2012

Automatic classification of human motions using doppler radar

Jingli Li
University of Wollongong

Automatic Classification of Human Motions using Doppler Radar

A thesis submitted in partial fulfillment of the requirements for the award of the degree

Master of Engineering by Research

by

Jingli Li

School of Electrical, Computer and Telecommunications Engineering

UNIVERSITY OF WOLLONGONG

August 2012

Statement of Originality

I, Jingli Li, declare that this thesis, submitted in partial fulfillment of the requirements for the award of Master of Engineering - Research, in the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Jingli LI

August 31, 2012

Contents

List of Figures

List of Tables

Acronyms

1-D	One dimensional
2-D	Two dimensional
2D-PCA	Two-dimensional principal component analysis
2D2-PCA	Two-directional two-dimensional principal component analysis
ANN	Artificial neural network
CW	Continuous wave
CWT	Continuous wavelet transform
DFT	Discrete Fourier transform
FMCW	Frequency modulated continuous wave
HICA	Hierarchical image classification architecture
ICA	Independent component analysis
LDA	Linear Discriminant analysis
LM	Levenberg Marquardt

m-D	Micro-Doppler signature
MDL	Main Doppler Line
MSE	Mean square error
PCA	Principal component analysis
RBF	Radial basis function
RF	Radio frequency
RP	Range profile
STFT	Short-time Fourier transform
SVM	Support vector machine
TFDS	time frequency distribution series
WV	Wigner-Ville distribution

Abstract

Traditionally, the Doppler radar is an effective tool for detecting the position and velocity of a moving target, even in adverse weather conditions and from a long range. In recent years, radar systems have been used to detect and identify targets of interest due to their various advantages. Classification, recognition, and identification of targets and motion kinematics based on micro-Doppler signatures have become an emerging research field with numerous civilian and military applications.

This project investigates the automatic classification system of human motions using a Doppler radar. The radar signals are obtained by using a frequency modulated continuous wave radar to scan moving targets. The short time Fourier transform is used to convert the radar signal into spectrogram to provide the time-varying frequency information. Window function types, sizes and overlapping rate of the short time Fourier transform are explored to provide higher resolution for the spectrogram. Intensity transformation and thresholding techniques are applied on the spectrograms to enhance the weak micro-Doppler signatures and remove the background noise.

To identify the movement of a target using a Doppler radar, extraction and

analysis of prominent micro-Doppler features from the spectrogram are important. Instead of processing the entire spectrograms, local windows are detected to reduce redundant information and provide features that are invariant to the target's speed. Local window alignment method is also investigated since misaligned images produce severe artifacts in scatter matrices of principal component analysis. Based on the local windows, the new two-directional, two-dimensional principal component analysis and GIST methods are performed to obtain feature vectors. The support vector machine with RBF kernel is used to classify the feature vectors into motion types. The proposed two-directional, two-dimensional principal component analysis and GIST approaches achieve classification rate of 97.8% and 98.5%, respectively. To compare with the proposed method, the traditional 1-D PCA and HICA are also tested on the same database, and they reach classification rates of 97.6% and 97.7%, respectively.

Acknowledgments

I would like to express my gratitude to my Parents and husband, who have supported me during my studies and research projects.

I also wish to give the deepest appreciation to my principal supervisor, Dr. Son Lam Phung, for all of his time, guidance, counsel, and technical support.

Special thanks also go to my co-supervisor Dr. Fok Hing Chi Tivive for all his guidance, assistance, and knowledge.

Moreover, I gratefully acknowledge the ongoing support of the staff of the School of Electrical, Computer and Telecommunications Engineering for giving me personal and professional support during my studies at the University of Wollongong.

Finally thanks to my fellow students and friends, who have helped me during my study at the University.

Chapter 1

Introduction

Chapter contents

1.1 Research objectives

Human movement classification based on radar has become an emerging research field with numerous civilian and military applications, such as surveillance, search and rescue, and health care [3, 4, 5]. Most sensors that have been applied to detect and recognize human motions operate only in constrained environments. For instance, among the most common approaches for human movement analysis, visual image sequences are dependent on distance, obstacles, variations in lighting, and deformations of clothing [6]. Furthermore, user privacy is compromised since images contain facial signatures.

Using radar technology to identify human motions provides a solution that avoids most of these problems. When a person is performing a physical activity, different parts of the human body have different movements. The movements of

various body parts give rise to micro-Dopplers, which can be clearly detailed in the time-frequency domain using time-frequency representations. Doppler radar systems based on the micro-Doppler phenomenon have numerous applications. For example, a radar system in smart vehicles can sense pedestrians in difficult environments, such as night times or foggy weather conditions. Unlike cameras, radar signals do not capture human faces and therefore, user privacy is protected. Additionally, radar signal can penetrate through most non-metallic materials, which enables it to detect human heartbeat and breathing for applications involving anti-terrorism and search-and-rescue [7].

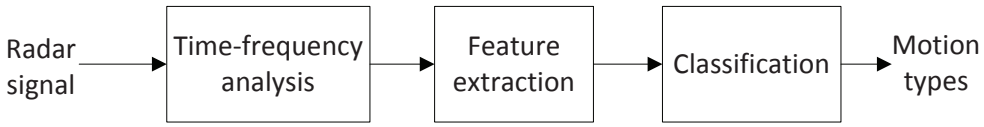


Figure 1.1: Overview of the proposed approach.

The overall goal of this project is to develop a new approach to automatically classify human motions using a Doppler radar. The proposed approach consists of three stages: time-frequency analysis, feature extraction, and classification, as shown in Figure 1.1. Radar signals are obtained by applying a frequency modulated continuous wave radar to scan the target. The first stage uses time-frequency analysis to preprocess the radar signals for revealing the characteristics of human motions. Then, feature extraction techniques are performed to obtain the feature vectors. As extracting salient features is vital to the system performance, developing an efficient and robust method for feature extraction is considered to be important in this project. In the last stage, the extracted features are classified into

corresponding categories using a classifier.

The specific aims of the project are to:

- Investigate algorithms to convert the 1-D radar signals into time-frequency representations, which provide the characteristic micro-Doppler signatures.
- Investigate techniques to enhance the micro-Doppler signatures and remove the background noise.
- Investigate methods to reduce the redundant information and extract features from the time-frequency planes.
- Evaluate the proposed approach and compare it with other existing methods.

1.2 Thesis organization

This thesis consists of six chapters:

- **Chapter 1** outlines the project background and objectives. It highlights the research contributions and publications.
- **Chapter 2** gives a literature review on the Doppler radar system. In this chapter, reviews of several existing feature extraction techniques based on micro-Doppler signatures for target movement classification are presented.
- **Chapter 3** introduces the radar equipment and database, and presents the signal preprocessing steps. The signal preprocessing includes time-frequency analysis, intensity transformation, and intensity thresholding.

- **Chapter 4** describes the proposed method for feature extraction and classification. The two-directional, two-dimensional principal components and GIST features are extracted from the time-frequency representations. Instead of processing the entire time-frequency distribution, features extracted from sliding local windows are proposed. The position, size and alignment of local windows are investigated in this chapter.

- **Chapter 5** presents the experimental results of the proposed feature extraction approaches. Several existing feature extraction methods are tested on the same database to compare with the proposed method.

- **Chapter 6** summarizes the research activities and provides the concluding remarks.

1.3 Contributions

The principal contributions of this thesis are listed as follows.

- We propose a human motion classification approach based on a frequency modulated continuous wave radar for applications in security and surveillance. The proposed approach consists of time-frequency analysis, intensity transformation and thresholding, local window identification and alignment, feature extraction, and classification techniques. The proposed approach is applied to classify three basic human movements: (i) walking normally with both arms swinging, (ii) walking while carrying an object in one hand, and (iii) walking while holding a heavy object with two hands.

- We propose two feature extraction approaches, namely two-directional, two-

dimensional principal component analysis and GIST methods. We build a radar dataset of three basic human motions and compare the proposed feature extraction approaches with several existing methods.

The publications arised from this Masters research project (August 2011 - August 2012) are listed as follows.

- J. Li, S. L. Phung, F. H. C. Tivive, and A. Bouzerdoum, "Automatic Classification of Human Motions using Doppler Radar", in The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1-6.

  Abstract: This paper presents a new approach to classify human motions using a Doppler radar for applications in security and surveillance. Traditionally, the Doppler radar is an effective tool for detecting the position and velocity of a moving target, even in adverse weather conditions and from a long range. In this paper, we are interested in using the Doppler radar to recognize the micro-motions exhibited by people. In the proposed approach, a frequency modulated continuous wave radar is applied to scan the target, and the short-time Fourier transform is used to convert the radar signal into spectrogram. Then, the new two-directional, two-dimensional principal component analysis and linear discriminant analysis are performed to obtain the feature vectors. This approach is more computationally efficient than the traditional principal component analysis. Finally, support vector machines are applied to classify feature vectors into different human motions. Evaluated on a radar data set with three types of motions, the proposed approach has a classification rate of 91.9%.

Chapter 2

Literature Review

Chapter contents

This chapter presents an overview of Doppler radar systems and the existing approaches for classification of target movements using Doppler radar. When the transmitted signal from a radar system is reflected back by a moving target, its carrier frequency is shifted depending on the target's relative velocity; this frequency shift is known as the Doppler effect. In addition, micro motions of the target (such as vibrations, rotations or swings) lead to frequency attenuations in the sidebands of the target's Doppler frequency, producing a micro-Doppler (m-D)

signature. The micro-Doppler signature is an intricate frequency modulation represented in the joint time and frequency domain that characterizes the object's movement distinctively [2].

Based on the micro-Doppler signatures, radar target classification and recognition have many applications including providing health care, countering terrorism, conducting urban military operations, providing urban border security, dealing with hostage situations, and detecting soldiers in dense foliage [3, 4, 5, 8, 9]. The advantages of applying radar technology for human motion classification are listed below:

- detection of position and velocity of a target,
- day and night usage,
- robustness to environmental conditions,
- penetration through most non-metallic materials,
- longer operation distance,
- less privacy violation,
- unobtrusive appearance.

This chapter is organized as follows. In Section 2.1, the mathematical theory of Doppler effect and important radar terminologies are introduced. In Section 2.2, the time-frequency analysis and two examples of micro-Doppler signatures are presented. In Section 2.3, various existing feature extraction approaches for radar based classification are described.

2.1 Overview of Doppler radar system

Radar is an electronic device that is usually used for detection and localization of moving objects based on the well-known Doppler effect. In this section, the mathematical theory of Doppler effect is described and a brief review of two types of radars is given.

2.1.1 Doppler effect

A radar operates by emitting electromagnetic waves and receiving echoes of that signal. The return signal contains a frequency offset in the carrier frequency. Consider the scenario shown in Figure 2.1, the transmitting antenna radiates a continuous wave with a carrier frequency f_c. A portion of the emitted signals is reflected against the target and then received by the antenna. The transmitted signal is denoted as $R_{tx}(t)$ and the received signal is denoted as $R_{rx}(t)$. Suppose that a human target at a distance R is moving with a changing velocity v towards the radar.

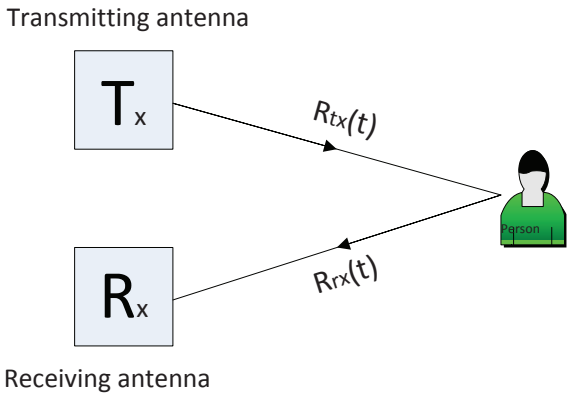


Figure 2.1: A person moving towards the radar.

The transmitted signal from the radar is given by

$$R_{tx}(t) = A\, e^{j(\omega t + \phi_0)}, \tag{2.1}$$

where A is the amplitude, ω is the angular frequency, and ϕ_0 is the initial phase. The reflected signal from the human target is represented as

$$R_{rx}(t) = a(t)\, e^{j[\omega t + \phi_0 + \phi(t)]}, \tag{2.2}$$

where $a(t)$ is the time-varying amplitude and $\phi(t)$ is the time-varying phase change. In a 'mono' radar, the returned signal is $a(t)\ \cos(\omega t + \phi_0 + \phi(t))$ [3]. The parity of the returned signal (cosinus function) prevents a 'mono' radar from determining the direction of the target: whether it is moving towards or backwards the radar. The 'stereo' radar resolves this problem by introducing the Q channel, which is obtained through making a 90 degrees out of phase from the received signal. The received signal is denoted as the I channel.

The total number of wavelengths λ included in the two-way path between the radar and the target is $2R/\lambda$. The unit of the wavelength λ and the distance R is meter. Hence, the total angular excursion $\phi(t)$ during the two-way path is computed as

$$\phi(t) = (2\pi)(2R/\lambda) = 4\pi R/\lambda \text{ radians}. \tag{2.3}$$

As frequency is defined as the change in $\phi(t)$ with respect to time, the Doppler angular frequency is computed by

The f_d denotes the Doppler frequency shift, which is given by

$$f_d = \frac{2v}{\lambda} = \frac{2vf_c}{c}, \tag{2.4}$$

where c is the speed of light and f_c is the carrier frequency of the radar. The frequency shift is positive when the target is moving towards the radar and is negative when the target is moving away from the radar. Equation (2.4) shows that the Doppler frequency increases with the carrier frequency. Therefore, a

better Doppler resolution can be achieved by choosing a higher carrier frequency f_c. The velocity of the target is expressed as

$$v = \frac{f_d c}{2 f_c}. \tag{2.5}$$

2.1.2 Radar types

Radars can be grouped into several categories according to the radar architectures. They differ from each other in the transmission waveforms, bandwidth requirements, implementation complexity, and the way to separate transmission and reception [10]. According to the transmission waveform, there are two main types of radars: continuous wave radar and pulse Doppler radar. These two types of radars are described in the following sections.

2.1.2.1 Continuous wave radar

Continuous wave radar transmits a continuous electromagnetic wave and then receives the echo from reflecting objects. There are two kinds of continuous wave radar, namely unmodulated continuous wave radar and frequency modulated continuous wave radar. In an unmodulated continuous wave radar, the arrival time of the reflected signal cannot be measured. In other words, the location of the target cannot be determined.

Frequency Modulated Continuous Wave (FMCW) radar has been developed to measure both the distance and velocity of moving target. In a FMCW radar, the frequency of the transmitted signal changes as a known function of time. This contributes to the measurement of the target's location. The principle of the FMCW radar is based on the carrier frequency modulation. Typical modulation waveforms are sinusoidal, linear sawtooth, and triangular. Among them, linear

sawtooth and triangular are the most popular. Using the linear sawtooth as shown in Figure 2.2, the frequency of the transmitting signal increases linearly with time. The same linear frequency change is reflected back to the receiver with a delay T_p, which is related to the range of the target. The received echo is mixed with a portion of the transmitted signal and generate a beat signal with frequency f_B, which is proportional to the delay T_p.

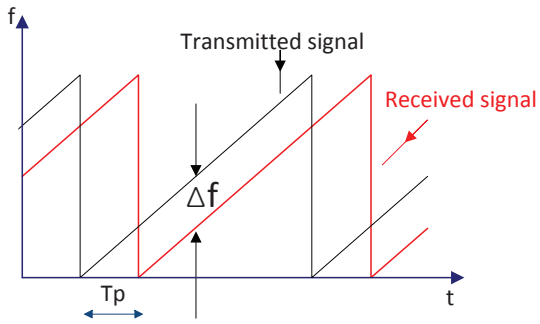


Figure 2.2: Principle of a frequency modulated continuous wave radar based on the linear sawtooth waveform.

With a linear sawtooth modulation waveform, the distance between the target and the radar is calculated as

$$R = \frac{c}{2} T \frac{f_B}{\Delta f}, \tag{2.6}$$

where f_B is the beat frequency, Δf is the frequency deviation, and T is the sawtooth repetition time period.

Although FMCW radar manages to determine the distance of the target, it is unable to differentiate between multiple targets. When more than one target are presented, the received signal of a FMCW radar is the sum of all the reflections from targets at diverse distances. Therefore pulse Doppler radar has been developed to resolve multiple targets.

2.1.2.2 Pulse Doppler radar

In pulse Doppler radars, short pulses are modulated and sent by the transmitting antenna. The arrival time of the reflected echo is evaluated to compute the distance from the radar to the target. Besides, the Doppler shift between the frequency of the emitted signal and the received signal enables the measurement of the speed of the target. The range resolution is the minimum distance at which two targets can be separated. The range resolution is proportional to the bandwidth of the pulse B. For mono-static radars, the range resolution is given by [11]

$$\Delta R = \frac{c}{2B}. \tag{2.7}$$

Equation (2.7) implies that a high bandwidth is required to detect closely spaced targets.

2.2 Micro-Doppler signatures

Generally, a moving target exhibits not only translational motion but also micro-motions, such as vibrations, rotations, and swings[12]. These micro-motions generate sidebands about the Doppler frequency. The frequency modulation induced by micro-motion is referred to as the micro-Doppler phenomenon [13].

Fourier transform has been widely used to analyze the frequency properties of a signal. The basic idea of Fourier transform is that any signal can be decomposed as a superposition of weighted sinusoidal functions at different frequencies. However, Fourier transform is unable to provide information of time-varying frequency [14].

In [1], two common methods were proposed for simultaneously describ-

ing a signal both in time and frequency domains, namely the instantaneous frequency analysis technique [15, 16, 17] and the joint time-frequency analysis technique [18, 19]. The instantaneous frequency analysis applies only on mono-component signals rather than signals with multiple components [1]. Considering that radar signal is comprised of various frequency components, the joint time-frequency analysis is usually used to examine the micro-Doppler shift.

2.2.1 Joint time-frequency signal processing

Most time-frequency transforms can be grouped into either linear or quadratic techniques. Short-time Fourier transform (STFT) and the continuous wavelet transform (CWT) are two well-known linear time-frequency transforms. The Wigner-Ville (WV) distribution and the time-frequency distribution series (TFDS) are two common examples of quadratic transforms. Short-time Fourier transform and the Wigner-Ville distribution are discussed in the following sections.

