A thesis for the Degree of
Master of Engineering

Target Detection and Classification Algorithm Using

Seismic Sensor and Pulse Doppler Radar

Youn Joung Kang

Department of Ocean System Engineering
GRADUATESCHOOL
JEJU NATIONAL UNIVERSITY

2014. 2.

Target detection and classification algorithm using seismic sensor and pulse doppler radar

Youn Joung Kang
(Supervised by professor Chong Hyun Lee)

A thesis submitted in partial fulfillment of the requirement
for the degree of Master of Engineering

2014. 2.

This thesis has been examined and approved by

Thesis director, Jinho Bae, Associate Professor, Dept. of Ocean System Engineering

---

Chong Hyun Lee, Associate Professor,Dept. of Ocean System Engineering

---

Young-Chol Choi, Senior Researcher, Korea Institute of Ocean Science and Technology

---

Date

GRADUATE SCHOOL
JEJU NATIONAL UNIVERSITY

Table of Contents

List of figures

List of tables

Abstract

This thesis research achieved its primary scientific objective to perform robust, automatic detection and classification of moving targets using seismic sensor and pulse doppler radar on a stationary platform. Detection and classification algorithm using seismic sensor and pulse doppler radar signal is a problem of current interest. The purpose is to detect and classify the moving target without human aid. The target classes were included human running, human walking and animal. Evaluation of theory on realistic experimental data is vital to the advancement of knowledge. The data can be used to rigorously evaluate new classification, detection, and feature selection algorithms. Computer simulation analysis also plays a crucial role in theoretical development. The experimental data collected by this thesis research can be utilized to improve the accuracy of computer models.

This thesis contributed a novel set of high-performance seismic and doppler based features. The Fisher used for selecting a feature set. In addition, the feature set included both statistical and the linear predictive coding (LPC) residual energy feature. The selected feature set was shown to perform well on the seismic sensor and doppler radar based target classification problem.

The design and detailed analysis of target classification algorithms based on support vector machine (SVM) and binary tree architecture (BTA) classifiers were designed to accomplish high-performance target classification. The importance of both classifier selection and feature selection was analyzed in detail.

In order to process the complex-valued signals from pulse doppler radar, complex-valued SVM classifier was derived. Complex-valued SVM classifier processes the complex-valued signals measured by PDR to identify moving targets from the background.

Chapter 1. Introduction

1.1 Background

Personnel detection deals with the prevention, detection, and response to unauthorized persons from crossing an established perimeter [1]. It is required in a variety of military and civilian situations. Personnel detection is an important aspect of intelligence, surveillance, and reconnaissance (ISR). It plays a vital role in perimeter and camp protection and in curtailing illegal border crossings by people from neighboring countries, to name few [2] [3]. All these applications involve deployment of sensors for a prolonged time and often camouflaged to avoid discovery by others. Due to the low power requirement, the sensors used consist of non-imaging sensors such as acoustic, seismic, magnetic, E-field, passive infrared, ultrasonic, and radar. If imaging sensors are used, they are used to take a snapshot of the target to corroborate the findings by other modalities. In this paper, we consider a subset of the sensors listed above, namely, acoustic, seismic [4] [5] [8], and ultrasonic sensors [6] [7]. It will be clear throughout the paper that these three sensors are adequate to detect and identify people and distinguish them from other targets such as animals. However, no single sensor is adequate for the job. Fusion of the outputs or features from these sensors is the key for detection and classification with high confidence.

Detection and classification of any target should be approached via phenomenology of the target and sensor's ability to capture the phenomenology properly. This implies that the characteristics of the sensor should be adequate to capture the phenomenon being observed. For example, using a microphone with 1 kHz bandwidth will not do justice to music with 20 kHz bandwidth. Selection of the features for classification should represent the phenomenon being observed.

The main focus of this thesis is to develop algorithms for detection of people, by understanding the underlying phenomenology of the signatures generated by humans and animals, and the de-

tection of these signatures using multiple sensor modalities. Furthermore we process the data obtained by different non-imaging sensors to extract the phenomenology based features and apply algorithms to detect personnel.

1.2 Thesis outline

This thesis is organized as follows: in chapter 2, target detection and classification using seismic sensor data. chapter 3 include the target detection and classification using pulse doppler sensor data. Sensors modalities and target phenomenology are discussed in each chapter. The paper is concluded in Section 4.

Chapter 2. Seismic sensor signal processing

Unattended ground sensors (UGS) are widely used in industrial monitoring and military operations. Such UGS are usually lightweight devices that automatically monitor the activities at a site, and transfer target detection and classification reports to some higher level processing center. Commercially available UGS systems make use of multiple sensing modalities (e.g., acoustic, seismic, passive infrared, magnetic, electrostatic, and video). Efficacy of UGS systems is often limited by high false alarm rates because the onboard data processing algorithms may not be able to correctly discriminate different types of targets (e.g., humans from animals) [9]. Acoustic and seismic sensors are the most common modalities used in UGS systems. In this chapter, seismic sensors are chosen for target detection and classification because they are less sensitive to Doppler effects (e.g., noise originating from moving vehicles), and atmospheric and terrain variations, as compared to acoustic sensors [10]. In a target detection and classification problem, the targets usually include human running, human walking , and animals(difference of size). Discriminating human footstep signals from other targets and noise sources is a challenging problem, because the signal to noise ratio (SNR) of footsteps decreases rapidly with the distance between the sensor and the pedestrian. Furthermore, the footstep signals vary greatly for different persons and environments. Recent literature has shown detection of heavy vehicles that radiate loud signatures [11]. However, the signatures of humans and light vehicles are usually weak and contaminated with the sensor noise.

2.1 Target detection using single seismic sensor

2.1.1 Preprocessing

In the real world, measurement signal has noise, so before the signal processing preprocessing should be done. In this section, for preprocessing the seismic sensor data, band pass filter and envelope detector are briefly explained.

2.1.1.1 Band pass filter

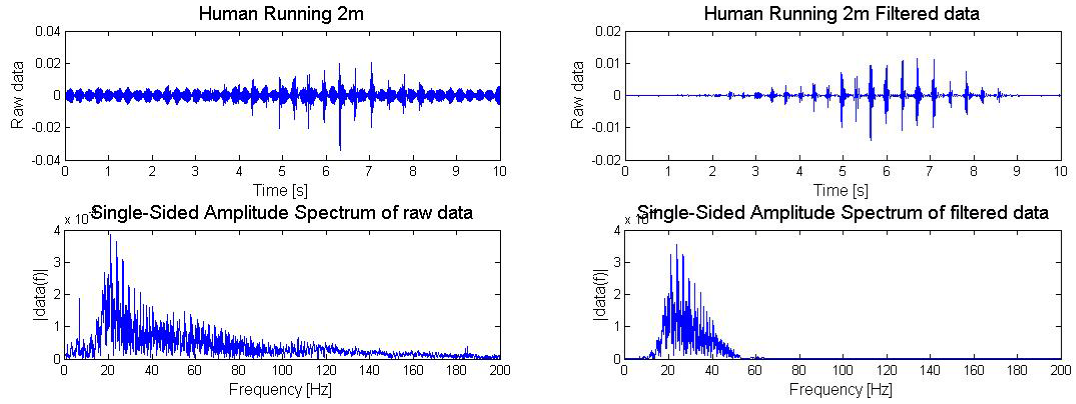
Every system has noise, especially measure data from real world. Figure 2.1 (a), (b), (c) show the raw data from seismic sensor. Each raw data occurred when human is running(a) and walking(b) across of sensor and dog passes the sensor(c). In the FFT of the raw data, there are two significant noise frequency. One is the power noise near 60Hz, and the other is unknown noise near 10Hz. In this thesis, every signal processing use the band pass filter which cut-off frequency are 10Hz and 50Hz for remove the noise. Figure 2.1 (d), (e), (f) show the band pass filtered data of (a), (b), (c).

2.1.1.2 Envelope detector

Most practical envelope detectors use either half-wave or full-wave rectification of the signal. For the higher signal level, use absolute value of filtered data. Figure 2.2 show the filtered signal and envelope detected signal. Filtered signal is thin line, and its envelope marked with thick line. The output of envelope detector is

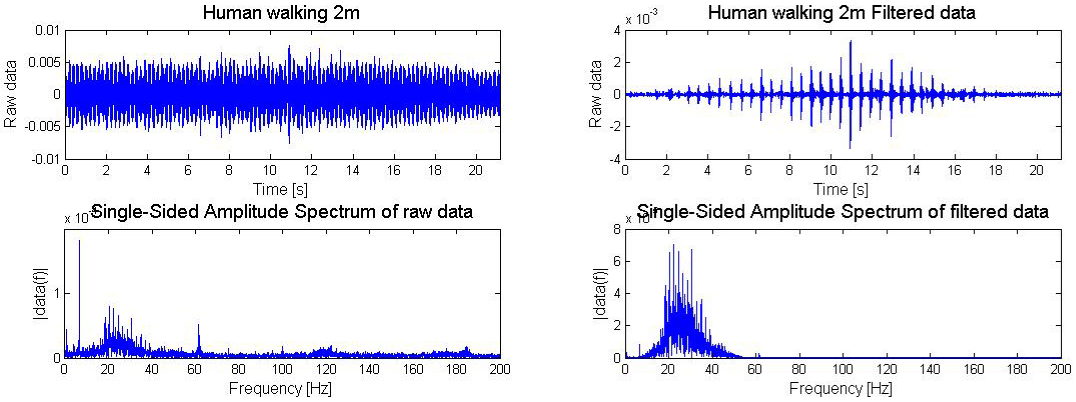
$$output = |input| \times \exp\left[-\frac{\Delta t}{RC}\right] \qquad 2.1$$

Where the $input$ is band pass filtered data, Δt is sampling time and RC is time constant.



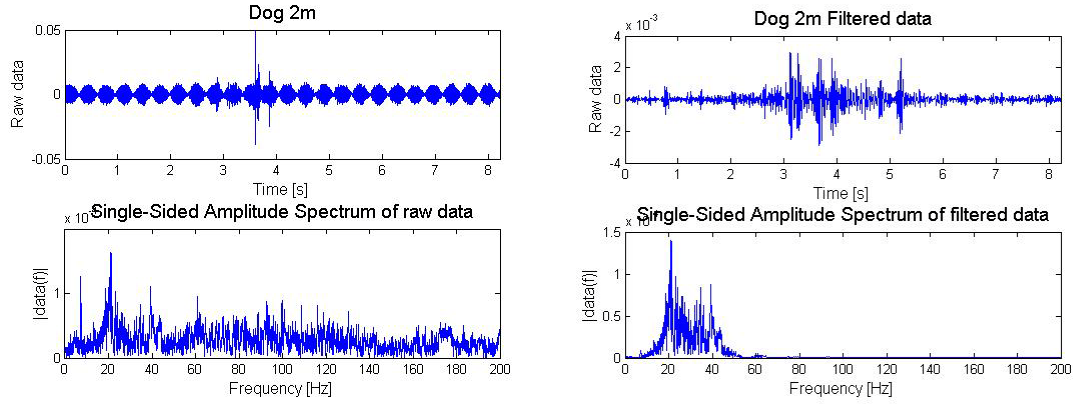
(a)

(d)



(b)

(e)

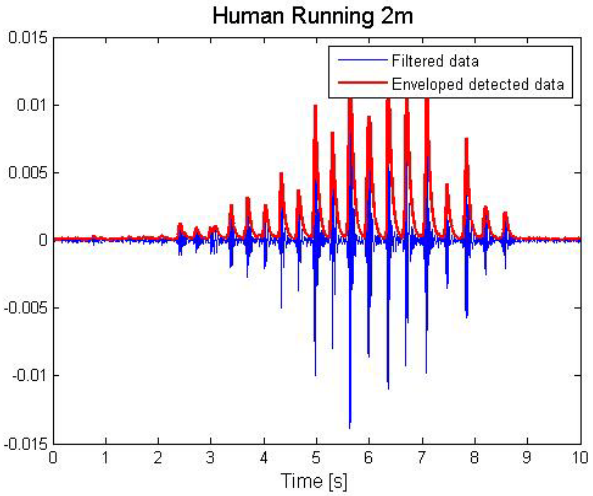


(c)

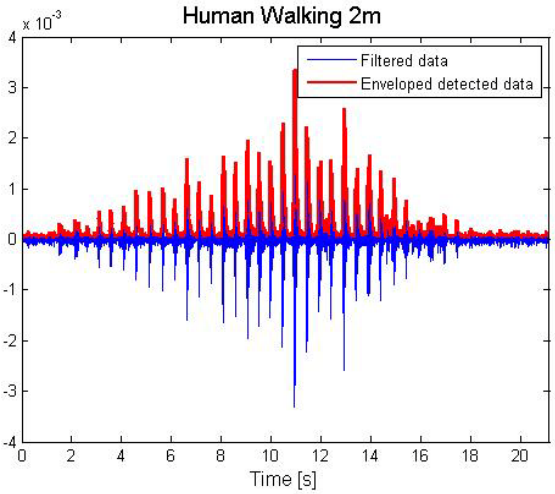
(f)

Figure 2. 1 Raw data from seismic sensor and band pass filtered data.

