)	
	2m	3m	2m	3m
SVD + ICA [25]	1.52	1.55	3.54	3.43
This work	0.22	0.45	1.91	2.49

Table 3.3 : Estimation errors for different motions

Methods	RR (RPM)		HR (BPM)	
	Periodical	Random	Periodical	Random
LSTM [60]	3.56	2.17	7.36	4.62
This work	1.44	1.17	6.61	5.5

3.8 Results

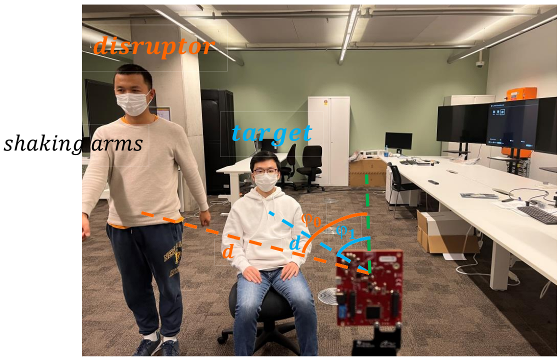
We test our system by conducting a series of experiments under different ranges between the target and radar, body motion types, number of moving disruptors, and different environments (i.e., office, meeting room and kitchen). We use an empirical Cumulative Distribution Function (CDF) of the absolute error to evaluate the overall system performance, shown in Fig.9. We can see that 90 percent of respiration experiments yield less than 0.5 RPM error. And 80 percent of heartbeat experiments can achieve less than 5 BPM errors. The median errors in respiration rate (RR) and heartbeat rate (HR) estimation are 0.19 RPM and 2.29 BPM, respectively.

Next, we compare our work with two state-of-the-art works based on SVD+ICA [25] and LSTM [60]. In fairness to these two works, the comparison is under the same range and body motion type. Table 3.1 indicates the overall error rates of the three works. It can be observed that our work achieves the lowest RR error rate, which is nearly 8 times less than that of SVD+ICA and 17 times less than that of the LSTM based works, respectively. Also, the HR error rate of our work is 25 percent and 31 percent less than those of SVD+ICA and LSTM based works.

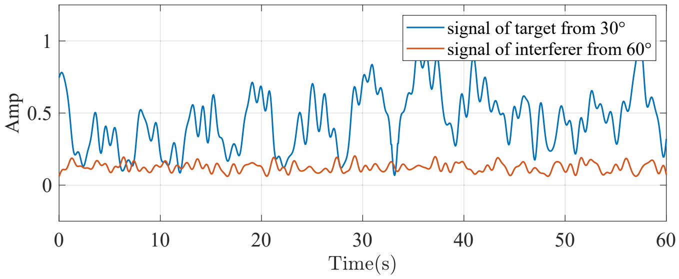
Furthermore, we conduct two fine-grained comparison experiments.

(1) We conduct the comparison at the distance of 2m and 3m between our work and SVD+ICA [25]. By observing Table 3.2, our method achieves over 7 times and 3 times improvement on RR errors. Also in HR comparison at the distance of 2m and 3m, our method outperforms SVD+ICA which reports 3.54 BPM and 3.43 BPM error rates, nearly 1.9 and 1.4 times higher than our methods. It should be pointed out that the maximum range of our work is 7m (it is presented later) and the SVD+ICA fails when the range is over 3.5m. Also, this work fails when the target contains body motion and other moving persons. This indicates that ICA based methods are more fragile when the signal contains interference from different sources due to the limited observation dimensions. In contrast, our proposed WMC-VMD is more effective.

(2) We further compare the performance using the same body motions with those for LSTM [60]. As shown in Table 3.3, our method still achieves nearly half of the error rates



(a) Beamforming Validation Scene



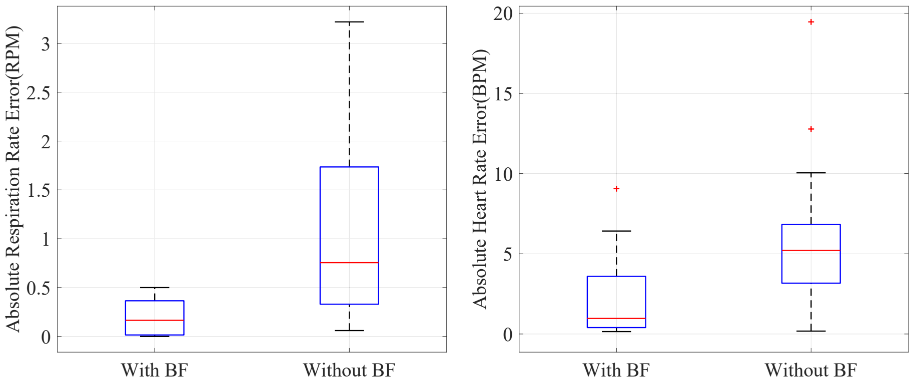
(b) Received Signal after Beamforming

Figure 3.9 : An Example of Beamforming

in RR estimation for both two motion types. In terms of HR, even though our method does not perform better than LSTM for random motion, it still achieves a 10 percent lower error rate than LSTM for periodical movement. The observation illustrates that LSTM can be more effective with random noise suppression since the deep learning based algorithm is more efficient in learning and fitting specific patterns from pseudo random observations. In contrast, our proposed VMD based methods perform better in estimating patterns from periodical signals. It is noted that LSTM requires a large amount of data for prior training while our method does not.

3.8.1 Impact of Beamforming

In order to validate the effect of beamforming, we design an experiment to evaluate whether the signal from the target can be extracted validly under dynamic interference. In this experiment, two people are in front of the radar. In Fig. 3.9a, one is the target person who keeps still, while another person keeps moving near the target and plays the role of



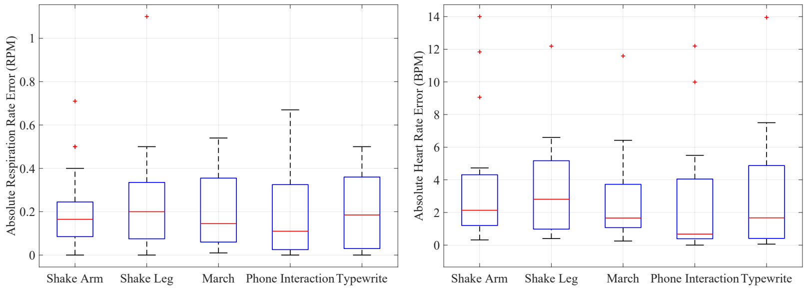
(a) Respiration (b) Heart

Figure 3.10 : Impact of Beamforming

an interferer. Both of them are located at the same distance (2 m) but at different angles. The AoAs of the target and interferer are in the directions of 30 degree and 60 degree from the mmWave radar, respectively. Besides, the target person breathes normally while the disruptor shakes the arms at a high speed to disturb the target.

Fig. 3.9b illustrates the filtered and normalized signal from these two persons using Tx and Rx beamforming. The blue line is the vital sign signal from the static person who is breathing stably. It can be observed that the waveform is periodic and superimposed by heartbeat. Another line shows the received signal collected from the interferer. Its frequency is higher while the signal amplitude is much weaker than the static target. This is contributed by Tx and Rx beamforming, which can amplify the signal from the direction of the target and attenuate the noise from the interferer. In practice, the signal amplitude caused by arm shaking should be much larger than the vital sign signal since the chest motion is very small. The result is in accordance with our expectations shown in Fig. 3.6d.

To further demonstrate the efficiency of beamforming, we compare the performance between enabling and disabling Tx- and Rx beamforming. One person is asked to walk around the target to induce interference. Moving closer to the target person is allowed. As shown in Fig. 3.10, when BF is enabled, both the heart and respiration rates can achieve much lower errors. The RR error is 0.17 RPM with beamforming but it reaches 0.76 RPM without beamforming. For HR estimation, the error without BF is 5.21 BPM, over 5 times



(a) Respiration (b) Heart

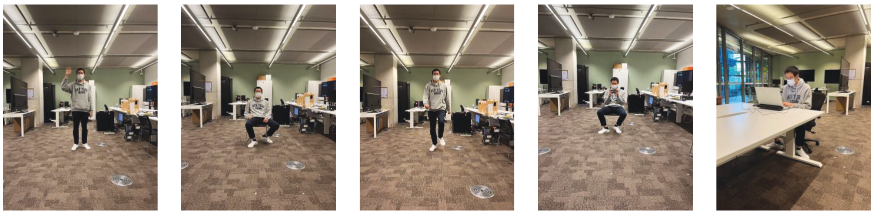


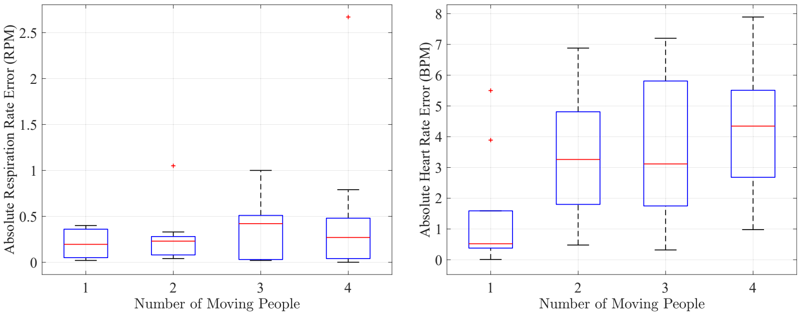
Figure 3.11 : Impact of body motion

higher than the beamforming-enabled mode.

3.8.2 Impact of Body Motion

To investigate the impact of different motion gestures, we ask the tracked person to perform five types of daily motions at the range of 3 m to the mmWave radar: shaking arm, shaking leg, march, phone interaction and typewriting. Specifically, the person waves their arms and shakes leg and this lasts in the first two motion types. Both arms and legs move periodically over the whole sampling period in march. In the gesture of phone interaction, the person uses the cellphone to do their intended interaction, such as typing messages, browsing information and playing games. This motion mainly contains the finger motion. Finally, the person adjusts their body gesture during typewriting. This gesture includes the random motion of fingers, arms and torso.

The result is shown in Fig. 3.11. The trends of respiration and heart rates error are similar. When the target person performs phone interaction, both HR and RR reach the lowest error (0.11 RPM and 0.68 BPM). Shaking leg contributes to the highest error rates (0.2 RPM and 2.81 BPM). The error rates are similar in the rest of the gestures but they can still adversely affect the accuracy.



(a) Respiration (b) Heart

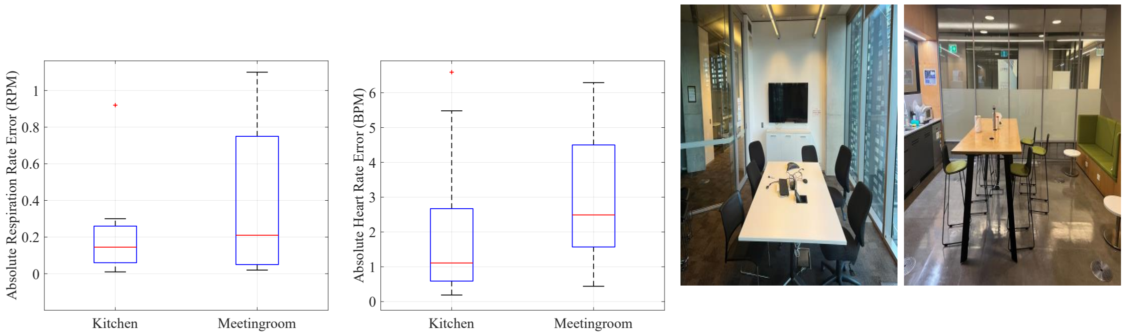
(c) Different Moving People Scenarios. In condition of four moving people, a cart is pushed by a person to induce more interference.

Figure 3.12 : Impact of Moving people

From the results, we have the following observations: **1) Phone interaction.** Since interacting with cellphone only contains the motion of fingers and the extent is much smaller than other types of motions, it has the lowest impact on the estimation accuracy. 2) **Shaking leg.** The frequency of the shaking leg is very close to the heartbeat, which interferes with estimation of heart rate. **3) Typewriting.** Typewriting affects more in respiration rate because there could be arm and torso motion with a close frequency to respiration rate. **4) Shaking arm.** Shaking arm is of the greatest extent. This contributes to both large errors in RR and HR. **5) Marching.** On the contrary, marching is performed over the sample period and the frequency is close to the heartbeat. This leads to a higher error rate in heart rate estimation but has less impact on respiration rate estimation.

3.8.3 Impact of Moving People

To study the impact of dynamic interference, four volunteers are asked to move in space. Fig. 3.12c shows the testing scenarios with one to four moving persons. In each experiment, each person is asked to walk past the target by turns.



(a) Respiration (b) Heart (c) Meeting room and kitchen

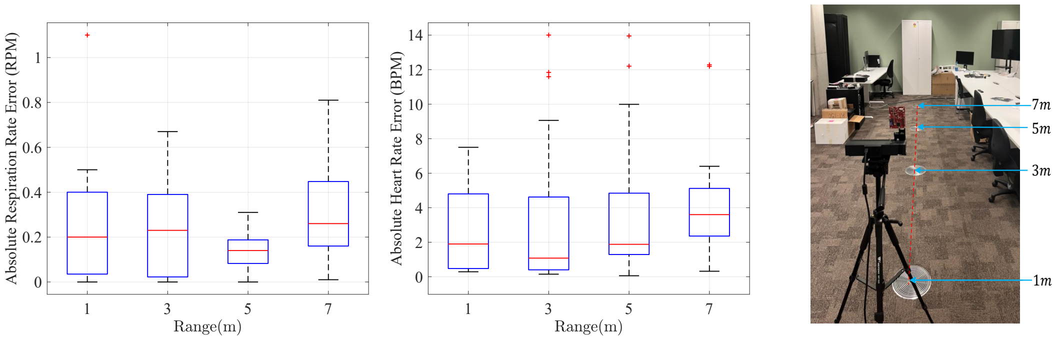
Figure 3.13 : Impact of Environments

Fig. 3.12a and 3.12b show the impact of moving people. The figures demonstrate that the error coarsely increases with increased moving individuals. We can see that the heart rate is easier to be interfered by moving people while respiration rate estimation is less affected. Our system can achieve the lowest error when there is one moving person (0.195 RPM for respiration rate and 0.52 BPM for heart rate). The respiration error reaches the highest (i.e., 0.42 RPM) in the three-people scenario. The highest error of heartbeat estimation is 4.35 BPM in the four-people scenario.

3.8.4 Impact of Environment

We also conducted the experiment in a meeting room and restaurant (shown in Fig. 3.13c) to evaluate the impact of the environment. In both two scenarios, the different individuals are allowed to enter and leave the room at any time. Also, the testing target is asked to randomly perform the intended gestures, such as typewriting in the meeting room and eating or drinking in the kitchen.

As shown in Fig. 3.13a, 3.13b, the accuracy of heart and respiration rates in the kitchen is always higher in comparison to the meeting room. This could be related to the size of these two places. The meeting room is more crowded. It implies that more interfering signals from passengers' movement are superimposed on the vital sign signal due to the short distance during walking.



(a) Respiration (b) Heart (c) Range

Figure 3.14 : Impact of different range. The plates on the ground are placed in 1, 3, 5 and 7m as the signs of different range.

Table 3.4 : Acceleration and Without Acceleration performance

	RR (RPM)	HR (BPM)	Time (ms)
Without Acceleration	0.33	5.02	25.35
With Acceleration	0.29	3.06	8.56

3.8.5 Impact of Range

This experiment verifies the accuracy of the heart and respiration rate estimated at different distances to the mmWave radar, i.e., 1m, 3m, 5m and 7m (shown in Fig. 3.14c). In each experiment, the target person repeats the same gestures and the interfering persons along the same routine. The experiment results are shown in Fig: 3.14a and 3.14b.

For the respiration rate, the accuracy goes down when the range increases. The minimum and maximum errors appear at the range of 1m and 7m, respectively. The error at 7 m is 0.26 RPM which is nearly double as the error (0.14 RPM) at 1 m. The heart rate is more sensitive to the variation of range. The maximum error is 3.6 BPM, which is more than three times the minimum error rate (1.08 BPM) at 3 m. When the range increases, the signal will be intensively attenuated. In this case, the weaker vital sign signal is more sensitive to interferences.

3.8.6 Impact of Acceleration scheme

We also evaluate the performance of our proposed acceleration strategy. Specifically, we choose the first 100 components on the whole spectrum to estimate the RR and HR. And we compare the error and time consumption between the application and nonapplication of the acceleration scheme. As shown in Table 4, when the acceleration scheme is applied, the average computation time is 8.56 ms, which is approximately 1/3 of the one without acceleration (25.25 ms).

It is surprising to see that the error of RR and HR (0.29 RPM and 3.06 BPM) with acceleration are both smaller than those in non-acceleration case (0.33 RPM and 5.02 BPM). The possible reason is as follow. The respiration and heartbeat signals are both low-frequency. When we choose all components in the spectrum, those high-frequency components often contribute as noise and adversely affect the decomposition. When we only choose the low-frequency components, those high-frequency noises are eliminated, which improves the decomposition accuracy.

Chapter 4

Sensing ECG, Pulse and Blood Pressure Simultaneously via Physics-driven Learning

In the last chapter, the vital signs detection of heartbeat and breath under the complex cases are investigated. However, there is a further demand that if it is possible to obtain a more fine-grained and higher-level vital signs, such as ECG, arterial pulse and blood pressure. These signals are usually contains more accurate physiological information that could reflect better healthy information for users. In this chapter, the methodology of ECG, arterial pulse and blood pressure is investigated using mmWave radar and physics-driven learning method, aimed to improve the sensing efficiency and interpretation.

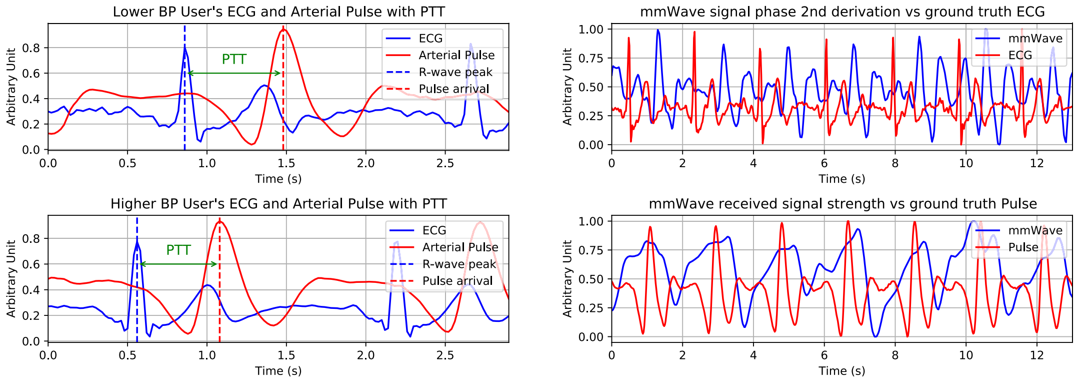
4.1 Preliminaries

This part discusses the feasibility of using mmWave signal to reconstruct the ECG and pulse, and using the relationship of ECG, pulse and blood pressure to estimate blood pressure.