2.2.1.1 Short-time Fourier transform

Short-time Fourier Transform is a widely used linear transform for converting time-varying signals into spectra-temporal representations. For a discrete signal $x[n]$, the STFT is defined as

$$X[n,k] = \sum_{r=-\infty}^{\infty} x[r]\, w[r-n]\, e^{-j2\pi rk/N}, \quad k = 0, 1, \ldots, N-1 \tag{2.8}$$

where n is the discrete time index, k is the discrete frequency index, N is the size of the required FFT and $w[n]$ is the window function. The idea of STFT is first multiplies the signal need to be transformed with a window function, then takes the Fourier transform of the resulting signal. By sliding the window along with

the signal, time-dependent frequency information are generated. The duration (window size) and overlapping rate of the window are selected to make the signal of interest stable.

As one of the most widely used time-varying spectral representations, the spectrogram of a signal is computed by taking the squared magnitude of the STFT of the signal. In [9, 12, 20, 21, 22, 23, 24, 25, 26], spectrogram are used as the time-frequency representation of radar signals. Micro-Dopplers of human motion revealed by spectrograms are able to determine the human's characteristics, such as size, gender, action, and speed [20]. It should be noted that the window size of spectrogram is a trade-off between frequency resolution and time resolution. A shorter window size provides higher time resolution at the cost of reduced frequency resolution, whereas a longer window size gives better frequency resolution but poorer time resolution.

2.2.1.2 Wigner-Ville distribution

Wigner-Ville (WV) distribution is a well-known quadratic transform that usually used to improve the resolution in the time-frequency domain. The mathematical formulation of the WV distribution in discrete domain is given by

$$W(n,k) = \sum_{m=-\infty}^{\infty} w(m)\, w(-m)\, x(n+m)\, x^*(n-m)\, e^{-j(4\pi/N)mk}, \tag{2.9}$$

where $x(n+m)$ is the signal shifted in time, $*$ indicates the complex conjugate, and $w(m)$ is a window function. However, the Wigner-Ville distribution is not suitable for the time-frequency analysis of multicomponent signals because of the cross-term interference [27]. The WV distribution of two signals' sum is not the

sum of their WV distribution. Given a signal $s = s_1 + s_2$,

$$W_s(n,k) = W_{s_1}(n,k) + W_{s_2}(n,k) + 2Re\{W_{s_1 s_2}(n,k)\}, \tag{2.10}$$

where $Re\{x\}$ represents the real part of x, and $W_s(n,k)$, $W_{s_1}(n,k)$, $W_{s_2}(n,k)$ are the WV distributions of signal s, s_1, and s_2 respectively. The last term of Equation (2.10) is the cross-WV distribution of s_1 and s_2. It is computed as

$$W_{s_1 s_2}(n,k) = \sum_{m=-\infty}^{\infty} w(m)\, w(-m)\, s_1(n+m)\, s_2^*(n-m)\, e^{-j(4\pi/N)mk}. \tag{2.11}$$

This term is the cross-term that reflects the correlation of two signal components. If a signal consists of two or more components, cross-terms will be displayed in its WV distribution. Thus, WV distribution is not applicable for radar signals with multi-components.

2.2.2 The micro-Doppler effect of a rigid body motion

In this section, we discuss the mathematical description of micro-Doppler produced by rotational motions of a rigid body. A rigid body is a non-deformable solid body. In other words, the distance between two points of a rigid body remains constant at all times. In coherent radar systems, the returned signal has a phase change caused by the rotations. It is possible to measure the rotation of a reflecting surface according to its corresponding phase change. Thus, the Doppler frequency shift, which represents the change of phase as a function of time, can be used to detect rotations of structures attached to the target [28].

From the mathematical model of micro-Doppler effects given in [8], a target can be represented as a set of point scatterers, where each point scatter produces a micro-Doppler shift. Considering one rotational point scatter in the returned

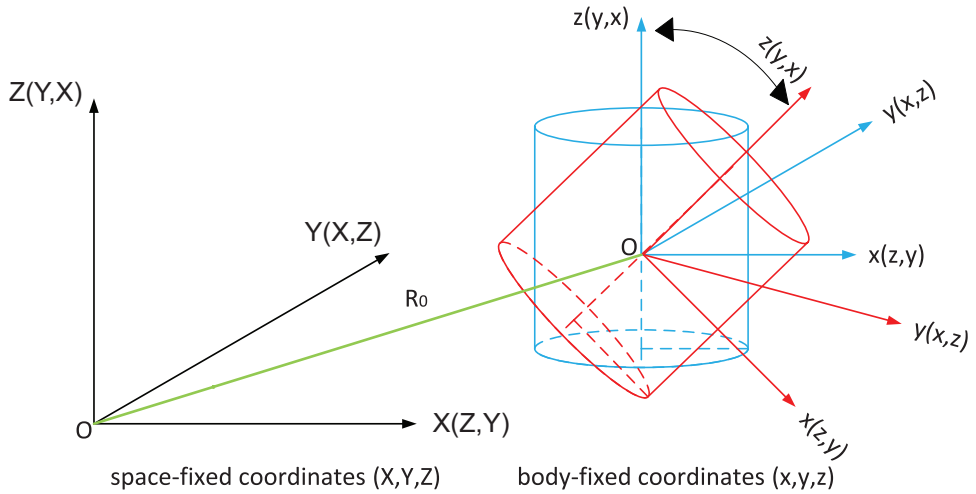


Figure 2.3: A rotating target [1].

radar signal, two coordinate systems are used to represent the motion of a rotating point scatter target, as shown in Figure 2.3. One is the space-fixed system (X, Y, Z) and the other is the body-fixed system (x, y, z), which is fixed to the target. Let the center of mass (CM) of the target be the origin of the body-fixed system. The distance between the origins of the space-fixed system and the body-fixed system is denoted as vector $\mathbf{R}_0$. The orientation of the axes of the body-fixed system that is relative to the axes of the space-fixed system is given by three independent angles. Denoting $\mathbf{r}$ as the position of a point P in the body-fixed system, the position of P in the space-fixed system is defined as $(\mathbf{R}_0 + \mathbf{r})$. Then, its velocity v_P is given by

$$v_P = \frac{d}{dt}(\mathbf{R}_0 + \mathbf{r}) = V + \Xi \times \mathbf{r}, \tag{2.12}$$

where V is the translation velocity of the target center, and Ξ is the angular velocity of the target rotation

Equation (2.12) shows a target motion is represented by both the translational motion and the micro-motion.

If the radar transmits a sinusoidal waveform with a carrier frequency f_c, the signal returned from the point scatterer P is a function of the range between the

radar and the point P, which is represented by $R = |\mathbf{R}_0 + \mathbf{r}|$:

$$s(t) = \rho(x, y, z) \exp\{j2\pi f_c \frac{2R}{c}\} = \rho(x, y, z) \ \exp\{j\phi(t)\}, \tag{2.13}$$

where $\rho(x, y, z)$ is the reflectivity function of the point scatter P. By taking the time derivative of the phase, the Doppler frequency shift induced by the target motion can be derived as

$$\begin{aligned} f_d &= \frac{1}{2\pi}\frac{d\phi(t)}{dt} && (2.14)\\ &= \frac{2f_c}{c}\frac{d}{dt}R && (2.15)\\ &= \frac{2f_c}{c}\ \frac{d}{dt}\{(\mathbf{R}_0 + \mathbf{r})^T \cdot \mathbf{n}\} && (2.16)\\ &= \frac{2f_c}{c}[V + \mathbf{\Xi} \times \mathbf{r}]^T \cdot \mathbf{n}, && (2.17) \end{aligned}$$

where $\mathbf{n} = (\mathbf{R}/|\mathbf{R}|)$ is the unit vector of the radial velocity for defining the direction of the radar line of sight. The Doppler frequency shift is then written as

$$\begin{aligned} f_D &= \frac{2f_c}{c}[V + \mathbf{\Xi} \times \mathbf{r}] && (2.18)\\ &= \frac{2f_c}{c}V + \frac{2f_c}{c}\ \mathbf{\Xi} \times \mathbf{r} && (2.19)\\ &= f_t + f_r && (2.20) \end{aligned}$$

where f_t is the Doppler shift caused by the translation, and f_r is the micro-Doppler due to the rotation. Therefore, the Doppler frequency shift is the sum of the frequency shift caused by translational motion and rotational motion.

2.2.3 The micro-Doppler effect of a nonrigid body motion

The nonrigid body is a deformable body. During the movement of a nonrigid body, the distance between two points of the body and the body shape could change. When analyzing the micro-Doppler effect caused by a nonrigid body

motion, the body can be regarded as several jointly connected rigid segments. In other words, we can treat a nonrigid body motion as the motion of multiple rigid bodies [2].

One of the important nonrigid motions is the human motion. Human motion is an articulated locomotion, the motion of limbs of the human bodies can be characterized by repeated periodic movements.

Among all the human movements, walking is a typical example of nonrigid body motion. In a walking motion, each foot moves from one position to the next position. The arms and legs swing periodically and the body's center of gravity moves up and down while walking. Although human walk has a general pattern, the individual human gait varies slightly from each other. This is why people can be recognized by the walking style [29].

Micro-Doppler signature of human gait displays composite Doppler spectra that include several components. When a person is moving, different body parts such as the torso, hands, and legs have different movements and velocities. The features of Doppler components caused by different body parts can be used to identify their movements [3].

Most methods on human movement modeling divide the human body into several body parts and model the velocity profile of these parts individually. In [2], the simulation of radar backscattering from a walking human is given. Figure 2.4 shows that different body parts have different speeds and induce additional frequency modulations on the returned radar signal. The micro-Doppler phenomenon can be used to identify different human motions, such as two-arms swing, no arm swing, running, skipping, punching, or kicking.

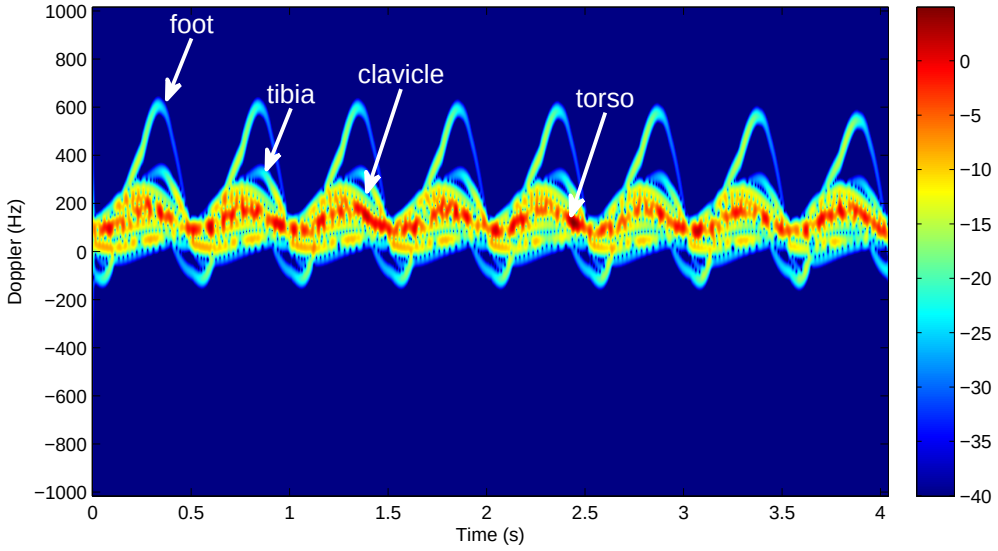


Figure 2.4: Simulation of the micro-Doppler signature of human walking [2].

2.3 Motion recognition using Doppler radar

Micro-Doppler features are regarded as a particular signature of an object that enables identification of motion dynamics and recognition of the objects [12]. In recent years, micro-Doppler signatures have been used to detect and identify targets of interest. Classification, and recognition of targets and motion kinematics of a target based on micro-Doppler signatures are important research that has been investigated [2]. To identify the movement of a target using a Doppler radar, it is essential to extract and analyze prominent micro-Doppler features. Existing approaches of feature extraction for radar based classification can be categorized into four main groups:

- parametric approaches,
- nonparametric approaches,
- envelope detection approaches,

- discriminant analysis approach.

2.3.1 Parametric approaches

For the parametric approach, specific parameters characterizing the micro-Doppler are extracted and sent to a classifier. The parameters are extracted from time-frequency distributions that reveal the prominent micro-Doppler features. Using selected parameters instead of the whole time-frequency distributions reduces the data and processing time.

Kim and Ling developed a parametric approach to distinguish seven human activities: running, walking, walking without moving arms, crawling, boxing, boxing while moving forward, and sitting relatively still [24, 25]. Radar signals of twelve human subjects are collected in the laboratory using a Doppler radar. STFT is used to convert the radar signals into spectrograms.

From the spectrograms of the seven activities, six Doppler features are extracted from the spectrogram: (1) the torso Doppler frequency, (2) the total bandwidth (BW) of the Doppler signal, (3) the offset of the total Doppler, (4) the bandwidth without micro-Dopplers, (5) the normalized standard deviation of the Doppler signal strength, and (6) the period of the limb motion. These features include information of the subject's torso speed, limb motions speed, bobbing motion of torso, and the swing rates of the limbs.

The feature (1) is calculated as the mean frequency value of the peak signal over time bins in one window. To compute features (2), (3), (4), and (6), two envelopes called high frequency envelope (E_H) and low frequency envelope (E_L) are utilized. Envelope E_H is formed by the highest Doppler frequency at each

time instance, while envelope E_L is formed by the lowest Doppler frequency. The feature (2) is obtained by averaging the difference between the biggest frequencies of the envelope E_L and the smallest frequencies in the envelops E_L. Feature (3) is computed as the mean value between the biggest frequencies of the envelope E_L and the smallest frequencies in the envelops E_L. Feature (4) is determined by averaging the difference between the smallest frequencies of envelope E_H and the largest frequencies from the envelope E_L. Feature (5) is generated by dividing the standard deviation of the intensity by the mean of the signal intensity of all the signals in the spectrogram. Feature (6) is the time period of the limbs' micro-Doppler signatures. Using these six micro-Doppler features, artificial neural network (ANN) and support vector machine (SVM) are applied to classify human activities.

Lei developed a parametric approach to classify four types of time-frequency distribution of micro dynamics including vibration, rotation, coning, and tumbling [22]. Simulated time-frequency distribution of the four dynamics are used. Through observation and analysis of the characteristics of the simulated time-frequency distributions, the main features that can be used to distinguish the four dynamics are determined. The features are chosen as symmetry, degree of slope, and availability of middle line.

Feature vectors are obtained by extracting the three features from each time-frequency distribution. Then, feature vectors are sent to five different classifiers, namely, Bayes linear classifier, quadratic classifier, support vector classifier, K-nearest neighbor rule classifier and neural network classifier. The experiment result shows that the method obtains high classification rates for different classi-

fiers both in training and testing stages.

The effectiveness of parametric approach depends on the chosen parameters. The approach can get promising result when representative parameters are selected. However, it may provide poor performance if there are errors in the selection of one or more parameters.

2.3.2 Nonparametric approaches

Instead of choosing a fixed set of features, the nonparametric approach uses portions or segments of the time-frequency distributions as features. Tivive and his colleagues developed a hierarchical classification architecture that extracts micro-Doppler features through learning [9]. Radar data of five persons are used in their approach. The STFT is adopted to convert radar signals into spectrograms. An image-based classification technique is applied on the spectrograms to classify three types of human walking motion: free-arm motion (two-arm swing), partial-arm motion (one-arm swing), and no-arm motion.

Their classification method has three processing stages. The first stage uses a set of directional nonlinear filters to extract motion energy and directional contrast from the spectrogram. The second stage comprises adaptive nonlinear filters for learning intrinsic features characterizing different classes of arm motions. The last stage uses linear SVM for classification. The coefficients of the adaptive filters are obtained through a training process.

The training process is based on the Levenberg-Marquardt (LM) algorithm [30]. Considering a training set of B input patterns $I_1, I_2, \cdots, I_B$. The desired output corresponding to each input is $c_1, c_2, \cdots, c_B$, where c_i is the desired output of the ith

input. The coefficients of the adaptive filter are obtained by minimizing the mean square error (MSE) between the desired outputs and the actual outputs. The MSE is calculated by

$$E_{mse} = \frac{1}{BQ} \sum_{b=1}^{B} \sum_{q=1}^{Q} (c_b{}^q - y_b{}^q)^2, \tag{2.21}$$

where $c_b{}^q$ and $y_b{}^q$ are the qth element of the desired output vector c_b and the actual output y_b, respectively, and Q is the number of classes. The LM algorithm is adopted to learn the optimum parameters of the adaptive filter.

Lyonnet and his colleagues proposed a time-frequency classifier to classify three types of human walking motion: one-arm swing, two-arms swing and no-arm swing [31]. Quadratic transform is used to convert radar signals into time-frequency representations. For each motion, the micro-Doppler signature of the entire time-frequency domain is employed in providing the distance measure. A distance-based classification method is used to measure the difference between the test data time-frequency distribution and the training average time-frequency distribution. Experiment results show that this classifier has a low probability of classification errors.

The nonparametric method generates satisfactory classification performance at the cost of a long training time. However, it does not utilize the periodic and evolving nature of the human gait in the classification process. These information can be used to increase the classification accuracy and reduce the classification time.

2.3.3 Envelope detection approaches

Orovic and his colleagues proposed an envelope detection approach to classify the human gait using distinction in the arm motions of the walking persons [32]. The proposed approach is developed to classify three types of motions: free-arm motion (two-arm swing), partial-arm motion (one-arm swing), and no-arm motion.

Time-frequency analysis is performed on the radar signals by using the multi-window S-method (also called Hermite S-method). To reveal the weak arm movements, support functions are defined to remove from the time-frequency distribution the main motion (such as torso and leg motions) and other relative strong components. The remaining pixels within the time-frequency distribution are divided into two sets: one contains positive frequency points and the other contains negative frequency points. The former are points that are above the main trajectory and the latter are points that are below the main trajectory. A smoothing technique is applied on the two sets of points to obtain two curves, called upper and lower envelopes. The upper and lower envelopes are described respectively by the positive and negative frequency points, with respect to the main trajectory.