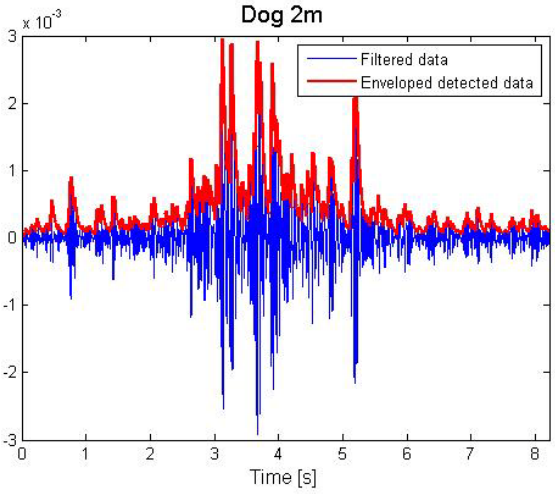
(a)raw data of human running (b)raw data of human walking (c)raw data of dog(d) filtered data of human running (e)filtered data of human walking (f) filtered data of dog



(a)



(b)



(c)

Figure 2. 2 raw data from seismic sensor and the envelope detected signal (a)human running (b)human walking (c)dog

2.1.2 Detection theory

2.1.2.1 Hypothesis

Signal detection is a classical problem of binary hypothesis testing. Under the null hypothesis H_0, the received signal $y(t)$ is composed of noise alone. The envelope of the random Gaussian noise is Rayleigh distribution.

$$H_0: \quad y(t) = n(t) \qquad \text{where} \quad n(t) \sim N(0, \sigma_0{}^2) \tag{2.2}$$

$$p_{X|H_0}\left(x \middle| H_0\right) = \frac{x}{\sigma_0{}^2} \exp\left[-\frac{x^2}{2\sigma_0{}^2}\right] \tag{2.3}$$

where $\sigma_0{}^2$ is the conditional variance and $p_{X|H_0}\left(x \middle| H_0\right)$ is the conditional probability density function of X given that the received signal is only noise (H_0). Equation(2.3) depends on the signal parameter $\sigma_0{}^2$. Estimation of facilitated by the following relation

$$\sigma_0 = \sqrt{\frac{2}{\pi}}\, E\left[X \middle| H_0\right] \tag{2.4}$$

where $E\left[X \middle| H_0\right]$ is the conditional expected value of X given H_0.

Under hypothesis H_1, the received signal $y(t)$ is the sum of the transmitted signal and noise. In the footstep signal, there are the noise n(t) and the signal s(t). In some cases, it is able to detect signals in the presence of noise by detecting the change in the mean of a test statistic. This was because the signal was assumed deterministic, and hence it's presence altered the mean of the received data. In some cases a signal is more appropriately modeled as a random process. In seismic

sensor, for which the waveform of a given footstep depends on the identity of the target, the context in which the running human, walking human, and dog. It is therefore, unrealistic to assume that the signal is know [12]. The noise and signal are the random Gaussian distribution in the seismic sensor. If both random signals n(t) and s(t) are statistically independent Gaussian distributed with the same zero mean and the different variance (σ_0^2, σ_1^2), then the variance of their sum equals the sum of their variance, i.e., if

$$n(t) \sim N(0, \sigma_0^2)$$
$$s(t) \sim N(0, \sigma_1^2)$$

then

$$H_1: \quad y(t) = s(t) + n(t) \qquad \text{where} \quad y(t) = n(t) + s(t) \sim N(0,\ \sigma_0^2 + \sigma_1^2) \sim N(0, \sigma^2)$$

$$p_{X|H_1}\left(x \middle| H_1\right) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right] \qquad 2.5$$

Here σ^2 is the conditional variance and $p_{X|H_1}\left(x \middle| H_1\right)$ is the conditional probability density function of X given that the received signal is a random signal plus noise. Figure 2.3 shows the probability density functions under H_0 and H_1.

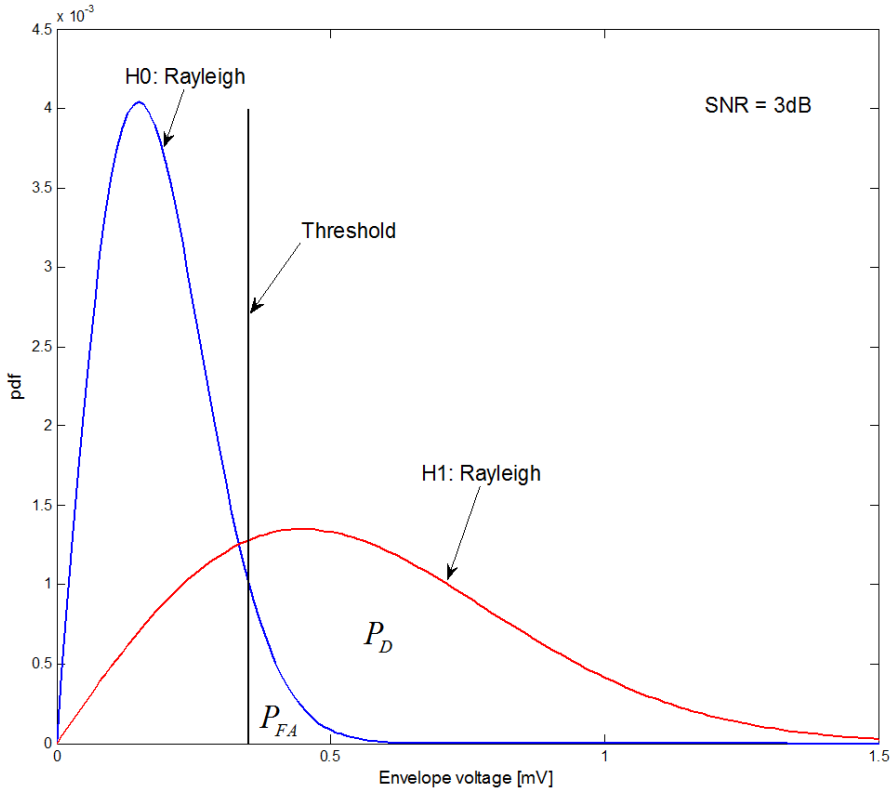


Figure 2. 3 Probability density functions of noise and signal-plus-noise

2.1.2.2 Likelihood ratio test

In statistics, a likelihood ratio test is a statistical test used to compare the fit of two hypothesis, one of which (H_0) is a special case of the other (H_1). The test is based on the likelihood ratio, which expresses how many times more likely the data are under one hypothesis than the other.

$$L(x) = \frac{p(x|H_1)}{p(x|H_0)} > \frac{P(H_0)}{P(H_1)} = \gamma \qquad 2.6$$

The function $L(x)$ is termed the likelihood ratio since it indicates for each value of x the likelihood of H_1 versus the likelihood of H_0. The threshold is determined by the prior probabilities. If, as is commonly the case, the prior probabilities are equal, LRT decide H_1 if

$$p(x|H_1) > p(x|H_0) \qquad 2.7$$

Using equation 2.3 and 2.4, solving for threshold yields

$$x \underset{H_0}{\overset{H_1}{\gtrless}} 2\alpha\sigma_0\sqrt{\frac{\ln\alpha}{\alpha^2-1}} = V_T \qquad 2.8$$

where $\alpha^2 = \dfrac{\text{variance of } s(t)+n(t)}{\text{variance of } n(t)} = \dfrac{\sigma^2}{\sigma_0{}^2} = \dfrac{\sigma_0{}^2+\sigma_1{}^2}{\sigma_0{}^2} = 1+SNR$.

2.1.2.3 Receiver operating characteristic

In signal detection theory, a receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the probability of detection(P_D) vs. the probability of false alarm(P_{FA}) at various threshold settings. In general, if both of the probability distributions for detection and false alarm are known, the ROC curve can be generated by plotting the Cumulative Distribution Function of the detection probability in the y-axis versus the Cumulative Distribution Function of the false alarm probability in x-axis.

The ROC is also known as a relative operating characteristic curve, because it is a comparison of two operating characteristics (P_D and P_{FA}) as the criterion changes [13].

Using equation 2.5, if the prior probabilities are not equal,

$$L(x) = \frac{p(x|H_1)}{p(x|H_0)} > \frac{P(H_0)}{P(H_1)} = \gamma \neq 1 \qquad 2.9$$

$$x \underset{H_0}{\overset{H_1}{\gtrless}} \sqrt{\frac{2\sigma_0{}^2\sigma^2}{\sigma^2-\sigma_0{}^2}\left(2\ln\frac{\sigma}{\sigma_0}+\ln\gamma\right)} = V_T \qquad 2.10$$

Or

$$x \underset{H_0}{\overset{H_1}{\gtrless}} \alpha\sigma_0\sqrt{\frac{2}{\alpha^2-1}\ln\left(\alpha^2\gamma\right)} \qquad 2.11$$

where $\alpha^2 = \dfrac{\text{variance of } s(t)+n(t)}{\text{variance of } n(t)} = \dfrac{\sigma^2}{{\sigma_0}^2} = \dfrac{{\sigma_0}^2+{\sigma_1}^2}{{\sigma_0}^2} = 1+SNR$

As shown in Figure 2.4, the probability of detection is the area under the signal-plus-noise curve above the detection threshold, and the probability of false alarm is the area under the noise-only curve above the threshold level V_T. Using equation 2.3, 2.5 and 2.10, P_D and P_{FA} are present as equation 2.12 and equation 2.13.

$$P_D = \int_{V_T}^{\infty} \frac{x}{\sigma^2}\exp\left[-\frac{x^2}{2\sigma^2}\right]dx = \exp\left[-\frac{{V_T}^2}{2\sigma^2}\right] = \left(\alpha^2\gamma\right)^{-\frac{1}{\alpha^2-1}} \qquad 2.12$$

$$P_{FA} = \int_{V_T}^{\infty} \frac{x}{{\sigma_0}^2}\exp\left[-\frac{x^2}{2{\sigma_0}^2}\right]dx = \exp\left[-\frac{{V_T}^2}{{\sigma_0}^2-1}\right] = \left(\alpha^2\gamma\right)^{-\frac{\alpha^2}{\alpha^2-1}} \qquad 2.13$$

Figure 2.4 plots the ROC curve solutions to equation 2.12 and 2.13 with various SNR. The circle in Figure 2.4 refer to the same prior probabilities, $\gamma = 1$. For calculating the threshold voltage, parameters of probability density function are needed. Figure 2.5 shows the detection result using the threshold. In the top graph, the blue line is envelope detected data of the seismic sensor data, and the red line is the calculated threshold voltage. The middle graph is the detection result. If the envelope detected signal is higher than the threshold, then the signal is detected. The last graph is the spectrogram of the seismic sensor signal.