4.1.1 The Relationship of ECG, Pulse and BP measurement

The ECG records the electrical activity of the heart and shows the initiation of each heartbeat. When the ECG generates the heart throbbing, blood pressure increases in the arteries, and a pulse wave propagates from the heart to peripheral sites. It initiates a pressure wave that travels through the arterial system, leading to a measurable pulse at distal sites, typically at the wrist or finger. Pulse Transit Time (PTT) measures the time between the R-wave in the ECG and the arrival of the pulse wave at a peripheral site. It provides insights into arterial stiffness and indirectly correlates with blood pressure, and its effectiveness of BP estimation based on PTT has been proven by [61].

An intuitive example that illustrates the correlation between BP and ECG versus ar-



(a) BP vs ECG vs Pulse. (b) mmWave signal vs ECG vs Pulse.

Figure 4.1 : The relationships of ECG, Pulse, BP and mmWave signals.

terial pulse is given in Fig. 4.1a, where it shows the ECG waves and the arterial pulse waves from two individuals that respectively has different blood pressures. It can be observed that the PTT from the user with higher BP is significantly shorter than lower BP user. Back to the RF sensing, if the ECG and arterial pulse can be sensed by the mmWave signal, it will be anticipated that the BP can be estimated from these RF sensed vital signs.

4.1.2 The Relationship of ECG, Pulse and mmWave Sensing

As abovementioned, the ECG is the bio-electrical activity, which means it is hard to be measured by the contactless sensing approaches. Fortunately, the heartbeat mechanism, caused by the heart throbbing triggered by ECG can be captured from the mmWave signal. Specifically, when the ECG activity happens in one cardiac cycle, the heart will generate mechanical systole and diastole once. On the body surface besides the heart position, it is reflected by a 0.2-0.5*mm* mechanical motion [62]. And this vibration can be reflected by the doppler variation in mmWave signal. Fig. 4.1b plots the ECG and the second derivation (the acceleration) of the doppler from the heart position on body surface. It can be observed that even through the mmWave doppler lags ECG, it is strictly consistent with the cardic periodicity shown in ECG. Further, each peak shown in doppler is also corresponding to each R peak in ECG. And this brings the feasibility to learn a mapping relationship between doppler and ECG.

Different with ECG, the pure electrical signal, the arterial pulse is generated by the pressure caused by the cardiac activity. It is usually measured at the sites of wrist or finger. When the pulse reaches, it will cause the dilation on the blood vessel, and then it shrinks at next time step. This blood vessel variation is a mechanical vibration and can be captured by the mmWave signal. However, the blood vessel is tightly wrapped by the muscles, and its motion is much weaker than cardiac mechanical motions. The phase extracted from mmWave doppler may lose its efficiency while the received signal strength could be more effective [63]. The lower diagram of Fig. 4.1b depicts the correlation between ground truth pulse and the received signal strength at the wrist site extracted from the mmWave signal. It can be easily observed that the mmWave signal strength is strictly corresponding to the pulse. They share same cycle period without any lead or lag. This observation enables to learn pulse from the mmWave signal.

In summary, based on the above observations, it is feasible to use mmWave signal to sense and learn ECG and pulse from the mechanical motions on the heart position on surface and wrist, then using the PTT from the recovered ECG and pulse to estimate the blood pressure.

4.2 Cardiac and Arterial Vibration Sensing

In this section, the scheme of localizing the wrist and chest, which further extract the heart and arterial mechanisms, is introduced. It first states how to model the mmWave signals and then focus on how to separate the location of wrist and chest by leveraging the cardiac periodicity.

4.2.1 mmWave Signal Modeling

An Frequency-Modulated Continuous Wave (FMCW) mmWave radar periodically transmits a series of linearly-increasing frequency signals, called chirps. The frequency starts from f_c and increases by its bandwidth B in a chirp period T_c. The transmitted signal could be formulated as:

$$x_T(t) = A_T e^{j2\pi(f_c t + \frac{B}{2T_c} t^2)}, \tag{4.1}$$

where A_T is the transmitted signal amplitude. For a target located at radial range R, after the round-trip delay $\tau = \frac{2R}{c}$ (c is speed of light), the received signal with strength A_R is described as

$$x_R(t) = A_R e^{j2\pi(f_c(t-\tau)+\frac{B}{2T_c}(t-\tau)^2)}. \tag{4.2}$$

After I/Q demodulation ($x_T x_R^*$) and omitting $\frac{B}{2T_c}\tau^2$ (small in practice) and using wavelength $\lambda = \frac{c}{f_c}$, the output intermediate signal $x(t)$ with strength $A = A_T A_R$ becomes:

$$x(t) = Ae^{j2\pi(\frac{2BR}{cT_c}t+\frac{2R}{\lambda})}. \tag{4.3}$$

For the mmWave radar with a uniform linear array (ULA) with L channels, the interval between neighboring antennas is d. If the target is located at the angle of θ with respect to the radar, the signal received by the l^{th} antenna is expressed as:

$$x(t,l) = Ae^{j2\pi(\frac{2BR}{cT_c}t+\frac{2R}{\lambda}+\frac{ldsin\theta}{\lambda})}. \tag{4.4}$$

As mentioned in Section 4.2, the cardiac and arterial activity will cause the vibration at body surface besides heart and wrist, so the range will have a slight offset leading to the doppler variation $R = R + \phi(t)$. Further, since the mmWave radar transmits the chirps periodically, the index of chirp m and the chirp period (this process can also called slow-time sampling) are introduced and leading to $R = R + \phi(mT_c)$. As the chirp duration time is small, the range offset in one chirp can be neglected and is mainly reflected on the slow-time axis. In our test scenario, the user sits in front of the radar with an arm stretched (shown in Fig. 4.7). Thus, there will be multiple reflected points in the field of view (FoV). Therefore, the signal in such scenario can be described by the combination of these P body surface points:

$$x(t,m,l) = \sum_{p=0}^{P-1} A_p e^{j2\pi(\frac{2BR_p}{cT_c}t+\frac{2R_p+2\phi_l(mT_c)}{\lambda}+\frac{ldsin\theta_p}{\lambda})}. \tag{4.5}$$

After sampling over the fast-time (t) with the sampling interval T_s, the signal is formulated by:

$$x(n,m,l) = \sum_{p=0}^{P-1} A_p e^{j\frac{4\pi BR_p}{cT_c}nT_s} e^{j\frac{4\pi R_p+4\pi\phi_p(mT_c)}{\lambda}} e^{j\frac{4\pi ldsin\theta_p}{\lambda}}. \tag{4.6}$$

It should be mentioned that the sampling period in one chirp is usually closed to chirp duration Tc, leading to $Tc \approx N \cdot T_s$ (N is the number of sampling point in one chirp). It can be observed that in order to obtain the doppler $\phi_{heart}(m)$ and the signal strength $A_{wrist}(m)$ caused by cardiac and wrist arterial vessel, we need to first estimate the locations (range and angle) of heart and wrist, respectively.

4.2.2 Cardiac Doppler Scan based Heart and Wrist Identification

To estimate the ranges and angles, the 2-D Fourier Transform (FT) is usually conducted over fast-time domain (t) and channel domain (k) to obtain the range-angle heatmap. However, it is still hard to identify which two parts are the location of heart and wrist as the user's entire body part including whole arm and chest are both contained in mmWave radar's FoV. To address this challenge, we propose the localization scheme based on the native cardiac frequency. First, the Discrete Fourier Transform (DFT) is performed over the fast-time domain to obtain the range spectrum (it is also called Range-FFT) on each chirp and channel:

$$X(k,m,l) = \sum_{n=0}^{N-1} x(n,m,l) e^{-j2\pi \frac{kn}{N}}, \tag{4.7}$$

where N is the number of samples during the chirp duration. As DFT is performed over fast-time, it will not affect slow-time and channel. Then using Eq. 4.6 to expand and reorder it, we can obtain:

$$\begin{aligned} &X(k,m,l) \\ &= \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} A_p e^{j\frac{4\pi B R_p}{cT_c} nT_s} e^{-j2\pi \frac{kn}{N}} e^{j\frac{4\pi R_p + 4\pi \phi_p(mT_c)}{\lambda}} e^{j\frac{4\pi l d sin\theta_p}{\lambda}}, \end{aligned} \tag{4.8}$$

where k is the index of the frequency bin. To change the frequency bin to the range bin, let $k = \frac{2BR}{c}$ (R is the range bin with the unit of *m*), and using $Tc \approx N \cdot T_s$ it becomes:

$$\begin{aligned} &X(\frac{2BR}{c}, m, l) \\ &= \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} A_p e^{j\frac{4\pi B}{c}(R_p \frac{Ts}{Tc} - R\frac{1}{N})n} e^{j\frac{4\pi R_p + 4\pi \phi_p(mT_c)}{\lambda}} e^{j\frac{4\pi l d sin\theta_p}{\lambda}} \\ &= \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} A_p e^{-j2\pi \frac{2B}{c} \frac{R-R_p}{N} n} e^{j\frac{4\pi R_p + 4\pi \phi_p(mT_c)}{\lambda}} e^{j\frac{4\pi l d sin\theta_p}{\lambda}}. \end{aligned} \tag{4.9}$$

Then canceling $\frac{2B}{c}$ on both sides ($\frac{c}{2B}$ is the range resolution) and using the pulse function $\delta(\cdot)$ to approximate the DFT result, the final result of Range-FFT becomes:

$$X(R,m,l) \simeq \sum_{p=0}^{P-1} A_p \delta(R - R_p) e^{j\frac{4\pi R_p + 4\pi \phi_p(mT_c)}{\lambda}} e^{j\frac{4\pi l d sin\theta_p}{\lambda}}. \tag{4.10}$$

For a certain range bin R if there are M chirps during the observation and L channels on the mmWave radar, the signal can be written in matrix $\boldsymbol{X}_R$ with the size $M \times L$.

From the equation, it can be observed that there are P peaks on the range spectrum and the next step is to identify which two peaks are corresponding to heart and wrist position.

There can be no doubt that the vibration of both heart and wrist position is a periodical signal and the frequency is equal to the cardiac frequency which is empirically known from 0.8*Hz* to 2*Hz*. This means the phase $\phi(m)$ on the ranges and angles of heart and wrist must be the periodic signal with the frequency from 0.8*Hz* to 2*Hz*. Based on this point, the core idea is to find out which two range bins contains the strongest periodic signal in this frequency range. First, we construct the Range-cardiac map. Specifically, if the observation period contains M chirps, the cardiac frequency basis is built and it forms the vector:

$$\boldsymbol{S}_f = [1, e^{j2\pi f T_c}, ..., e^{j2\pi f m T_c}, ..., e^{j2\pi f (M-1) T_c}]^T, \tag{4.11}$$

where $f \in [0.8, 2]$ *Hz* is the cardiac frequency index (The cardiac frequency interval is 0.01). Then we use the principle of Multiple Signal Classification (MUSIC) to construct the Range-Cardiac spectrum. The covariance matrix of $\boldsymbol{X}_R$ is constructed over all the channels to form L snapshots and followed by an eigen-decomposition:

$$\boldsymbol{X}_R \boldsymbol{X}_R^H = \boldsymbol{U}_R \boldsymbol{\Lambda} \boldsymbol{U}_R^H, \tag{4.12}$$

Where $\boldsymbol{\Lambda}$ is the diagonal matrix with the diagonal elements as its eigenvalues sorted in descending order and $\boldsymbol{U}_R$ is the corresponding eigenvector matrix. Given the number of cardiac frequency K (it is set to be 1 in our experiments as the cardiac frequency is generally fixed in a short time), the last $M - K$ columns of the eigenvector matrix is selected to be the subspace $\widetilde{\boldsymbol{U}}_R$. Then the power at range R and cardiac frequency f can be calculated by:

$$P_{range-cardiac}(R, f) = \frac{1}{\left| \boldsymbol{S}_f^H \widetilde{\boldsymbol{U}}_R \widetilde{\boldsymbol{U}}_R^H \boldsymbol{S}_f \right|}. \tag{4.13}$$

By traversing all the range bins and cardiac frequency, the range-cardiac map can be constructed. And the range bins contain heart and wrist can be acquired by searching the maximum two peaks on it.

An example is provided in Fig. 4.2, where it can be found that there are two light areas (with same cardiac frequency) in different ranges (R_0 and R_1) which are corresponding

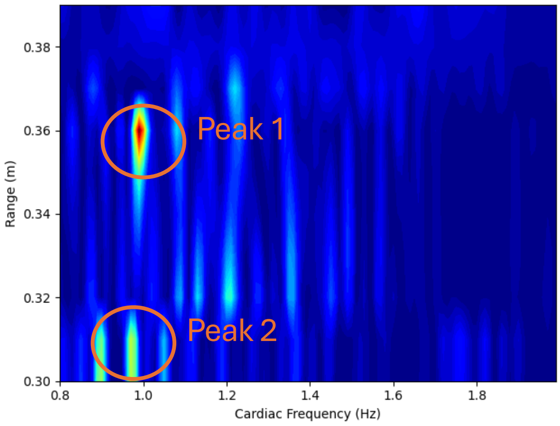


Figure 4.2 : Range-Cardiac Heatmap

to the heart and the wrist. It is noted that it is still unknown which ranges is heart or wrist because the cardiac frequency from both heart and wrist are same.

Once the range bins contain heart and wrist are identified, then we start to estimate the angles. For one of the range bin R_0, we can directly select the corresponding signal matrix $\boldsymbol{X}_{R_0}$ (its dimension is $M \times L$) for angle-cardiac map construction. First, the steering vector is built:

$$\boldsymbol{S}_\theta = [1, e^{\frac{j2\pi dsin\theta}{\lambda}}, ..., e^{\frac{j2\pi ldsin\theta}{\lambda}}, ..., e^{\frac{j2\pi(L-1)dsin\theta}{\lambda}}], \tag{4.14}$$

where $\theta \in [-90, 90]$ degree. Then the 2-D angle-cardiac basis matrix is constructed and flattened by:

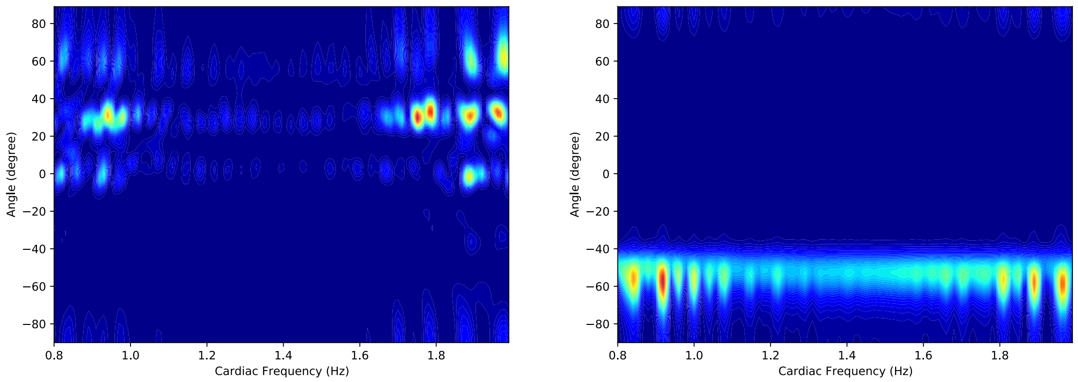
$$\boldsymbol{S}_{f,\theta} = Flatten(\boldsymbol{S}_f \cdot \boldsymbol{S}_\theta). \tag{4.15}$$

Then $\boldsymbol{S}_{f,\theta}$ becomes $1 \times ML$ with each element $e^{j2\pi(fmT_c + \frac{ldsin\theta}{\lambda})}$ ($m = 0, 1, ..., M-1$ and $l = 0, 1, ..., L-1$). Also $\boldsymbol{X}_{R_0}$ is flattened to $1 \times ML$ to match with the dimension of $\boldsymbol{S}_{f,\theta}$ and denoted by $\underline{\boldsymbol{X}}_{R_0}$. Similar with range-cardiac map, the MUSIC is used to calculate the angle-cardiac map by traversing $f \in [0.8, 2]$ *Hz* and $\theta \in [-90, 90]$ degree using:

$$P_{angle-cardiac}(\theta, f) = \frac{1}{\left| \boldsymbol{S}_{f,\theta}^H \underline{\widetilde{\boldsymbol{U}}}_{R_0} \underline{\widetilde{\boldsymbol{U}}}_{R_0}^H \boldsymbol{S}_{f,\theta} \right|}, \tag{4.16}$$

where $\underline{\widetilde{\boldsymbol{U}}}_{R_0}$ is the subspace obtained by selecting the last $ML - K$ columns of the eigenvector matrix from eigen-decomposition of $\underline{\boldsymbol{X}}_{R_0} \underline{\boldsymbol{X}}_{R_0}^H$. It should be mentioned that angle-cardiac calculation is slightly different with range-cardiac map as it requires to traverse

θ and f at the same time for the selected two range bins R_0 and R_1, which means the basis vector is 2D thereby it is 2D MUSIC process. But in range-cardiac map, the range is obtained by Range-FFT in prior, so it only needs to scan f for a certain range bin (its basis vector is 1D) and then use these basis to scan over all range bins.



(a) Angle-Cardiac Heatmap on R_0.　　(b) Angle-Cardiac Heatmap on R_1.

Figure 4.3 : Angle-Cardiac Heatmap.

Fig. 4.3 shows the example angle-cardiac maps in two candidature range bins ($R_0 = 0.31m$ and $R_1 = 0.36m$). It can be found a peak at the same cardiac frequency respectively at θ_0 and θ_1, aligning with the location in the test scenario (Fig. 4.7). It should mentioned that due to the tested arterial pulse is from left arm, so the wrist is always on the left of the heart in the test scenario. Therefore, the two angles of heart and wrist (θ_{heart} and θ_{wrist}) can be directly corresponded to θ_0 and θ_1 by finding out which angle is on the left and on right. Then we can use these two angles to identify the corresponding range bin R_{heart} and R_{wrist} from R_0 and R_1.