Then the features are extracted based on the envelopes of the arms in the time-frequency plane. The distance between the two envelopes is used to determine if the person is walking with or without arm swing. The distance between the envelopes of walking with arm-swing is larger than the walking without arm-swing. The shape of the envelopes is used to differentiate a person walking with one-arm or two-arm swing. For a person walking with two-arm swing, the upper and lower envelops are symmetrical across the torso motion component.

The envelope based approach utilizes the periodic information of the human gait. However, the performance of this method depends on how accurate the envelope is extracted.

2.3.4 Discriminant analysis approaches

There are several discriminant analysis approaches for feature extraction, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) [33]. Among them, PCA is the best-known unsupervised linear algorithm for feature extraction. It is a linear mapping that reduces the dimensionality of training samples by using the eigenvectors with the largest eigenvalues.

In [34], the feature vectors are obtained through Fourier transform and PCA. The Thales Man Portable Surveillance and Tracking Radar is used to collect radar data for a wheeled vehicle, a tracked vehicle, and personnel. The radar data is preprocessed to remove the clutter and background noise. Each preprocessed signal is divided into short frames. The Discrete Fourier Transform (DFT) is used to calculate the power spectrum of each frame. The power spectrum is normalized by applying a circular shift to move the Main Doppler Line (MDL) to the center frequency bin and is converted in decibel (dB) to reduce the variation in the value of each frequency bin.

The PCA is applied to reduce the dimension of the feature vector. The directions of maximum variance based on the training set is estimated by solving the following equation: $\mathbf{C}P = \lambda P$, where $\mathbf{C}$ is the covariance matrix calculated from the normalized power spectra, P is a matrix where each column is an eigenvector

of $\mathbf{C}$ and λ is an eigenvalue. A Bayes classifier is used to classify each frame into three classes.

In [23], three types of human walking motions: free-arm motion (two-arm swing), partial-arm motion (one-arm swing), and no-arm motion are classified. The spectrogram is used as the feature extraction domain. The training samples are time or frequency slices of the spectrograms. A time slice is a snapshot of the spectrogram capturing Doppler frequencies at a given time point. A frequency slice is a snapshot of the spectrogram at one Doppler frequency for all the time points. Then PCA is performed on the snapshots to obtain feature vectors. Then the feature vectors are sent to a minimum distance classifier to differentiate the three human motions.

The shortcoming of PCA is that when the spectrograms have high dimensions, a large covariance matrix is generated. This leads to a time-consuming computation of the covariance matrix and its eigenvectors.

In [35], a classification method is proposed using a combination of range profile (RP) and time-frequency image. The RP is obtained by taking the absolute value of range-compressed radar signal. The time-frequency image is obtained by performing the STFT on the range-compressed radar signal. Two-dimensional principal component analysis (2D-PCA) is applied on the time-frequency image to reduce the redundant information and to extract useful features. The time-frequency image is denoted as matrix $\mathbf{A}$. Unlike PCA, the matrix $\mathbf{A}$ does not need to be converted into vectors, it is directly projected onto a matrix $\mathbf{X}$ to obtain the feature matrix $\mathbf{Y}$.

$$\mathbf{Y} = \mathbf{AX}. \tag{2.22}$$

The projection matrix **X** is obtained by maximizing the generalized total scatter criterion. After the projection, a new dimension reduced matrix is obtained. A combination of the RPs and feature matrix **Y** is used as the extracted features to the nearest neighbor classifier. Although 2D-PCA is more efficient in computation, it demands more coefficients to represent a feature matrix.

2.4 Chapter summary

Significant research has been conducted for radar classification system. The weakness of existing systems are summarized as follows:

- Existing radar systems for human movement classification are based on simulated data. Most of the systems use radar data collected in an indoor environment. Analyzing long-range radar signals obtained in an outdoor environments is a challenging task.

- Several radar classification systems are based on observation of the micro-Doppler signatures. The micro-Doppler signatures from different motions are observed to identify a set of parameters that reveal the most representative characteristics. Errors in the selection of one or more parameters lead to a poor classification rate.

- The radar classification systems based on envelope detection demand a time-frequency representation with high resolution. Besides, these systems generate poor classification result when the motions have similar envelopes.

- The radar classification systems based on nonparametric analysis usually involve a long training time. The features processed in these systems consist

of redundant information, which increases the dimension of the signals.

- Although the radar classification systems based on discriminant analysis reduce the dimension of the feature vector, they suffer from long computational times and large memory requirements.

Important research directions in radar classification systems can be listed as follows:

- Building a radar dataset of long-range human motions for outdoor environments.
- Investigating time-frequency analysis technique to provide enough resolution for the feature extraction and classification.
- Proposing feature extraction algorithms that yield feature vectors of relatively small sizes. The algorithm should provide high classification rate within a relatively short processing time.

Chapter 3

Radar Signal Acquisition and Preprocessing

Chapter contents

3.1 Doppler radar data

In this project, we aim to investigate Doppler radar signals for different motions in outdoor environments. The outdoor environment not only enables long range recording but also provides more realistic radar signals. The radar data were collected at University of Wollongong, Australia, in outdoor environments. We

used a FMCW radar operating at 24 GHz to scan moving targets (people). It is a 'stereo' radar that generates both in-phase and quadrature signals, i.e., $I(t)$ and $Q(t)$.

3.1.1 Radar equipment

The FMCW radar is shown in Figure 3.1. The circuit board (green color) is a ST200 evaluation kit. It is a data acquisition and processing system, which transmits and receives radar signals of radio frequency (RF). The radar transceiver is attached to the ST200 board, which is linked to the computer via the USB cable. With the software installed on the computer, the ST200 board is able to transfer radar signals to the computer.

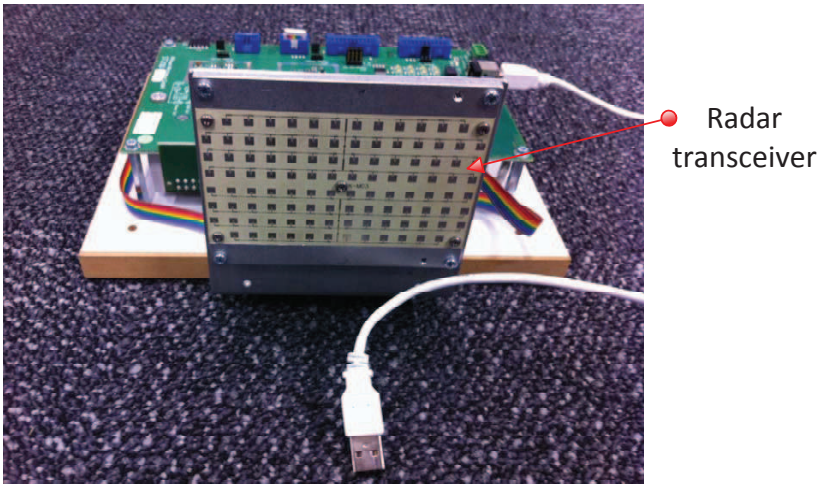


Figure 3.1: The FMCW radar equipment.

The operation range of the radar transceiver is from 1 to 70 meters. The radar transceiver consists of an RF low noise amplifier and two Intermediate Frequency (IF) preamplifiers for both I and Q channels. This reduces the need for external analogue electronics. The board is very small and easy to carry since it has the size of 10.5 by 8.5 cm. The unique function of wakeup within $4\mu s$ makes this

module suitable for battery-operated surveillance systems.

3.1.2 Radar signal database

The radar signal database comprises 20 subjects, including 8 females and 12 males. Each subject performed three types of motions: (i) walking normally with both arms swinging (2-AM), (ii) walking while carrying an object in one hand (1-AM), and (iii) walking while holding a heavy object with two hands (0-AM), see Figure 3.2. Recognizing these types of human motions has useful applications in surveillance and security.

(a) (b) (c)

Figure 3.2: Photos taken during radar signal acquisition: (a) normal walking; (b) walking while carrying an object in one hand; (c) walking while holding an object with two hands.

Each subject was walking towards the radar along the line of sight. The distance between the subject and the radar varied from 1 to 55 meters. Radar signal was recorded for 30 seconds and digitized at a sampling rate of 125 KHz. Table 3.1 lists the number of collected radar traces. Each subject repeated the three types of motions three times. Thus, for each type of motion, we have 60 radar signals in total.

For each radar signal, the first and the last 2 seconds are discarded to remove the acceleration and deceleration of the subject. In other words, each signal

Table 3.1: Summary of the radar data collected.

	0-AM	1-AM	2-AM
Trial 1	20	20	20
Trial 2	20	20	20
Trial 3	20	20	20
Total signals	60	60	60

represents a human walking at a relatively constant speed. As the frequency of the micro-Doppler is only a few hundred hertz, the radar signal is down-sampled by a factor of 16, which leads to a new sampling rate of 7812 samples/second. This reduces the amount of data to be processed. For both I and Q channels, the radar signals are subtracted by the mean value to remove the DC bias.

The measured radar signal contains background noise that has to be removed before time-frequency analysis. The radar returns from stationary background objects are normally near zero Doppler frequency with a small bandwidth, whereas the returns from moving targets are offset from the zero Doppler frequency due to their radial velocities [2]. Therefore, a notch filter can be used to suppress the background clutter. In this project, a band-reject filter with a notch around zero is applied to suppress the background noise.

3.2 Time-frequency analysis

To capture the time-varying micro-motions of the targets, the measured complex-valued signal is converted into spectrogram. Spectrogram has been widely used to describe the dynamics of frequency contents of a signal over time. The spectrogram $S(n,k)$ of a 1-D discrete signal $x[n]$ is computed by taking the squared

magnitude of its short-time Fourier transform:

$$S(n,k) = |X[n,k]|^2. \tag{3.1}$$

For a signal $x[n]$, the STFT is defined as

$$X[n,k] = \sum_{r=-\infty}^{\infty} x[r]\, w[r-n]\, e^{-j2\pi rk/N}, \quad k = 0, 1, \ldots, N-1, \tag{3.2}$$

where n is the discrete time index, k is the discrete frequency index, N is the FFT size, and $w[n]$ is a window function. The basic idea of STFT is first multiplies the signal $x[r]$ with window function $w[n]$ and then computes the Fourier transform of the product $x[r]\, w[r-n]$. Because of the short time duration of the window function, the Fourier transform of $x[r]\, w[r-n]$ reflects the signal's local frequency properties. By sliding $w[n]$ and repeating the aforementioned process, we know how the signal's frequency contents change over time.

The width of the window function $w[n]$ governs the resulting time and frequency resolutions. A shorter $w[n]$ contributes to a better time resolution and a worse frequency resolution, and vice-versa. The windows could be overlapped or disjointed. A bigger overlap of windows leads to a smoothed spectrogram with higher time resolution. Higher frequency resolution is achieved by zero padding the windows.

Figure 3.3 presents the spectrograms for three types of human motions: normal walking, walking while carrying an object in one hand, and walking while holding an object with two hands. The spine of each spectrogram represents the main translational motion of the person. In this case, the spectrogram has only positive Doppler as the person is moving towards the radar. Negative Doppler shifts caused by background noise may also exists in the spectrogram. In one gait cycle,

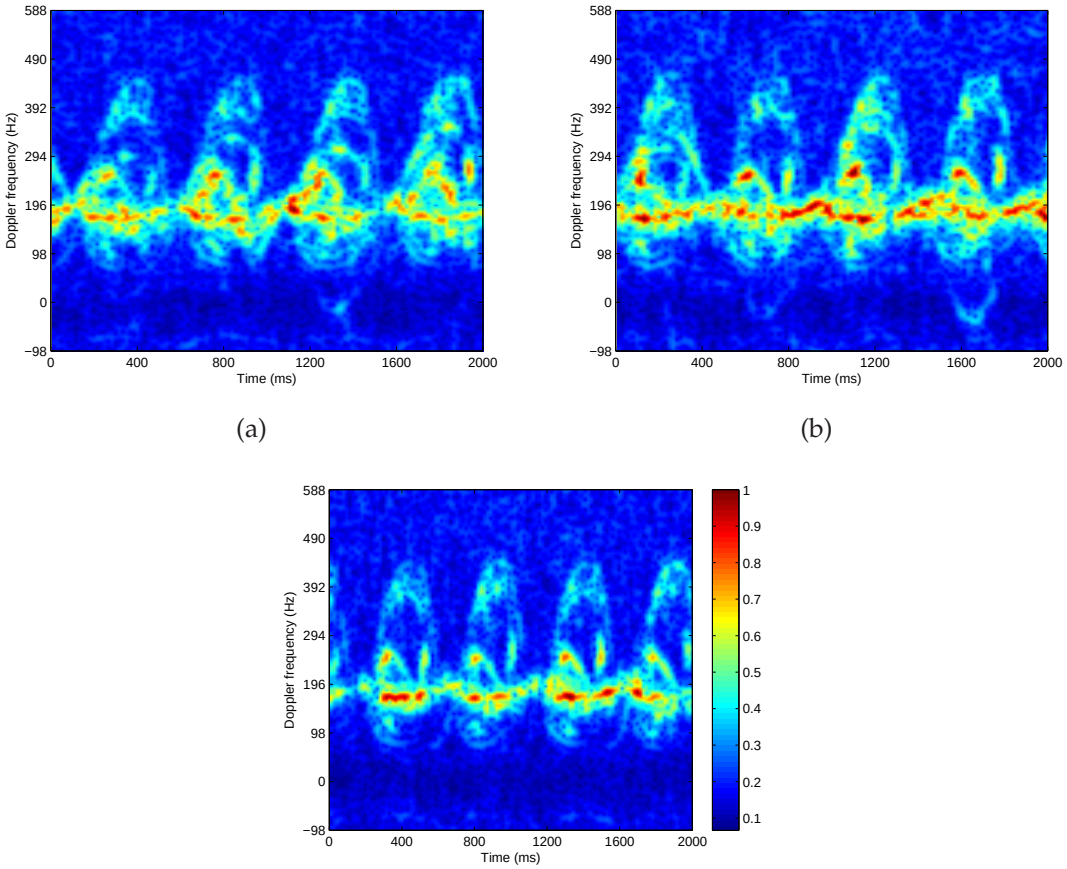


Figure 3.3: Spectrograms of three different types of human motions: (a) normal walking; (b) walking while carrying an object in one hand; (c) walking while holding an object with two hands. See the electronic color figure.

the arm movements cause a positive and a negative frequency shift while the foot movements produce only positive Doppler. The periodic peaks denote the arms, legs, and feet motions of a walking person. These spectrograms show that different human motions exhibit different micro-Doppler signatures, which can be utilized for human movement classification.

3.2.1 Type of window function

For a window function, all the values outside of a chosen interval are zero or go rapidly toward zero [36, 37]. Various window functions have been developed for

spectral analysis and filter design. The selection of window function types is a trade-off of the spectrogram. A smoothing window function reduces the spectral leakage at the cost of lower resolution of the spectrogram. The spectral leakage distorts the spectrogram by spreading energy from a given frequency component to adjacent frequency lines or bins [38]. Thus, choosing a window function that is able to suppress the spectral leakage is very important. In the following sections, several common window functions are presented.

As the simplest window, the rectangular window maintains the best resolution but has a high spectral leakage. Figure 3.4 shows the rectangular window function with a width of 64. For our situation, a high resolution is needed to reduce the blurring of the backscatters from different body parts. On the other hand, less leakage is also desirable because we do not want the signal intensities to be overestimated. Therefore, a smoother window function need to be explored.

The Hanning window has good frequency resolution and less spectral leakage [38, 39]. The Hanning window has the form $w[n] = 0.5 - 0.5\cos(\frac{2\pi n}{N-1})$ for n between 0 and $N-1$, as shown in Figure 3.5. Although the Hanning window has various strengths, it performs poorly when separating spectral components that are closely-spaced.

The Hamming window has a similar form to the Hanning window function: $w[n] = 0.538 - 0.462\cos(\frac{2\pi n}{N-1})$ for n between 0 and $N-1$ (see Figure 3.6). Hamming window is a well-known window function with an good balance of spectral leakage and resolution [10]. In addition, it gives a better performance than the Hanning window at separating a spectral component of small magnitude near to a large spectral component.

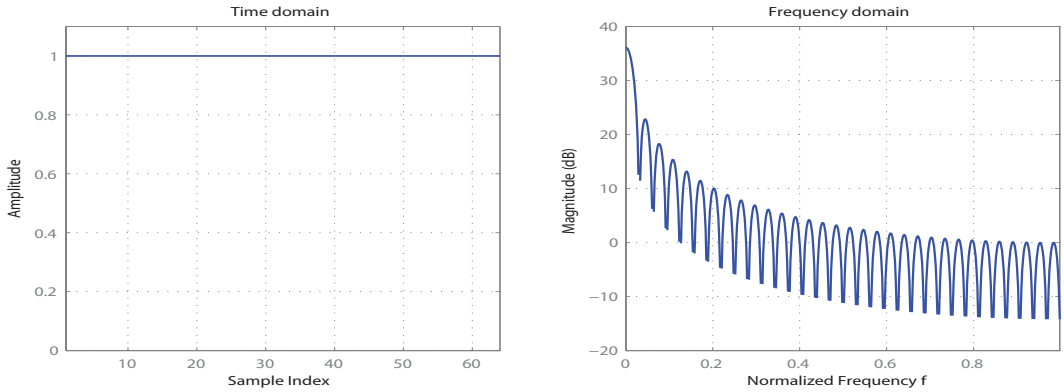


Figure 3.4: The rectangular window.