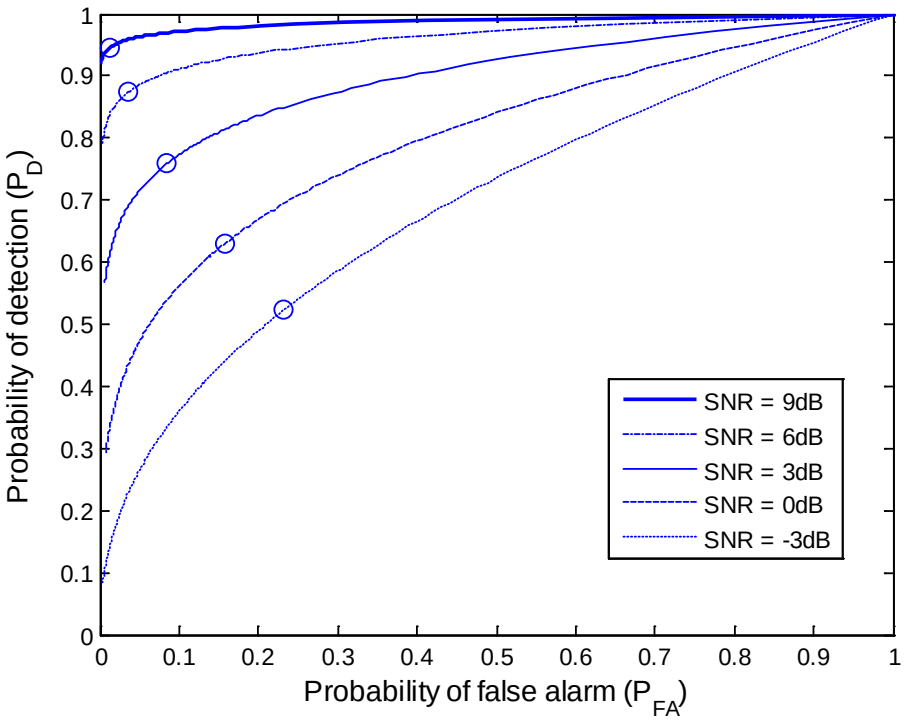


Figure 2. 4 Receiver operating characteristic curve

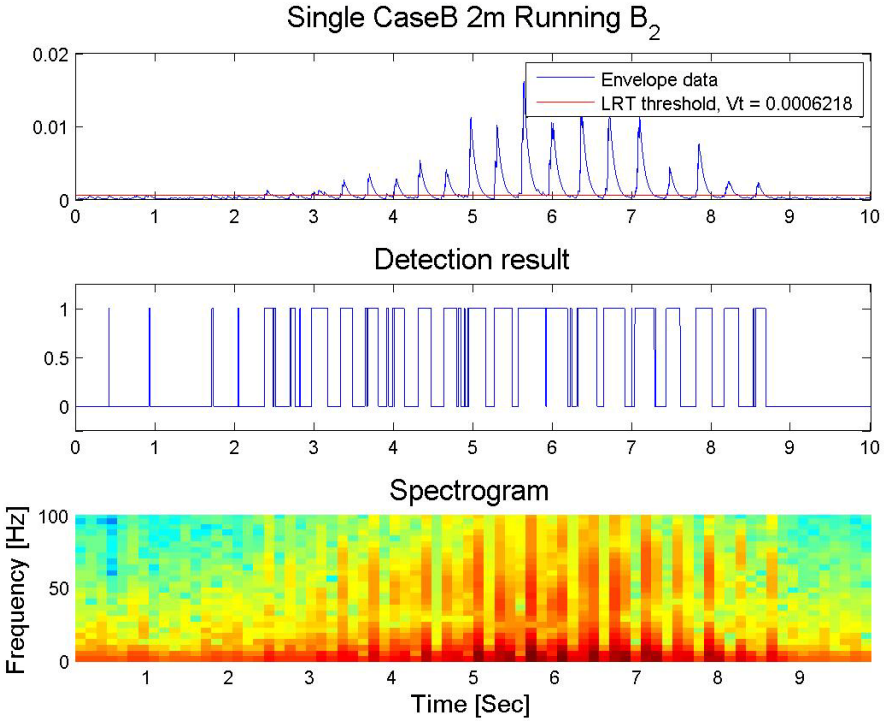


Figure 2. 5 Envelope detector result, detection result and spectrogram of the signal

2.2 Background adapted threshold

An accurate analysis of a system's detection performance is often difficult to obtain for several reasons. Predicting performance under operational conditions add additional complexity due to wide variation in target and noise statistics. Especially, performance of the seismic sensor depends on the characteristic of background. This section will focus on reduce the false alarm by considering the characteristics of the random noise on the detection system [18].

The theory developed in Chapter 2.1 provides acceptable results for target detection. In this section the Neyman-Pearson detection criterion was used to calculate the threshold level. In the Neyman-Pearson detector, the probability of false alarm p_{FA} can fixed. The detection threshold is easily obtained once the probability of false alarm is set. The false alarm happens when no signal is present but the noise level exceeds the detection threshold. As seen in Figure 2.3, the probability of false alarm is the area under the noise-only curve above the threshold level V_T.

$$P_{FA} = \int_{x=V_T}^{\infty} p_{X|H_0}(x|H_0)\,dx = \int_{x=V_T}^{\infty} \frac{x}{\sigma_0^{\,2}} \exp\left(-\frac{x^2}{2\sigma_0^{\,2}}\right) dx = \exp\left(-\frac{V_T^{\,2}}{2\sigma_0^{\,2}}\right) \quad 2.14$$

Solving for V_T yields

$$V_T = \sigma_0 \sqrt{2\ln\left(\frac{1}{P_{FA}}\right)} \quad 2.15$$

As shown in Figure 2.3, the probability of detection is the area under the signal-plus-noise curve above the detection threshold.

$$P_D = \int_{x=V_T}^{\infty} p_{X|H_1}(x|H_1)\,dx = \int_{x=V_T}^{\infty} \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)dx = \exp\left(-\frac{V_T^{\,2}}{2\sigma^2}\right) \qquad 2.16$$

where $\sigma^2 = \sigma_0^{\,2} + \sigma_1^{\,2}$, $\sigma_0^{\,2}$ is the variance of the noise and $\sigma_1^{\,2}$ is the variance of the signal. Solving for SNR gives below equation.

$$SNR = \frac{\text{variance of signal}}{\text{variance of noise}} = \frac{\sigma_1^{\,2}}{\sigma_0^{\,2}} = \frac{\sigma^2}{\sigma_0^{\,2}} - 1 \qquad \cdot 2.17$$

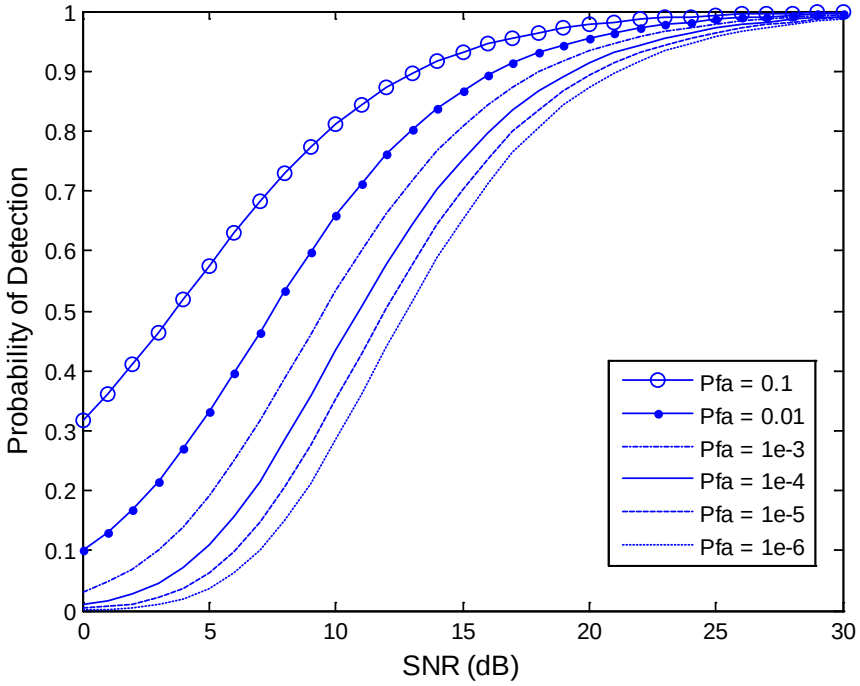
Figure 2.6 shows the probability of detection versus SNR and probability of false alarm. Figure 2.6(a) provides a quick view impact of varying system requirements on the required SNR. Lowering the false alarm rate results in higher required SNRs for the same probability of detection. Also, if the required probability of detection is reduced while maintaining the same false alarm rate, lower SNRs are required.

In this thesis we choose the 10% false alarm, using equation 2.14

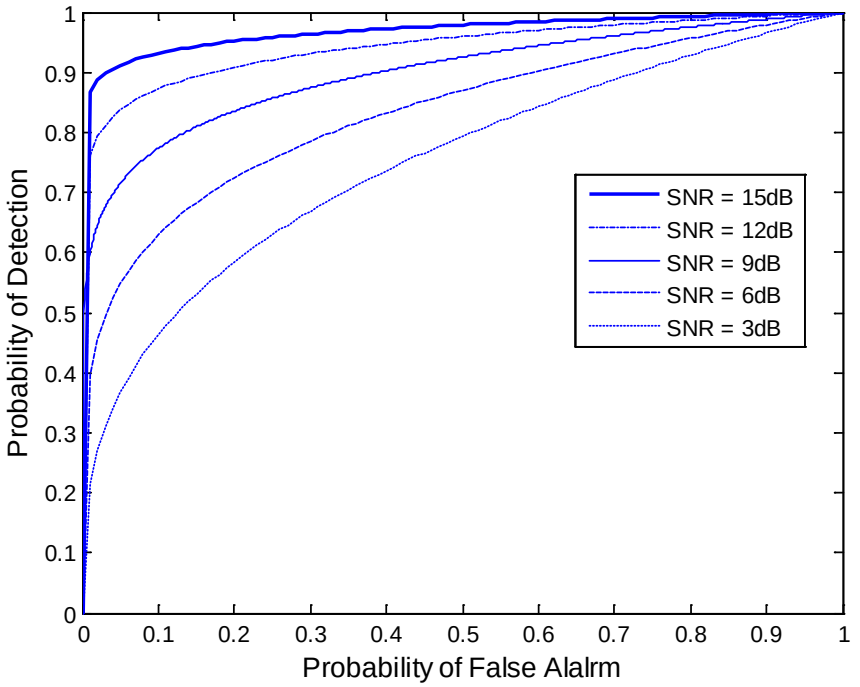
$$p_{FA} = \int_{x=V_T}^{\infty} \frac{x}{\beta^2}\exp\left(-\frac{x^2}{2\beta^2}\right)dx = \exp\left(-\frac{V_T^{\,2}}{2\beta^2}\right) = 0.1 \qquad \cdot 2.18$$

$$V_T = 2.145966\beta \simeq 2.146\beta \qquad 2.19$$

Figure 2.8 shows the data flow chart. Adaptive threshold is calculated using Equation 2.19.



(a)



(b)

Figure 2. 6 (a)Probability of detection versus SNR, and (b)receiver operation curve

If the noise statistics are not time-varying, the threshold V_T maintains the specified probability of false alarm. However, the noise statistics depend on the component of soil. And even the same location, hence the same soil component, the noise statistics varying with temperature and humidity of the soil and even the air. Figure 2.X shows that the statistical characteristic of background noise depends on the temperature and humidity of the air.