4.2.3 Beamforming and Refinement

Once the θ_{heart} and θ_{wrist} have been identified, the beamforming on these angles are conducted separately:

$$\boldsymbol{X}_{(heart,wrist)} = \boldsymbol{w}^H_{(heart,wrist)} \boldsymbol{X}_{R_{(heart,wrist)}}, \tag{4.17}$$

where $\boldsymbol{w}_{(heart,wrist)} = [1, e^{\frac{j2\pi d sin\theta}{\lambda}}, ..., e^{\frac{j2\pi l d sin\theta}{\lambda}}, ..., e^{\frac{j2\pi(L-1)d sin\theta}{\lambda}}]$ with $\theta = \theta_{heart}$ and θ_{wrist}. The second derivation of the phase of $\boldsymbol{X}_{heart}$ can used to represent the the cardiac vibration $\boldsymbol{s}_{heart}$:

$$s_m = \frac{(\phi_{m-3} + \phi_{m+3}) + 2(\phi_{m-2} + \phi_{m+2}) - (\phi_{m-1} + \phi_{m+1}) - 4\phi_m}{16T_c^2}, \tag{4.18}$$

where ϕ_m is the m_{th} phase of $\boldsymbol{X}_{heart}$. Also the arterial pulse vibration $\boldsymbol{s}_{wrist}$ is extracted by the signal strength at the angle θ_{wrist} and range bin R_{wrist} of the wrist, denoted by $|\boldsymbol{X}_{wrist}|$.

4.3 Physics-driven Learning and Inference of ECG, Pulse and BP

Different with purely data-driven methods that directly build the deep neural networks (DNNs) to predict the ECG or arterial pulse, we propose to embed the physical considerations into the DNNs, and use these physics to replace a part of structures in the DNNs. Compared with the pure data-driven methods, like the mainstreamed time series reconstruction of long-short-term memory (LSTM) and time-convolution network (TCN)., embedding physics into the learnable neural networks provides a more straightforward guidance while training, making the neural networks easier to converge and obtain a better performance. Furthermore, the physics-driven method could improve the interpretation, making the strategy less like a 'black box'.

4.3.1 Physical Dynamics of ECG and Pulse

Dynamics of ECG

The synthetic ECG signal is generated by modeling a trajectory in a 3D state space defined by the coordinates (x, y, z). The quasi-periodicity of the ECG is captured as the trajectory revolves around a unit-radius limit cycle in the (x, y) plane. Then it forms the ODE [64]:

$$\begin{aligned}
\frac{dx}{dt} &= \alpha x - \omega y, \\
\frac{dy}{dt} &= \alpha y + \omega x, \\
\frac{dz}{dt} &= -\sum_{i \in \{P,Q,R,S,T\}} a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - (z - z_0),
\end{aligned} \tag{4.19}$$

where $\alpha = 1 - \sqrt{x^2 + y^2}$, $\Delta\theta_i = (\theta - \theta_i)$, $\theta = atan2(y, x)$ and ω is the angular velocity of the trajectory as it moves around the limit cycle.

The rest of parameters are shown in Table. 4.1. By solving this ODE, the solution z is the synthetic ECG signal.

Table 4.1 : Parameters of the ECG Dynamics

Index (i)	P	Q	R	S	T
Time (secs)	-0.2	-0.05	0	0.05	0.3
θ_i (radians)	$-\frac{\pi}{3}$	$-\frac{\pi}{12}$	0	$\frac{\pi}{12}$	$\frac{\pi}{2}$
a_i	1.2	-5.0	30.0	-7.5	0.75
b_i	0.25	0.1	0.1	0.1	0.4

Physics of Arterial Pulse

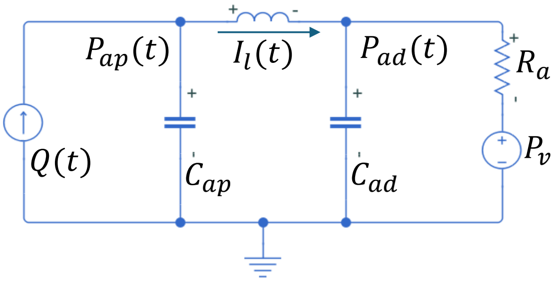


Figure 4.4 : Modified Windkessel model

The physical dynamics of arterial pulse can be approximated to widely-recognized Modified Windkessel model (MWK) (shown in Fig. 4.4) [65].

C_p represents the elastic capacitance of large arteries close to the heart, while C_{ad} represents that of muscular arteries further away from the heart. L represents the inertance of the flowing blood. R_a represents the peripheral resistance and P_v represents a constant venous pressure component.

Given that the model is composed of three energy storage elements, a state space model with three states is sufficient to fully describe the dynamics of the system. Using

Kirchhoff's voltage and current laws, we obtain the following three equations:

$$\begin{aligned} Q(t) &= C_{ap} \cdot \frac{dP_{ap}(t)}{dt} + I_1(t) \\ \frac{dI_1(t)}{dt} &= \frac{P_{ap}(t)}{L} - \frac{P_{ad}(t)}{L} \\ I_1(t) &= C_{ad} \cdot \frac{dP_{ad}(t)}{dt} + \frac{P_{ad}(t) - P_v}{R_a} \end{aligned} \tag{4.20}$$

where $P_{ap}(t)$ is the voltage across C_{ap}, $P_{ad}(t)$ the voltage across C_{ad}, and $I_1(t)$ the current through the inductor, L. 4.2. The input to the system is $Q(t)$. Then the arterial pulse is obtained by solving this ODE and pick up the solution $P_{ad}(t)$. The nominal value of each parameters is shown in Table.

Parameter	**Nominal value**
C_{ad}	0.15 ml/mmHg
C_{ap}	1.45 ml/mmHg
R_a	1.0 mmHg·s/ml
L	0.025 mmHg/(ml·s)

Table 4.2 : Nominal parameter values for the system.

4.3.2 Physics-Informed Neural ODEs of ECG and Pulse

Neural ODEs model continuous-time dynamical systems by parameterizing the system's derivatives with neural networks. The core idea involves representing the evolution of a system as a continuous dynamical process governed by ordinary differential equations (ODEs), where the neural network defines the vector field of the system. Given an initial value problem (IVP), Neural ODEs compute the trajectory of the system at time step t_p by numerically solving the ODE:

$$\mathbf{h}(t_p) = \mathbf{h}(t_0) + \int_{t_0}^{t_p} NN(\mathbf{h}(t), t, \boldsymbol{\theta})\, dt, \tag{4.21}$$

where NN represents the neural network parameterized by $\boldsymbol{\theta}$, and $\mathbf{h}$ represents the hidden states. It is noted that ODEs are usually solved by established ODEs solvers (such as the widely recognized Runge-Kutta methods) in practice.

Another key fact can be observed that the format of Neural ODEs aligns with the typical format of a dynamical system, when the hidden state $\mathbf{h}(t)$ is interpreted as a state vector representation (for example, the $[x, y, z]$ in ECG dynamics and $[P_{ap}(t), I_1(t), P_{ad}(t)]$ in arterial pulse physics). Another notable consideration is that most real-world dynamics problems are forced vibration (For example, the periodical blood flow caused by heartbeat can be treated as the input force to trigger the arterial pulse) rather than only IVPs. Therefore, it is necessary to introduce the $\mathbf{u}(t)$ to describe the ODEs input.

Most importantly, with the intention of incorporating physical knowledge into the architecture of Neural ODEs, the physical dynamics $f_{phy}(\cdot)$ are embedded into the Neural ODEs with the input $\mathbf{u}(t)$ and forms:

$$\frac{d\mathbf{h}(t)}{dt} = NN(\mathbf{h}(t), \mathbf{u}(t), t, \boldsymbol{\theta}) + f_{phy}(\mathbf{h}(t), \mathbf{u}(t), t). \tag{4.22}$$

By embedding physics-informed term, it can be observed that the neural networks play the role of 'residual' between the physics and practice. Hence, the learning system no longer needs the considerable amount of data to learn the generation from scratch, it will focus more on the 'discrepancy' such as the difference between ubiquitous physics and different individuals cases.

An example explanation is given: In purely data-driven tasks that learns ECG/pulse from the mmWave signals, they usually build a very complicated and deep network model to learn the generation of ECG/pulse from mmWave. The core of this approach is to 'fit' the mmWave signals with the ground truth ECG/pulse. In contrast, the proposed approach is to use the periodical mmWave signal as the input to feed into the physics-informed learning system, the basic reference ECG/pulse is generated by physic dynamics and the neural networks take the states and mmWave signal as inputs and it outputs the new modified derivatives of the states of the real-case. It only needs to learn how to transfer from standard ECG/pulse to the true user's ECG referring to the mmWave signal sampled from the particular user.

Finally, the solutions of the ECG/pulse-physics-informed neural ODEs: $\tilde{z}$ for ECG and $\tilde{P}_{ad}$ are respectively obtained and used for final ECG and pulse reconstruction. First, the Eq. 4.19 and Eq. 4.20 are reformatted by the state space forms. Then the mmWave signals are introduced as the system input (mmWave phase from heart position $\boldsymbol{s}_{heart}(t)$

and signal strength from wrist $\boldsymbol{s}_{wrist}(t)$ obtained in Section 4.2) in these two ODEs. Afterwards, the neural networks with the input of all the states and mmWave signals are incorporated to form the physics-informed Neural ODEs. Last, the solutions are obtained by solving these two physics-informed Neural ODEs:

$$
\begin{aligned}
\begin{bmatrix} \tilde{x}(t_p) \\ \tilde{y}(t_p) \\ \tilde{z}(t_p) \end{bmatrix} = \begin{bmatrix} \tilde{x}(0) \\ \tilde{y}(0) \\ \tilde{z}(0) \end{bmatrix} &+ \int_0^{t_p} \Bigg(NN_1 \left(\begin{bmatrix} \tilde{x}(t) \\ \tilde{y}(t) \\ \tilde{z}(t) \end{bmatrix}, \boldsymbol{s}_{heart}(t) \right) \\
&+ \begin{bmatrix} \alpha & -\omega & 0 \\ \omega & \alpha & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \tilde{x}(t) \\ \tilde{y}(t) \\ \tilde{z}(t) \end{bmatrix} \\
&- \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \left(\sum_{i \in \{P,Q,R,S,T\}} a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2} \right) + \boldsymbol{s}_{heart}(t) \right) \Bigg) dt,
\end{aligned} \tag{4.23}
$$

for ECG task and

$$
\begin{aligned}
\begin{bmatrix} \tilde{P}_{ap}(t_p) \\ \tilde{I}_1(t_p) \\ \tilde{P}_{ad}(t_p) \end{bmatrix} = \begin{bmatrix} \tilde{P}_{ap}(0) \\ \tilde{I}_1(0) \\ \tilde{P}_{ad}(0) \end{bmatrix} &+ \int_0^{t_p} \Bigg(NN_2 \left(\begin{bmatrix} \tilde{P}_{ap}(t) \\ \tilde{I}_1(t) \\ \tilde{P}_{ad}(t) \end{bmatrix}, \boldsymbol{s}_{wrist}(t) \right) \\
&+ \begin{bmatrix} 0 & -\frac{1}{C_{ap}} & 0 \\ \frac{1}{L} & 0 & -\frac{1}{L} \\ 0 & \frac{1}{C_{ad}} & -\frac{1}{R_a C_{ad}} \end{bmatrix} \begin{bmatrix} \tilde{P}_{ap}(t) \\ \tilde{I}_1(t) \\ \tilde{P}_{ad}(t) \end{bmatrix} + \begin{bmatrix} \frac{\boldsymbol{s}_{wrist}(t)}{C_{ap}} \\ 0 \\ \frac{P_v}{R_a C_{ad}} \end{bmatrix} \Bigg) dt,
\end{aligned} \tag{4.24}
$$

for arterial pulse task.

In the above, some parameters needs to be obtained in prior. In ECG task, ω directly determines the heartbeat rate during the observation period and it can be estimated by performing FT on the mmWave phase $\boldsymbol{s}_{heart}(t)$ from heart position and finding the corresponding frequency; the initial values $[\tilde{x}(0), \tilde{y}(0), \tilde{z}(0)]$ are set to be $[0, -1, 0]$ since these three points are corresponding to the start point of the ECG period [64]. To synchronize mmWave phase and ECG dynamics, the mmWave phase will be set to start from the first valley point as this point is corresponding to the start point of ECG period. In arterial pulse task, $[\tilde{P}_{ap}(0), \tilde{I}_1(0), \tilde{P}_{ad}(0)]$ are set to be $[1.25, 0, 1.25]$ by to manually adjustment according to the solution of Eq. 4.20. It should be noticed that in the experiments, the selection of the initial values is not very strict, as the neural network is still capable to 'compensate' the offset of the error caused by initial values so the performance will be less affected. The neural networks in both ECG and arterial pulse tasks, $NN_1(\cdot)$ and $NN_2(\cdot)$ are composed by three fully connected (FC) layers with 4 input and 3 output and activated by the leaky ReLU function.

4.3.3 ECG and Pulse Learning Model and BP Inference

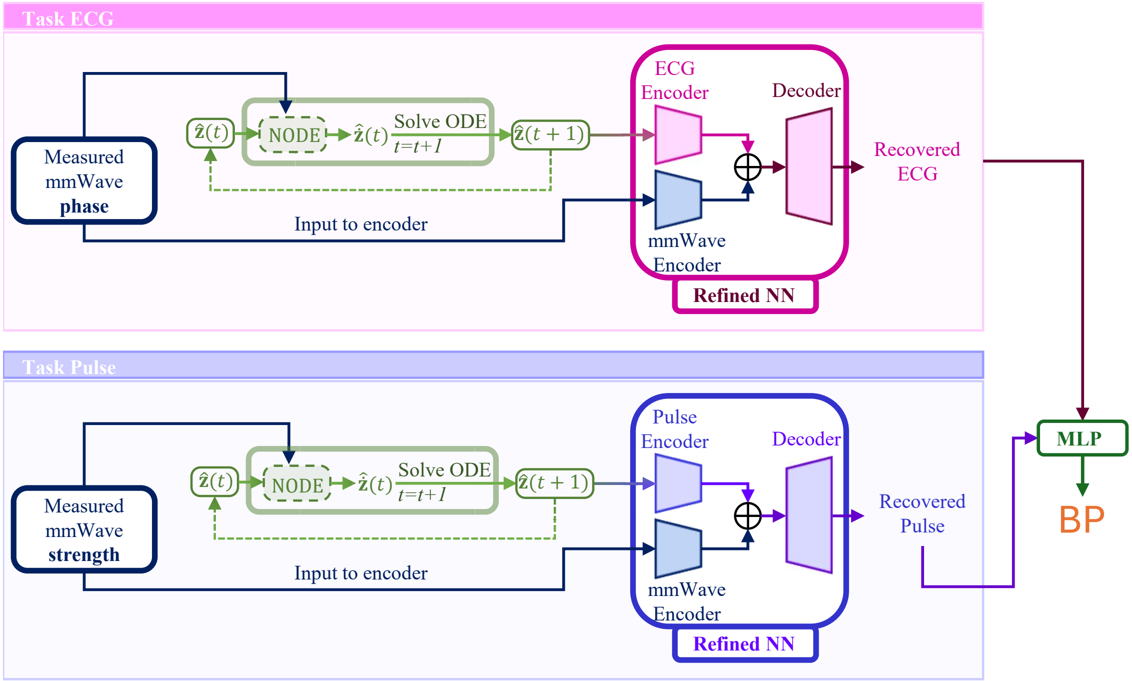
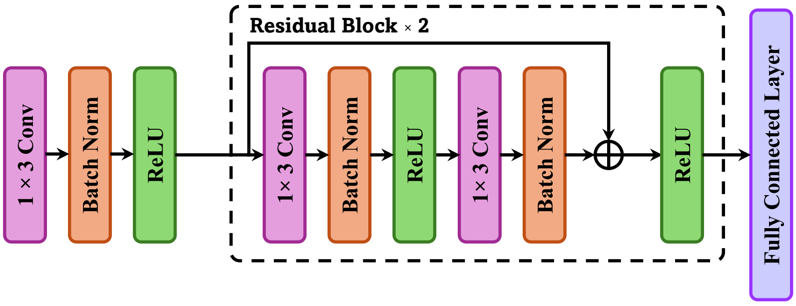


Figure 4.5 : Overview of the learning model.



(a) The encoder/decoder module.

Figure 4.6 : Physics-driven Learning structure of ECG and Pulse recovery and BP Estimation.

The deep learning model is shown in Fig. 4.5. Specifically, the mmWave signals: phase and strength are respectively fed into the physics-informed Neural ODEs of ECG and pulse tasks. Then the solutions are obtained by Runge-Kutta ODE solver. After solving the Neural ODEs of both ECG and pulse tasks, the output states $\tilde{z}$ and $\tilde{P}_{ad}$ are selected as they are corresponding to the ECG and pulse wave respectively.

To refine the prediction and increase and temporal correlation, $\tilde{z}$ and $\tilde{P}_{ad}$ paired with mmWave phase $\boldsymbol{s}_{heart}$ and $\boldsymbol{s}_{wrist}$ are input to the simple encoder-decoder. This encoder-decoder is used as the complementary part of the Neural ODEs. One function of this part is to make the predicted ECG/pulse strictly aligned with the ground truth ECG/pulse cycle, thereby it requires to input the mmWave signals as a reference because the extracted mmWave phase/strength from the user contain the truth cycles. Further even the user is asked to be static, it is still inevitable that sometimes there might be some tiny motion on the tested arm and the amplitude is usually greater than the cardiac and pulse mechanical vibration, but the neural network (encoder-decoder in this chapter) is capable of suppress this type of interference [34]. It is noted that the coarse ECG and pulse have already been learned from the previous physics-informed Neural ODEs, and this encoder-decoder is only targeted on aligning the cycles and suppressing the interference. Thus its architecture is not very deep and complicated (its architecture is shown in Fig. 4.6a), it only contains two 1-D CNN blocks, which is much simpler than the purely data-driven approaches.

Once the reconstructed ECG and pulse are acquired, the PTT are extracted by find the period of pulse peaks and R peaks. Then BP is estimated by building a linear regression model.

4.4 Implementation and Evaluation

4.4.1 Implementation

Unless stated otherwise, the experimental results reported in this section are based on the setup.

Data Collection. 30 volunteers (18 males and 12 females) without the known health condition related to the evaluation are invited for data collection. This research is granted

Figure 4.7 : Experimental Setup

by the institute's ethics approval. The ages of volunteers range from 20 to 54 years. During the test, they are asked to sit with the supported backs and reach out their left arm. The Ti IWR1843BOOST mmWave radar is deployed to collected the mmWave signals. The ground truths of ECG and arterial pulse are collected by NUL-218 electrocardiogram logger sensor and NUL-208 pulse logger sensor that provides Python API and data synchronization USB interface. The blood pressure is captured by MKB0805 blood pressure module that allows to read the data from UART in Python. In total approximately 100,000 cardiac cycles are collected. The sampling duration is set to be 60 secs for one time.

Experiment Setup. The mmWave radar is placed and fixed in front of the user. The field of view (FoV) is adjusted to cover the user's chest and the left arm. The data collection and testing setup are shown in Fig. 4.7. In the experiments, two transmitting antenna 1 and 3, and all the receiving antenna are enabled that forms an uniform linear array (ULA) with 8 elements. The detailed configuration of the mmWave radar is shown in Table. 4.3.

Metrics. The root mean square error (RMSE) and Pearson correlation coefficient (PCC) are used to evaluate the similarity between reconstructed waveforms and the ground truths waveforms of ECG and arterial pulse. The mean error (ME) and the standard deviation of mean error (STD) between the estimated BP and the corresponding ground truth BP are utilized to evaluate the BP estimation.