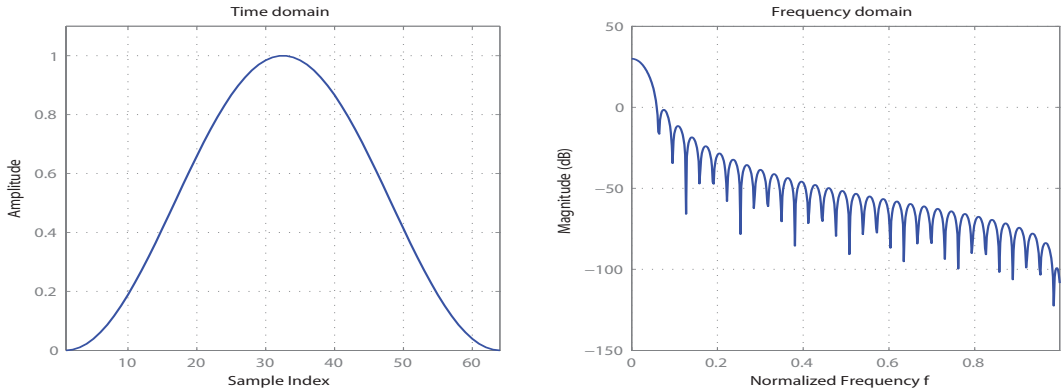


Figure 3.5: The Hanning window $w[n] = 0.5 - 0.5\cos(\frac{2\pi n}{N-1})$.

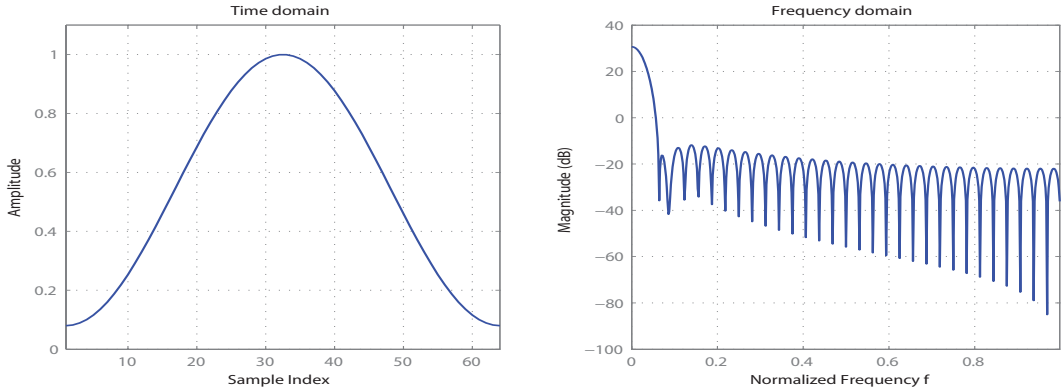
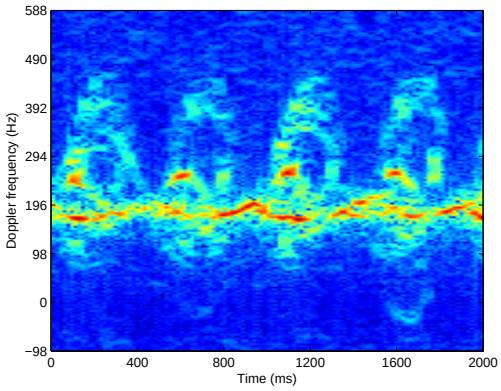
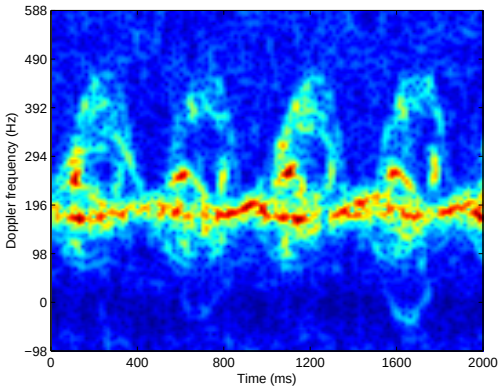


Figure 3.6: The Hamming window $w[n] = 0.538 - 0.462\cos(\frac{2\pi n}{N-1})$.

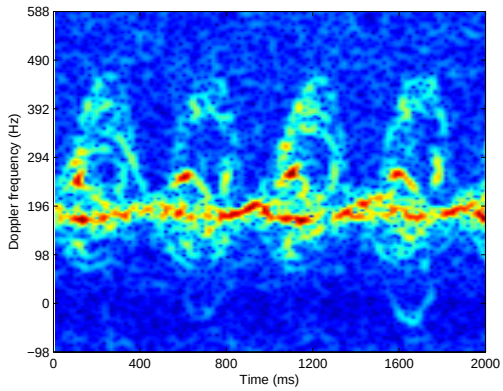
Figure 3.7 displays the spectrograms created using rectangular window, Hanning and Hamming window. The resolution of spectrogram obtained using rectangular window is the highest. However, the envelopes of the micro-Doppler signatures in Figure 3.7(a) are not as clear as Figure 3.7(c) and Figure 3.7(b)



(a) Using rectangular window



(b) Using Hanning window



(c) Using Hamming window

Figure 3.7: Spectrograms of a human gait using rectangular window, Hanning window and Hamming window.

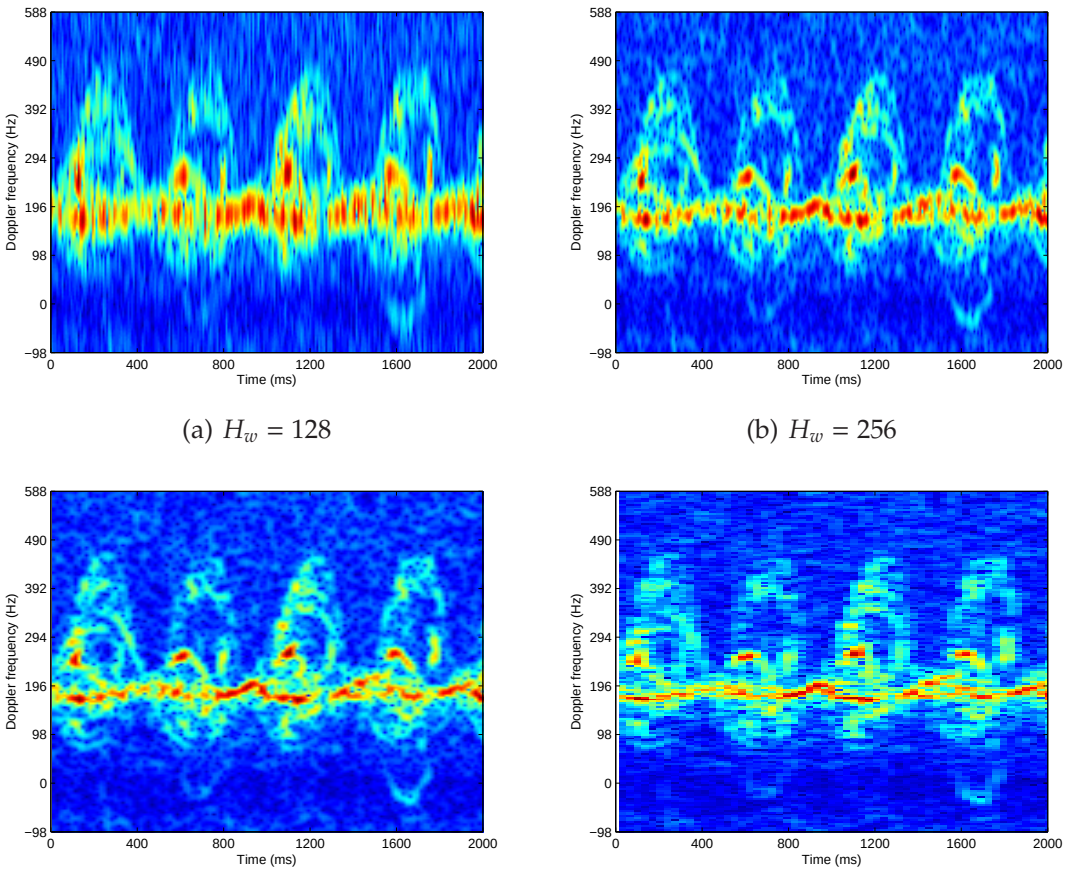
because of the discontinuities of the rectangular window. The micro-Doppler signatures obtained using Hanning window and Hamming window do not differ from each other too much. However, with careful comparison, some regions of Figure 3.7(b) look more blurring than Figure 3.7(c). Therefore, the Hamming window is adopted in the proposed method.

3.2.2 Size of window function

An experiment is conducted to investigate the effects of using different window widths to compute the spectrogram. A Hamming window is used in this experiment. The Hamming window width H_w ranges from 128 to 1024, i.e. the time

duration of the STFT window is from 16.4 ms to 131.1 ms with the sampling rate of 7812 samples/second.

Figure 3.8 shows that a wide window leads to a better frequency resolution and a worse time resolution, whereas a narrow window leads to a better time resolution and a worse frequency resolution. This means a balance of time and frequency resolution is required to obtain a good spectrogram resolution. The analysis of spectrograms in the following sections uses $H_w = 512$ only for visualization. In Section 5.1.1, experiments will be conducted to find the window size that leads to the best classification performance.



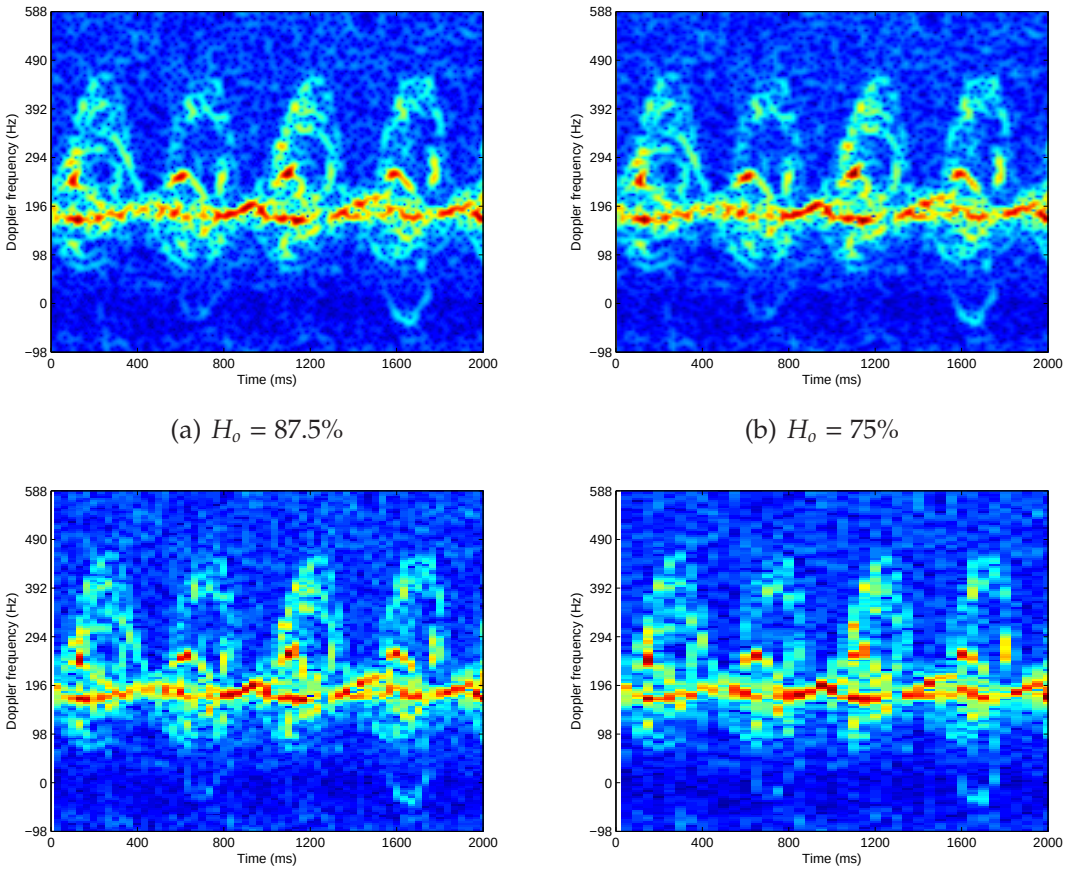
(a) $H_w = 128$ (b) $H_w = 256$

(c) $H_w = 512$ (d) $H_w = 1024$

Figure 3.8: Spectrograms of a human gait using different window sizes.

3.2.3 Overlapping of window function

When we perform STFT, the sliding windows can be overlapped or disjointed. Zero padding windows produces a smooth spectrogram in the frequency axes because it increases the number of frequency points. To reduce the computational complexity, we do not use zero padding windows. If a higher overlap between windows is adopted, a smooth spectrogram in the time axis is generated as more time points are added to the spectrogram.



(a) H_o = 87.5%

(b) H_o = 75%

(c) H_o = 50%

(d) H_o = 25%

Figure 3.9: Spectrograms of a human gait using different window overlaps.

For a Hamming window width of H_w = 512, different values of window overlap are adopted to examine the spectrogram. The values of overlapping rate H_o examined here are 87.5%, 75%, 50%, and 25%.

Figure 3.9 shows that increasing the overlap of successive Hamming windows leads to a spectrogram with higher time resolution. We should note that a bigger overlap also increases the size of the spectrogram. This way, the amount of data to be processed will be increased. For each value of overlap H_o, the number of frequency points in the spectrogram is 1024. The number of time points to represent a 2-second signal with overlapping of 87.5%, 75%, 50% and 25% are 237, 119, 60 and 40 respectively. To achieve a reasonable resolution while maintaining an appropriate size of the spectrogram, we choose 50% as the overlap of the Hamming window.

3.3 Spectrogram processing

The spectrograms in Section 3.2 are obtained by intensity normalization and intensity transformation for better visualization. In these spectrograms, we can clearly see the micro-Doppler signatures induced by the limb motions. Figure 3.10 displays an original spectrogram in which the micro-Doppler is very weak, the intensities of pixels in the spectrogram vary from 9.6274×10^{-7} to 0.0353. Here, we consider spectrograms as 2-D images so that image processing techniques can be used for enhancing the weak micro-Doppler signature. Intensity transformation is one of the most common methods to increase the contrast between the target regions and background [40]. The intensity transformation works directly on pixels of an image.

Although a notch filter is applied to the radar signals for removing the background clutter, there is still background noise in the spectrograms. Thus, noise removal techniques need to be used. Intensity thresholding is one of the most

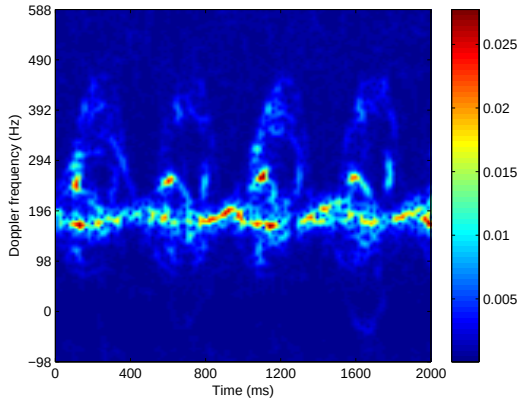


Figure 3.10: The original spectrogram without enhancement.

popular methods of background noise removal because it is intuitive, fast, and easy to implement [41]. In the following sections, intensity transformations and intensity thresholding are discussed.

3.3.1 Intensity transformation

In this section, three different intensity transformations are discussed. Firstly, the intensities of the spectrograms are normalized to the range $[0, 1]$ by dividing the maximum value, and then an intensity transform technique is used. A pixel transformation can be described as $b = T(a)$, where a is the original pixel value, b is the new pixel value and $a, b \in [0, U-1]$.

3.3.1.1 Gamma transformation

Gamma transformation is selected because it is a simple and effective technique for contrast enhancement. It is widely used in various devices for image capture, enhancement, and display. Gamma transformation is defined by the following power-law expression:

$$b = ca^{\gamma}, \tag{3.3}$$

where $c = (U-1)^{1-\gamma}$ and γ are positive constants, the input and output values are

non-negative real values. Plots of b versus a for several values of γ are shown in Figure 3.11.