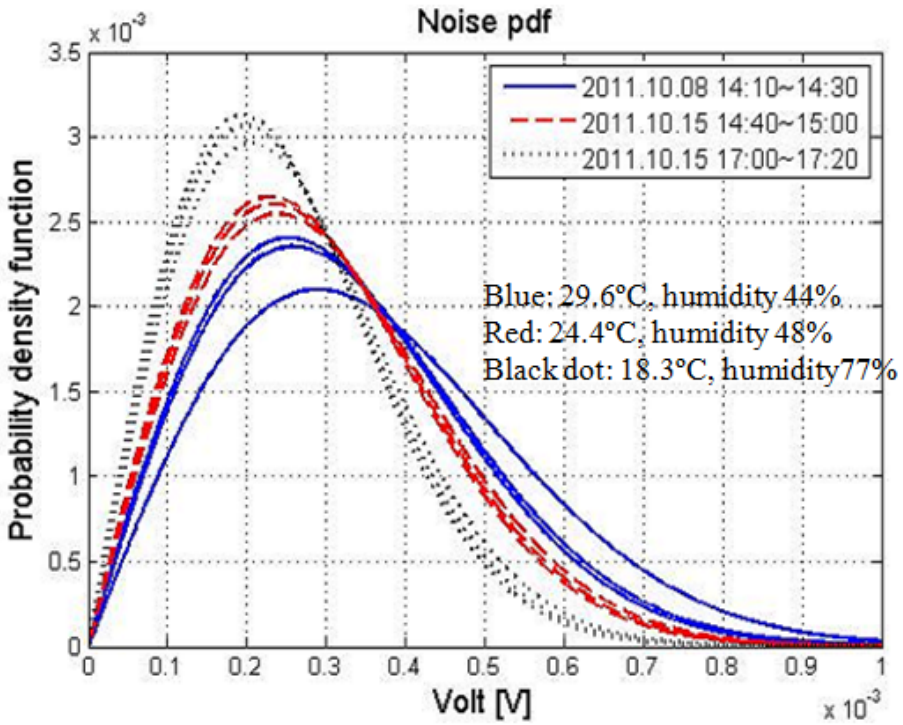


Figure 2. 7 Probability density function of noise with various temperature and humidity

(-: 29.6℃, humidity 44% ， --: 29.6℃, humidity 44% ， dot: 29.6℃, humidity 44% ）

In order to maintain a constant probability of false alarm in the presence of non-stationary noise, an adaptive threshold method is required. In this section, for guarantee the performance of detection, the algorithm update the parameter of background noise every 1000 samples and recalculate the threshold.

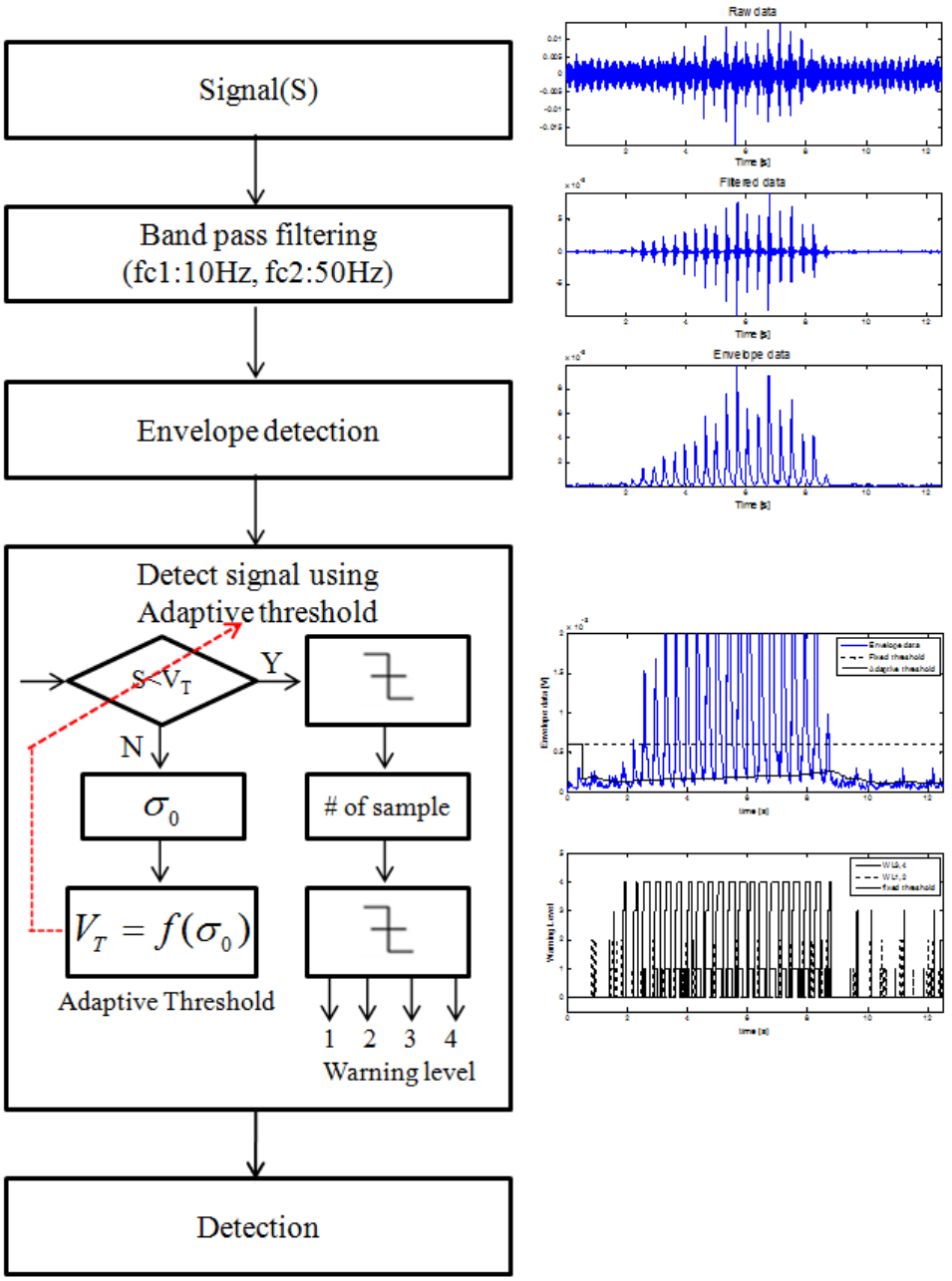
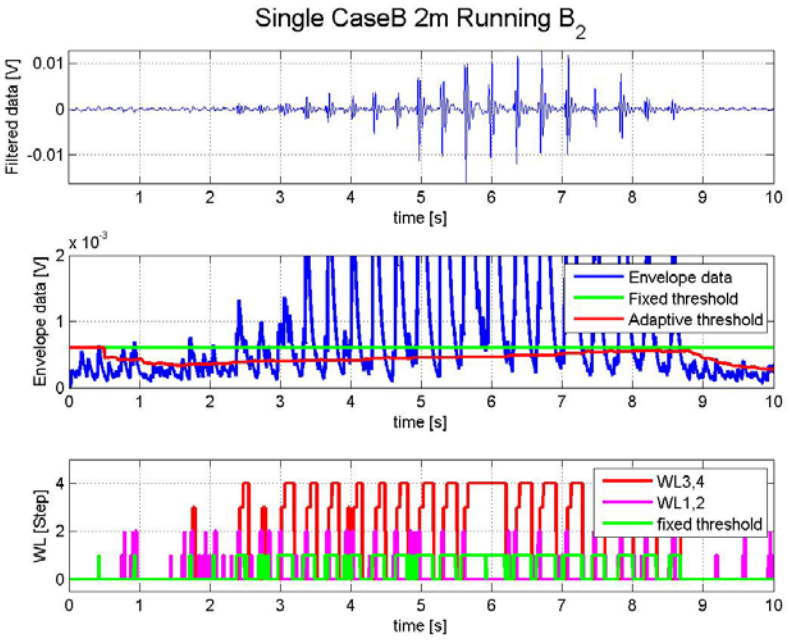
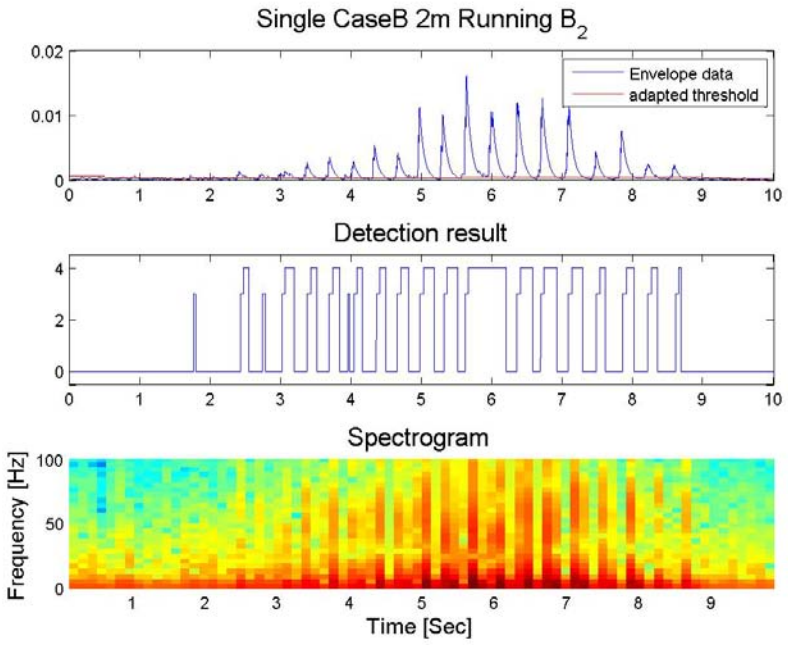


Figure 2. 8 Flow chart of the algorithm using background adapted threshold



(a)



(b)

Figure 2. 9 Detection result (a)using background adapted threshold (b)using fixed threshold

2.3 Target classification using seismic sensor

Classification is the process of separating objects into groups by comparing their attributes. The human body is an example of a complex classification system. Our eyes, skin, tongue, ears and nose constantly deliver raw information to our brains about the world around us. The brain is tasked with sorting through these raw signals and extracting important features for a given classification task. Due to the ability of the human brain to excel at classification under widely varying conditions, many classification systems today still incorporate a human operator at some point in the system. In fully-automatic classification systems, a computer performs all steps of the classification procedure with no assistance from a human operator. This dissertation involves the design of fully-automatic target classifiers. Figure 2.10 illustrates the procedural block diagram of a typical classification system.

The classification procedure begins with the acquisition of a frame of raw data. Basic preprocessing steps such as band pass filtering are performed in this stage. The data are then fed into a target detection algorithm. If the detection algorithm decides that a target is present, the target features are extracted from the data. In the feature extraction stage, the data is processed to quantify the various target features used for classification. In the training phase, feature selection would follow feature extraction. After feature extraction, the classification stage utilizes statistical models and thresholds created during the training procedure for discrimination. The classification step produces the estimated target class, and then the cycle repeats for subsequent data.

Data acquisition and target detection were discussed in earlier sections of this dissertation. Feature extraction, feature selection, classifier training, and classifier performance evaluation will be the topics of the remainder of this dissertation.