Implementation Platform. Our scheme is implemented using Python, with PyTorch

Table 4.3 : Parameters of mmWave Setup

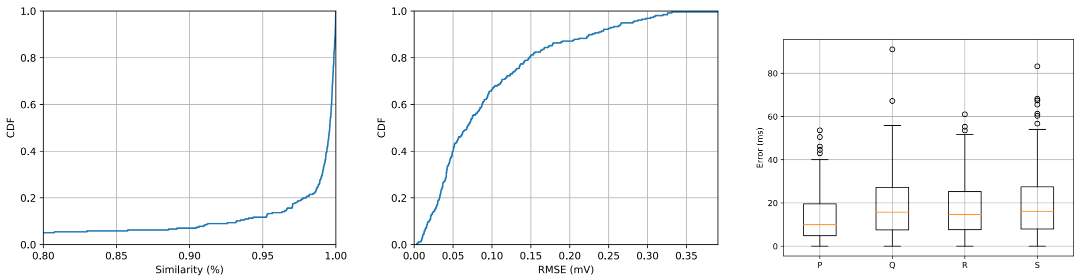
Parameters	Value
Start frequency	77 GHz
Frequency slope	70 MHz/us
Idle time	43.10 us
Sample Number	256
Ramp end time	57us
Sample rate	5.5 MHz
Frame periodicity	20 ms

utilized for learning model construction, training and testing. And we employed the python interface for radar data capture. The system operated on a desktop computer equipped with an Intel(R) Core CPU i7-9700 (3.6 GHz A4) and 32GB of memory. Model training is accelerated using a GPU RTX6000.

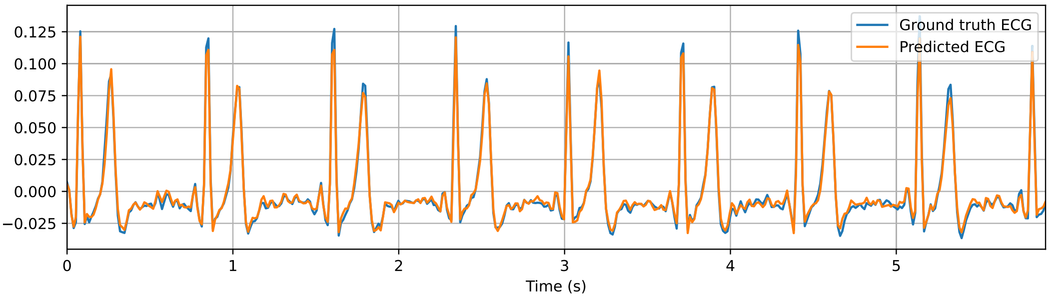
4.4.2 Overall Performance

ECG Recovery

The ECG morphology is evaluate by PCC and RMSE first. Referring to Fig. 4.8d, the example shows that the ECG waveform is well reconstructed and the cycles of the ground truth and the reconstructed waveform are aligned. Also the cardiac activities of Q, R, S, and T waves are also recovered and close to the ground truth. Fig. 4.8a illustrates the morphology similarity performance by cumulative distribution function (CDF). It can be observed that it 80 percent samples achieves over 95 percent of similarity, and the most of test samples achieves over 0.85 PCC. The CDF of the RMSE is also depicted in Fig. 4.8b, where the RMSE of all samples are below 0.35 mV and 90-percentile RMSE is 0.25 mV. Besides, the timing error of the cardiac events is also shown in Fig. 4.8c. The median errors of Q, R, S, and T waves are respectively 16.6 ms, 9.6 ms, 13.9 ms and 15.7 ms. The R wave achieves the best performance as the amplitude of R peak is larger than other three cardiac activities, which makes its feature more remarkable and easier to be learned.



(a) The CDF of overall PCC. (b) The CDF of overall RMSE. (c) The ECG events normalized timing error across 4 physiological status.

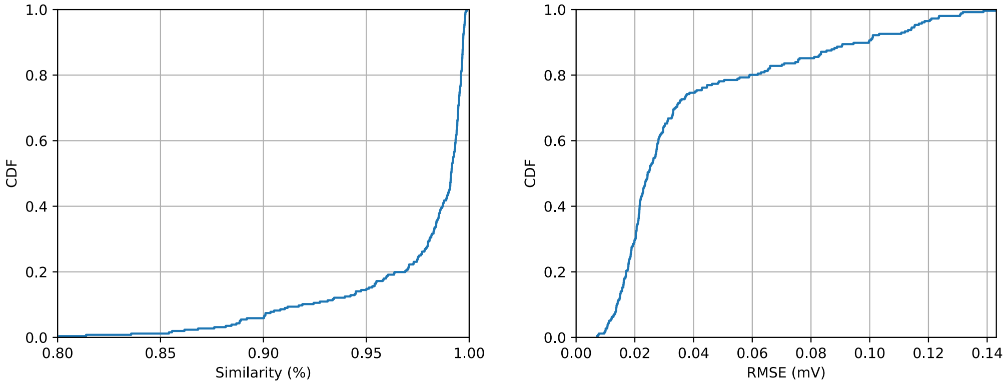


(d) The example of ECG waveform recovery.

Figure 4.8 : The performance of ECG waveform recovery.

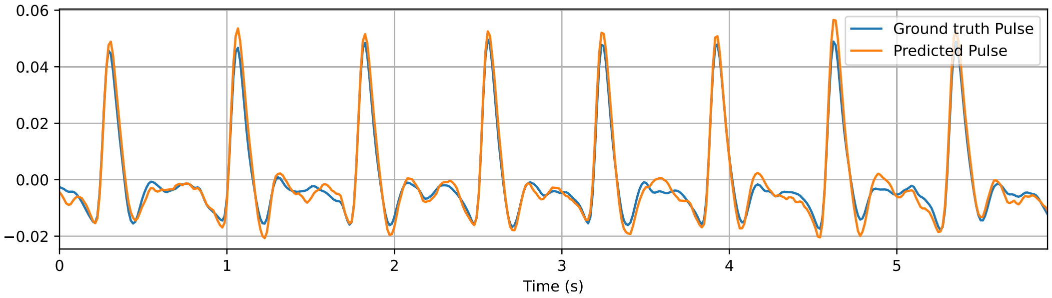
Pulse Recovery

Similar to ECG evaluation, a pulse waveform recovery example is shown in Fig. 4.9c. The result shows that the recovered the pulse waveform could well aligned with the ground truth pulse. Despite of the discrepancy at the low amplitude part, the pulse cycle is almost consistent with the reference, which reveals the effect of the proposed physics-embedded neural networks and the encoder-decoder method. The CDF charts of the PCC and the RMSE are listed in Fig. 4.9a and Fig. 4.9b to evaluate the overall performance of the similarity and error. It can be observed that most of the testing samples achieve about 90 percent of similarity, and the median correlation is 93 percent. Besides, about 80 percent of samples achieves less than 0.06 mV RMSE and the median RMSE reaches 0.03 mV.



(a) The CDF of overall PCC.

(b) The CDF of overall RMSE.

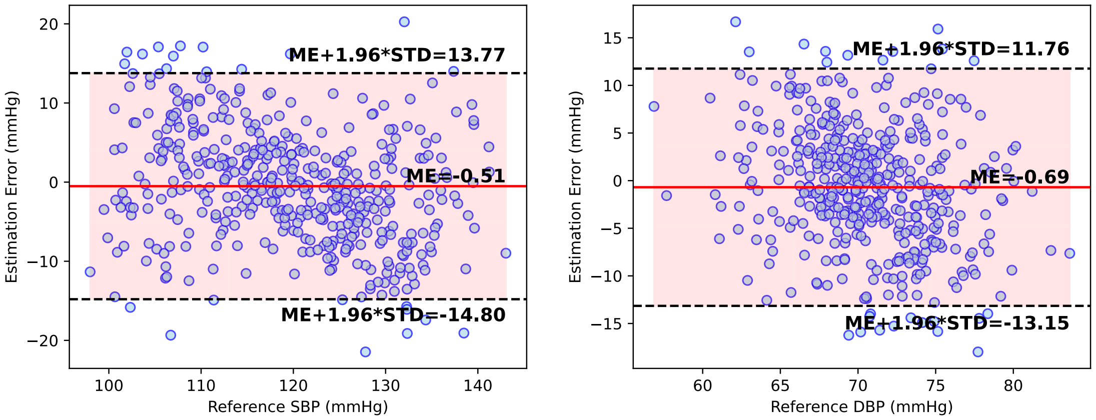


(c) The example of pulse waveform recovery.

Figure 4.9 : The performance of pulse waveform recovery.

BP Estimation

The performance of BP estimation is shown in Fig. 4.10. The Bland-Altman diagram, which charts the estimation error and reference BP on the vertical and horizontal axis, is utilized to visualize the performance. For the DBP estimation, the ME and STD are -0.69 mmHg and 6.71 mmHg. The ME and STD of the SBP are -0.51 and 7.76 mmHg. The proposed method slightly outperforms the minimum requirements of the FDA's AAMI [66] for a medical BP device (i.e., ME $\leq$ 5 mmHg, STD $\leq$ 8 mmHg). In addition, the limits of agreement (LOA) of DBP and SBP are $[-13.15, 11.46]$ mmHg and $[-14.8, 13.77]$, and over 95 percent of data are within the this area. Besides, the median errors of DBP and SBP are 4.59 mmHg and 5.18 mmHg. This overall performance shows an acceptable measurement results, even though there is still a discrepancy between the proposed non-contact sensing scheme and the standard BP measurement.



(a) SBP.

(b) DBP.

Figure 4.10 : Bland-Altman diagram of BP estimation results in SBP and DBP.

4.4.3 Comparison

The comparison with the state-of-the-arts is conducted. It should be noted that many of these works collect data individually, resulting in variations in data collection environments, testers characteristics, and device setups. Additionally, datasets in this domain are rarely open-sourced. Hence, the comparison cannot represent the proposed the scheme is better than other listed methods.

Table 4.4 : Comparison of ECG recovery.

Method	PCC	RMSE
RF-ECG [67]	0.864	0.1
CTL-ECG [62]	0.89	0.081
AirECG [68]	0.95	0.1
Proposed method	0.947	0.16

Table 4.4, 4.5 and 4.6 illustrate the comparison of performance on ECG, pulse wave recovery and BP estimation, respectively. In the task of ECG recovery, the proposed method achieves 0.947 of the median correlation, which is almost consistent with the best one (0.95). Meanwhile, our method achieves the best performance (0.93 PCC) on pulse wave recovery out of the other two SOTA methods. Even though the STD does

Table 4.5 : Comparison of arterial pulse recovery.

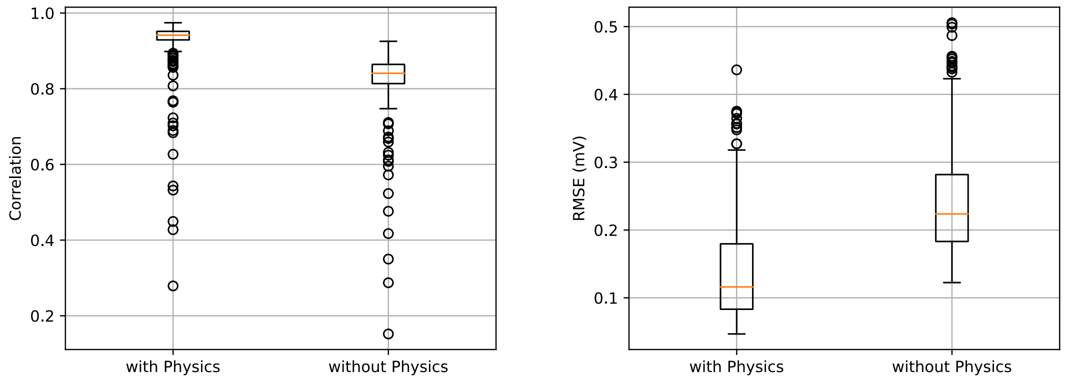
Method	PCC
Geng *et al.* [69]	0.84
Hu *et al.* [70]	0.903
Proposed method	0.93

Table 4.6 : Comparison of BP estimation.

Method	ME: DBP / SBP	STD: DBP / SBP	Correlation: DBP / SBP
Hu *et al.* [70]	0.12 / -1.8	5.47 / 6.82	0.85 / 0.86
hBP-Fi [71]	1.99 / -2.05	6.3 / 6.83	0.8 / 0.8
AirBP [63]	-0.23 / -0.3	3.79 / 4.80	0.94 / 0.94
LSTM-based [63]	0.19 / 0.688	5.33 / 5.07	–
Proposed method	-0.69 / -0.51	7.71 / 7.76	0.82 / 0.84

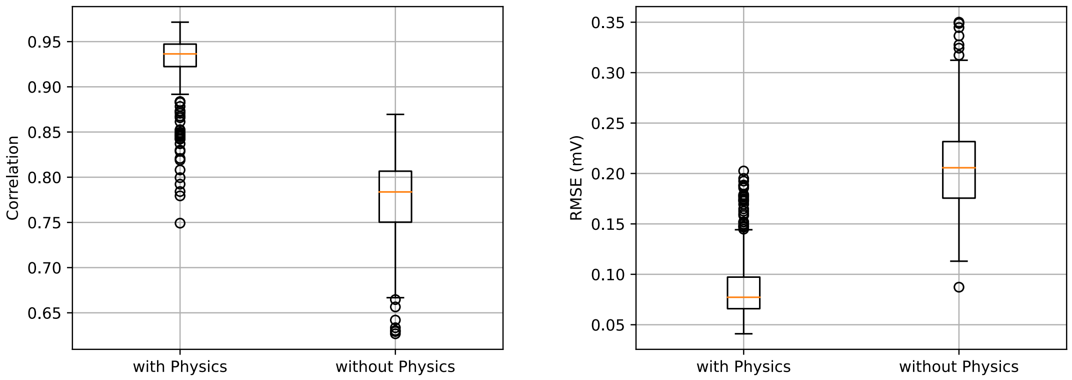
not achieve the best performance in BP estimation, the ME is still comparable with the current best work [63], and the correlation ranks the second over all the compared work.

4.4.4 Effectiveness of Physics Driving



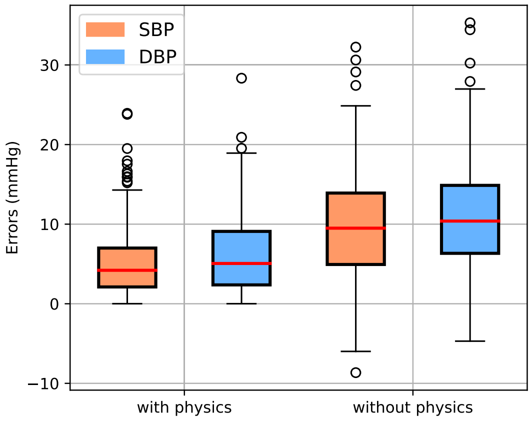
(a) ECG Waveform Similarity.

(b) ECG Waveform RMSE.



(c) Pulse Waveform Similarity.

(d) Pulse Waveform RMSE.

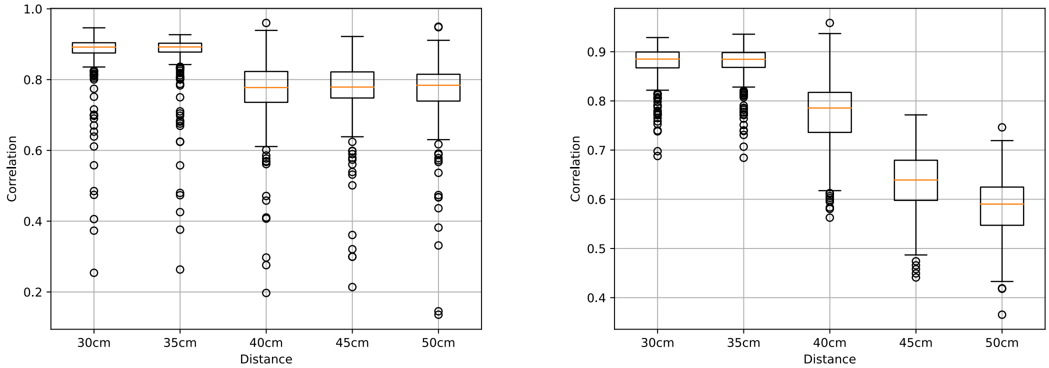


(e) BP Estimation Errors.

Figure 4.11 : Impact of Physics driving.

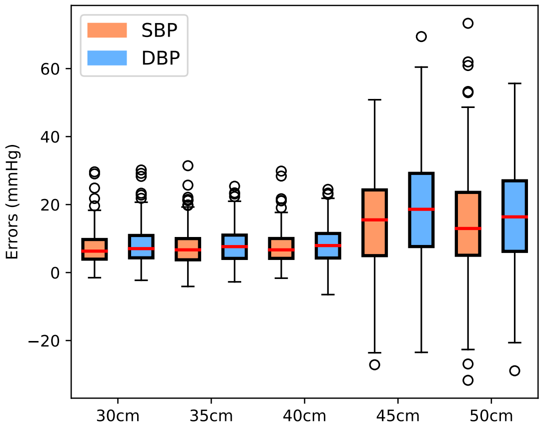
In this section, the ablation is conducted to evaluate the performance of physics driving. We compare the performance of the proposed method and the purely data-driven method that remove the physics Neural ODEs in the learning model. The results are depicted in Fig. 4.11. Regarding of waveform recovery, the similarities of both ECG and pulse with physics are higher than the case without physics. Especially, a significant gap of nearly 15 percent PCC could be observed in pulse waveform recovery, as shown in Fig. 4.11c. Also in RMSE, the cases without physics have larger error than physics-driven cases. Further, the errors on BP estimation is lower with the assistance of physics (shown in Fig. 4.11e), and this could be caused by that the ECG and pulse waveform recovery performs better with physics driving, which contributes to a more accurate localization of key cardiac events (R peaks on ECG and peaks on pulse waveform), thereby achieving a better inference of BP.

4.4.5 Impact of Distance



(a) ECG Waveform Similarity.

(b) Pulse Waveform Similarity.



(c) BP Estimation Errors.

Figure 4.12 : Impact of Distance Variation.

The results in different distance between the testers and the mmWave radar is shown in Fig. 4.12. It can be observed that with the increase of the distance, the performance of ECG, pulse waveform recovery and BP estimation drops differently. As shown in Fig. 4.12a, even the distance increased to 50cm, the PCC is still closed to 0.8. However, the similarity declines more remarkably in pulse, the similarity drops to 0.78 PCC in 40cm and decreases to about 0.6 PCC at the distance of 50cm (shown in Fig. 4.12b). Such a phenomenon could be induced by that the signal from the wrist is usually weaker than the heartbeat from the chest, thereby the signal-to-noise-ratio decrease more than heartbeat. Therefore, the performance pulse recovery is more sensitive to the distance increase than

ECG recovery. Besides, in the BP estimation error, the error also grows when the distance is over 40cm (shown in Fig. 4.12c), which fulfill the observation from the ECG and pulse, as it depends on the key cardiac events in both recovered waveforms.