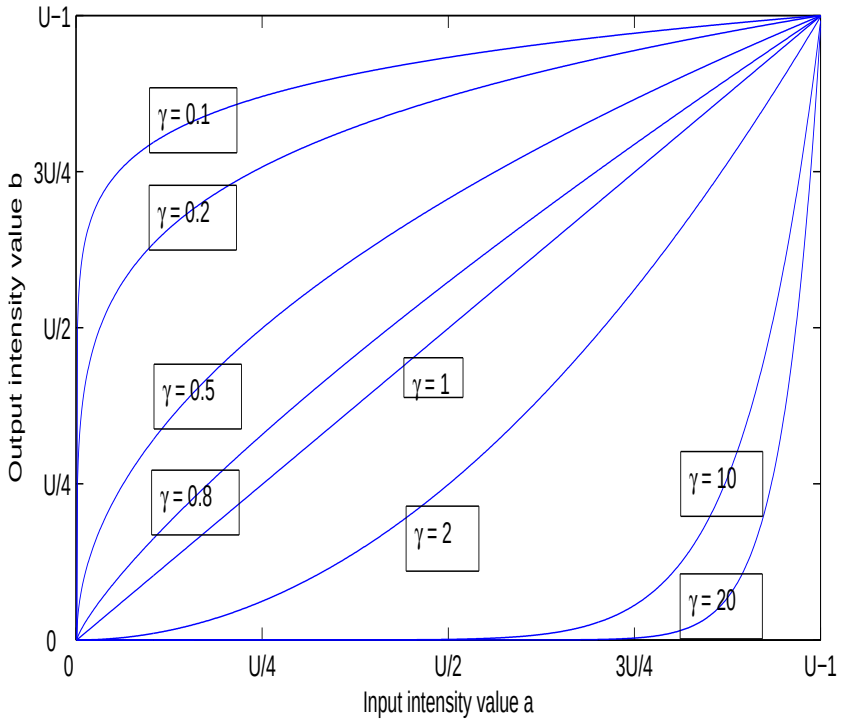
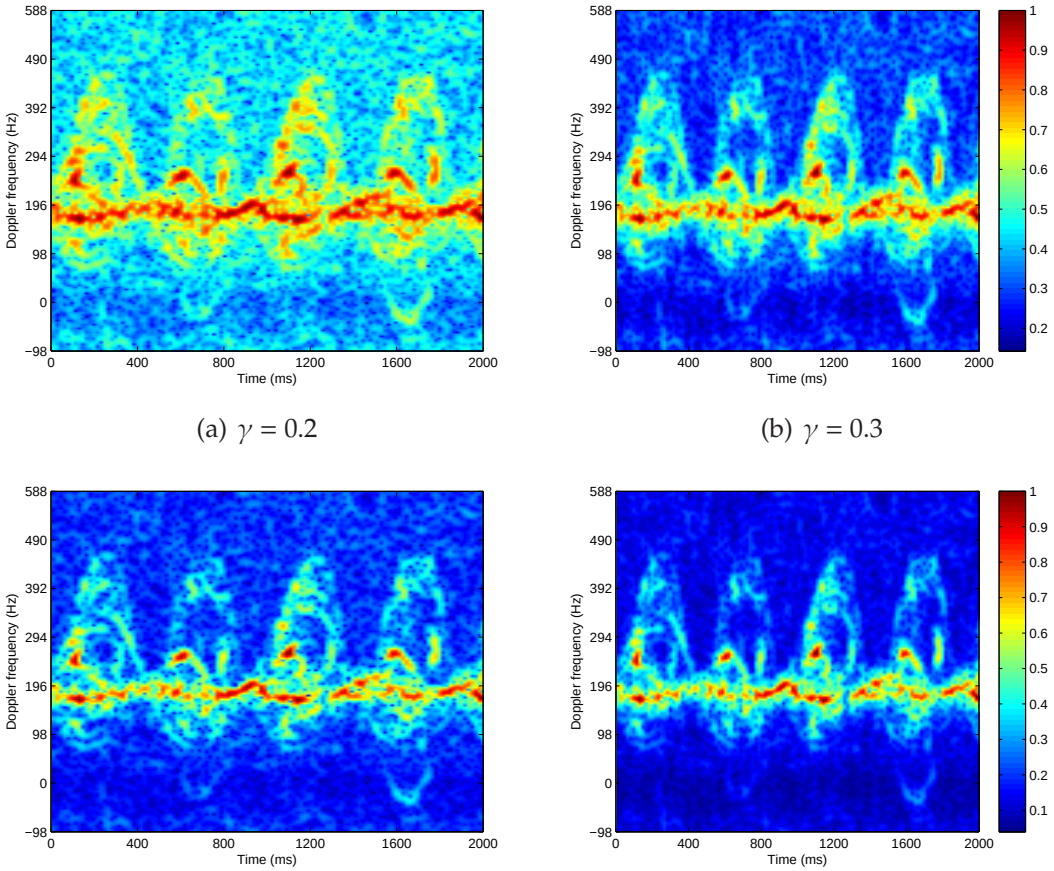


Figure 3.11: Plots of the Gamma transformation using different γ values.

In the case $\gamma = 1$, Equation (3.3) becomes an identity transformation. The curves that we got when $\gamma < 1$ show that they map a narrow range of small intensity input values into a wider range of output values. In other words, it enhances the input pixels with a lower intensities while suppresses the input pixels with higher intensities. The curves of $\gamma > 1$ have the opposite effect as those of $\gamma < 1$. The micro-Doppler signatures that we want to extract from the spectrograms usually have relatively low intensities. Thus, a small value of γ is used to reveal the micro-Doppler signatures. It should be noted that the background of the spectrogram also has a low intensity, which means a smaller value of γ will enhance the background too.



(a) $\gamma = 0.2$ (b) $\gamma = 0.3$

(c) $\gamma = 0.4$ (d) $\gamma = 0.5$

Figure 3.12: Spectrogram of a human gait using different γ values.

In Figure 3.12, γ ranges from 0.2 to 0.5 are used to enhance the spectrograms. As we can see, more details of the micro-Doppler signatures and the background noise became visible when the value of γ is decreasing. In Section 5.1.2, experiments will be conducted to find the value of γ that leads to the highest classification rate.

3.3.1.2 Histogram equalization

Histogram equalization enhances the contrast of an image using its histogram [42]. It changes the distribution of the intensity values. The most frequent intensity values are spread through the process so that the contrast of the image is improved.

Assume that the image to be processed has L intensity levels $\{0, 1, \cdots L-1\}$

and the number of pixels with intensity value of i is h_i. The cumulative image histogram c_i is calculated by

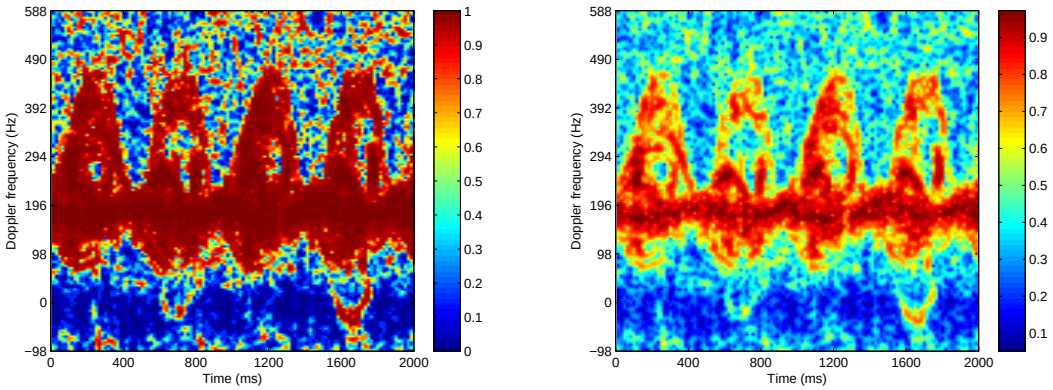
$$c_0 = h_0, \tag{3.4}$$

$$c_i = c_{i-1} + h_i, \text{ for } i = 1, 2, \cdots L-1. \tag{3.5}$$

The gray-scale transformation is defined as

$$b = \text{round}[(L-1) \times c_a/Q], \tag{3.6}$$

where Q is the total number of image pixels. Applying the transformation we can obtain the histogram equalized image. Figure 3.13(a) shows the spectrogram after using histogram equalization.



(a) by histogram equalization (b) by Naka-Rushton transformation

Figure 3.13: Spectrograms enhanced by two types of intensity transformation technique.

3.3.1.3 Naka-Rushton transformation

The Naka-Rushton transform [43] is usually used to perform enhancement on images. It is given by

$$b = \frac{a}{a + \bar{a}}, \tag{3.7}$$

where $\bar{a}$ is the mean value of all the intensities of pixels in one image. This transform is computational efficient and easy to implement. Figure 3.13(b) shows the spectrogram after using Naka-Rushton transformation to enhance the contrast of the image.

3.3.2 Intensity thresholding

After intensity transformation, intensity thresholding is applied on the spectrogram to remove the background noise. Given an image $f(x, y)$ that consists of light object on a dark background, the intensity values of the object and the background pixels can be grouped into two dominant classes. The way to extract the object from the background is to select a threshold t to divide the image into two classes. The thresholded image $g(x, y)$ is defined as follows

$$g(x,y) = \begin{cases} f(x,y) & \text{if } f(x,y) > t, \\ 0 & \text{if } f(x,y) \leq t. \end{cases} \tag{3.8}$$

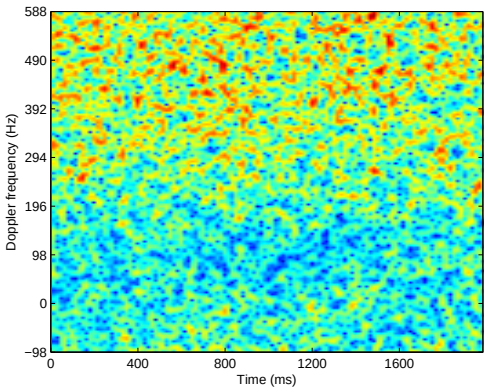
Any point (x, y) in the class of $f(x, y) > t$ is called an object point, the others are called background points. The key issue of image thresholding is to find the optimal threshold that separates the background noise and the regions of interest. Three ways to determine the optimal threshold are listed below:

- Fixed threshold,
- Otsu's method,
- Entropy-based method.

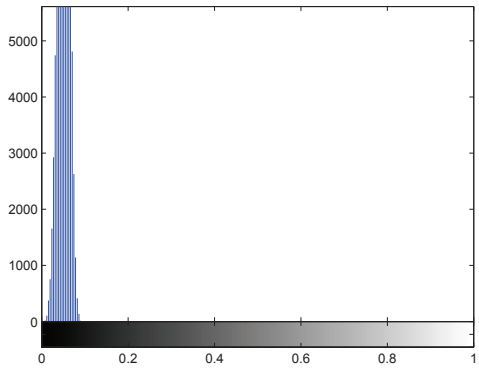
3.3.2.1 Fixed threshold

In this method, the optimal threshold is obtained by comparing the histograms of the spectrogram and the background signals. Background signals are collected

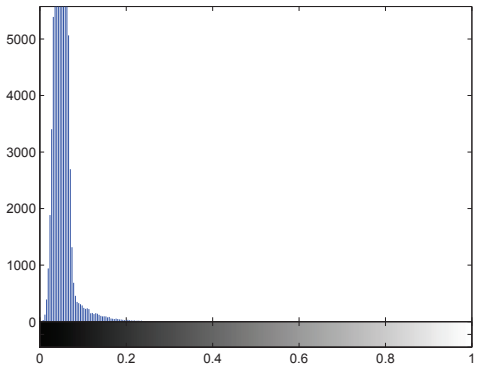
by facing the radar antenna on the same background that no target presents. Figure 3.14(a) shows the spectrogram of a background signal after using Gamma transformation. The histogram of the background signal is shown in Figure 3.14(b), which implies the background signal has a Gaussian-like distribution.



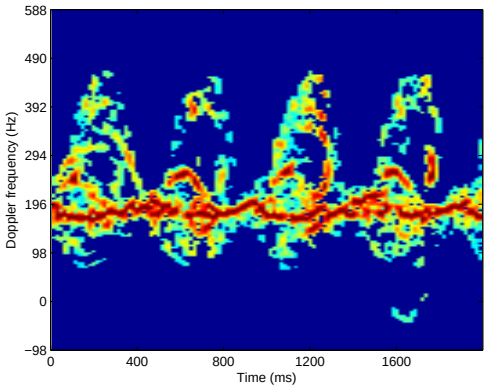
(a) spectrogram of the noise



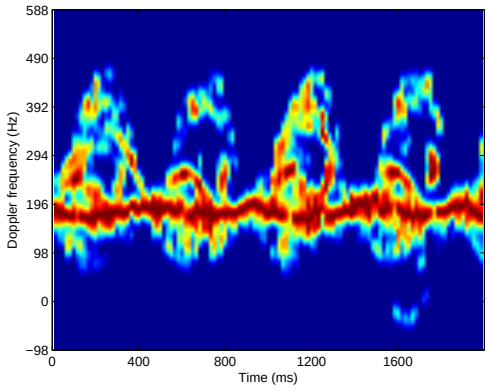
(b) histogram of the noise spectrogram



(c) histogram of the target return and noise



(d) de-noised spectrogram



(e) de-noised spectrogram after smoothing

Figure 3.14: An example of the fixed threshold approach.

Figure 3.14(c) shows a histogram when the human Doppler signal is present. The lowest power level at which the signal histogram starts to deviate from the Gaussian-like noise distribution can be used as the noise threshold. By comparing Figure 3.14(b) with Figure 3.14(c), the noise threshold is determined as 0.08. By using this threshold, the spectrogram of a human motion is processed and shown in Figure 3.14(d). Image smoothing is used to improve the thresholding. Here, we use a Gaussian filter of size 5×5 pixels and standard deviation $\sigma = 2.5$ to smooth the spectrogram; the result is shown in Figure 3.14(e).

Figure 3.14 shows that most of the background noises are removed after this process. The threshold that we get through observation is not precise and is dependent on the scene. Therefore, we need to find a method to automatically calculate the threshold of the spectrogram image.

3.3.2.2 Otsu's method

Otsu's method is a well-known and frequently used method for thresholding [44]. Its idea is to search for the optimum threshold that separates the two classes while keeping the within-class variance minimal.

Given an image that has L distinct intensity levels $\{ 0, 1, \ldots L-1 \}$. Define h_i as the number of pixels with intensity value of i. The probability of intensity level i is given as

$$p_i = h_i / Q, \tag{3.9}$$

where Q is the total number of image pixels, $Q = \sum_{i=0}^{L-1} h_i$. Using a threshold t, the image pixels are divided into two classes:

- Class C_1 contains pixels in the range $[0, t]$.

- Class C_2 includes pixels in the range $[t+1, L-1]$.

The probability density of class C_1 and C_2 are calculated as

$$W_1 = \sum_{i=0}^{t} p_i, \tag{3.10}$$

$$W_2 = \sum_{i=t+1}^{L-1} p_i. \tag{3.11}$$

For every intensity level i, the within-class variance is defined as the weighted sum of variances of the two classes:

$$\sigma^2(i) = W_1\, \sigma_1^2(i) + W_2\, \sigma_2^2(i), \tag{3.12}$$

where σ_1^2 and σ_2^2 are the variances of the the two classes. The intensity level that leads to the minimum within-class variance is considered as the optimal threshold. The within-class variance is defined as

$$\sigma_b^2(i) = W_1 W_2\, (\mu_1 - \mu_2)^2, \tag{3.13}$$

where $\sigma_b^2(i)$ is the between-class variance, μ_1 and μ_2 are the mean pixel intensities of the two classes. For any given intensity levels, the total variance is the sum of the within-class variances and the between-class variance, which is the sum of weighted squared distances between the class means and the grand mean. As the total variance is constant and independent of i, minimizing the within-class variance is the same as maximizing between-class variance [44]. The intensity level that leads to the maximum between-class variance is considered as the optimal threshold.

3.3.2.3 Entropy-based method

Entropy-based method performs the automatic thresholding based on the concept of Shannon's entropy [45, 46, 47]. Entropy measures the uncertainty of its information contained in a source. Kapur [48] developed an entropy-based thresholding method by considering an image's histogram as a probability distribution and selects an optimal threshold value that yields the minimum error of thresholding.

Given the same image with L intensity levels, the total entropy E of the image is given by

$$E = -\sum_{i=0}^{L-1} p_i \ln p_i. \tag{3.14}$$

Suppose t is the threshold that can divide the image into the foreground class and the background class. Then the probability distributions of the background class and the object class are listed as follows:

Background class: $\frac{p_0}{p}, \frac{p_1}{p}, \ldots, \frac{p_t}{p}$ and Object class: $\frac{p_{t+1}}{1-p}, \frac{p_{t+2}}{1-p}, \ldots, \frac{p_{L-1}}{1-p}$

where $p = \sum_{i=0}^{t} p_i$. The entropy associated with the background distribution is given by

$$E_b(t) = -\sum_{i=0}^{t} \frac{p_i}{p} \ln \frac{p_i}{p}. \tag{3.15}$$

The entropy associated with the object distribution can be calculated as

$$E_f(t) = -\sum_{i=t+1}^{L-1} \frac{p_i}{1-p} \ln \frac{p_i}{1-p}. \tag{3.16}$$

The optimal threshold is the value of t that can maximize the entropy of the two classes, that is,

$$t = \arg\max \{E_b(t) + E_f(t)\}. \tag{3.17}$$

This way the maximum information between the foreground and background classes is obtained.

3.4 Chapter summary

In this chapter, we described the radar signal acquisition and several preprocessing techniques. First, descriptions of the radar equipment and the data collection process are given. Second, STFT is applied on the 1-D radar signals to obtain spectrograms. The window function types, sizes and overlapping rates of STFT are investigated to provide the highest time-frequency resolution. Then, by casting the spectrograms as images, image processing techniques are used to enhance the weak micro-Doppler signatures and remove the background noise. The experiments and analysis will be presented in Chapter 5.

Chapter 4

Feature Extraction and Classification

Chapter contents

4.1 Local window extraction

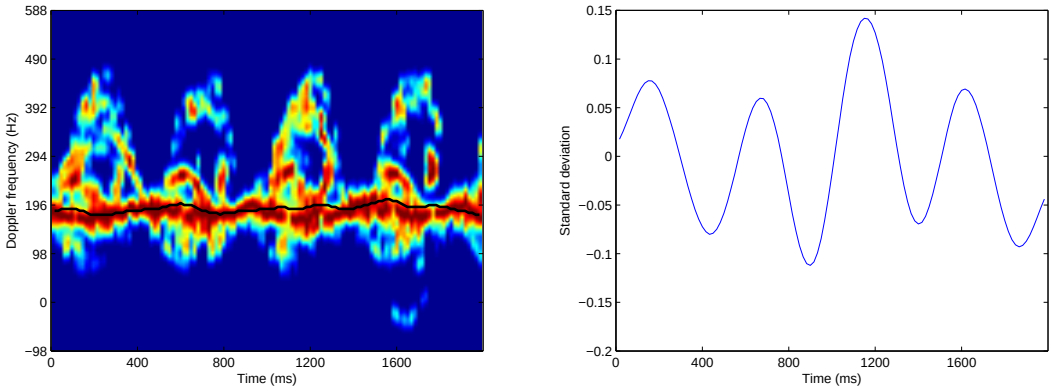
In our approach, the pre-processed spectrogram is considered as an image, and image features are extracted to characterize the micro-Doppler signatures. Instead of processing the entire spectrogram, we propose to extract features from sliding local windows to reduce the amount of data. By using local windows centered at the torso frequency, features that are invariant to the target's speed are obtained. The position of local windows should be selected so that each local window contains most of the micro-Doppler signatures and less redundant background.

The size of local window is another important factor of the proposed approach. A small window size lacks salient features for classification while a large window size produces significantly more noise. The position and size of local windows are explored in the following sections.

4.1.1 Window position

As most of the micro-Dopplers are found around the torso frequency, a suitable way is to locate the local windows centered at the torso frequency. Firstly, we locate the spectrogram spine by computing the maximum power spectrum at each time instance and then applying a median filter. Here median filtering is used as a smoothing technique, which also removes noise in smooth patches. The dark line in Figure 4.1(a) represents the smoothed torso energy on a de-noised spectrogram. After identifying the spine, local windows are centered at the spine frequency points. Because the vertical positions of local windows are determined as the location of the spine, the horizontal positions of local windows still need to be explored.