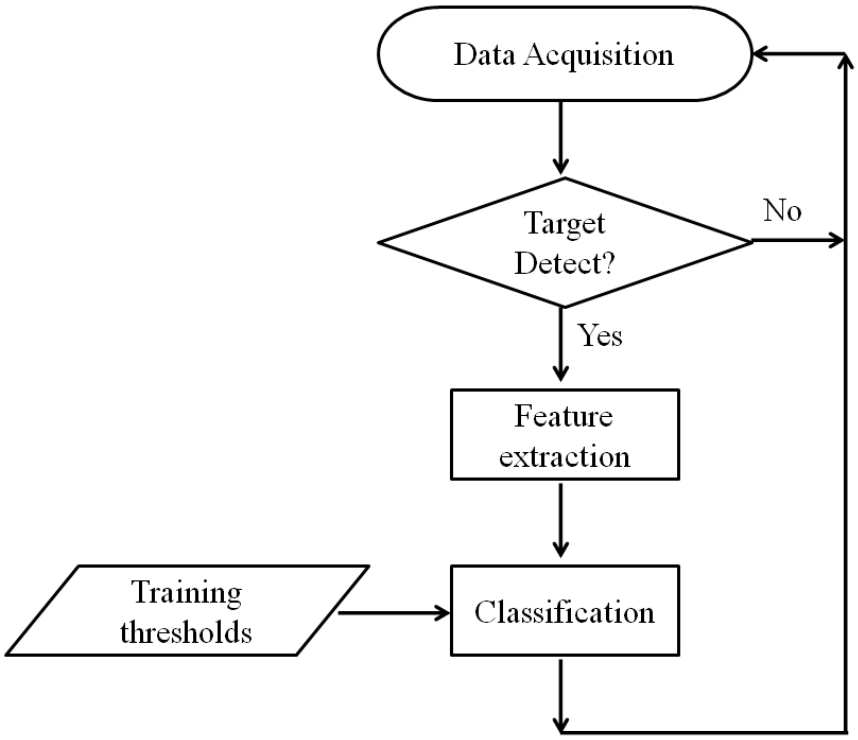


Figure 2. 10 Data flow chart for the target classification

2.3.1 Data acquisition

In order to measure the data of moving object, we design the movement path as shown in figure 2.11. At the first, we set movement as horizontal movement. Distance between the target and the sensors(Seismic sensors and PDR sensors) is 2m to 6m each 2meters. Type of movement are human running, human running, and the animals which are difference kinds of dogs. data measured date is 15th October in 2011.

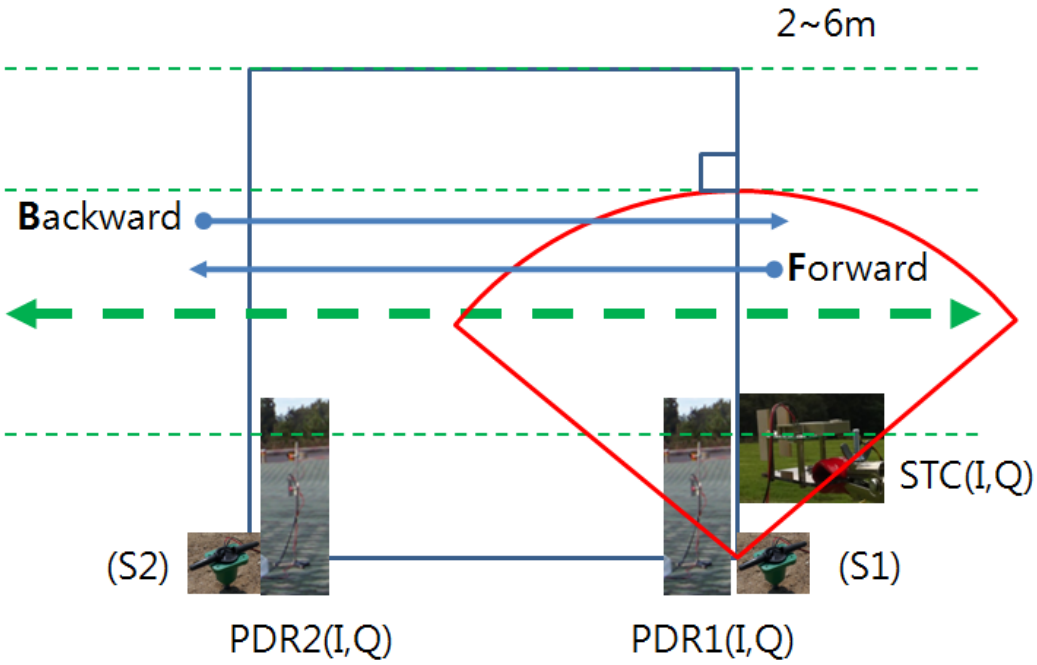


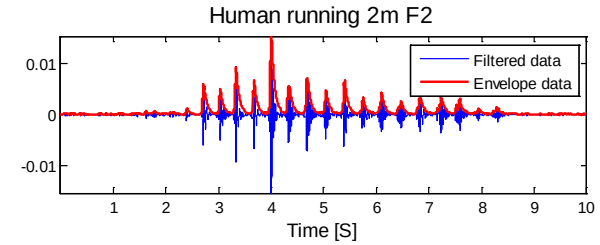
Figure 2. 11 Data measurement path

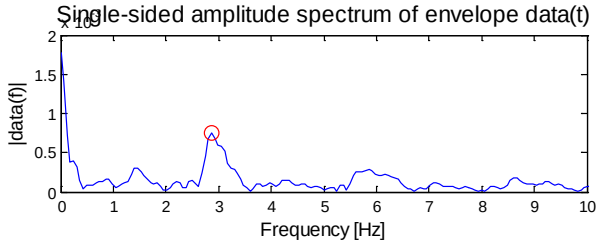
2.3.2 Feature extraction

The goal of this section is to introduce the features and feature extraction method for cadence analysis of temporal gait patterns. Obtaining robust feature vectors through the process of feature extraction and selection is vital to the design of any classification system. The chosen feature set must maintain reasonable class separability under adverse conditions (low SNR, high interference, background variation, system configuration changes, etc.) to be of practical use. Unfortunately, there is no single feature set that performs well (or is appropriate) for all classification tasks. In contrast to statistical learning theory, the mathematical framework of feature extraction and selection is quite limited. The majority of feature extraction and selection algorithms in practice today are based on ad-hoc, heuristic methods. In this section, features are extracted by footstep interval and the statistical method. After that, select the important features using Fisher's score.

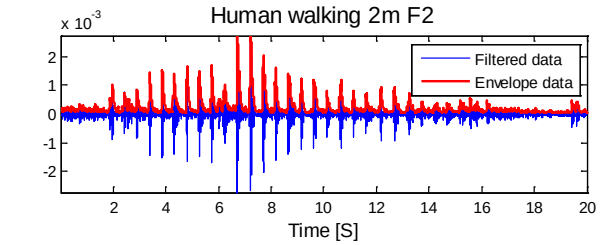
2.3.2.1 Frequency of footstep intervals

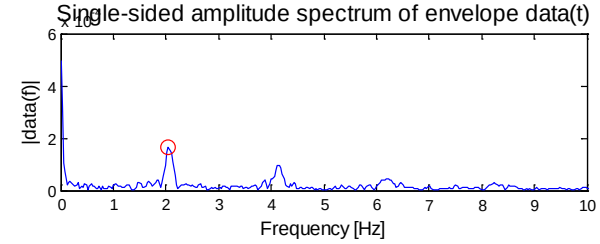
Whenever a person or an animal walks, the footfalls make audible sounds. One can analyze the signatures of human and animal footfalls and classify them into respective classes. It is estimated that the cadence of the humans walking lies between 2 to 3 Hz while the cadence of animals walking is around 2.5 - 3 Hz. figure 2.12 shows the generated frequency for each human(act running and walking) and animal(dog). Moreover, these footfalls are impulsive in nature and result in several harmonics. Even if many people are walking in a file (on a path), they tend to synchronize their stride with others and walk more or less at the same cadence. This gives a way to estimate the cadence and then classify it. Cadence estimation and classification is similar to the algorithm for seismic data and is presented in the seismic data analysis section.



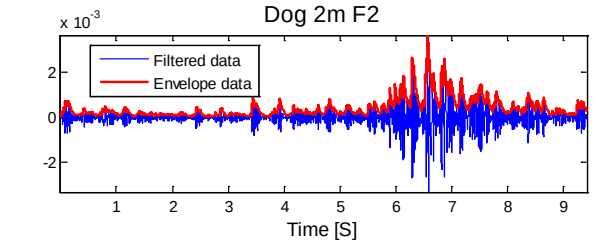


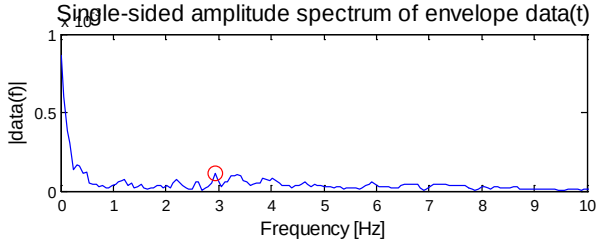
(a)





(b)





(c)

Figure 2. 12 Seismic signal and frequency generated by (a)human running (b)human walking (c)dog

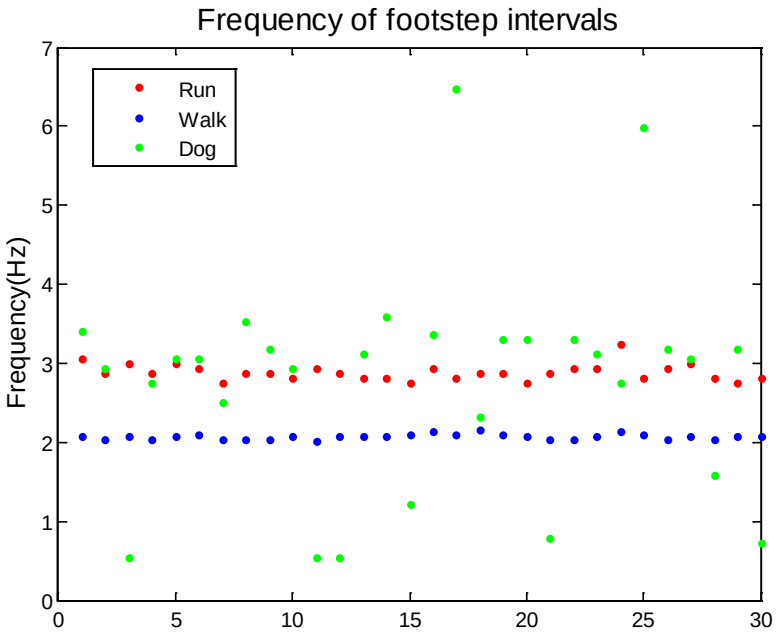


Figure 2. 13 Foot step interval for human running, human walking and dog

Figure 2.13 shows the frequency generated by human running, human walking and dog. Each of them have the specific frequency. When person is running or walking, it is 100% to classify. However, frequency of dog has bad behavior. Next section discuss about the other features like statistical features from one each pulse of the seismic data.

2.3.2.2 Single footstep

This section details the statistical features in one gait of targets. Seismic signals have specific characteristic when the target is moving. Figure 2.14 shows the one pulse for human running, human walking, and dog moving. Compare figure 2.14 (a) and (b), the waveform in time domain have different because of the gait. When human is walking, 2 kinds of vibration are occurred. One is heel drop vibration and the other is toe off vibration. Otherwise, when human is running, only toe off vibration is present. Shown in figure 2.14 (b), one pulse of walking has more vibration than running in figure2.14 (a). Dog has 4 legs, means there are more vibration than human. Even we know there are difference among the one pulse signal of targets, it should be visualized. In this section, feature extracted from one pulse signal each target using statistical features proposed by Ridge [14].

The statistic feature using amplitude is not good feature, because it depend on the distance between targets and the sensor. The goal of this thesis is classification of the moving target with any distance, so frequency related statistical features are extracted. 6 kinds of feature are extracted such as total power over all frequency, the weighted mean of the frequency of the event, measure of the frequency bandwidth, standard deviation of the power in frequency, skew of the power in frequency, kurtosis of the power in frequency. The features is in equation 2.20-2.25.

Total power: $P = \sum_{f} p_f$ 2.20

where p_f denotes the power at frequency f.

Mean frequency: $\overline{f} = \frac{\sum_f f \cdot p_f}{P}$ · 2.21

where P denotes the total power.

RMS Bandwidth: $B = \sqrt{\frac{\sum_f f^2 * p_f}{P} - \overline{f}^2}$ 2.22

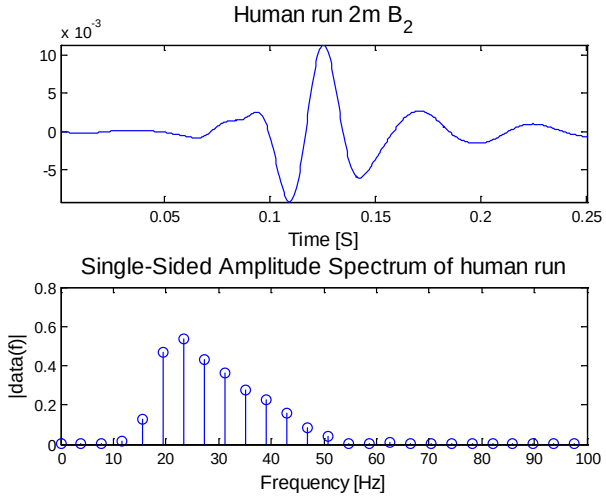
where * indicated element-wise Multiplication.