Chapter 5

mmWave Imaging based User Identification

Despite that the vital signs could be captured in a robust and accurate level, there is still a question which user is currently being monitored. This goal is usually obtained by using camera or other users' unique features like fingerprint or iris, but during these process, the sensitive information always be recorded, which might cause privacy concern. In this chapter, the method of user identification by purely using mmWave modal is studied, and it focus on learning the features extracted by imaging the users' body profile via mmWave radar.

5.1 Adaptive Imaging for Moving Person

This chapter describes the radio imaging part of the proposed scheme. It will first introduce the basic principle from the view of signal model. The challenges under the practical scenarios will be discussed. Following this, the proposed ISAR-based imaging approach will be elaborated.

5.1.1 Signal Model for Human Body Imaging

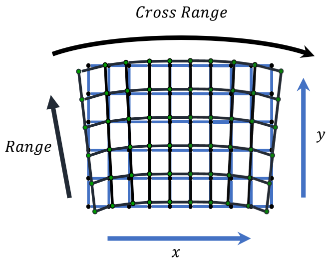


Figure 5.1 : Approximation in non-uniform grid (polar coordinate) to uniform grid (Cartesian coordinate) under the short-time translational motion.

In practical scenarios, an individual's motion trajectory may exhibit variations, potentially following either a curved or a straight-line path within a brief time interval of several

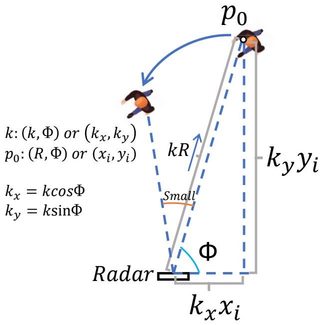


Figure 5.2 : The diagram of signal re-modeling.

seconds. The person motion can be considered as the translational motion from one point to another. In this case, we can place the radar at one side of their moving routines and image each moving person' body profile. We then make the following assumptions: (1) the doppler frequency of each scatter on the person is close, as the body parts tends to have the relatively small motion compared with the while translational movement; (2) the object person contain no body rotation when moving; (3) the coherent processing interval is short in each mmWave frame (this is available in most of the COTS mmWave radar); (4) The observation angle from the view of mmWave radar is limited since the observation period is set to be short, 3-4 seconds in our scheme. Therefore, the non-uniform signal space (radial range and angle) can be approximated by a uniform grid space shown in Fig. 5.1 (An example is provided: when the observation angle θ is from $-20°$ to $20°$ and the radial range r is about $2m$, the trajectory is a curve (the polar coordinate). Using $x = r \cdot \sin(\theta)$ and $y = r \cdot \cos(\theta)$ to calculate the Cartesian coordinate, we can obtain $(-0.52, 1.41)$, $(0.52, 1.41)$ and $(0, 1.5)$ for $-20°$, $20°$ and $0°$. The translated trajectory can be approximated to a line that starts from $(-0.52, 1.455)$ to $(0.52, 1.455)$, and the maximum approximation error is 0.045 m, which is much smaller than the size of a real human body.).

The proposed radio imaging technique is derived from the ISAR algorithm [72]. To elaborate it, we begin by creating a uniform 2D xy spatial plane using Cartesian coordinates to represent the signal, as illustrated in Fig. 5.2. For simplicity, we disregard the vertical axis z (or elevation in polar coordinates) in this initial stage, with an in-depth exploration of imaging along the z-axis planned in this section. The person is located at

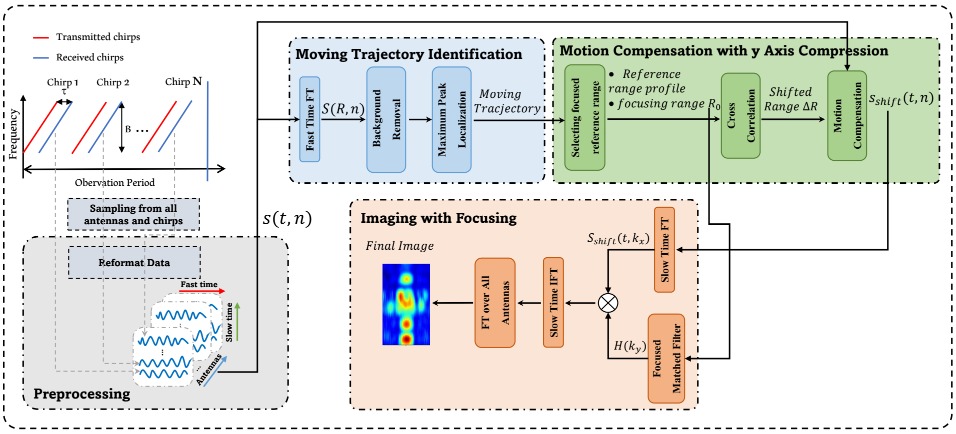


Figure 5.3 : The workflow of human body imaging.

the range of R and the angle of ϕ with respect to a mmWave radar.

Typically, a frequency modulated continuous wave (FMCW) radar transmits a series of pulses periodically. The transmitted signal can be expressed as:

$$s_T(t,n) = A_0 e^{j2\pi(f_c t + \frac{B}{2T}t^2)} \mathrm{rect}(\frac{t-nT}{T}), \tag{5.1}$$

where A_0 is the signal amplitude, f_c is the carrier frequency, B represents the bandwidth and T denotes the chirp duration time, and n denotes the pulse index. In addition, rect(x) is the rectangular function defined as:

$$\mathrm{rect}(x) = \begin{cases} 1, & \text{if } |x| \leq 0.5, \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

The received signal at an antenna from the i^{th} scatterer at location (x_i, y_i) of the person at the range R can be expressed by:

$$s_R(t,n) = p_0 e^{j2\pi(f_c(t-\tau) + \frac{B}{2T}(t-\tau)^2)} \mathrm{rect}(\frac{t-\tau-nT}{T}), \tag{5.3}$$

where p_0 is the received signal amplitude, $\tau = 2R/c$ is the round-trip delay and c is the speed of light.

To extract the target's range and Doppler information, the received signal is mixed with a reference signal from the local oscillator (LO) and the output is passed through a low-pass filter to remove the high-carrier-frequency component. The resulting baseband signal can be expressed as:

$$s(t,n) = p_0 e^{-j2\pi(f_c\tau + \frac{B}{T}\tau t)} \text{rect}(\frac{t-\tau-nT}{T}). \tag{5.4}$$

Since n denotes the pulse index, which corresponds to the sequential order of transmitted pulses over time. This time axis, which usually increments at the step of the pulse repetition interval (PRI), is commonly referred to as 'slow time'. The received signal for each pulse is often stacked to form a 2-D matrix, with t and n representing fast time and slow time, respectively. Since τ is usually much smaller than the pulse duration T, the rectangular function can be neglected. Letting $f = f_c + \frac{B}{T}t$,$t \in [0, T]$, the signal at a slow-time step is given by:

$$s(f) = p_0 e^{-j2\pi f \frac{2R}{c}}. \tag{5.5}$$

Defining the wave number $k = \frac{2\pi f}{c}$, the signal in Eq. 5.5 is rewritten as:

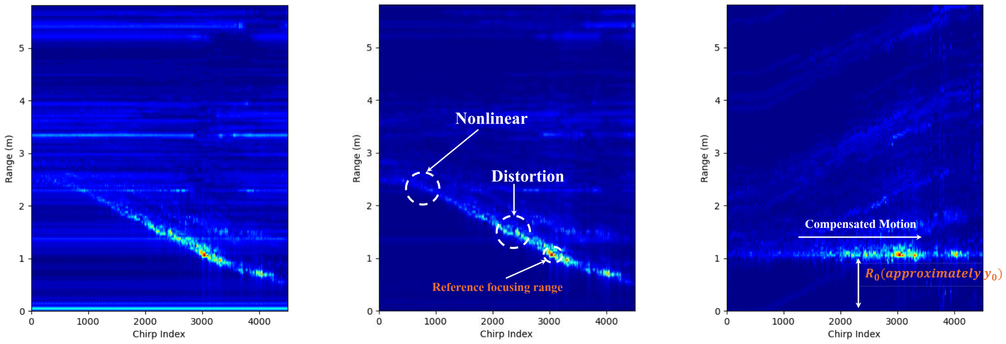
$$s(k) = p_0 e^{-j2kR}. \tag{5.6}$$

Next, we convert the scalar models from 1D (range) to a 2D plane by decomposing the signals into the range (y-axis) and slow-time (x-axis) dimensions. Considering the signal propagation direction ϕ, the wave vector components (k_x, k_y) can be expressed as: $(k_x, k_y) = (k\cos\phi, k\sin\phi)$ (shown in Fig. 5.2). Here, k_x and k_y describe the spatial frequency components of the radar signal in the horizontal and vertical dimensions, respectively. Then the signal is expressed as:

$$s(k_x, k_y) = p_0 e^{-j2(k_x x_i + k_y y_i)}. \tag{5.7}$$

This expression shows that the scatterer at location (x_i, y_i) contributes a phase shift proportional to its position, which is essential for resolving the target's spatial shape. When the observation time is short, and the person is moving at a low speed, the changes in the radial range R and slow-time can be approximated as closely following the variations along the y and x axes. Consequently, $s(k_x, k_y)$ can be effectively approximated within the Range-Doppler domain by applying a 2D Discrete Fourier Transform (DFT) across the fast time and slow time dimensions, as previously discussed in [73]:

$$s(k_x, k_y) \simeq \text{DFT}_{2D}(s(t,n)). \tag{5.8}$$



(a) Range profile with background

(b) Range profile without background

(c) Range profile after compensation

Figure 5.4 : An example of trajectory tracking and motion compensation.

After obtaining $s(k_x, k_y)$, a matched filter $H(k_x, k_y) = e^{j2(k_x x_i + k_y y_i)}$ is applied to focus on the scatter point at (x_i, y_i). This filter compensates for the phase delays introduced by the wave propagation. Finally, the 2-D Inverse Discrete Fourier Transform (IDFT) is performed to recover the final image. The whole image recovery process can be summarized by:

$$\text{Image}(x, y) = \text{IDFT}_{2D}(\text{DFT}_{2D}(s(t, n))H(k_x, k_y)). \tag{5.9}$$

5.1.2 Challenges in Real Scenarios

However, there remain several challenges in practical scenarios for human body imaging:

1) Moving Person Focusing: Precisely focusing on a moving person presents a challenge, primarily due to interference from other objects in the surrounding space. This interference can result in the mis-localization of the person.

2) Complex Motion Patterns: The motion of the person involves nonlinear trajectories, including distortion, deceleration, and acceleration. These complex motion patterns make it difficult to both focus on and accurately identify the person's reflection position. In this case, designing an effective matched filter becomes a challenge as it necessitates acquiring the reflector's position beforehand.

3) Computationally Intensive Image Recovery: The process of image recovery requires traversing all spatial positions in the $x - y$ domain, demanding significant computational resources and time.

Hence, there is a pressing need for a scheme that can effectively address the following key objectives: (1) estimating the person position automatically and accurately (2) compensating the nonlinear motion patterns and (3) reducing the computation cost, preferably, compressing person's movement to one axis (x or y) instead of scanning in both two axes (x and y) is expected to be proposed. And it will be discussed in the next section.

5.1.3 Imaging with Focusing and Nonlinear Motion Compensation

The proposed scheme is depicted in Fig. 5.3. It will be divided into four parts to clearly illustrate the process.

Trajectory Identification. After reformatting the data, we perform a DFT on $s(t, n)$ across the fast time domain, resulting in the range profile. This transformation yields the range-cross range domain, as illustrated in Fig. 5.4a. It is worth noting that in environments containing other stationary objects, these objects will manifest themselves in the range profile at every cross-range index. However, by subtracting the mean value across the slow time, the static objects can be filtered out, thereby highlighting the moving person. The rationale behind it lies in the fact that the phase in the signal contributed by a moving person varies with time, while the phase of static objects remains constant.

Moreover, the Doppler interference also becomes visible after mean subtraction. It is known that the Doppler is caused by the variation of range when the object is moving. The cross-range domain, formed by stacking all range profiles over time, illustrates this variation, revealing Doppler information. An example is presented in Fig. 5.4b. After applying the mean subtraction, the motion variation of the person becomes clearly discernible. Notably, the person's position varies at different distances within the range profile across cross-range, generating Doppler. Furthermore, the trajectory exhibits nonlinear attributes, including distortion in the midsection and deceleration toward the end. It is noteworthy that in cases of constant object speed (indicating linear range variation), Doppler remains constant. Hence, nonlinear factors such as distortion and deceleration contribute to Doppler interference. Motion compensation methods are discussed next.

Motion Compensation. Next we carry out the motion compensation to eliminate the

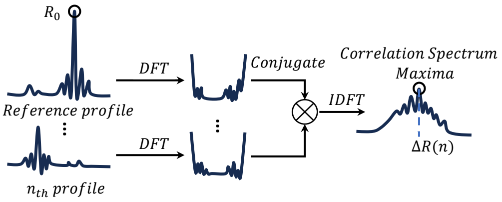


Figure 5.5 : Cross-correlation in frequency domain.

range shift caused by these negative motion factors. In this process, the trajectory can be shifted to a straight line in y axis (approximated to range axis). The reference range profile which contains highest power is recommended to be chosen and the range of the peak is the selected focused range.

The focused reference distance is calculated by performing DFT in the fast-time domain after mean subtraction over all the observation chirps and finding the maxima:

$$R_0 = \underset{n\in[0,N-1],R\in[0,R_m]}{\text{argmax}} |\text{DFT}_t(s(t,n) - \frac{1}{N}\sum_{n=0}^{N-1} s(t,n))|,$$

where R is the index of range bin after DFT in the fast-time domain, and R_m denotes the total number of range bins ($R_m = \frac{(K-1)c}{2B}$ and K is the DFT size.). As depicted in Fig. 5.4b, the range of the person undergoes changes over the slow time due to its movement. Therefore, when a reference range R_0 is chosen, the person's range can be expressed as $R_0 + \triangle R(n)$. Referring back to Eq. 4.6 and disregarding the $rect(\cdot)$ term, we can express the signal at the n_{th} slow time index as:

$$s(t,n) = p_0 e^{-j2\pi\frac{2R_0+\triangle R(n)}{c}(f_c+\frac{B}{T}t)}. \tag{5.10}$$

After reordering the equation, we obtain:

$$s(t,n) = p_0 e^{-j2\pi\frac{2R_0}{c}(f_c+\frac{B}{T}t)} e^{-j2\pi\frac{\triangle R(n)}{c}(f_c+\frac{B}{T}t)}. \tag{5.11}$$

Similar to Eq. 5.5 and Eq. 5.6 in Section 5.1.1, where $f = f_c + \frac{B}{T}t$ and $k = \frac{2\pi f}{c}$, we can express the signal model as:

$$s(k,n) = p_0 e^{-j2\pi k R_0} e^{-j2\pi k \triangle R(n)}. \tag{5.12}$$

It can be observed that to compensate the range shift, we use $e^{j2\pi k\triangle R(n)}$ and cancel the slow-time-variant part $e^{-j2\pi k\triangle R(n)}$ in the equation above. In this case, we initially estimate $\triangle R(n)$ across the slow time.

The cross-correlation is used to estimate $\triangle R(n)$. In practice, noise in the system may cause false peaks and large range estimation errors. Such errors can be mitigated by using the cross-correlation method that considers the similarity of the whole range profile. It is performed between the selected range profile and all of the profiles. To make this process faster, this can be carried out in frequency domain [72]. DFT is firstly performed on both reference profile and other to be shifted. Then the conjugate reference profile is multiplied with each one and this is followed by IDFT to transfer back to the range domain and obtain the cross spectrum. At last, $\triangle R(n)$ can be acquired by searching the peak index from the correlation spectrum (shown in Fig. 5.5):

$$\triangle R(n) = \underset{R \in [0, R_{max}]}{\text{argmax}}\ cor(S_{ref}(R), S(R, n)), \tag{5.13}$$

where $cor(\cdot)$ represents the cross-correlation operation, and R_{max} denotes the maximum range of range profile. It is noted that this parameter is controlled by the bandwidth B, chirp duration time T, sampling frequency f_s and the number of samples in fast time of mmWave radar configuration (shown in Table. 5.1).

Then the compensation of the motion is performed refer to $\triangle R(n)$ in the cross spectrum. The compensation factor $e^{j2k\triangle R(n)}$ is used to multiply with the original signal, which can be expressed by:

$$s_{\text{shift}}(t, n) = s(t, n)e^{j2k\triangle R(n)}. \tag{5.14}$$

MIMO Imaging with Focusing. Fig. 5.4b and Fig. 5.4c depict the range profile before and after motion compensation. It can be observed that after conducting the compensation, the range profile becomes more linear and whole motion pattern is projected and compressed to x-axis (only in cross range and focusing on fixed R_0). Here, it should be noted that the movement is compensated to the selected reference range R_0 (radical range in polar coordinates) but y_0 is the distance in Cartesian coordinate. According to the assumption in Section 5.1.1, y_0 is very closed to R_0 in the valid field of view (FoV) (it is usually less than 10 cm difference), so R_0 can be approximated to y_0, where y_0 in the following article is identified by R_0 and used for imaging.

Given that the position along the x-axis is identified, there is no necessity for a 2D FT on the matrix. Instead, we conduct a DFT over the slow time, expressed as:

$$s_{\text{shift}}(t, k_x) = \text{DFT}_n(s_{\text{shift}}(t, n)). \tag{5.15}$$

Due to the fixed y_0, the exhaustive scanning of the entire xy plane for imaging and matched filter design becomes unnecessary. The matched filter can be limited to a single dimension, specifically in terms of k_x. By leveraging the relationship $(k_x, k_y) = (k \cos\phi, k \sin\phi)$, which yields $k_y = \sqrt{k^2 - k_x^2}$, the matched filter is simplified as:

$$H(k_x) = e^{j2y_0\sqrt{k^2-k_x^2}}. \tag{5.16}$$

Then the image at (x, y_0) can be obtained by perform 1-D IDFT over k_x domain:

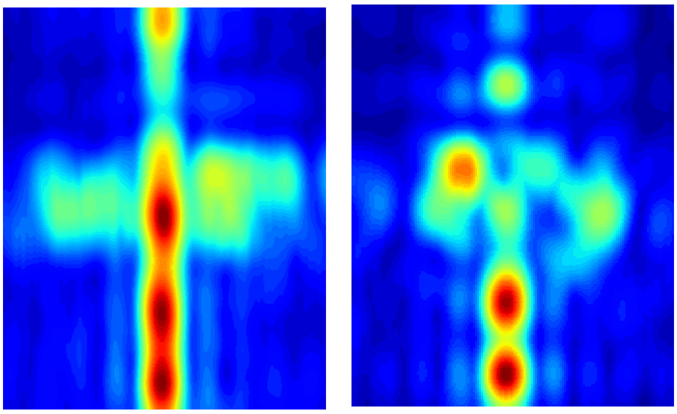
$$\text{Image}(x, y_0) = \text{IDFT}_{k_x}(s_{\text{shift}}(k_x)e^{j2y_0\sqrt{k^2-k_x^2}}). \tag{5.17}$$

Since the MIMO radar is used in our scheme and the antenna array is placed vertically (shown in Fig. 5.9), the AoA θ of elevation (equivalent to z axis) can be obtained by perform FT over the antenna arrays to obtain the final image, and this can be represented by:

$$\text{Image}(x, y_0, \theta) = \text{DFT}_l(\text{IDFT}_{k_x}(s_{\text{shift}}(k_x, l)e^{j2y_0\sqrt{k^2-k_x^2}})), \tag{5.18}$$

where l denotes the antenna index.