By finding the starting point of each local window along time axis, the horizontal position of every local window is determined. Local window alignment method is investigated since misaligned images produce severe artifacts in the scatter matrices of PCA analysis. Here, we align the local windows from the start of each gait cycle to improve the performance of 2D2-PCA feature extraction. The start of a gait cycle is identified when the spectrogram has the lowest standard deviation along the vertical direction. The reason is a low standard deviation indicates that the data points are very close to the mean, whereas a high standard



(a) the identified torso in a de-noised spectrogram

(b) the standard deviation

Figure 4.1: The example of the identified torso in a de-noised spectrogram and the smoothed standard deviation of each column: (a) the identified torso in a de-noised spectrogram; (b) the smoothed standard deviation of each column within the spectrogram.

deviation indicates that the data points are spread out over a large range of values.

From each column of the spectrogram, the standard deviation σ is computed. Let σ_a be the average value of all the standard deviations. A new set of standard deviations σ_n is computed as $\sigma - \sigma_a$. A Gaussian smoothing filter is applied on the σ_n to obtain a smooth set of standard deviations σ_s. Figure 4.1 displays the spectrogram of a walking human and the smoothed standard deviation of each column. The figure shows that the start of each gait cycle corresponds to a local minimum of σ_s. Thus, by detecting the local minima in the smoothed standard deviation, we can align the local windows.

4.1.2 Window size

After determining the position of local windows, the next step is to select the window size. The height of the local windows L_h should be big enough to cover most of the micro-Doppler signatures. The number of gait cycles included in a local window is determined by the width of the local window L_w. The

width of local window is empirically selected in Section 5.1.1 to provide the highest classification rate. The height of local window is explored in the following paragraph.

The largest Doppler shift is caused by the feet motion since the velocity of feet is the highest among all the body parts. The average velocity of a walking human is around $v = 1.3$ m/s. According to [3], the velocity of the feet is estimated to be $v_{feet} = 2.6 \times v = 3.4$ m/s. Based on the carrier frequency of the radar and velocity of the body parts, we can use Equation (2.4) to calculate the frequency shift of the body parts. The frequency shift caused by the feet motion is around 544 Hz and the Doppler shift of human torso is around 208 Hz. The frequency difference between feet and torso is 336 Hz, which corresponds to approximately 45 frequency points when using 1024 FFT points. Thus, the height of the local window is determined as 90 pixels, which is equivalent to 90 frequency points. Considering that different people have different walking speed, we choose $L_h = 180$ pixels as the height of the local window. Compared with the height of the entire spectrogram, which is 1024 pixels, using local window reduces significantly the redundant information.

4.2 Two-Directional, Two-Dimensional PCA features

In this section, two-directional, two-dimensional form of principal component analysis is presented as one of the proposed feature extraction techniques. Principal component analysis (PCA) is a well-known technique for feature extraction; it has been widely used in pattern recognition and signal processing [49, 50, 51]. When dealing with images, the 2-D matrices have to be converted into vectors before PCA can be applied. With high-dimensional feature vectors, the compu-

tation of the covariance matrix is intensive and requires a large amount of data space[52].

Unlike PCA, another technique named two-dimensional principal component analysis (2D-PCA) is based on 2-D matrices rather than vectors. With this technique, 2-D matrices are directly used to compute the covariance matrix, the size of the covariance matrix is reduced. Although 2D-PCA has better computational efficiency than PCA, it uses more coefficients[52, 53].

Considering that 2D-PCA is projecting the matrix in the row direction, an alternative 2D-PCA can be developed to project in the column direction. By projecting the original matrix onto both the row and column directions simultaneously, two-directional, two-dimensional principal component analysis (2D2-PCA) is developed. We propose to use two-directional, two-dimensional form of principal component analysis (2D2-PCA) to extract features from the local windows. Next, we present a description of 2D2-PCA.

Consider a set of P training samples where each sample is an image of size $M \times N$ pixels: $\mathbf{A}_i \in \mathbb{R}^{M \times N}$, $i = 1, 2, \cdots, P$. In our approach, every local window extracted from the spectrogram is an image sample. The mean of all the samples is computed as

$$\overline{\mathbf{A}} = \frac{1}{P} \sum_{i=1}^{P} \mathbf{A}_i. \tag{4.1}$$

The i-th training sample can be represented as a set of row vectors of N elements:

$$\mathbf{A}_i = [a_i^1 \;\; a_i^2 \;\; \cdots \;\; a_i^M]^T, \tag{4.2}$$

where a_i^m denotes the m-th row of $\mathbf{A}_i$. For the horizontal direction, let $\mathbf{\Phi} \in \mathbb{R}^{N \times D}$ be a projection matrix with unitary columns. An image $\mathbf{A}$ is projected onto $\mathbf{\Phi}$ to

yield a $M \times D$ matrix: $\mathbf{Y} = \mathbf{A}\mathbf{\Phi}$. An optimal projection matrix is determined by using the total scatter of the projected samples.

The image covariance matrix $\mathbf{H}$ in the horizontal direction for the given training set is defined as follows:

$$\mathbf{H} = \frac{1}{P}\sum_{i=1}^{P}(\mathbf{A}_i - \overline{\mathbf{A}})^T(\mathbf{A}_i - \overline{\mathbf{A}}). \tag{4.3}$$

The feature vectors obtained by projecting the image onto the eigenvectors of $\mathbf{H}$ have the minimal mean-square reconstruction error. The optimal projection matrix $\mathbf{\Phi} = \{\phi_1, \phi_2, \ldots, \phi_D\}$ is obtained by meeting the orthogonal constraints as well as maximizing the generalized total scatter $J(\mathbf{\Phi}) = \mathbf{\Phi}^T\mathbf{H}\mathbf{\Phi}$. In other words, the following criteria is adopted:

$$\begin{cases} {\phi_i}^T\phi_j = 0,\ i \neq j,\ i, j = 1, 2, \ldots, D, \\ \{\phi_1, \phi_2, \ldots, \phi_D\} = \arg\max J(\mathbf{\Phi}). \end{cases}$$

The optimal projection matrix $\{\phi_1, \phi_2, \ldots, \phi_D\}$ can be chosen as the eigenvectors of $\mathbf{H}$ with the first D largest eigenvalues. The value of D is selected so that

$$\frac{\sum_{i=1}^{D}\lambda_i}{\sum_{i=1}^{M}\lambda_i} \geq \theta, \tag{4.4}$$

where θ is a predefined threshold, and λ_i are the eigenvalues of $\mathbf{H}$ sorted in the descending order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$.

Similarly for the vertical direction, let $\mathbf{\Omega} \in \mathbb{R}^{M \times E}$ be a projection matrix with unitary columns. An image $\mathbf{A}$ is projected onto $\mathbf{\Omega}$ to yield a $E \times N$ matrix: $\mathbf{X} = \mathbf{\Omega}^T\mathbf{A}$. By treating a training sample $\mathbf{A}_i$ as a set of column vectors of M elements we can get:

$$\mathbf{A}_i = [a_i^1\ a_i^2\ \ldots\ a_i^N], \tag{4.5}$$

where a_i^n denotes the nth column of $\mathbf{A}_i$. The image covariance matrix in the vertical direction is computed:

$$\mathbf{V} = \frac{1}{P}\sum_{i=1}^{P}(\mathbf{A}_i - \overline{\mathbf{A}})(\mathbf{A}_i - \overline{\mathbf{A}})^T. \tag{4.6}$$

The projection matrix $\mathbf{\Omega} = \{\omega_1, \omega_2, \ldots, \omega_E\}$ is calculated by meeting the orthogonal constraints as well as maximizing the generalized total scatter $J(\mathbf{\Omega}) = \mathbf{\Omega}^T\,\mathbf{V}\,\mathbf{\Omega}$. That is, the following criteria is adopt:

$$\begin{cases} {\omega_i}^T\omega_j = 0,\ i \neq j,\ i, j = 1, 2, \ldots, E, \\ \{\omega_1, \omega_2, \ldots, \omega_E\} = \arg\max J(\mathbf{\Omega}). \end{cases}$$

The columns of the optimal projection matrix $\mathbf{\Omega}$ are the eigenvectors of $\mathbf{V}$ that correspond to the E largest eigenvalues. The value of E is determined similarly to Equation (4.4).

For a given input image $\mathbf{A}$ of size $M \times N$, a feature matrix $\mathbf{C}_{\text{PCA}}$ of size $E \times D$ can be obtained by applying projection matrices $\mathbf{\Phi}$ and $\mathbf{\Omega}$ simultaneously:

$$\mathbf{C}_{\text{PCA}} = \mathbf{\Omega}^T\mathbf{A}\mathbf{\Phi}. \tag{4.7}$$

The local windows of the spectrogram are projected to the feature space using 2D2-PCA technique. The resulting feature matrices $\mathbf{C}_{\text{PCA}}$ are then converted to feature vectors for classification.

Given an image of size $M \times N$ pixels, the sizes of covariance matrix, projection matrix and feature vectors using PCA, 2D-PCA, and 2D2-PCA are listed in Table 4.1. It should be noted that for 2D2-PCA, there are two covariance matrices and two projection matrices, one for the row direction and the other for the column direction. In the table, the parameters of G, F, E and D denote the sizes of the matrices and vectors used in classification stage.

Table 4.1: Comparison of PCA, 2D-PCA and 2D2-PCA.

	Matrix size			
Method	Image	Covariance matrices	Projection matrices	Feature matrix
PCA	$M \times N$	$MN \times MN$	$MN \times G$	$1 \times G$
2D-PCA	$M \times N$	$N \times N$	$N \times F$	$M \times F$
2D2-PCA	$M \times N$	$M \times M$ and $N \times N$	$M \times E$ and $N \times D$	$E \times D$

The 2D2-PCA has been shown to be more computationally efficient than the traditional PCA [54]. Moreover, 2D2-PCA preserves the spatial topology of the 2-D input. Compared to the conventional PCA, 2D2-PCA uses covariance matrices of small size, which can be computed more efficiently and accurately, even on a small training set. The high dimensionality of PCA usually results in singularity of the covariance matrix, which makes it difficult to calculate the projection axes [55].

4.3 GIST features

In this section, GIST method is presented as another proposed feature extraction technique. The amount of information comprehended from a real-world scene at a glance is refer to as The GIST of a scene [56, 57]. The GIST of a scene consists of several levels: low-level (e.g., color and contours), intermediate-level (e.g., shapes and texture) and high-level (e.g., activation of semantic knowledge). Here, we apply the GIST descriptor to extract GIST features from each local window.

The GIST descriptor first filters an image in multiple scales and orientations using a set of Gabor filters. The outputs of these filters are then weighted to obtain the GIST features. A 2-D Gabor function $g(x, y)$ and its Fourier transform $G(u, v)$

are defined as [58]

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp[-\frac{1}{2}(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2})+2\pi jW_x], \tag{4.8}$$

$$G(u,v) = \exp\{-\frac{1}{2}[\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}]\}, \tag{4.9}$$

where $\sigma_u = 1/2\pi\sigma_x$, $\sigma_v = 1/2\pi\sigma_y$ and W is the modulation frequency.

Let $g(x,y)$ be the mother wavelet, a set of Gabor filters are generated through dilation and rotation. The generating function is

$$g_{mn}(x,y) = a^{-m}g(x',y'),\ a>1,\ m,n=\text{integer} \tag{4.10}$$

$$x' = a^{-m}(x\cos\theta + y\sin\theta), \tag{4.11}$$

$$y' = a^{-m}(-x\sin\theta + y\cos\theta), \tag{4.12}$$

where $\theta = n\pi/N_o$, N_o is the total number of orientations, m and n represent the scales and orientations respectively.

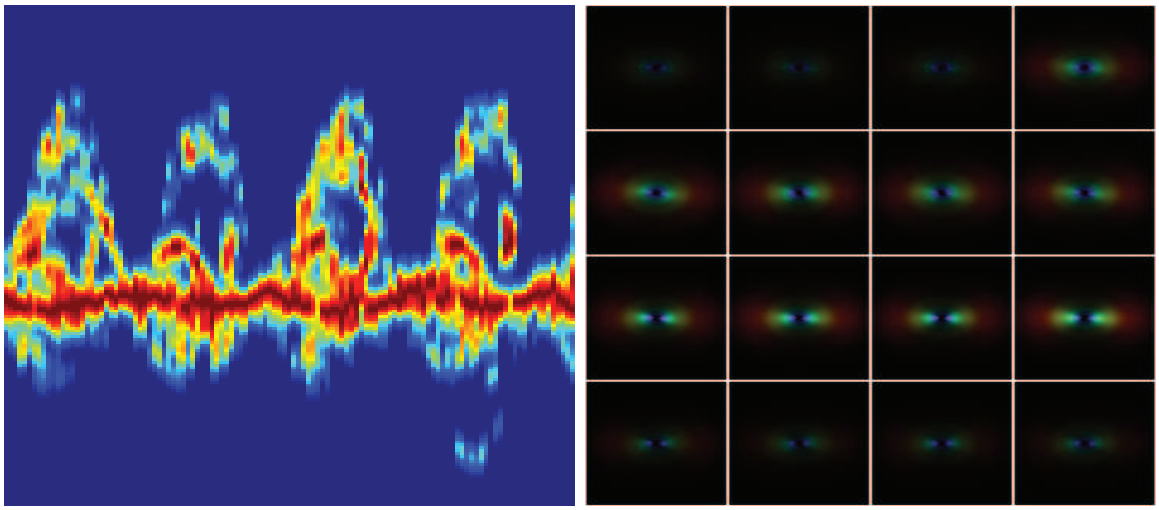
Images need to be preprocessed before sending to the Gabor filters. The preprocessing includes four steps: (i) padding images to reduce boundary artifacts, (ii) whitening, (iii) local contrast normalization, and (iv) cropping output images to have the same sizes as the input images.

Given an image A, the output of one Gabor filter is

$$O(x,y) = \sum_{x_1}\sum_{y_1} A(x_1,y_1)\otimes g(x-x_1,y-y_1), \tag{4.13}$$

where $\otimes$ indicated the 2-D convolution. All filtered output images are evenly partitioned into N_b non-overlapping blocks. Let μ_i be the mean value of each block.

$$\mu_i = \frac{1}{N_xN_y}\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} \mid O(x,y)\mid, i=1,2,\cdots,N_b, \tag{4.14}$$



(a) Input image (b) Descriptor

Figure 4.2: Example of a 2 seconds spectrogram and its output from one Gabor filter: (a) input image; (b) magnitude of the filtered image on polar plot.

where $N_x \times N_y$ is the size of each block. The GIST feature vector is constructed using μ_i as components. Figure 4.2 shows the output of one Gabor filter. In the figure, the output image is split into 16 blocks, the mean value of each block is a feature. If the Gabor filters has N_s scales and N_o orientations, the GIST feature size is $N_s \times N_o \times N_b$.

4.4 Classification

In the classification stage, support vector machines are used to discriminate the various types of human motions. In machine learning, SVM is a widely used classifier due to its superior performance [59]. For a two-class problem, the key idea of SVM is to determine the separating hyperplane with the highest margins to the two classes. A margin is the distance from the hyperplane to the nearest input vector. Support vectors are the input vectors that lie closest to the separating hyperplane.

Consider a set of P training feature vectors $\mathbf{x}_i \in \mathbb{R}^d$ and the corresponding

labels $y_i \in [+1, -1]$, where $i = 1, 2, \ldots, P$. SVM training can be expressed as an optimization problem:

$$\text{minimizing } \{\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{P}\xi_i\} \tag{4.15}$$

subject to

$$y_i\,(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i\,, \text{ and } \xi_i \geq 0 \text{ for all } i, \tag{4.16}$$

where $\mathbf{w}$ is the vector normal to the hyperplane, b is a bias term, ξ_i is a non-negative slack variable, and C is a penalty parameter representing the trade-off between maximizing the margin and minimizing the training error. This optimization has a dual form in which $\mathbf{w}$ is represented as $\mathbf{w} = \sum_{i=1}^{M}\alpha_i\,y_i\,\mathbf{x}_i$:

$$\text{maximizing } \Gamma(\alpha) = \sum_{i=1}^{P}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{M}\alpha_i\,\alpha_j\,y_i\,y_j\,\mathbf{x}_i^T\mathbf{x}_j \tag{4.17}$$

subject to

$$\sum_{i=1}^{P}\alpha_i\,y_i = 0, \text{ and } 0 \leq \alpha_i \leq C \text{ for all } i. \tag{4.18}$$

Once the optimal α_i has been computed, the decision function for SVM can be written as

$$y_p(\mathbf{x}) = \text{sgn}\{\sum_{i=1}^{P}\,y_i\,\alpha\,\psi(\mathbf{x}_i, \mathbf{x}) + b\}, \tag{4.19}$$

where $\psi(\cdot,\cdot)$ is a kernel function that maps the input feature vectors to a higher dimensional space (to improve class separation). If $y_p(x)$ is positive, the feature vector $\mathbf{x}$ belongs to class +1 and vice versa.

The linear kernel is defined as

$$\psi(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i{}^T\mathbf{x}_j. \tag{4.20}$$

To deal with the case when the relationship between feature vectors and class

labels is nonlinear, a nonlinear kernel that maps samples into higher dimensional space is needed. Three basic nonlinear kernels are listed as follows:

$$\text{polynomial} \quad : \quad \psi(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i{}^T \mathbf{x}_j + r)^d, \gamma > 0. \tag{4.21}$$

$$\text{radial basis function(RBF)} \quad : \quad \psi(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2), \gamma > 0. \tag{4.22}$$

$$\text{sigmoid} \quad : \quad \psi(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i{}^T \mathbf{x}_j + r). \tag{4.23}$$

Here, γ, r and d are kernel parameters. Among the three nonlinear kernels, the RBF kernel is a good choice. Compared with the polynomial kernel, the RBF kernel has fewer parameters. The number of parameters influences the complexity of the classifier [60]. The sigmoid kernel is not valid under some parameters that the kernel matrix is not positive semi-definite [59] [61]. In addition, the RBF kernel has fewer numerical difficulties. For example, the RBF kernel value ranges from 0 to 1 ($0 < \psi(\mathbf{x}_i, \mathbf{x}_j) \leq 1$), while the polynomial kernel value goes from 0 to infinity. In the proposed method, we use both linear kernel and RBF kernel.