Power SDF: $P\_SDF = \frac{F\sum_f p_f{}^2}{P^2} - 1$ 2.23

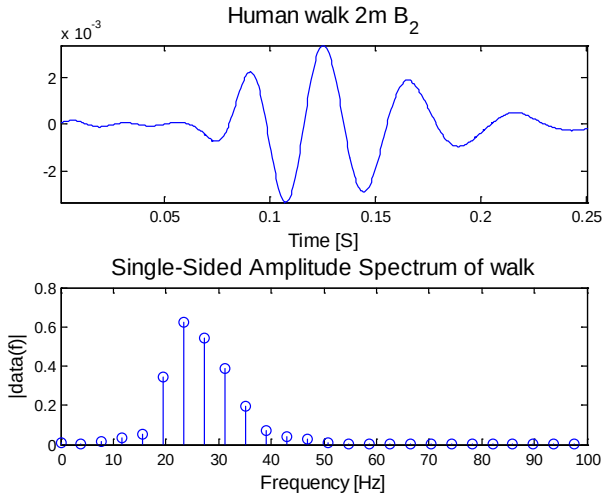
where F denotes the maximum frequency.

Power skewF: $PSF = \frac{\frac{1}{F}\sum_f \left(p_f - \overline{P}\right)^3}{\overline{P}^3}$ 2.24

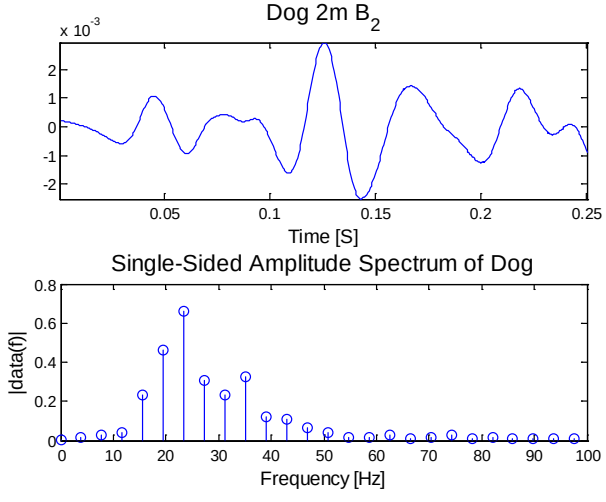
Power kurtosisF: $PSF = \frac{\frac{1}{F}\sum_f \left(p_f - \overline{P}\right)^4}{\overline{P}^4}$ 2.25



(a)



(b)



(c)

Figure 2. 14 single footstep waveform and fft spectrum of (a)human running, (b)human walking, (c)dog. (spectrum is normalized by the energy) NFFT=512, Fs=2000Hz

2.3.3 Feature selection

Feature selection is the process of pruning features acquired in the feature extraction stage into an efficient set for classification. Due to the inescapable limitations of finite sample size and finite computational resources, feature selection is a critical step in the classification process. The larger the selected feature set becomes, the harder it is for a classifier to accurately model a given class with a fixed training set size. This "curse of dimensionality" effect must be balanced with the need to include enough classification features for high-performance discrimination. Feature selection may be motivated by either the need to reduce the computational burden on the feature extractor or the desire to optimize classification performance.

This dissertation utilizes a Fisher score [15] for the feature selection. The fisher score is defined as the ratio of the between-class scatter matrix $\mathbf{S}_B$ to the average within-class scatter matrix $\mathbf{S}_W$. For the k-class problem, the average within-class scatter matrix is defined as

$$\mathbf{S}_W = \frac{1}{k}\sum_{i=1}^{k}\mathbf{S}_i \qquad 2.26$$

Where

$$\mathbf{S}_i = \sum_{\mathbf{x}\in D_i}(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T \qquad 2.27$$

and

$$\mathbf{m}_i = \frac{1}{n_i}\sum_{\mathbf{x}\in D_i}\mathbf{x} \qquad 2.28$$

where n_i is the size of the feature vector set D_i that makes up class i. Likewise, the between-class scatter matrix is defined as

$$\mathbf{S}_B = \sum_{i=1}^{k} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \qquad 2.29$$

and

$$\mathbf{m} = \frac{1}{n} \sum_{\mathbf{x}} \mathbf{x} \qquad \cdot 2.30$$

where n is the total number of feature vectors and m is the mean vector of the entire feature set. The one-dimensional Fisher score is then

$$\text{Fisher score (1D)} = \frac{diag(\mathbf{S}_B)}{diag(\mathbf{S}_W)} \qquad \cdot 2.31$$

where diag(.) extracts the diagonal elements of a matrix. The vital feature extracted using calculated Fisher's score. Figure 2.15 shows the Fisher score of the 6 extracted features. The peak in the Fisher score is from the RMS bandwidth feature and second peak is total power in frequency. In table 2.1 indicate the feature number, feature name, and the details.

Figure. 2.16 shows the RMS bandwidth and total power in frequency of 3 targets. Dots are RMS bandwidth and total power when the target is dog, cross are human walking, and stars are human running. Shown in figure 2. 16 dog is separable with human, but human running and human walking is non-separable. In figure 2. 13, human running and human walking can divide using footstep interval or frequency of gait. Figure 2.17 shows that combine 2 kinds of feature from Fisher score and footstep interval.

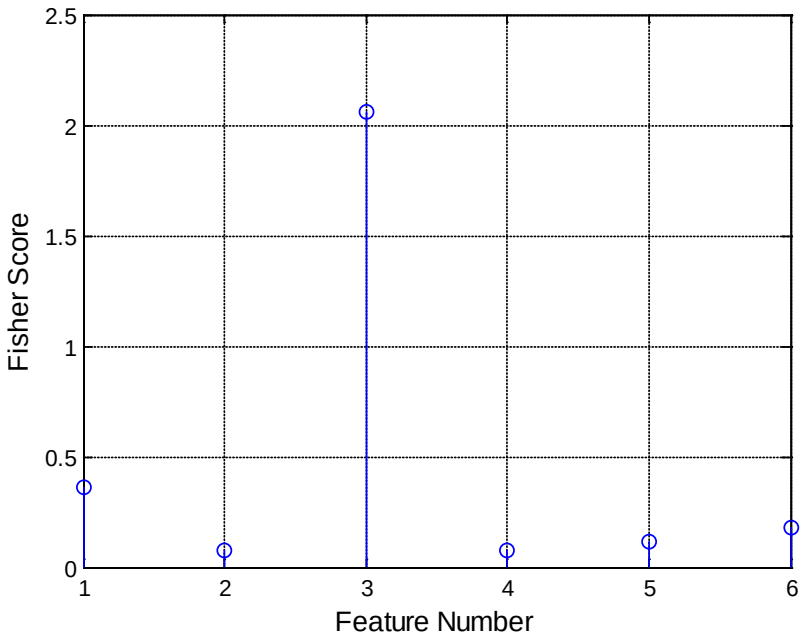


Figure 2. 15 Fisher's score of 6 kinds of statistical features

Table 2. 1 Feature number and feature name

Feature Number	*Feature name*	
1	Total power	Total power over all frequency.
2	Mean frequency	The weighted mean of the frequency of the event.
3	RMS bandwidth	A measure of the bandwidth.
4	Power SDF	Standard deviation of the power in frequency.
5	Power SkewF	Skew of the power in frequency.
6	Power KurtF	Kurtosis of the power in frequency.

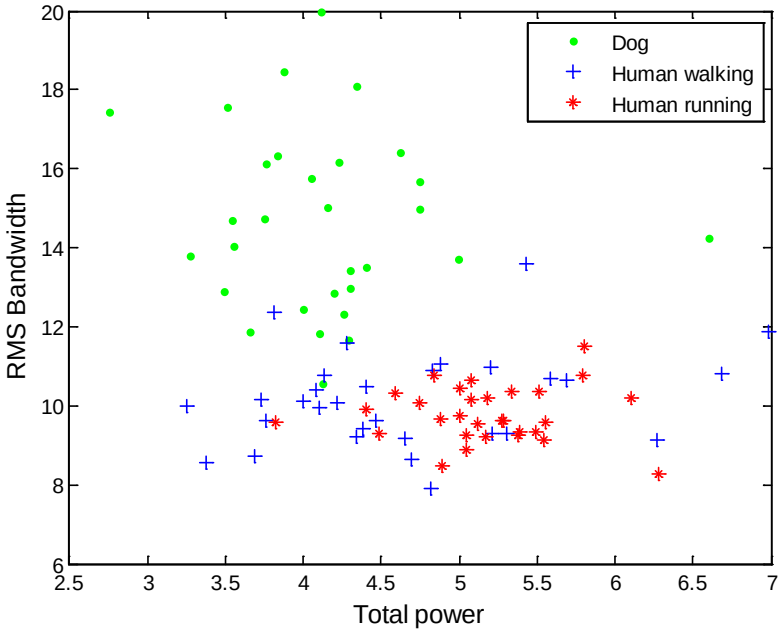


Figure 2. 16 Total power and RMS bandwidth

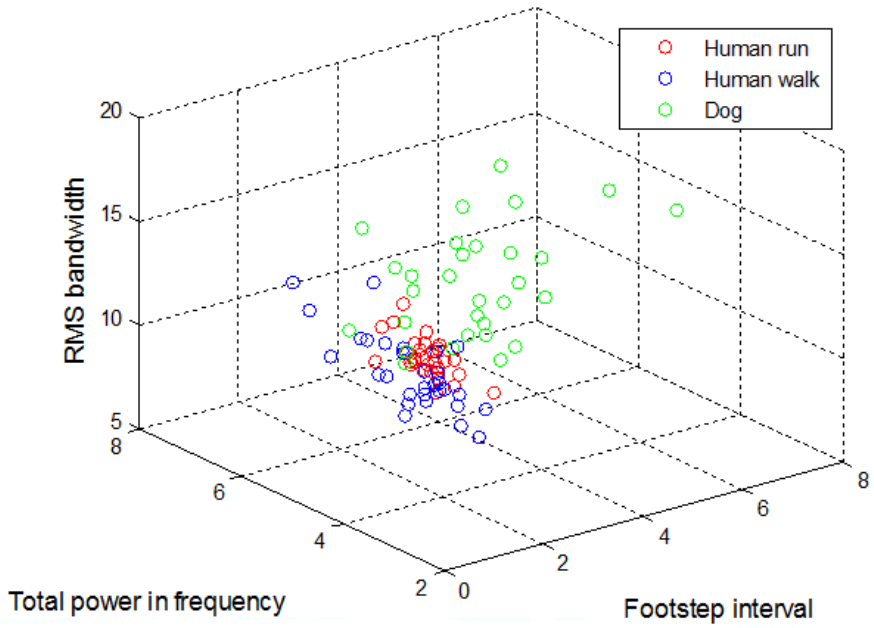


Figure 2. 17 Total power, footstep interval and RMS bandwidth

2.3.4 Classifier training

This section discuss about the classification. In the previous section, feature extraction and Fisher score are discussed. Every data analysis previous procedures, i.e band pass filtering, feature extract and dimension reduction using Fisher score. The data flow chart in Figure 2.18.