Computational Complexity. Compared to the original imaging method, the proposed approach exhibits reduced computational complexity. Assuming the number of DFTs performed on both the fast and slow time domains is denoted as M and N respectively, and for simplicity, considering that the spatial bins of the xy plane are set to be M for the x-axis and N for the y-axis. In the proposed approach, a 2D DFT is performed once initially. Subsequently, it requires the matched filter to search all spatial bins, followed by IDFT each time. This results in a computational complexity of approximately $MN(MN+1)\log(MN)$ for the original method. In contrast, with the motion compensation effectively compressing motion to the x-axis, there is no need for a full 2D plane scan to design the matched filter. Additionally, the proposed scheme only employs 1D DFT and IDFT once. Even when considering the computation required for trajectory localization and motion compensation, the overall complexity remains lower at approximately $MN(3 \log M + 2 \log N)$.



(a) Unprocessed ISAR. (b) Our Imaging scheme.

Figure 5.6 : Comparison of imaging performance.

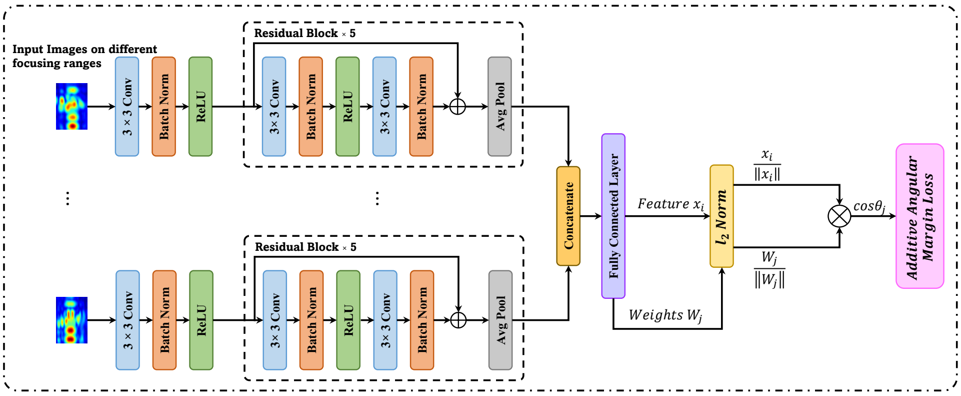


Figure 5.7 : ResNet based multi-range fusion Re-ID network architecture.

5.1.4 Sensing Performance and Comparison

Here, we present some results in Fig. 5.6 to demonstrate the imaging improvement that can be achieved by our proposed imaging scheme, compared to the unprocessed ISAR method. In the unprocessed result shown in Fig. 5.6a, the image is defocused and blurred because of the lack of phase error correction. In contrast, the imaging result obtained by the proposed method shown in Fig. 5.6b accurately portrays the user's body profile.

5.2 Identification Model

In this section, the learning based methodology is introduced for training and user identification. The main contents contain the feature extraction neural networks and the training scheme of arc-margin loss.

5.2.1 Feature Extraction Network

As discussed in the previous section, the reconstructed human body image is expected to represent unique body characteristics for different users. Given that an individual's body can be influenced by various factors such as height, weight, figure, and more, these images hold the potential for user identification and authentication. Extensive research efforts have been directed towards feature extraction from these images, with deep convolutional neural network (DCNN) based methods emerging as a proven and dominant approach. These methods have gained traction not only in computer vision but also in the domain of radio images.

ResNet [74], is one of the primary and efficient DCNN for feature extraction. ResNet's pivotal innovation lies in its residual block, which plays a crucial role in reducing overfitting during the training process and enhancing inference performance. The architecture of the residual block employed in this chapter is depicted in Fig. 5.7. It consists of two 3×3 convolution (Conv) layers, two Batch Normalization (BN) layers, and an activation function, Rectified Linear Unit (ReLU). In our model, the input part comprises a 3×3 convolution layer, followed by BN and ReLU activation. A total of 5 residual blocks are applied to form the complete network architecture. Towards the end of the network, an average pooling layer (Avg pool) and a fully connected network (FCN) are employed to map the features into a 1×512 vector.

The mmWave radar usually has high range resolutions and the scattering points from human body could change during the movement, potentially leading to slightly varying imaging results for the same user. To improve the robustness of the scheme, the imaging results from the adjacent range bins of the reference are input to the neural network. Our approach involves applying multiple matched filters (in our work, we use 5 filters with 5 cm interval distances) to obtain multiple images around the selected range y_0. It is reasonable to expect that using features from multiple images captured at adjacent ranges can be beneficial for classification, as it provides more information. Therefore, our neural network (NN) must be capable of effectively fusing these images. In conventional image classification tasks, RGB images are processed by convolutional layers that operate on the three color channels. However, in our task, the images generated by radio imaging exhibit causality in terms of ranges. These multiple images cannot be straightforwardly placed into separate channels. Instead, they are individually fed into the network, and the output features from these different images are concatenated before being input into a final FCN for classification. The overall architecture of our model is visualized in Fig. 3.4.

The number of input radio images is set to be five corresponding to performing five different matched filters at the ranges of y_0, $y_0 \pm 5$ *cm* and $y_0 \pm 10$ *cm*. Since the mmWave radar is placed at 45° with respect to the wall, the width of the users, therefore, will be compressed. And ± 20 *cm* is enough to represent most of the users. It should be noticed that increasing the number of input images entails conducting more matched filtering at smaller intervals around the reference range y_0, necessitating the addition of more parallel residual blocks. This can potentially improve the performance, but it could inevitably increase the computation time, GPU memory usage and training duration. Therefore, the balance between the accuracy and the cost of computation resources need to be considered.

5.2.2 Training Strategy and Loss Function

In conventional classification tasks, the training strategy often involves using the softmax function as a loss to train the network to distinguish between multiple classes. While softmax is known for its high performance in classification tasks, it has a critical limitation that cannot be overlooked: it lacks the ability to effectively handle intra-class appearance variations. In scenarios where data is collected from different ranges, places, and times (as in our work, spanning over three months), even for the same individual, the features extracted from data collected in the first month may differ from those collected in the third month. If the model is trained using softmax, its identification capabilities could deteriorate, making it less efficient in such situations. To address this challenge, we employ the Additive Angular Margin Loss [75], which has been demonstrated to effectively handle the problem of large intra-class variation. The core principle of this strategy is to maximize the cosine similarity among different classes while minimizing it for the same class.

Specifically, recalling the softmax loss function and expanding the last FCN layer by

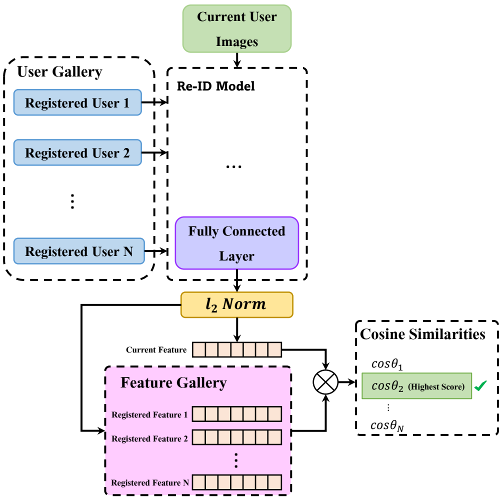


Figure 5.8 : Identification Procedure in Test.

$W^T x + b$, it becomes:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}}, \tag{5.19}$$

where $x_i \in \mathbb{R}^d$ is the concatenated feature of i-th sample belonging to y_i-th class and the feature dimension d is set to be 1024. W_j is the j-th column of the weight $W \in \mathbb{R}^{d\times n}$ (n and N are the number of classes and batch size, respectively) and $b_j \in \mathbb{R}^n$ is the bias. By using the product between two vectors, it yields $W_j^T x_i = \|W_j\|\,\|x_i\|\cos\theta_j$ and θ_j is the angle between W_j and x_i. By using l_2 normalization to remove the magnitudes and adding a re-scale factor s and penalty m, the final Additive Angular Margin Loss becomes:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}}{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}}. \tag{5.20}$$

By employing normalization and the cosine function, our approach compresses the feature space into a bounded hypersphere, with the boundary defined by the rescaling factor s. This means that even in scenarios with significant intra-class variation in the data, the features are confined within this hypersphere due to the finite amplitude of the cosine function. As a result, these features tend to cluster around the center of their respective classes, leading to improved recognition performance and classification robustness.

5.2.3 Identification

After training the model, the cosine similarity is utilized for user identification. The cosine margin ensures that features belonging to the same class exhibit a higher cosine similarity, while features from different classes have a lower similarity score. In this case, user identification is transformed into a matching procedure (shown in Fig. 5.8). Specifically, known classes' features are employed as the gallery of identities (selected from the training/testing samples). When the Re-ID task is initiated, the current user's images are fed into the model, and the output from the FCN is treated as the input feature. Subsequently, an l_2 norm is applied to both the input feature and all the gallery features. These normalized features are then used to calculate the cosine similarity between the input feature and all the gallery identities. Finally, the pair with the highest cosine similarity is selected as the Re-ID result. It's worth noting that this method isn't a recognition process; rather, it's a pure matching procedure that aligns with the definition of Re-ID.

5.3 Implementation and Results

In this section, we will conduct a series of experiments to validate our scheme.

5.3.1 Implementation

Unless stated otherwise, the experimental results reported in this section are based on the setup.

Data Collection. The datasets contain 25.6k samples from ten individuals. The volunteers consist of 15 males and 15 females, and their heights range from 155 *cm* to 188 *cm* and the weights are between 42 *kg* to 90 *kg*. Therefore, the datasets are collected from individuals who exemplify typical human characteristics. During the data collection, participants were instructed to move freely from one side of a corridor scene to the other. The participants had the autonomy to determine their own moving patterns, postures, and walking speeds, with no specific requirements. We set the maximum sampling period of the mmWave radar to 4 seconds for each imaging. The data collection period spanned approximately three months. This diverse and extended data collection duration allowed us to capture a wide range of human movement scenarios and variations.

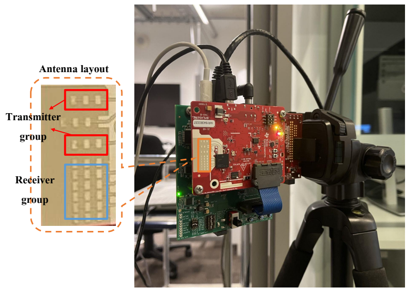


Figure 5.9 : mmWave radar placement. The radar is placed at approximately $45°$ with respect to the glass wall.

Experiment Setup. We utilize COTS mmWave hardware, the TI IWR1843BOOST radar, along with the data capture board TI DCA1000EVM. Two transmitting and four receiving directional antennas are used, shown in Fig. 5.9. Therefore, a virtual an 8-element uniform linear array (ULA) is formed in the vertical domain, with the vertical angular resolution of 15 degrees. The radar was positioned at a $45°$ angle to the wall and fixed at a height of 1.2 meters on a tripod. To ensure that the antenna was perpendicular to the ground, the radar was rotated by $90°$ shown in Fig. 5.9. The radar configuration details are provided in Table. 5.1. Notably, the sampling rates significantly impact the range resolution. With a fixed number of sample points, increasing the sampling rate extends the maximum range but compromises range resolution, leading to less precise estimations and complicating phase shift compensation (Doppler) during user movement. Balancing the configuration of sampling rates and sample numbers requires careful consideration of specific scenarios and computing resources.

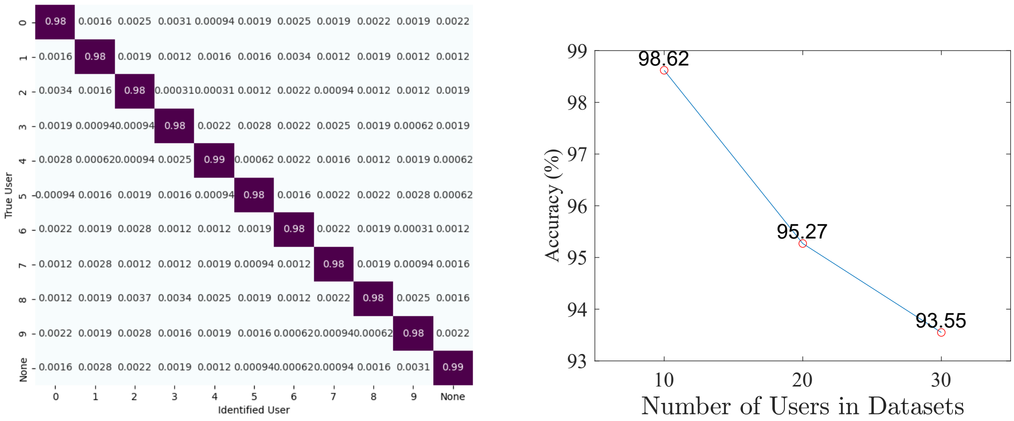
Metrics. (1) The Re-ID accuracy is the primary evaluation metric and represents the probability of the current individual being correctly identified by RImID. (2) The confusion matrix can provide more detailed information. In the confusion matrix, the rows represent the ground truth ID, while the columns denote the predicted identities. The value at the i^{th} row and j^{th} column indicates the percentage of person i being identified as person j. This matrix offers a comprehensive view of the identification performance.

Implementation Platform. Our scheme is implemented using Python, with PyTorch utilized for model learning and Cupy for accelerating the radio imaging process. And we employed the MATLAB interface for radar data capture. The system operated on a

Table 5.1 : Parameters of mmWave Setup

Parameters	**Value**
Start frequency	77 GHz
Frequency slope	70 MHz/us
Idle time	43.10 us
Sample Number	256
Ramp end time	57us
Sample rate	5.5 MHz
Frame periodicity	40 ms

desktop computer equipped with an Intel(R) Core CPU i7-9700 (3.6 GHz A4) and 32GB of memory. Model training is accelerated using a GPU RTX6000.



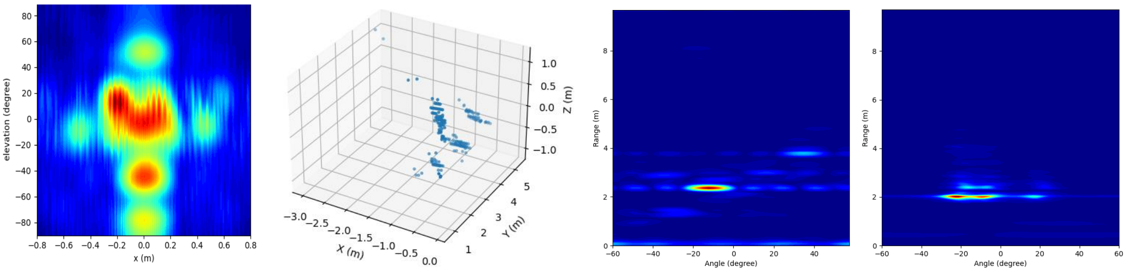
(a) The confusion matrix of 10 users Re-ID.

(b) The performance of larger datasets with more users.

Figure 5.10 : Re-ID performance.

5.3.2 User Re-ID Performance

The neural network comprises 5 radio images from adjacent ranges of reference distance (y_0, $y_0 \pm 5$ *cm* and $y_0 \pm 10$ *cm*) as input and is trained using a dataset consisting of 25.6k samples. This dataset is divided into 17,070 samples for training and 8,530 samples for testing, maintaining a ratio of 2:1. Following a training period of 4 hours, the model accuracy converges, achieving optimal performance as detailed next.



(a) Human body profile.

(b) Human point clouds.

(c) DFT based heatmap.

(d) MVDR based heatmap.

Figure 5.11 : Sensing results and comparison between the proposed scheme and other common methods.

10 Users Re-ID. We evaluate our scheme under the condition where 10 persons walked in front of the mmWave radar. They are allowed to move freely, without any specific restrictions on their movement patterns. The data collected during this evaluation is not a part of the training process. The confusion matrix, as shown in Fig. 5.10a, illustrates the results, and it is evident that the accuracy for all users exceeds 98 percent. The similarity score threshold is set to be 0.1 and is used to first determine whether the user is existed or not. If the similarity score is below the threshold, the current label will be directly recognized as non-user. This demonstrates the effectiveness of our approach in correctly identifying individual persons in a real-world scenario with diverse movements.

Larger Scale Users Re-ID. We have also tested our scheme for more people without changing the used neural network. The results are shown in Fig. 5.10b. The performance degrades, as expected. However, it remains strong as the number of individuals increases. Specifically, we can still achieve 95 percent and 93 percent accuracy with 20 and 30 users, respectively. It is important to highlight that the complexity of neural networks generally needs to increase with larger datasets to enhance learning capacity. Remarkably, in our experiments with 20 and 30 users, we maintained the same neural network configuration and learning strategy without adding new layers. This suggests that there is potential for even greater performance improvements by scaling up the complexity of the neural networks.

5.3.3 Comparison with state-of-the-arts

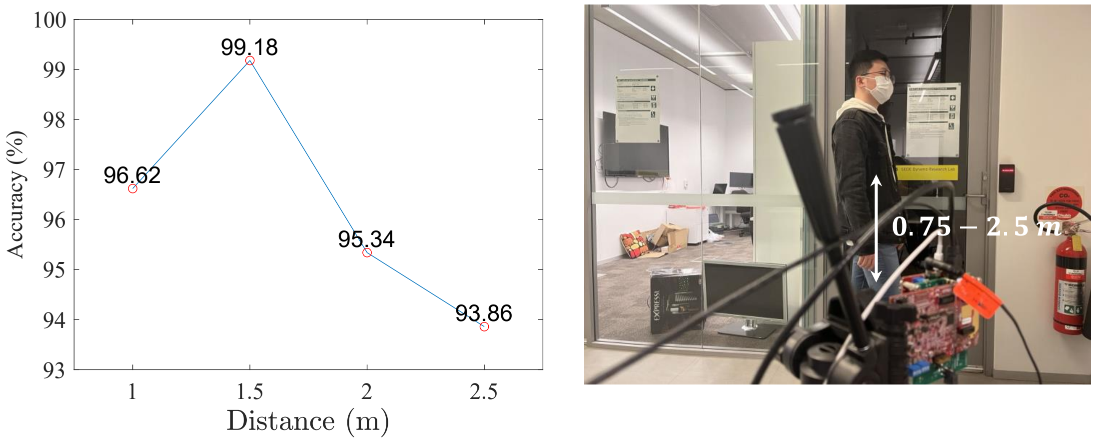
As far as our knowledge, there have been limited prior works that focus on Re-ID using radio imaging, and most of the state-of-the-art (SOTA) Re-ID methods are based on gait analysis or point cloud data. Furthermore, many of these works collect data individually, resulting in variations in data collection environments, testers characteristics, and device setups. Additionally, datasets in this domain are rarely open-sourced.