For a multi-class problem involving K classes, we apply the one-versus-one approach that requires $K(K-1)/2$ two-class SVM classifiers, each classifier is trained on two classes. The final classification output is obtained by combining all two-class SVM classifiers with a majority voting scheme.

4.5 Chapter summary

In this chapter, we described the proposed feature extraction and classification methods. Local windows instead of the entire spectrograms are used to reduce the amount of data and produce features that are invariant to the speed of target.

The position and size of local windows are explored to create suitable inputs for feature extraction. Two proposed feature extraction approaches are presented to extract features from local input windows, namely 2D2-PCA and GIST methods. Then, SVM classifier with linear and RBF kernels are used to classify feature vectors into human motion categories. The experiments and analysis will be presented in Chapter 5.

Chapter 5

Results and Analysis

Chapter contents

In this chapter, we evaluate the proposed methods for feature extraction and classification on the acquired Doppler radar signal database. The chapter is organized as follows. Section 5.1 analyzes the parameters of the proposed method in terms of classification performance. Section 5.2 compares the proposed method with several existing methods.

The classification rate is calculated as the percentage of test samples that are correctly classified. Five fold cross-validation is applied to evaluate the classification performance. The entire radar database are partitioned into five subsets.

Then, for each fold, four subsets are used for training and the remaining subset is kept for testing. The final classification rate is obtained by averaging the results from all five folds. The optimal parameters of the proposed method are determined by dividing the training set of the first fold into training and validation set. Based on the validation set, the parameters are selected when the proposed method achieves the maximum classification rate.

5.1 Parameter Analysis

In this section, we search for the parameters of the proposed method that lead to the best classification performance. Section 5.1.1 explores the Hamming window size and the input window size. Section 5.1.2 investigates the intensity transform and thresholding methods to enhance the micro-Doppler signature. Section 5.1.3 investigates the thresholds of the proposed 2D2-PCA approach. Section 5.1.4 investigates the configuration the proposed GIST method.

5.1.1 Window sizes

To capture the micro-Doppler signatures, a Hamming window was used in the STFT and the overlap was set to 50%. Since the time-frequency resolution of the spectrogram depends on the width of the Hamming window, experiments were conducted to explore the window width that yields the highest classification result.

Let H_w be the width of the Hamming window, and L_w be the width of the local window. The segment of radar signal to be classified is given by $t_x = [L_w\,(H_w - N_o) + N_o]/F_s$ (seconds), where N_o is the number of overlapping

samples in the STFT and F_s is the sampling rate. The preliminary experiments are conducted to search for the optimum widths of Hamming window and local window as follows.

- The Hamming window we adopted has a width H_w varied from 256 to 768 with a step of 128. That is, the length of the STFT window is varied from 32.8 to 98.3 ms.
- The local window has a height (along the frequency axis) of 180, and a width (along the time axis) varing from 50 to 250 with a step of 50.
- For each pair of (H_w, L_w), two SVM classifiers with the linear kernel and RBF kernel are used. For the linear SVM and the RBF SVM, the parameters are obtained through grid-search approach. The parameters that lead to the best classification accuracy is used.

Table 5.1: Classification rates (%) of 2D2-PCA method on the validation set for different sizes of Hamming window and local window using linear SVM.

	Hamming window S_h				
Local window S_l	256	384	512	640	768
50	59.7	69.6	71.5	72.3	70.3
100	61.8	69.1	73.5	**74.6**	72.0
150	63.3	68.9	**73.7**	72.6	66.7
200	63.4	70.0	70.7	67.1	66.7
250	63.2	69.9	69.4	66.1	65.9

Table 5.1 shows the classification rate for each window pair (H_w, L_w) with linear SVM classifier used, while Table 5.2 lists the classification rate for each window pair (H_w, L_w) with RBF SVM classifier applied. Both tables show that the top two classification rates were achieved by using window pairs $(H_w = 640, L_w = 100)$ and

Table 5.2: Classification rates (%) of 2D2-PCA method on the validation set for different sizes of Hamming window and local window using RBF SVM ($\gamma = 0.125$).

	Hamming window S_h				
Local window S_l	256	384	512	640	768
50	71.2	77.9	80.8	83.6	83.3
100	72.9	78.7	83.2	**85.9**	83.5
150	72.8	80.8	**84.1**	84.0	81.4
200	73.2	82.2	83.5	82.6	81.7
250	76.1	80.8	81.7	81.1	79.5

($H_w = 512, L_w = 150$). The lengths of the acquired signals for ($H_w = 640, L_w = 100$) and ($H_w = 512, L_w = 150$) are $t_x = 4.1$ s and $t_x = 4.9$ s, respectively. These two pairs of window size are used for evaluating the proposed method on the test set. Since the RBF SVM achieves better classification performance, it is used as the classifier in the following sections.

5.1.2 Intensity transformation and thresholding

Intensity transformations are investigated first to enhance the weak micro-Doppler signatures in the spectrogram and then intensity thresholding methods are used to remove the background noise. Considering different classification rates may obtain by using various combinations of intensity transformation and intensity thresholding techniques, experiments are conducted to search for the combination that leads to the highest classification result. Table 5.3 lists the six combinations of applying different intensity transformation and intensity thresholding techniques.

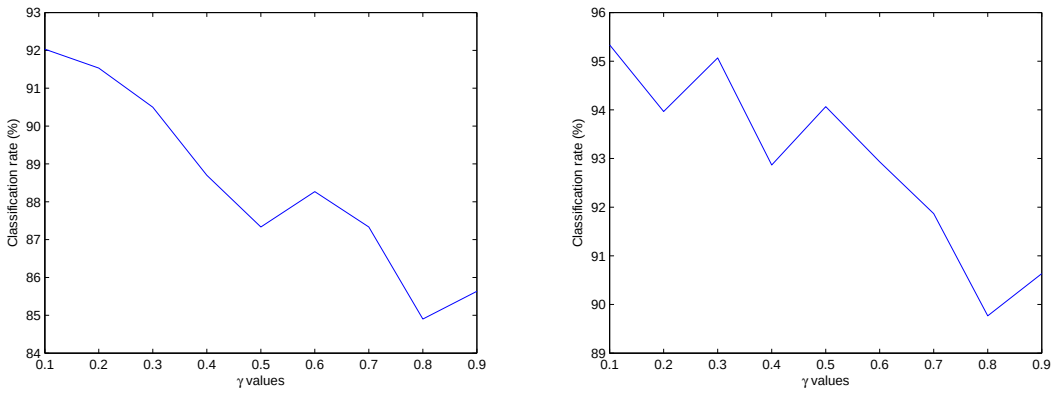
The two chosen pairs of Hamming and local window sizes, ($H_w = 512, L_w = 150$) and ($H_w = 640, L_w = 100$), are used to determine the appropriate intensity transformation and thresholding techniques.

Firstly, we compare the classification rate of the six intensity processing tech-

Table 5.3: Combinations of different intensity transformations and intensity thresholding methods.

Combination	Intensity transformation	Intensity thresholding
1	Gamma transformation	Entropy based method
2	Gamma transformation	Otsu's method
3	Naka-Rushton transformation	Entropy based method
4	Naka-Rushton transformation	Otsu's method
5	Histogram equalization	Entropy based method
6	Histogram equalization	Otsu's method

niques using window size ($H_w = 512, L_w = 150$). For the first two combinations listed in Table 5.3, Gamma transformation is used to enhance the weak pixels that represent the micro-Dopplers. Several γ values are applied to find the best classification performance.



(a) Entropy-based method (b) Otsu's method

Figure 5.1: Classification rates of 2D2-PCA method for different value of γ when using entropy-based method and Otsu's method: (a) Gamma transformation with entropy-based method; (b) Gamma transformation with Otsu's method.

.

Here, γ ranging from 0.1 to 0.9 with a step of 0.1 is adopted to enhance the weak pixels of the spectrogram. Figure 5.1(a) shows the classification rate for different γ values when the entropy-based method is utilized for image thresholding. The highest classification rate of 92.0% is achieved when $\gamma = 0.1$. Figure 5.1(b) shows the classification rates of using Otsu's method for thresholding, and we get the

best performance of 95.3% when $\gamma = 0.1$. Both figures show that a relative small value of γ is more applicable since it enhances the weak micro-Doppler signatures better.

To get a precise value of γ, further experiments are conducted to test the performance of the proposed system. Here, the value of γ ranges from 0.01 to 0.2 with a step of 0.01. The highest classification rate of 92.9% using entropy-bases method for thresholding is achieved when $\gamma = 0.12$. When Otsu's method is adopted, the best performance of 95.8% is obtained by using $\gamma = 0.07$. Although a smaller γ value is desired, the highest classification result is not generated by the smallest γ value that we applied. The reason is that a smaller value of γ enhances the weak pixels better, and the weak pixels consist of not only micro-Doppler signatures but also background noise.

Table 5.4: Classification rate of different combinations of intensity transformations and thresholding methods for window size ($H_w = 512, L_w = 150$).

	Intensity transformation	Intensity thresholding	Classification rate (%)
1	Gamma transformation	Otsu's method	95.8
2	Gamma transformation	Entropy based method	92.9
3	Naka-Rushton transformation	Otsu's method	94.6
4	Naka-Rushton transformation	Entropy based method	95.1
5	Histogram equalization	Otsu's method	89.1
6	Histogram equalization	Entropy based method	95.6

The classification rates using the two optimum γ values are compared with the other four combinations of intensity processing techniques. The classification rates of the six combinations are listed in Table 5.4. The best classification result is achieved when using Gamma transformation and Otsu's method.

For the window size ($H_w = 640, L_w = 100$), we conduct the same experiment to find the optimum combination of intensity transformation and thresholding tech-

niques. Different values of γ are used to get the optimum one. When $\gamma = 0.27$, we get the highest classification rate of 94.7%, using entropy-based method for thresholding. With Otsu's method, the best performance of 97.6% is achieved when $\gamma = 0.15$. The classification rates of the six combinations are listed in Table 5.5.

Table 5.5: Classification rate of different combinations of intensity transformations and thresholding methods for window size ($H_w = 640, L_w = 100$).

	Intensity transformation	Intensity thresholding	Classification rate (%)
1	Gamma transformation	Otsu's method	97.6
2	Gamma transformation	Entropy based method	94.7
3	Naka-Rushton transformation	Otsu's method	97.5
4	Naka-Rushton transformation	Entropy based method	96.8
5	Histogram equalization	Otsu's method	89.0
6	Histogram equalization	Entropy based method	97.0

Table 5.4 and Table 5.5 show that the combination of Gamma transformation and Otsu's method outperforms other techniques. In the following sections, we use Gamma transformation and Otsu's method as the intensity processing method.

5.1.3 2D2-PCA threshold

In 2D2-PCA, the size of the feature matrix is controlled by two predefined thresholds. Let θ_C be the threshold that determines the column size D of the feature matrix and θ_R be the threshold that controls the row size E of the feature matrix. Table 5.6 lists the sizes of feature matrices with respect to the classification performances by using various thresholds based on window size ($H_w = 512, L_w = 150$). Table 5.7 lists the sizes of feature matrices and classification rates by using various thresholds based on window size ($H_w = 640, L_w = 100$).

In Tables 5.6 and 5.7, the first row represents the θ_C values while the first column represents the θ_R values. The second row and the second column represent

Table 5.6: Feature matrix size generated by 2D2-PCA and the classification rates for window size ($H_w = 512, L_w = 150$).

	θ_C	0.4	0.5	0.6	0.7	0.8	0.9	0.95
θ_R		2	3	4	6	9	13	18
0.4	6	86.8	94.2	96.0	95.3	96.0	96.5	96.8
0.5	11	92.7	95.0	95.8	96.3	96.2	**97.2**	95.7
0.6	21	95.2	96.2	96.1	96.7	96.0	96.1	96.4
0.7	38	95.3	96.5	96.7	96.7	96.1	95.9	96.8
0.8	64	94.1	95.2	95.9	96.0	96.1	96.5	94.9
0.9	101	94.2	95.0	94.1	95.5	95.7	94.8	95.3
0.95	123	94.2	94.6	94.2	95.2	95.4	95.4	95.4

Table 5.7: Feature matrix size generated by 2D2-PCA and the classification rates for window size ($H_w = 640, L_w = 100$).

	θ_C	0.4	0.5	0.6	0.7	0.8	0.9	0.95
θ_R		2	3	4	6	8	12	16
0.4	5	91.7	95.0	96.6	96.3	95.7	97.8	97.1
0.5	8	92.7	97.5	97.9	97.6	97.9	**98.4**	98.4
0.6	15	95.4	97.0	97.5	97.6	97.8	98.2	98.2
0.7	26	96.8	97.1	97.3	97.8	97.9	97.7	97.7
0.8	43	97.7	97.4	97.7	98.0	97.3	97.7	97.7
0.9	67	97.1	97.4	97.1	95.9	96.8	97.2	97.2
0.95	82	96.3	97.4	96.9	96.0	96.6	97.3	97.5

the column size D and the row size E of the feature matrix, respectively. Table 5.6 and 5.7 show that increasing the thresholds generates larger feature matrix. The generalization of the classifier is dependent on the size of the features. A relatively small size of feature vector with a large number of feature vectors leads to better generalization and higher classification result.

When adopting the window size ($H_w = 512, L_w = 150$), we achieve the highest classification rate of 97.2% by using $\theta_H = 0.9$, and $\theta_V = 0.5$. The feature matrix size is $E \times D = 11 \times 13$, while the original image size is 180×150. The highest classification rate of 98.4% is achieved by using $\theta_H = 0.9$, and $\theta_V = 0.5$ based on window size ($H_w = 640, L_w = 100$). The feature matrix size is $E \times D = 8 \times 12$ in this case. In the following sections, we use the window size ($H_w = 640, L_w = 100$)

since it gives better classification performance.

5.1.4 GIST configuration

The size of GIST features is controlled by two main parameters. One is the number of Gabor filters N_g, and the other is the number of features N_b from the output image of each Gabor filter. Through dilating and rotating a mother wavelet, a set of Gabor filters is generated. Assume the Gabor filters have N_s scales and N_o orientations, we can get $N_g = N_s \times N_o$. The output of each filter is partitioned into N_b blocks and the mean value of each block is taken as a single feature. Thus, the size of GIST feature is given by $N_g \times N_b = N_s \times N_o \times N_b$.

The values of N_s and N_o are explored first to obtain the highest classification result. Table 5.8 lists the classification rates of using different numbers of scales and orientations to generate the Gabor filters. The highest classification rate is achieved when the Gabor filters have four scales and eight orientations, leading to 32 Gabor filters.

Table 5.8: Classification rates of GIST descriptor for different numbers of scales and orientations.

	Number of scales N_s			
Number of orientations N_o	2	4	8	16
4	86.3	96.6	97.1	96.8
8	90.6	**97.4**	96.2	97.1
16	94.0	97.2	97.3	97.0
32	95.3	96.5	97.0	96.9

Table 5.9: Classification rates of GIST for different numbers of features from each filtered image.

N_b	2×2	4×4	8×8	16×16
Feature size	128	512	2048	8192
Classification rates (%)	95.8	97.1	**97.8**	97.3

The next parameter to be determined is the number of partitions, which is, in

turn, related to the dimension of the feature vector. Table 5.9 lists the classification rates and GIST feature sizes of using different numbers of features from each filtered image.

Table 5.9 shows the highest classification rate is reached by $N_b = 8 \times 8 = 64$, where the number of features is 2048. Similar classification rate is obtained when $N_b = 4 \times 4 = 16$, and the number of features is only 512. To improve the computational efficiency and reduce the memory demand of the proposed method, we adopt $N_b = 16$ partitions of each filtered image.

5.2 Comparison of classification approaches

The proposed approaches are compared with other feature extraction methods: (i) traditional 1-D PCA and (ii) Hierarchical Image Classification Architecture [9]. SVM with RBF kernel is used as the classifier for evaluating these two approaches. The five fold cross-validation is used for each approach and the classification rate is obtained by averaging the results of all five folds. For each fold, 4000 local rectangular windows centered on the torso spine are chosen from each class for training and the remaining 1000 local windows are used for testing. The same techniques of intensity transformation and intensity thresholding are applied on the local windows.

5.2.1 The traditional 1-D PCA

For the traditional PCA, the local windows reshaped as vectors are used. This way, for each class we get $5,000$ vectors with size of $V_s = 180 \times 100 = 18,000$. The covariance matrix has a size of $18,000 \times 18,000$, which is about 324 million entries.

Similarly as the 2D2-PCA, the traditional PCA has the parameter threshold θ that controls the size of the feature vectors. Here, we also conduct experiment to search for the optimal threshold for the traditional PCA based on the same dataset that we used for parameter analysis in Section 5.1. Table 5.10 lists the feature vector sizes and classification rates of PCA method for different values of thresholds θ.

Table 5.10: Classification rates of PCA method for different values of threshold.

Threshold	0.5	0.6	0.7	0.8	0.9	0.95
Feature size	24	60	145	320	676	1052
Classification rates (%)	95.9	**98.0**	97.9	97.7	97.4	96.9

Table 5.10 shows as the threshold increases, the size of the feature vector increases as well. The classification rate reaches the highest value when $\theta = 0.6$, where the feature vector has a length of 60. Therefore, we use $\theta = 0.6$ to obtain the classification rates of the five fold cross-validation.

5.2.2 The Hierarchical Image Classification Architecture

We also adopt the method proposed in [9] for comparison, where the Hierarchical Image Classification Architecture (HICA) is used to extract Doppler features. The local window size that we selected is too big for HICA, which increases the complexity of the HICA. In [9], the chosen window size is 56×56. Thus, we down-sample our local window size by a factor of 2, which yields a window I_d of size 90×50. Since the first two stages of HICA involves down-sampling procedures, both the row and column sizes of window I_d should be divisible by 4 to complete the feature extraction process. By removing the last two rows and two columns of the window I_d, a new window I with size of 88×48 is obtained.

In the experiment, we use the same configuration as described in [9]. The first

stage consists of six directional first-order Gaussian derivative filters of size 9×9 with standard deviation $\sigma = 1.5$. The second stage consists of twelve adaptive filters with convolution masks of size 5×5. The coefficients of the adaptive filters are obtained through a training process. The third stage of the HICA is a linear SVM classifier.