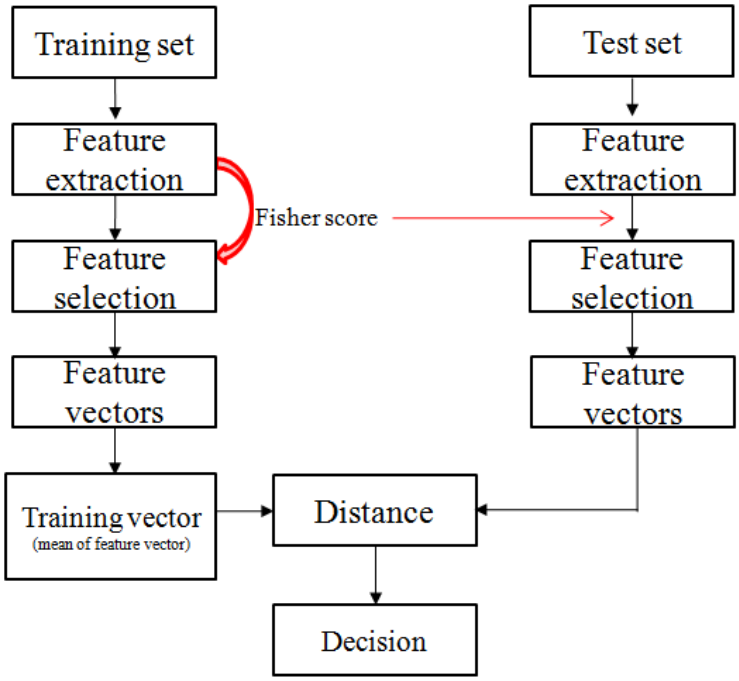


Figure 2. 18 Data flow chart for classification

In Figure2.18, the training vector for the classify is the mean value of each feature vectors. There are 2classes, human running, human walking and dog. In this section, the simplist classifying algorithm is used. The test data which belongs to a class determination of whether the average of the feature vectors of each class that you want to test and compare the distance to the feature vectors of the class represented by the minimum distance. Wherein one element of the feature vector by calculating the distance by determining the most frequent class classifies the high class.

2.3.5 Classifier performance

In this section, we present experimental results. There are 3 classes, each class has 30 data which distance between the sensor and the target is 2m, 4m and 6m. Class1 is the running action, class2 is walking action and class3 is a dog. Human cannot control the activity of dog, means activity of dog is walking or running. The classification result is in table 2.2.

Table 2. 2 Classification result with 7 features

		Chosen class			
		Human running	Human walking	Dog	Total
Actual class	Human running	27	3	0	30
	Human walking	14	15	1	30
	Dog	12	6	12	30

Table 2. 3 Classification result with 3 features using Fisher score

		Chosen class			
		Human running	Human walking	Dog	
Actual class	Human running	27	3	0	30
	Human walking	10	20	0	30
	Dog	9	2	19	30

Table 2.2 is the classification result with 7 kinds of features. When use all features, the accuracy is 90% from human running, 50% from human walking, 40% from dog and totally 60%. After feature dimension reduced using Fisher score, use 3 kind of features, the accuracy of human running is same as 90%, human walking is 66% , dog is 63.33%. Totally the accuracy is 73.33%, it improve than use all features. The result is in table 2.3.

Chapter 3. PDR signal processing

3.1 Introduction

Chapter 2 dealt the problem of detection and classification using seismic sensor. This chapter uses the Doppler information for detection and classification using pulse doppler radar(PDR). In this chapter, a support vector machine (SVM) is proposed for classifying dog, running human and walking human subject using Doppler information. SVM is a binary classifier that is popularly used in a machine learning area to construct a maximal-separating hyperplane. SVM has been used extensively for many diverse classification problems for its superior performance over other classification methods, such as the Fisher linear discriminator and the Bayesian decision method. In the radar signal-processing community, SVM has been used mostly for the target recognition in synthetic aperture radars [19–22]. The considered classified classes in this study are dog, running human and walking human. In order to recognize the classes, the time-varying Doppler signatures are examined knowing that different motions generate different Doppler signatures. The features of Doppler information are extracted from spectrogram. For the generation of training data set of SVM, 3 subjects are measured. After a training process, the SVM classifies its class based on the input features. A multi level SVM is implemented using a decision tree structure and binary SVM. The training process and the resulting classification accuracy are reported.

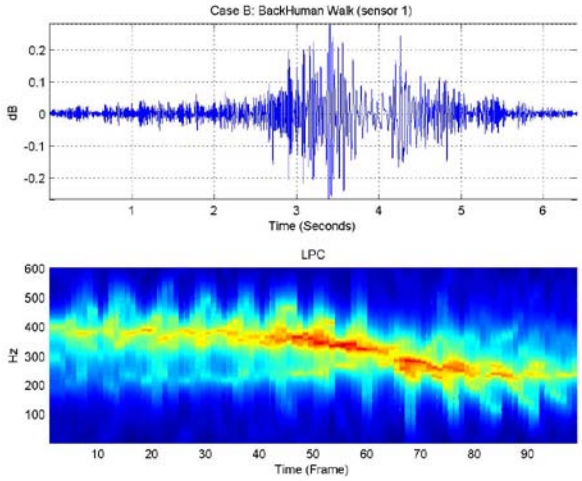
In PDR applications the used signals are complex-valued. In order to process them, the classifier should deal with complex data. Last of this chapter, for the advanced topic, support vector machines in the complex plane is developed [23, 24].

3.2 Pulse Doppler Radar signal

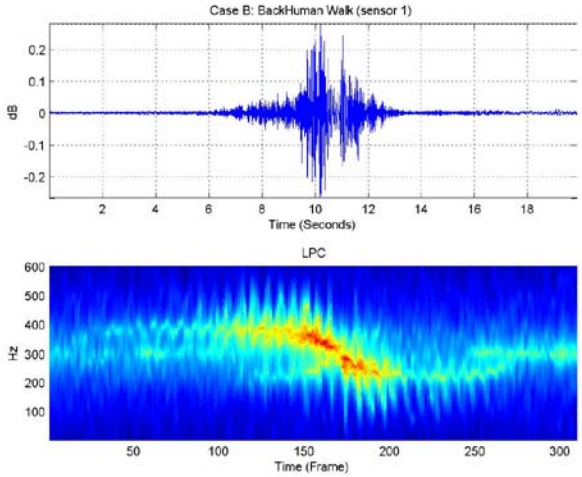
The basic Doppler radar transmits a signal at a single frequency and when the wave is reflected off a moving object, the frequency of the transmitted signal is Doppler shifted. The resulting frequency received by the radar can be used to determine the radial velocity of the moving target. For this reason, Doppler-based radars are excellent for detecting movement while suppressing any stationary clutters in the background. Low-cost Doppler radars using off-the-shelf components or even single-chip modules are available due to the recent advances in microelectronics. Unique and interesting aspect of Doppler returns from humans is the appearance of microDoppler features. Micro Dopplers are generated from the limb motions of humans, and they contain valuable information related to human motions like running, walking and clawing[16-18].

3.3 Target detection using PDR signal

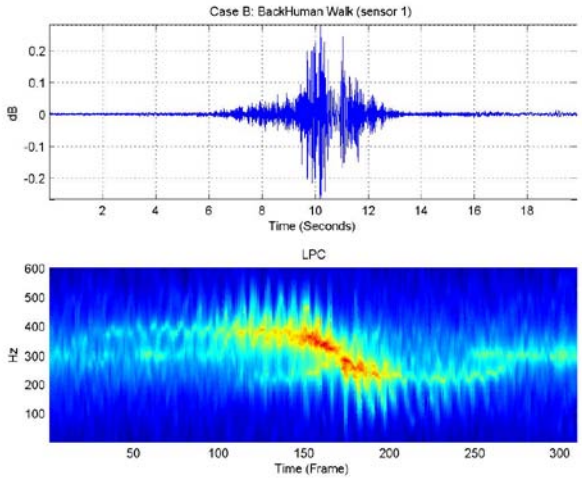
Figure 3.1 is a spectrogram of human target walking across the pulse doppler radar at a range of 2m. In Figure 3.1 (a) the detection threshold is set to -15dBm. No pixels false alarm within this detection, but numerous missed detections occurs. In Figure 3.1 (c) a much higher threshold of -37dBm is used. This threshold dramatically increase the false alarm pixels but considerably decreases the number of missed detection. The threshold of -23dBm used in Figure 3.1 (b) achieves an acceptable balance between the number of false alarm pixels and missed detection pixels.



(a) -15



(b)-23



(c)-30

Figure 3. 1 Waveform and spectrogram for human target walking detection threshold level: (a)-15dBm, (b)-23dBm, (c)-30dBm

3.4 Target classification using PDR

3.4.1 Feature extraction

Obtaining robust feature vectors through the process of feature extraction and selection is vital to the design of any classification system. The chosen feature set must maintain reasonable class separability under adverse conditions to be of practical use. Unfortunately, there is no single feature set that performs well for all classification tasks. In contrast to statistical learning theory, the mathematical framework of feature extraction and selection is quite limited. The majority of feature extraction and selection algorithms in practice today are based on ad-hoc, heuristic methods. One could potentially use the time-frequency (spectrogram) coefficients directly as a feature set. Although spectrograms contain fine detail that is useful for visual classification, they fail to provide a compact data representation for efficient computation. A typical spectrogram-based feature vector might contain around 16,000 elements—far too large for small or even moderately large training sets to appropriately model the resulting high-dimensional feature space.

The discrete-time real cepstrum is defined as the inverse discrete Fourier transform of the log-magnitude spectrum of sequence, $c[n]$.

$$c[n] = \mathrm{IDFT}\left[\log\left| \mathrm{DFT}\left(x[n]\right) \right| \right] \qquad 3.1$$

Filter-bank methods and linear predictive coding(LPC) are the two primary approaches to calculating the cepstral coefficients. Linear predictive coding is an all-pole system that models a signal $x[n]$ as the linear combination of the previous p samples

$$\hat{x}[n] \cong \sum_{k=1}^{p} a_k x[n-k] \qquad 3.2$$

where a_k are the LPC coefficients. In this thesis, 16th order of LPC coefficients are used as feature vector as shown in Figure 3.2.

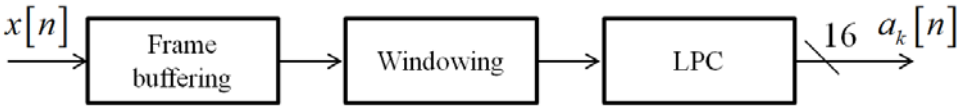


Figure 3. 2 Feature extraction using 16th order LPC

3.4.2 Classifier

3.4.2.1 Support vector machine

This section will introduce the basic ideas behind support vector machines (SVMs) as developed by V. Vapnik, et al. The approach will follow the developments in the classical SVM work [25] and the excellent tutorial in [26]. Support vector machines (SVMs) are quite general learning machines introduced to solve problems in the fields of pattern recognition, regression estimation, and density estimation. The development presented here will focus on the pattern recognition problem.

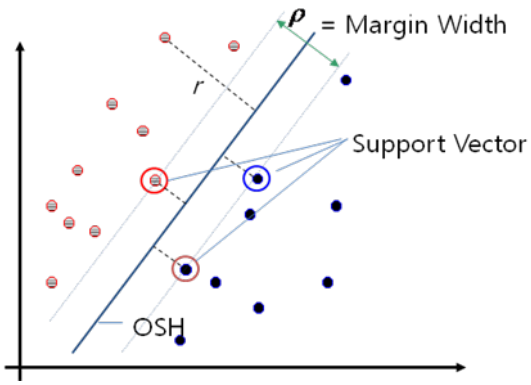


Figure 3. 3 Support vector machines

Figure 3.3 illustrates the basic concepts of support vector machines. Support vector machines take a set of input features that are typically not linearly separable and transform the features into a higher dimensional feature space. In this new space, the data are separated by a linear hyperplane. In Figure 3.3, the support vectors are the feature vectors lying on the two dashed hyperplanes. The support vectors provide the most information for the classification task but they are also the feature vectors that are hardest to classify. Many hyperplanes can separate the data in Figure 3.3. However, the goal of support vector machines is to find the *optimal* hyperplane so that the classifier will perform well on both the training set and unseen test samples.