Here, we provide a brief overview of mmWave radar-based and camera-based SOTA methods, along with their techniques and results. (1) Gait-based methods [79], [81], and [84] rely on micro-Doppler patterns induced by gait for classification. (2) Point cloud-based methods [78], [82] classify point cloud data generated by the entire person during motion. (3) Range-Angle heatmap-based methods [77], [90] generate range-angle heatmaps through 2-D spectrum analysis on the fast-channel domain, followed by recognition. (4) Raw radar data [85] based method first organize the data into a 3D array in fast, slow and channel axes, and directly feed it into a well-designed neural network, to learn the pattern signals. (5) Optic photos [86], [87], [88], [89] based methods use RGB cameras to capture the images and use the sophisticated networks to learn the texture features for Re-ID. Table. 5.2 shows our work offers several advantages, such as the ability to work over the long term (discussed in detail later), and it achieves higher accuracy compared to the prior radar based methods. The optic photos based methods can achieve higher accuracy than radar, because the optic cameras can always provide much higher resolution, especially fine-grained texture features (distinct facial, physique images), than radar. However, it may encounter problems with insufficient illumination and privacy issues that typically do not arise in radio based Re-ID measures.

As shown in Fig. 5.11, we also compare the range-angle heatmaps and point clouds generated by human bodies, as they can be used for direct comparison with our body imaging scheme. From Fig. 5.11b, it is evident that due to the sparsity of the antennas, point clouds are unable to represent the fine-grained features of the human body accurately. Similarly, the range-angle heatmap generated using the DFT over the antenna domain, as shown in Fig. 5.11c, struggles to capture the object person effectively. Zero padding is often employed to address antenna limitations, but this leads to spatial

Table 5.2 : State-of-the-arts of Re-ID

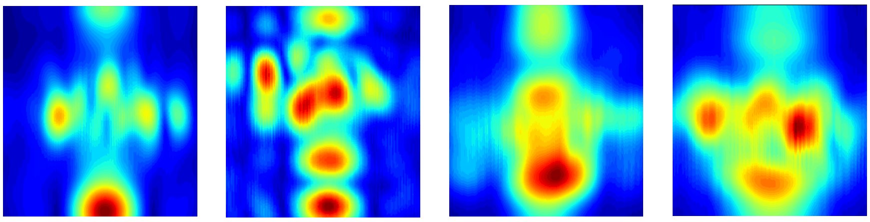
Method	Feature	Accuracy	Long-term
Meng *et al.* [76]	Gait from Pointclouds	90%	-
Fan *et al.* [77]	Range-Azimuth Range-Elevation Heatmaps	96.5%	Yes
mID [78]	Pointclouds	89%	-
GaitCube [79]	Gait from Micro-Doppler	98.3%	-
MmSense [80]	Vital Signs and body Curve from Beam Scan	93%	-
Pegoraro *et al.* [81]	Gait from Micro-Doppler	98%	-
Cheng *et al.* [82]	Pointclouds	78.46%	-
MU-ID [83]	Gait from Range-Doppler	97%	-
Vandersmissen *et al.* [84]	Gait from Micro-Doppler	78.46%	-
CubeLearn [85]	Raw radar cube	98.06%	-
Gu *et al.* [86]	Optic camera images	99.8%	Yes
SPOT [87]	Optic camera images	96.44%	-
DSCNet [88]	Optic camera image	99.77%	-
3DSL [89]	Optic camera image	98.3%	-
Li *et al.* [90]	Range-Azimuth Range-Elevation Heatmaps	95.9%	-
Ours	**Radio Body Image**	98.62%	Yes



(a) Distance impact.

(b) Range test scenario.

Figure 5.12 : The impact of range. The range is the minimum distance when the person passes over the radar.



(a) 1m.

(b) 1.5m.

(c) 2m.

(d) 2.5m.

Figure 5.13 : Imaging Results showing range impact.

frequency leakage, and the person is only unclearly represented by a light area. The Minimum Variance Distortionless Response (MVDR) [91] method, known for its super-resolution capacity in estimating angles, still faces challenges in identifying person features, as illustrated in Fig. 5.11d. In contrast, our proposed method provides a fairly high-resolution representation of the person. As shown in Fig. 5.11a, the proposed sensing scheme can generate the body profiles of the users, which provide more direct and salient feature information of the user's figure. Although it may not be as clear as an optical camera due to the person's movement, it still captures essential features like arms, head, and torsos in the imaging result. This demonstrates the advantage of our approach in generating detailed body images from radio imaging data.

5.3.4 Impact of Distance for Imaging and Re-ID

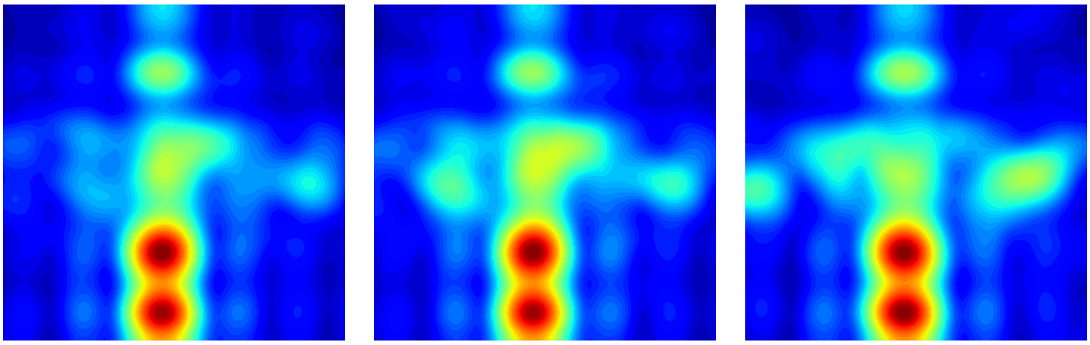
To investigate the impact of distance, we conducted experiments where the person is asked to move at different ranges: 1m, 1.5m, 2m, and 2.5m. The maximum distance is set at 2.5m, reaching the wall on the other side (shown in Fig. 5.12b). It's important to note that as the person moves from far to near the radar, the cross-range also changes during this period. The range refers to the distance when the person moves to the right in front of the radar. The results, as shown in Fig. 5.12a, indicate that as the range increases, the Re-ID accuracy decreases. At the maximum range of 2.5m, the accuracy drops to 93.86 percent. Even so, it can still demonstrate the robustness of our scheme even at longer distances from the radar.

The quality of imaging results is indeed more sensitive to the distance of the object person from the radar. At ranges of 2m and 2.5m, the images become more ambiguous, as shown in Fig. 5.13c and Fig. 5.13d. The best imaging quality appears to be at 1.5m, as depicted in Fig. 5.13b. Furthermore, when the person moves too close to the mmWave radar, certain body parts may go beyond the FoV. This can be observed in Fig. 5.13a, where the person's head is not entirely depicted in the image at a distance of 1m, as both the head and lower body are outside the valid sensing scope. This highlights the importance of maintaining an appropriate distance between the person and the radar for optimal imaging results.

The radio images exhibit diversity across different ranges, even when originating from the same user. This variability may stem from noise propagation, variations in Angle of Arrival, and fluctuations in received signal strength between far and close fields. To mitigate these effects, all images of a single user captured at different ranges are assigned the same label and subsequently fed into the network. This approach enables the network to learn common features shared by images of the same individual, which are less influenced by range variations.

Interestingly, the accuracy at a range of 1m is lower than at 1.5m. This suggests that at very close ranges, the person may go out of the antenna FoV, causing the imaging process to miss the top and bottom parts of the person. This limitation can lead to a decline in performance.

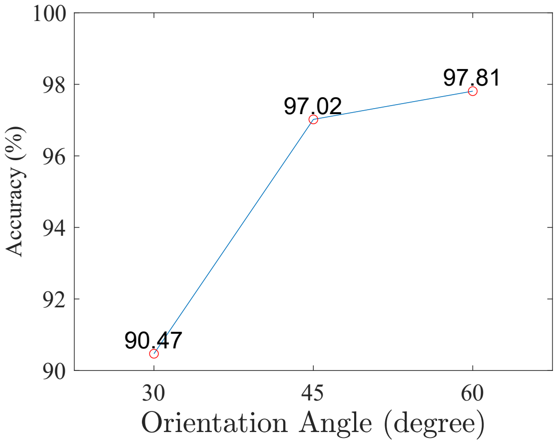
5.3.5 Impact of Orientation Angle



(a) 30°.

(b) 45°

(c) 60°.



(d) Angle Impact.

Figure 5.14 : Impact of Orientation Angle.

To evaluate the impact of the orientation angle on system performance, we conducted experiments for 30°, 45° and 60°, and present results in Fig. 5.14. It can be observed that when the angle increases, the body figure becomes clearer as the mmWave radar beam has a better projection of the body movement. When the angle is very small, only the lateral side image can be obtained, and the image for the body profile cannot be obtained.

5.3.6 Impact of Moving Speed for Imaging and Re-ID

The impact of person movement speed is evaluated in this section, considering low (0.5-1m/s), medium (1-1.5m/s), and high ($>$ 2m/s) velocities. After training on the entire dataset, a new dataset collected at different speeds is used for evaluation, ensuring that the model is not trained on this specific dataset. The results are presented in Fig. 5.15.

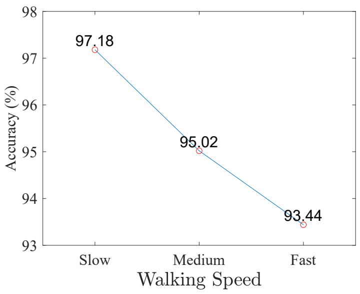
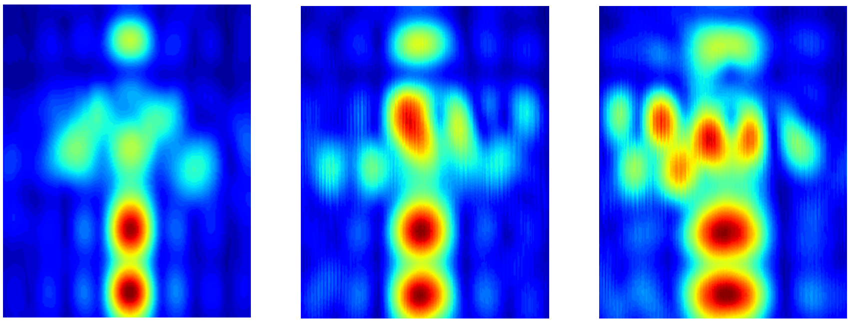


Figure 5.15 : Re-ID Results showing moving speed impact.

As expected, the highest accuracy of 97.18 percent is achieved when the person's speed is low. Conversely, as the speed increases, the accuracy tends to decrease. However, even at the highest speed, the scheme still maintains a high level of accuracy, exceeding 90 percent. This demonstrates the robustness of our approach in recognizing individuals across a range of walking speeds, making it suitable for real-world scenarios where walking speeds can vary. It should be mentioned that to address discrepancies among images captured from the same user at varying speeds and ranges, we treat all the images from the particular user as one label. And during the training process, all the images are fed into the neural network together. Notably, speed is not provided as additional information or labels to the network, with the objective of enabling the network to focus on learning features less influenced by changes in range and speed.

The results reveal the following insight. Due to the fixed sampling period, the collected data frames remain the same regardless of the person's speed. However, when the person moves at a faster speed, it covers a longer cross-range distance. Since the valid cross-range of the radar is fixed, the person exits the valid field earlier, resulting in fewer valid data frames. Consequently, the imaging quality is adversely affected, leading to a drop in the model's Re-ID performance. In this case, the imaging results provide a more intuitive explanation for this phenomenon. In Fig. 5.16a, the best imaging quality is observed at slower speeds. With an increase in the person's moving speed, the image quality deteriorates. Since the number of transmitted pulses is fixed, it implies that the sample rate in the slow-time domain remains constant. When the speed is faster, the Doppler frequency becomes higher. Under the fixed sample rate, this can lead to overlapping in the slow-time domain, known as speed ambiguity, as illustrated in Fig. 5.16b and Fig.



(a) Slow speed. (b) Medium speed. (c) Fast speed.

Figure 5.16 : Imaging Results showing moving speed impact.

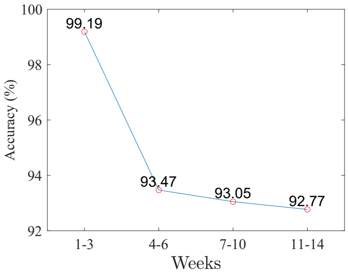


Figure 5.17 : Results showing long-term impact.

5.16c. This overlapping makes it more challenging for the model to distinguish persons, further impacting Re-ID performance.

The walking speed of the user also affects compensation. As the user's speed increases, the phase shift becomes more unstable, resulting in greater range variation and subsequently amplifying Doppler interference. The estimation accuracy may decrease, resulting in compensation less effective. Consequently, the quality of the radio image generated may deteriorate, potentially leading to less accurate Re-ID.

5.3.7 Impact of Time Period

The effect of time periods poses a significant challenge for person Re-ID. In reality, even within a short period, such as two consecutive days, a person's appearance can change. For instance, the person might have a different hairstyle or wear clothes with varying colors or materials one day compared to the previous day. Over a longer term, these variations can become more pronounced. For example, as the seasons change and it becomes winter, the person might start wearing a coat, leading to a noticeable change in appearance. These variations can result in significant intra-class changes, leading to a decline in the model's performance. This phenomenon is often referred to as model decline.

In this section, we explore the impact of time periods on person Re-ID. The dataset is collected over a span of 3 months, approximately 14 weeks, which we split into four parts. The training dataset comprises data collected in the first three weeks, while the remaining data from weeks 4-6, 7-10, and 10-14 are used to test the Re-ID accuracy. The results, as shown in Fig. 5.17, demonstrate that the highest accuracy (99.19 percent) is achieved during the nearest weeks, as expected. However, even after two to three months, the accuracy only decreases slightly to its lowest point, 92.77 percent. This suggests that as the time period becomes longer, the accuracy experiences only a minor drop. This behavior can be attributed to the nature of radio imaging, which provides fairly clear body information of moving persons, and this information remains relatively constant over a term. Unlike optical cameras that capture detailed appearance changes, such as hairstyle variations, radio imaging focuses on consistent body structure. By using this consistent body information as input and leveraging training loss functions and strategies, high-dimensional features can be extracted, and intra-class changes are constrained, resulting in consistently high accuracy.

Table 5.3 : The comparison of Softmax and Additive Angular Margin used in this work

Method	Performance on Accuracy(%)			
	Week 1-3	Week 4-6	Week 7-10	Week 11-14
Softmax	94.81	86.35	79.38	70.43
This work	99.19	93.47	93.05	92.77

To verify the effectiveness of the classification loss function used in our framework, we compare the performance of standard softmax loss and Additive Angular Margin loss in our ISAR-based person re-identification task considering the long-term impact. Both models use the same backbone architecture and training configurations, differing only in the loss function. As shown in Table. 5.3, although their results do not contain large discrepancy in the first 3 weeks, the performance of the softmax drops significantly along the period. In last collection period of week 11-14, the Additive Angular Margin loss outperforms the softmax by approximately 22 percent accuracy. It could demonstrate that Additive Angular Margin is capable of handling with the long-term Re-ID and suppress the intra-class change from the radio images.

Chapter 6

Conclusion and Future Work

In this chapter, we present a summary of our work in this thesis and discuss potential future work.

6.1 Conclusions

The majority of this thesis has been focused on mmWave vital signs sensing and user identification. Regarding vital sign sensing, we conducted two studies: 1) We investigated the multi-modal sensing approach that fusing computer vision and the mmWave sensing scheme for user localization in the complex environment and the interference suppression of adjacent passengers and users themselves. 2) We investigated the integrated higher-level vital signs sensing that contains a robust localization of heartbeat and wrist, and a physics-driven deep learning approach for ECG and pulse recovery and BP estimation.

As for user identification sensing, we study the mmWave imaging approach to improve the sensing fineness and long-term robustness. We proposed an enhanced mmWave ISAR imaging method which could obtain the user's stable body profile and addressed the challenges of low spatial resolution in mainstream mmWave devices. We proposed an end-to-end network architecture combined with a sophisticated training and identification strategy, which could be capable of solving the problem of the performance decline in long-term user ID.

We summarize our main contributions as follows:

- We proposed a hybrid mmWave and camera method to sense the breath and heartbeat in complex environment. To localize the users of interest, we utilized the object detection algorithms based on YOLO to acquire their angles and using the range-angle obtained from MVDR to estimate their ranges. Then we proposed to use

Tx-Rx beamforming and introduced an weighted-multi-channel VMD algorithms that could eliminate the spatial and body motion noises. Moreover, this is the first work that addresses the three challenges of accurate localization, suppressing dynamic environment interference and the users' body motion at the same time. Our approach was evaluated on more than ten users and different scenarios. The experimental results show that vital signs sensing strategy exhibitsthe robust and superior performance compared to other mainstream approaches.

- We proposed an integrated higher-level vital signs sensing scheme for recovering ECG and pulse waveform and BP estimation. We proposed to leverage the cardiac frequency band and combine with the range and angle spectrum to obtain the accurate localization of heat and wrist. Then we introduced the neural-ODEs embedded learning method that not only improves the explainability but also increase the accuracy. The experimental results reveal the robustness and the flexibility. Our method was evaluated more than 20 users with diverse ages, and achieved a promising performance. To the best of our knowledge, this is the first work that could sensing three different higher-level vital signs: ECG, pulse and BP simultaneously.
- We proposed a novel mmWave user ID scheme. By using the mmWave ISAR imaging, more virtual apertures could be generated thereby the spatial resolution is significantly improved, which contributes to the more fine-grained user body profile images. Further, we leveraged the arcmargin loss and cosine similarity as the training loss function and identification metric, addressing the challenges of intra-class change in long-term user ID. The experimental results on 30 users over 100 days demonstrate the high effectiveness and robustness. This work is also the first research that using mmWave imaging for user ID.

6.2 Future Work

While much progress in both mmWave vital signs sensing and user identification has been made, there remain a number of directions in our future research, some of which we'll discuss below.