For a fair comparison, the RBF SVM classifier is used in the third stage of the HICA. A two-step training process is adopted to integrate the RBF SVM with the HICA. First, based on the LM algorithm, a linear classifier is used in the third stage to obtain the coefficients of the adaptive filters in the second stage. The training process continues until the minimum MSE is achieved on the validation set. Then, with the features extracted by the trained filters, we train the 3-class RBF SVM to classify the features into categories.

5.2.3 Feature extraction comparison

The classification rates obtained using the five fold cross-validation for the four methods are shown in Table 5.11. The proposed 2D2-PCA method has an average classification rate of 97.8%, whereas GIST has a classification rate of 98.5%. Compared to the proposed 2D2-PCA method, the traditional 1-D PCA achieves a similar classification rate. However, the large size of covariance matrix leads to a long computational time as well as a large memory demand. Although the HICA has a classification rate of 97.7%, the training of the filters in the second layer is very time consuming, which may take several hours when the training set is large.

For further analysis, the confusion matrix for the 2D2-PCA approach is shown

Table 5.11: Classification rates on the test set of different methods.

Method	Average classification rate (%)
2D2-PCA	97.8 ± 0.5
GIST	98.5 ± 0.9
Traditional 1-D PCA	97.6 ± 0.7
HICA [9]	97.7 ± 0.7

in Table 5.12. The main diagonal entries are the correct classification rates for each human motion. The off-diagonal entries indicate the misclassification rates. Results for 0-AM (no arm swing), 1-AM (one arm swing) and 2-AM (two arm swing) show that 98.42% discrimination between the 1-AM class and other two classes is achieved with the proposed approach. Small errors occur when differentiating the 0-AM and 2-AM classes as their micro-Dopplers are quite similar. The overall result indicates that the proposed 2D2-PCA method is effective in classifying human motions based on micro-Doppler signatures.

Table 5.12: Confusion matrix for the 2D2-PCA approach. The entry at (row r, column c) is the percentage of human motion r that is classified as human motion c.

	0-AM	1-AM	2-AM
0-AM	97.14	0.62	2.16
1-AM	0.64	98.42	0.94
2-AM	1.58	0.62	97.80

Table 5.13 shows the confusion matrix for the GIST method. Similarly as 2D2-PCA, higher discrimination rate is achieved between the 1-AM and other classes, small errors occur for the 0-AM and 2-AM classes. The overall result also indicates that based on micro-Doppler signatures, the proposed GIST method provides promising result in classifying human motions.

Several experiments are conducted to further compare the speed and robustness of each method. To compare the processing speed, one simple way is to

Table 5.13: Confusion matrix for the GIST approach. The entry at (row r, column c) is the percentage of human motion r that is classified as human motion c.

	0-AM	1-AM	2-AM
0-AM	97.68	0.14	2.18
1-AM	0.18	99.78	0.04
2-AM	1.82	0.06	98.12

measure the time for each method to obtain one feature vector. For each method, the total time needed to extract features from 1000 local windows is recorded. Dividing the total time by the number of local windows, the average time used for each method to extract features from one local window is obtained.

Table 5.14: Average feature extraction time of different methods.

Method	Feature extraction time (*ms*)
2D2-PCA	0.15
GIST	292.90
Traditional 1-D PCA	0.27
HICA [9]	8.80

Table 5.14 lists the average time used for each approach to extract features from one local window. Among all the methods, the proposed 2D2-PCA has the shortest feature extraction time. The proposed GIST method takes a longer time to extract features from one local window because it contains several preprocessing steps. The traditional 1-D PCA has a short feature extraction time after a long time for computing the projection matrix. Noted that for the HICA method, except for the training and testing time listed in Table 5.14, it also contains a long training time of filters in the second layer.

To compare the robustness of all the methods, we generate several data sets with different levels of noise. White Gaussian noise is added to the radar signals, with signal-to-noise ratio (SNR) ranging from 5 dB to 30 dB. The experiment

applies the five fold cross-validation and the classification rates are averaged.

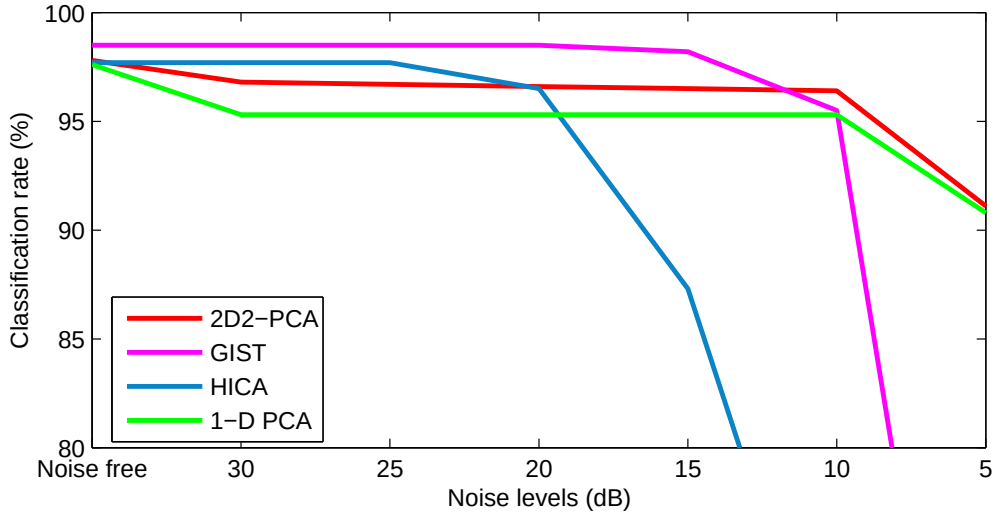


Figure 5.2: Classification performance of different methods with respect to the SNR of the input signal (dB).

Figure 5.2 shows the classification results of the four methods for each SNR level. The proposed GIST method outperforms other methods when the SNR is between 15 dB and 30 dB. The proposed 2D2-PCA method achieves the highest classification rate when the SNR is between 10 dB and 5 dB. Noted that the 2D2-PCA and traditional 1-D PCA reach classification rates of more than 90%, even when the SNR value is small. The projection matrices of 2D2-PCA and 1-D PCA are obtained from eigenvectors associated with largest eigenvalues. Therefore, the feature vectors of 2D2-PCA and 1-D PCA preserve the pixels with highest energy, which are mainly formed by the Doppler signatures and are less influenced by the noise.

The classification results that we obtained so far are based on different radar signals of same subjects. That is, for each person, the four-fifths radar signals are used for training while the remaining one-fifth signals are used for testing. Here, we apply another training and testing scheme, which uses all the data from

16 subjects for training and the data from the remaining 4 subjects for testing. The new scheme is more realistic since it classifies the motions from unknown humans, based on the data from known subjects.

Table 5.15: Classification rates of different methods when the training and testing sets contain different subjects.

Method	Average classification rate (%)
2D2-PCA	82.5
GIST	90.9
Traditional 1-D PCA	80.9
HICA [9]	89.7

Table 5.15 displays the classification rates of the four methods when the training and testing sets contain different subjects. The classification rate of 90.9% is obtained by the proposed GIST method. The overall result shows the proposed GIST method provides a higher classification rate than the other methods.

5.3 Chapter summary

In this chapter, we presented the experimental results of the proposed feature extraction approaches on the acquired radar database. Various parameters for the proposed approaches were explored to select the optimum ones. Different feature extraction approaches were compared in terms of classification accuracy and processing time. The proposed GIST method achieved the highest classification rate of 98.5% with the optimum parameters selected. The proposed 2D2-PCA approaches provides promising result in terms of computational efficiency and robustness in noisy environment.

Chapter 6

Conclusion

Chapter contents

Human movement classification using radars has become an emerging research field with numerous civilian and military applications, such as surveillance, search and rescue, and health care. To automatically classify human motions using Doppler radar, we propose an approach that integrates the time-frequency analysis, image processing, feature extraction, and classification techniques. The radar signals captured by a FMCW radar are used for analysis and classification.

This chapter is organized as follows: Section 6.1 summarizes the research contributions of the thesis; Section 6.2 outlines the future work and research directions; Section 6.3 draws conclusion for the thesis.

6.1 Research summary

The research activities have been documented in several chapters. They are summarized as follows.

- We provided the literature review of radar systems and the existing approaches for target or movement classification using Doppler radar. The time-frequency analysis is adopted to capture the micro-Doppler signature of a motion dynamic. The existing approaches to extract features from various time-frequency distributions are analyzed.

- We constructed a radar dataset of several human motions using a FMCW radar. Clutter suppression technique is explored to reduce the noise caused by stationary background objects. Micro-Doppler signatures are represented using STFT and enhanced by 2-D image enhancement techniques.

- We proposed two feature extraction algorithms. Instead of extracting features from the entire time-frequency distributions, we used local windows to reduce the redundant information and provide features that are invariant to the target's speed. The 2D2-PCA and GIST methods are applied on the local windows to extract feature vectors. The suitable parameters for both feature extraction approaches are selected.

- We compared the proposed feature extraction algorithms with several existing approaches. The classification rate, processing time and anti-noise ability of all the approaches are compared.

6.2 Future work

Possible research directions can be summarized as follows:

- Finding another time-frequency distribution to improve the time-frequency resolution.

- Completing the radar database for more types of human motions, such as skipping, walking, punching, and running.
- Exploring methods to combine the 2D2-PCA and GIST methods to improve the performance and efficiency of the proposed approach.

6.3 Conclusion

This thesis has presented a new approach to automatically classify human motions using a Doppler radar for applications in security and surveillance. The proposed approach consists of three main parts: time-frequency analysis, feature extraction, and classification. Two feature extraction algorithms are proposed and evaluated on the acquired radar dataset. The proposed 2D2-PCA approach achieves 97.8% classification rate while the proposed GIST method reaches 98.5% classification rate. The traditional 1-D PCA and HICA are also applied on the same database to compare with the proposed method, and they reach classification rate of 97.6% and 97.7%, respectively.

References

[1] V. C. Chen, G. E. Smith, K. Woodbridge, and C. J. Baker, *Radar micro-Doppler signatures for characterization of human motion*. CRC Press, 2010, pp. 502–537.

[2] V. C. Chen, *The micro-Doppler effect in radar*. Artech House, 2011.

[3] A. Boissier and C. D. Loisy, "Urban surveillance use of continuous wave radar for monitoring human activity," Department Surface Radar, Technical Report, 2009.

[4] S. Singh, Q. Liang, D. Chen, and L. Sheng, "Sense through wall human detection using UWB radar," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, p. 20, 2011.

[5] J. W. Palmer, K. F. Bing, A. C. Sharma, and E. F. Greneker, "Detecting concussion impairment with radar using gait analysis techniques," in *IEEE Radar Conference*, 2011, pp. 222–225.

[6] M. S. Nixon and J. N. Carter, "Automatic recognition by gait," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 2013–2024, 2006.

[7] S. I. Ivashov, V. V. Razevig, A. P. Sheyko, and I. A. Vasilyev, "Detection of hu-

man breathing and heartbeat by remote radar," in *Progress in Electromagnetic Research Symposium*, 2004.

[8] T. Thayaparan, S. Abrol, and E. Riseborough, "Micro-Doppler radar signatures for intelligent target recognition," Defence Research and Development Canada, Technical Memorandum, 2004.

[9] F. H. C. Tivive, A. Bouzerdoum, and M. G. Amin, "A human gait classification method based on radar Doppler spectrograms," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–13, 2010.

[10] B. E. Godana, "Human movement characterization in indoor environment using GNU radio based radar," Master thesis, 2009.

[11] M. I. Skolnik, *Introduction to radar systems*. McGraw-Hill Book Company, 2001.

[12] V. Chen, "Doppler signatures of radar backscattering from objects with micro-motions," *IET Signal Processing*, vol. 2, no. 3, pp. 291–300, 2008.

[13] V. C. Chen, F. Li, S. S. Ho, and H. Wechsler, "Micro-Doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 2–21, 2006.

[14] T. Thayaparan, S. Abrol, E. Riseborough, L. Stankovic, D. Lamothe, and G. Duff, "Analysis of radar micro-Doppler signatures from experimental helicopter and human data," *IET, Radar, Sonar and Navigation*, vol. 1, no. 4, pp. 289–299, 2007.

[15] M. S. Gupta, "Definition of instantaneous frequency and frequency measurability," *American Journal of Physics*, vol. 43, pp. 1087–1088, 1975.

[16] S. Kay, "A fast and accurate single frequency estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 12, pp. 1987–1990, 1989.

[17] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Transactions on Information Theory*, vol. IT-20, pp. 591–598, 1974.

[18] L. Cohen, *Time-Frequency Analysis*. Prentice Hall PTR, 1995.

[19] S. Qian and D. Chen, *Joint time-frequency analysis: Methods and applications*. PTR Prentice Hall, 1996.

[20] M. S. Andric, B. P. Bondzulic, and B. M. Zrnic, "Feature extraction related to target classification for a radar Doppler echoes." 18th Telecommunications forum TELFOR.

[21] V. C. Chen, "Spatial and temporal independent component analysis of micro-Doppler features," in *IEEE International Radar Conference*, 2005, pp. 348–353.

[22] J. Lei, "Pattern recognition based on time-frequency distributions of radar micro-Doppler dynamics," in *Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, and First ACIS International Workshop on Self-Assembling Wireless Networks.*, 2005, pp. 14–18.

[23] B. G. Mobasseri and M. G. Amin, "A time-frequency classifier for human

gait recognition," in *SPIE Defense, Security and Sensing*, vol. 7306, 2009, pp. 730 628–730 628–9.

[24] Y. Kim and L. Hao, "Human activity classification based on micro-Doppler signatures using an artificial neural network," in *IEEE Antennas and Propagation Society International Symposium*, 2008, pp. 1–4.

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

[26] J. L. Geisheimer, E. F. G. III, and W. S. Marshall, "High-resolution Doppler model of the human gait," in *SPIE Radar Sensor Technology and Data Visualization*, vol. SPIE 4744, 2002.

[27] T. Thayaparan, L. Stankovi, and I. Djurovi, "Micro-Doppler-based target detection and feature extraction in indoor and outdoor environments," *Journal of the Franklin Institute*, vol. 345, no. 6, pp. 700–722, 2008.

[28] V. C. Chen and H. Ling, *Time-frequency transform for radar imaging and signal analysis*. Boston: Artech House, 2002.

[29] J. E. Cutting and L. T. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, vol. 9, no. 5, pp. 353–356, 1977.

[30] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[31] B. Lyonnet, C. Ioana, and M. G. Amin, "Human gait classification using microDoppler time-frequency signal representations," in *2010 IEEE Radar Conference*, 2010, pp. 915–919.

[32] I. Orovic, S. Stankovic, and M. Amin, "A new approach for classification of human gait based on time-frequency feature representations," *Signal Processing*, vol. 91, no. 6, pp. 1448–1456, 2011.

[33] V. C. Chen, "Evaluation of Bayes, ICA, PCA and SVM methods for classification," RTO SET Symposium on Target Identification and Recognition Using RF Systems, Technical Report, 2004.

[34] G. E. Smith, K. Woodbridge, and C. J. Baker, "Naive Bayesian radar micro-Doppler recognition," in *International Conference on Radar*, 2008, pp. 111–116.

[35] S. K. Han, H. T. Kim, S. H. Park, and K. T. Kim, "Efficient radar target recognition using a combination of range profile and time-frequency analysis," *Progress in Electromagnetics Research*, vol. 108, pp. 131–140, 2010.

[36] E. Weisstein, *CRC Concise Encyclopedia of Mathematics*, 2nd ed. Chapman and Hall/CRC, 2003.

[37] C. Cattani and J. Rushchitsky, *Wavelet and Wave Analysis As Applied to Materials With Micro Or Nanostructure*. World Scientific Publishing Company, 2007.

[38] H. Wang, "Evaluation of various window functions using multi-instrument," Virtins Technology, Technical Report, 2009.

[39] M. Viquesney, "Windowing: Optimizing FFTs using window functions." National Instruments, 2011.

[40] R. C. Gonzalez and R. E. Woods, *Digital image processing*, 3rd ed. Prentice Hall, 2008.

[41] P. K. Sahoo, S. Soltani, and A. K. C. Wong, "A survey of thresholding techniques," *Computer Vision, Graphics, and Image Processing*, vol. 41, no. 2, pp. 233–260, 1988.

[42] F. Miller, A. Vandome, and J. McBrewster, *Histogram Equalization*. VDM Publishing House Limited, 2011.

[43] K. I. Naka and W. A. H. Rushton, "S-potentials from luminosity units in the retina of fish," *The Journal of Physiology*, vol. 185, pp. 587–599, 1966.

[44] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[45] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[46] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

[47] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.

[48] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, pp. 273–285, 1985.

[49] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[50] I. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.

[51] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990.

[52] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[53] D. Zhang, S. Chen, and J. Liu, "Representing image matrices: Eigenimages versus eigenvectors," in *Proceedings of the Second international conference on Advances in neural networks - Volume Part II*. Springer-Verlag, pp. 659–664.

[54] D. Zhang and Z. Zhou, "(2D)2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition," *Neurocomputing*, vol. 69, no. 13, pp. 224–231, 2005.

[55] D. Zhang, X. Jing, and J. Yang, *2D Image Matrix-Based Discriminator*. IGI Global, 2006, pp. 258–286.

[56] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, pp. 145–175, 2001.

[57] ——, *Building the gist of a scene: the role of global image features in recognition*, ser. Progress in Brain Research, 2006, vol. 155, pp. 23–36.

[58] P. Tao, C. He, C. Qian, and H. Sun, "A novel spatial-scale weighted GIST descriptor for SAR image retrieval," in *Proceedings of the 2008 First International Conference on Intelligent Networks and Intelligent Systems*. IEEE Computer Society, pp. 539–542.

[59] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1999.

[60] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," *Bioinformatics*, vol. 1, no. 1, pp. 1–16, 2010.

[61] H. T. Lin and C. J. Lin, "A study on Sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods," Department of Computer science, National Taiwan University, Technical Report, 2003.