- We have shown that our research are effective to obtain the heartbeat and breath from the user with body motion in dynamic environment. In the future, we will focus on more complicated scenarios where the users contain larger-amplitude motion, making it more robust in practice. Once the accurate target positions are obtained, advanced learnable algorithms can be further used to achieve vital sign recognition, by suppressing the impact of large body movement.
- Similarly, the more efficient method will be investigated to obtain the higher-level vital signs when the user contains the larger body part motion. And it is expected to expand to multiple user scenarios.
- The integrated sensing research will be conducted. It will focus on combining the dynamic vital signs sensing, higher-level vital signs sensing and user identification together, forming an integrated sensing system that could achieve these goals at the same time.

Future work could explore embedding the proposed model into a distributed sensing network, potentially leveraging federated learning or edge inference frameworks to enable collaborative, privacy-aware, and adaptive sensing systems that continuously learn from user interactions and context dynamics.

By extending this research into the broader field of ubiquitous sensing, the system could evolve into a foundational element of next-generation intelligent environments, offering seamless integration between human physiological state, identity, and surrounding infrastructure.

Bibliography

[1] Y. Wang, W. Wang, M. Zhou, A. Ren, and Z. Tian, “Remote monitoring of human vital signs based on 77-GHz mm-wave FMCW radar,” *Sensors*, vol. 20, no. 10, p. 2999, 2020.

[2] P. Cao, W. Xia, M. Ye, J. Zhang, and J. Zhou, “Radar-id: human identification based on radar micro-doppler signatures using deep convolutional neural networks,” *IET Radar, Sonar & Navigation*, vol. 12, no. 7, pp. 729–734, 2018.

[3] H. Yin, S. Yu, Y. Zhang, A. Zhou, X. Wang, L. Liu, H. Ma, J. Liu, and N. Yang, “Let IoT knows you better: User identification and emotion recognition through millimeter wave sensing,” *IEEE Internet of Things Journal*, 2022.

[4] X. Xu, J. Yu, C. Ma, Y. Ren, H. Liu, Y. Zhu, Y.-C. Chen, and F. Tang, “mmECG: Monitoring human cardiac cycle in driving environments leveraging millimeter wave,” in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 90–99.

[5] T. K. Vodai, K. Oleksak, T. Kvelashvili, F. Foroughian, C. Bauder, P. Theilmann, A. Fathy, and O. Kilic, “Enhancement of remote vital sign monitoring detection accuracy using multiple-input multiple-output 77 GHz FMCW radar,” *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, 2021.

[6] Y. Li, D. Zhang, J. Chen, J. Wan, D. Zhang, Y. Hu, Q. Sun, and Y. Chen, “Towards domain-independent and real-time gesture recognition using mmwave signal,” *IEEE Transactions on Mobile Computing*, 2022.

[7] O. R. Fogle and B. D. Rigling, “Micro-range/micro-doppler decomposition of human radar signatures,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3058–3072, 2012.

[8] F.-K. Wang, C.-J. Li, C.-H. Hsiao, T.-S. Horng, J. Lin, K.-C. Peng, J.-K. Jau, J.-Y. Li, and C.-C. Chen, “A novel vital-sign sensor based on a self-injection-locked oscillator,” *IEEE transactions on microwave theory and techniques*, vol. 58, no. 12, pp. 4112–4120, 2010.

[9] F.-K. Wang, C.-H. Fang, T.-S. Horng, K.-C. Peng, J.-Y. Li, and C.-C. Chen, “Concurrent vital sign and position sensing of multiple individuals using self-injection-locked tags and injection-locked i/q receivers with arctangent demodulation,” *IEEE transactions on microwave theory and techniques*, vol. 61, no. 12, pp. 4689–4699, 2013.

[10] F.-K. Wang, P.-H. Juan, S.-C. Su, M.-C. Tang, and T.-S. Horng, “Monitoring displacement by a quadrature self-injection-locked radar with measurement-and differential-based offset calibration methods,” *IEEE Sensors Journal*, vol. 19, no. 5, pp. 1905–1916, 2018.

[11] Y. Yuan, C. Lu, A. Y.-K. Chen, C.-H. Tseng, and C.-T. M. Wu, “Multi-target concurrent vital sign and location detection using metamaterial-integrated self-injection-locked quadrature radar sensor,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 12, pp. 5429–5437, 2019.

[12] T.-Y. J. Kao, Y. Yan, T.-M. Shen, A. Y.-K. Chen, and J. Lin, “Design and analysis of a 60-ghz cmos doppler micro-radar system-in-package for vital-sign and vibration detection,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 4, pp. 1649–1659, 2013.

[13] T. Zhang, J. Sarrazin, G. Valerio, and D. Istrate, “Estimation of human body vital signs based on 60 ghz doppler radar using a bound-constrained optimization algorithm,” *Sensors*, vol. 18, no. 7, p. 2254, 2018.

[14] J.-K. Park, Y. Hong, H. Lee, C. Jang, G.-H. Yun, H.-J. Lee, and J.-G. Yook, “Noncontact rf vital sign sensor for continuous monitoring of driver status,” *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 3, pp. 493–502, 2019.

[15] Y. Xu, Q. Li, and Z. Tang, “Accurate and contactless vital sign detection in short time window with 24 ghz doppler radar,” *Journal of Sensors*, vol. 2021, 2021.

[16] Y. Yang, J. Cao, X. Liu, and X. Liu, “Multi-breath: Separate respiration monitoring for multiple persons with uwb radar,” in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1. IEEE, 2019, pp. 840–849.

[17] H. Lee, B.-H. Kim, J.-K. Park, S. W. Kim, and J.-G. Yook, “A resolution enhancement technique for remote monitoring of the vital signs of multiple subjects using a 24 ghz bandwidth-limited fmcw radar,” *IEEE Access*, vol. 8, pp. 1240–1248, 2019.

[18] Y. Wang, W. Wang, M. Zhou, A. Ren, and Z. Tian, “Remote monitoring of human vital signs based on 77-ghz mm-wave fmcw radar,” *Sensors*, vol. 20, no. 10, p. 2999, 2020.

[19] W. Lv, W. He, X. Lin, and J. Miao, “Non-contact monitoring of human vital signs using fmcw millimeter wave radar in the 120 ghz band,” *Sensors*, vol. 21, no. 8, p. 2732, 2021.

[20] T. Wang, D. Zhang, Y. Zheng, T. Gu, X. Zhou, and B. Dorizzi, “C-FMCW based contactless respiration detection using acoustic signal,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–20, 2018.

[21] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, “Vital sign and sleep monitoring using millimeter wave,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 2, pp. 1–32, 2017.

[22] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, “Smart homes that monitor breathing and heart rate,” in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 837–846.

[23] P. Nguyen, X. Zhang, A. Halbower, and T. Vu, “Continuous and fine-grained breathing volume monitoring from afar using wireless signals,” in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[24] S. Yue, H. He, H. Wang, H. Rahul, and D. Katabi, "Extracting multi-person respiration from entangled RF signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, pp. 1–22, 2018.

[25] M. Mercuri, Y. Lu, S. Polito, F. Wieringa, Y.-H. Liu, A.-J. van der Veen, C. Van Hoof, and T. Torfs, "Enabling robust radar-based localization and vital signs monitoring in multipath propagation environments," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 11, pp. 3228–3240, 2021.

[26] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Monitoring vital signs using millimeter wave," in *Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing*, 2016, pp. 211–220.

[27] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu, "ViMo: Multiperson vital sign monitoring using commodity millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1294–1307, 2020.

[28] Z. Xu, C. Shi, T. Zhang, S. Li, Y. Yuan, C.-T. M. Wu, Y. Chen, and A. Petropulu, "Simultaneous monitoring of multiple people's vital sign leveraging a single phased-MIMO radar," *arXiv preprint arXiv:2110.08401*, 2021.

[29] S. M. Islam, N. Motoyama, S. Pacheco, and V. M. Lubecke, "Non-contact vital signs monitoring for multiple subjects using a millimeter wave FMCW automotive radar," in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*. IEEE, 2020, pp. 783–786.

[30] 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 Conference (RadarConf18)*. IEEE, 2018, pp. 1450–1455.

[31] U. Ha, S. Assana, and F. Adib, "Contactless seismocardiography via deep learning radars," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.

[32] P. Zhao, C. X. Lu, B. Wang, C. Chen, L. Xie, M. Wang, N. Trigoni, and A. Markham, "Heart rate sensing with a robot mounted mmwave radar," in *2020 IEEE Interna-*

tional Conference on Robotics and Automation (ICRA). IEEE, 2020, pp. 2812–2818.

[33] U. Ha, S. Assana, and F. Adib, “Contactless seismocardiography via deep learning radars,” in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.

[34] Z. Chen, T. Zheng, C. Cai, and J. Luo, “MoVi-Fi: Motion-robust vital signs waveform recovery via deep interpreted rf sensing,” in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 392–405.

[35] T. Zheng, Z. Chen, S. Zhang, C. Cai, and J. Luo, “MoRe-Fi: Motion-robust and fine-grained respiration monitoring via deep-learning UWB radar,” in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 111–124.

[36] S. Ishizaka, K. Yamamoto, and T. Ohtsuki, “Non-contact blood pressure measurement using doppler radar based on waveform analysis by lstm,” in *ICC 2021-IEEE International Conference on Communications.* IEEE, 2021, pp. 1–6.

[37] R. Kawasaki and A. Kajiwara, “Continuous blood pressure estimation using millimeter wave radar,” in *2022 IEEE Radio and Wireless Symposium (RWS).* IEEE, 2022, pp. 135–137.

[38] L. Singh, S. You, B. J. Jeong, C. Koo, and Y. Kim, “Remote estimation of blood pressure using millimeter-wave frequency-modulated continuous-wave radar,” *Sensors*, vol. 23, no. 14, p. 6517, 2023.

[39] Z. Shi, T. Gu, Y. Zhang, and X. Zhang, “mmbp: Contact-free millimetre-wave radar based approach to blood pressure measurement,” in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, 2022, pp. 667–681.

[40] Z. Wang, B. Jin, F. Zhang, S. Li, and J. Ma, “Uwb-enabled sensing for fast and effortless blood pressure monitoring,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 2, pp. 1–26, 2024.

[41] Q. Zhang, D. Li, R. Zhao, D. Wang, Y. Deng, and B. Chen, “Rfree-id: An unobtrusive human identification system irrespective of walking cofactors using cots rfid,” in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2018, pp. 1–10.

[42] X. Huang, N. Patel, and K. P. Tsoi, “Application of mmwave radar sensor for people identification and classification,” *Sensors*, vol. 23, no. 8, p. 3873, 2023.

[43] L. Xu, K. Wang, C. Gu, X. Guo, S. He, and J. Chen, “Gestureprint: Enabling user identification for mmwave-based gesture recognition systems,” in *2024 IEEE 44th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2024, pp. 1074–1085.

[44] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, “mid: Tracking and identifying people with millimeter wave radar,” in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2019, pp. 33–40.

[45] ——, “Human tracking and identification through a millimeter wave radar,” *Ad Hoc Networks*, vol. 116, p. 102475, 2021.

[46] T. Wang, Y. Zhao, M.-C. Chang, and J. Liu, “Open-set occluded person identification with mmwave radar,” *IEEE Transactions on Mobile Computing*, 2025.

[47] Y. Xiang, A. Mu, L. Tang, X. Yang, G. Wang, S. Guo, G. Cui, and L. Kong, “Person identification method based on pointnet++ and adversarial network for mmwave radar,” *IEEE Internet of Things Journal*, vol. 11, no. 6, pp. 10 104–10 114, 2023.

[48] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, “Enabling identification and behavioral sensing in homes using radio reflections,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[49] L. Fan, T. Li, R. Fang, R. Hristov, Y. Yuan, and D. Katabi, “Learning longterm representations for person re-identification using radio signals,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 699–10 709.

[50] M. Z. Ozturk, C. Wu, B. Wang, and K. R. Liu, “Gaitcube: Deep data cube learning for human recognition with millimeter-wave radio,” *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 546–557, 2021.

[51] J. Pegoraro, F. Meneghello, and M. Rossi, “Multiperson continuous tracking and identification from mm-wave micro-doppler signatures,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 2994–3009, 2020.

[52] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, “Mu-id: Multi-user identification through gaits using millimeter wave radios,” in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 2589–2598.

[53] G. M. Brooker *et al.*, “Understanding millimetre wave FMCW radars,” in *1st international Conference on Sensing Technology*, vol. 1, 2005.

[54] G. Jocher, K. Nishimura, T. Mineeva, and R. Vilariño, “yolov5,” *Code repository https://github. com/ultralytics/yolov5*, 2020.

[55] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[56] M. Alizadeh, G. Shaker, J. C. M. De Almeida, P. P. Morita, and S. Safavi-Naeini, “Remote monitoring of human vital signs using mm-wave FMCW radar,” *IEEE Access*, vol. 7, pp. 54 958–54 968, 2019.

[57] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE transactions on signal processing*, vol. 62, no. 3, pp. 531–544, 2013.

[58] T. Zheng, Z. Chen, C. Cai, J. Luo, and X. Zhang, “V2ifi: In-vehicle vital sign monitoring via compact rf sensing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1–27, 2020.

[59] C. E. Shannon, “A mathematical theory of communication,” *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[60] J. Gong, X. Zhang, K. Lin, J. Ren, Y. Zhang, and W. Qiu, “RF vital sign sensing under free body movement,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1–22, 2021.

[61] D. Buxi, J.-M. Redouté, and M. R. Yuce, “Blood pressure estimation using pulse transit time from bioimpedance and continuous wave radar,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 917–927, 2016.

[62] J. Chen, D. Zhang, Z. Wu, F. Zhou, Q. Sun, and Y. Chen, “Contactless electrocardiogram monitoring with millimeter wave radar,” *IEEE Transactions on Mobile Computing*, vol. 23, no. 1, pp. 270–285, 2022.

[63] Y. Liang, A. Zhou, X. Wen, W. Huang, P. Shi, L. Pu, H. Zhang, and H. Ma, “airbp: Monitor your blood pressure with millimeter-wave in the air,” *ACM Transactions on Internet of Things*, vol. 4, no. 4, pp. 1–32, 2023.

[64] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, “A dynamical model for generating synthetic electrocardiogram signals,” *IEEE transactions on biomedical engineering*, vol. 50, no. 3, pp. 289–294, 2003.

[65] S. E. Francis, “Continuous estimation of cardiac output and arterial resistance from arterial blood pressure using a third-order windkessel model,” Ph.D. dissertation, Massachusetts Institute of Technology, 2007.

[66] G. S. Stergiou, B. Alpert, S. Mieke, R. Asmar, N. Atkins, S. Eckert, G. Frick, B. Friedman, T. Graßl, T. Ichikawa *et al.*, “A universal standard for the validation of blood pressure measuring devices: Association for the advancement of medical instrumentation/european society of hypertension/international organization for standardization (aami/esh/iso) collaboration statement,” *Hypertension*, vol. 71, no. 3, pp. 368–374, 2018.

[67] Z. Wang, B. Jin, S. Li, F. Zhang, and W. Zhang, “Ecg-grained cardiac monitoring using uwb signals,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 4, pp. 1–25, 2023.

[68] L. Zhao, R. Lyu, H. Lei, Q. Lin, A. Zhou, H. Ma, J. Wang, X. Meng, C. Shao, Y. Tang *et al.*, “Airecg: Contactless electrocardiogram for cardiac disease monitoring via mmwave sensing and cross-domain diffusion model,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 3, pp. 1–27, 2024.

[69] F. Geng, Z. Bai, H. Zhang, C. Liu, P. Wang, Z. Li, L. Du, X. Chen, and Z. Fang, “Non-contact stable arterial pulse measurement using mmwave array radar,” *Bioengineering*, vol. 11, no. 12, p. 1203, 2024.

[70] Q. Hu, Q. Zhang, H. Lu, S. Wu, Y. Zhou, Q. Huang, H. Chen, Y.-C. Chen, and N. Zhao, “Contactless arterial blood pressure waveform monitoring with mmwave radar,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 4, pp. 1–29, 2024.

[71] Y. Cao, S. Zhang, F. Li, Z. Chen, and J. Luo, “hbp-fi: Contactless blood pressure monitoring via deep-analyzed hemodynamics,” in *IEEE INFOCOM 2024-IEEE Conference on Computer Communications*. IEEE, 2024, pp. 1211–1220.

[72] V. C. Chen, *Inverse Synthetic Aperture Radar Imaging; Principles*. Institution of Engineering and Technology, 2014.

[73] K. Qian, Z. He, and X. Zhang, “3D point cloud generation with millimeter-wave radar,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–23, 2020.

[74] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[75] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.

[76] Z. Meng, S. Fu, J. Yan, H. Liang, A. Zhou, S. Zhu, H. Ma, J. Liu, and N. Yang, “Gait recognition for co-existing multiple people using millimeter wave sensing,” in

Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 01, 2020, pp. 849–856.

[77] L. Fan, T. Li, R. Fang, R. Hristov, Y. Yuan, and D. Katabi, "Learning longterm representations for person re-identification using radio signals," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 699–10 709.

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

[79] M. Z. Ozturk, C. Wu, B. Wang, and K. R. Liu, "Gaitcube: Deep data cube learning for human recognition with millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 546–557, 2021.

[80] T. Gu, Z. Fang, Z. Yang, P. Hu, and P. Mohapatra, "Mmsense: Multi-person detection and identification via mmwave sensing," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, 2019, pp. 45–50.

[81] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 2994–3009, 2020.

[82] Y. Cheng and Y. Liu, "Person reidentification based on automotive radar point clouds," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[83] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "Mu-id: Multi-user identification through gaits using millimeter wave radios," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 2589–2598.

[84] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor person identification using a low-power fmcw radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.

[85] P. Zhao, C. X. Lu, B. Wang, N. Trigoni, and A. Markham, “Cubelearn: End-to-end learning for human motion recognition from raw mmwave radar signals,” *IEEE Internet of Things Journal*, 2023.

[86] X. Gu, H. Chang, B. Ma, S. Bai, S. Shan, and X. Chen, “Clothes-changing person re-identification with rgb modality only,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1060–1069.

[87] C. Chen, M. Ye, M. Qi, J. Wu, J. Jiang, and C.-W. Lin, “Structure-aware positional transformer for visible-infrared person re-identification,” *IEEE Transactions on Image Processing*, vol. 31, pp. 2352–2364, 2022.

[88] Y. Zhang, Y. Kang, S. Zhao, and J. Shen, “Dual-semantic consistency learning for visible-infrared person re-identification,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1554–1565, 2022.

[89] J. Chen, X. Jiang, F. Wang, J. Zhang, F. Zheng, X. Sun, and W.-S. Zheng, “Learning 3d shape feature for texture-insensitive person re-identification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8146–8155.

[90] T. Li, L. Fan, Y. Yuan, and D. Katabi, “Unsupervised learning for human sensing using radio signals,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 3288–3297.

[91] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.