Support vector machines are a supervised learning technique. The supervisor provides a set of l labeled training data vectors (each of dimension d)

$$(\mathbf{x}_i, y_i) \quad i = 1, \cdots, l \quad \mathbf{x}_i \in \mathbf{R}^d \quad y_i \in \{-1, 1\} \tag{3.3}$$

where the are the training data vectors and the $\mathbf{x}_i$, y_i are the class labels assigned by a supervisor (note the use of bold font for vectors). The general goal of SVMs is to find the optimal hyperplane (decision rule) that separates the training data well and also generalizes appropriately for unseen test data.

The support vectors lie on planes defined by the equations $y_i(\mathbf{x}_i \cdot \mathbf{w} + b) = 1$. Thus, the "hard" margin is equal to $\frac{2}{\|\mathbf{w}\|}$. Support vector machines provide the optimal separating hyperplane by maximizing the margin. In the case of separable data, the hard margin is maximized which involves minimizing $\|\mathbf{w}\|$, minimizing the norm of the hyperplane's normal vector. For non-separable data, the concept of a "soft" margin is introduced [19]. A soft margin is created by adding the non-negative slack variables ζ_i. The slack variables represent the shortest distance between an incorrectly classified training vector and its correct classification region $\zeta_i = |y_i - (\mathbf{x}_i \cdot \mathbf{w} + b)|^2$. Slack variables help the SVM to minimize training set errors.

Maximizing the soft margin requires minimizing the following objective function

$$L_p = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} \zeta_i \tag{3.4}$$

subject to the following constraints

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq 1 - \zeta_i \quad \text{for} \quad y_i = 1 \tag{3.5}$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq 1 - \zeta_i \quad \text{for} \quad y_i = -1 \tag{3.6}$$

$$\zeta_i \geq 0 \quad \forall i \tag{3.7}$$

The first term on the right hand side of equation 3.4 is the margin maximizing term responsible for the well-known generalization capability of SVMs. The second term on the right hand side of equation 3.4 is related to the empirical risk. Minimizing the empirical risk is tantamount to minimizing the training set error. The number of errors on the training set can be brought down to zero by utilizing a highly complex classifier. However, overly complex classifiers rarely perform well on unseen test samples. The cost parameter *C* controls the number of support vectors used to model the decision boundary. The greater the number of support vectors, the more complex the decision boundary. As the number of support vectors gets smaller, the decision boundary becomes progressively smoother. A balance must be achieved between using not enough support vectors to model the inherent complexity of the problem, and using too many support vectors which will over-train the model and perform poorly on unseen test data. Obtaining the optimal model complexity is the subject of structural risk minimization as developed in [25].

Returning now to the optimization problem, constraint equations 3.5 and 3.6 can be combined into a single equation

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \zeta_i \qquad 3.8$$

The method of Lagrange multipliers will be used to perform this inequality constrained optimization problem. The primal Lagrangian formed from objective function 3.4 and constraint equations 3.7 and 3.8 is

$$L_p = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\alpha_i\left\{y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \zeta_i\right\} - \sum_{i=1}^{n}\beta_i\zeta_i \qquad 3.9$$

where α_i and β_i are the non-negative Lagrange multipliers. The primal Lagrangian is now minimized with respect to $\mathbf{w}$. To find the minimizer $\mathbf{w}$, start by solving the following set of simultaneous equations

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0: \quad \frac{\partial L_{pd}}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i \qquad \cdot 3.10$$

$$\frac{\partial L_p}{\partial b} = 0: \quad \frac{\partial L_{pd}}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0 \qquad \Rightarrow \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \qquad 3.11$$

$$\frac{\partial L_p}{\partial \zeta_i} = 0: \quad C - \alpha_i - \beta_i = 0 \qquad \Rightarrow \quad C = \beta_i + \alpha_i \qquad 3.12$$

The last equation constrains the α_i to a hypercube $0 \le \alpha_i \le C$. The dual Lagrangian can now be formed by substituting equations 3.10~3.12 back into the primal Lagrangian. The following dual Lagrangian and its associated constraints are found by substitution of equations 3.10~3.12.

$$L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \qquad 3.13$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \quad \alpha_i \ge 0 \text{ for } \forall \alpha_i \qquad 3.14$$

Note that (by a clever choice) the slack variables ζ_i do not show up in the dual Lagrangian. To minimize the objective function of equation 3.4, the dual Lagrangian must be maximized with respect to the α_i.

The resulting decision function to be applied to test data is

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \qquad 3.15$$

where

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i \qquad b = y_k - \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i^T \mathbf{x}_k \text{ for any } \alpha_k > 0. \qquad 3.16$$

3.4.2.2 Non-linear support vector machines for non-separable training data

So far, only linear decision functions have been considered. The key idea of support vector machines is to transform the *input space* (which is likely non-separable) into a higher dimensional *feature space* where the training data are linearly separable. The linear decision functions in the feature space are typically non-linear in the input space.

Define a mapping Φ which maps the d-dimensional input space into a higher-dimensional (potentially infinite-dimensional) feature space H.

$$\Phi : \mathbf{R}^d \rightarrow H \tag{3.17}$$

With this mapping, the inner products $\mathbf{x}_i \cdot \mathbf{x}_j$ in input space become $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ in the feature space. Defining the precise form for Φ is often difficult and/or costly. Thus, we seek a kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ so that defining Φ is unnecessary. Mercer's Theorem provides us the answer. Mercer's Theorem states that the mapping and the kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \sum_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ exist if and only if [26]

$$\forall g(\mathbf{x}) \text{ with } \int g(\mathbf{x})^2 d\mathbf{x} < \infty \tag{3.18}$$

then

$$\int K(\mathbf{x}, \mathbf{y}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0 \tag{3.19}$$

Therefore, for all pairs of mappings and kernels that satisfy Mercer's Theorem, the kernel function $K(\mathbf{x}_i, \mathbf{x}_y)$ takes the place of $\mathbf{x}_i \cdot \mathbf{x}_j$ in the non-linear optimization problem. Thus, the

results from linear SVMs can immediately be extended to the non-linear case as follows

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \qquad 3.20$$

Again, note that only the support vectors contribute to the sums. The support vector approach allows us to perform the inner products in high-dimensional feature space using a kernel with complexity determined by the number of support vectors.

There is no known method for selecting the most appropriate kernel function for a given classification task. SVM practitioners typically begin by trying kernels that have been known to produce good results. Some of the most common kernel functions are

Polynomial: $K(\mathbf{x}, \mathbf{x}_i) = \left[\gamma(\mathbf{x} \cdot \mathbf{x}_i) + \delta\right]^d$ 3.21

Radial Basis Function(RBF) or Gaussian: $K(\mathbf{x}, \mathbf{x}_i) = \exp\left[-\gamma \left\|\mathbf{x} \cdot \mathbf{x}_i\right\|^2\right]$ 3.22

Sigmoid: $K(\mathbf{x}, \mathbf{x}_i) = \tanh\left[\gamma(\mathbf{x} \cdot \mathbf{x}_i) - \delta\right]$. 3.23

The kernel function parameters γ, d and δ are typically chosen by heuristic methods.

Although SVMs were presented here for solving pattern recognition problems, SVMs have shown promise in solving the more general problems of regression estimation and density estimation. Readers who wish to further investigate the theory of support vector machines should consult the classic work [25] and the comprehensive tutorial in [26].

3.4.3 Classifier performance

The radial basis function (RBF) was chosen for the support vector machine kernel. The RBF kernel is well-behaved mathematically, uses only a single free parameter γ, and performs well on many classification tasks. When using the RBF kernel, the SVM designer must provide appropriate values for the cost parameter *C* and the RBF kernel parameter γ.

In this section, binary classifier discuss after that multi-level SVM using binary tree architecture. First, 3kinds of binary classifier (dog, human walking), (dog, human running), and (human running, human walking) is discussed and then (human running, human walking, dog) using binary tree architecture and binary SVM classifiers.

3.4.3.1 Binary classifier

This section will analyze the performance 3 kinds of binary classifier (dog, human walking), (dog, human running), and (human running, human walking). The feature set utilized in this section is 16th order LPC coefficient and window size is 128 samples. For the training 20% of data was used. SVM parameters are $C = \infty$ and $\gamma = 0.5$.

Figure 3.4 shows the training result when the classes are dog and human walking. The classification result is in table 3.1. Even training 2~8m distance, the classification results are 100% with 16m distance for dog and 10m distance for human walking. Table 3.1 indicated that dog and human walking can be separate perfectly using SVM classifier.

Dog, human running case is in figure 3.5 for the training result and the classification result is in table 3.2. The classification performance is around 85% for dog and 70% for human running.

For classifying human running and human walking, the training result is in figure 3.6 and the classification result is in table 3.3. The target is human but the activity is different as running and walking. Performance of the classifier is 100% for human walking and over 95% for human running.

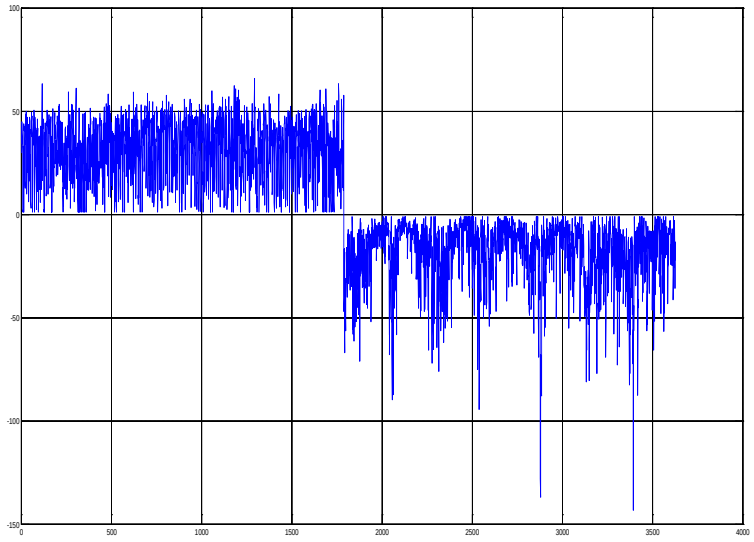


Figure 3. 4 Training result, Dog=1 and Human Walking=-1

Table 3. 1 Classification result of Dog versus Human walking

Test data	*Dog*	*Human walking*
Trained distance (2,4,6m B/F, 8m B)	35/35 (100%)	35/35 (100%)
Total distance (D 2:2:16, H: 2:2:10m)	80/80 (100%)	50/50 (100%)

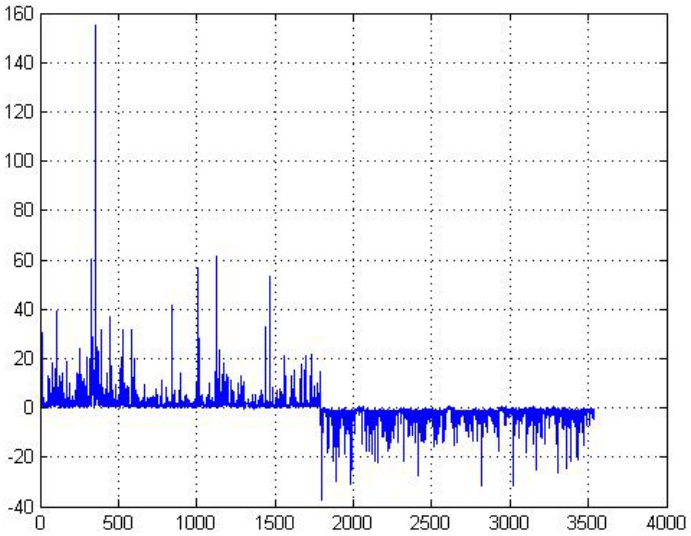


Figure 3. 5 Training result, Dog = 1and Human running = -1

Table 3. 2 Classification result of Dog versus Human running

Test data	*Dog*	*Human running*
Trained distance (2,4,6m B/F, 8m B)	30/35 (85.71%)	25/35 (71.43%)
Total distance (D 2:2:16, H: 2:2:10m)	69/80 (86.25%)	35/50 (70%)

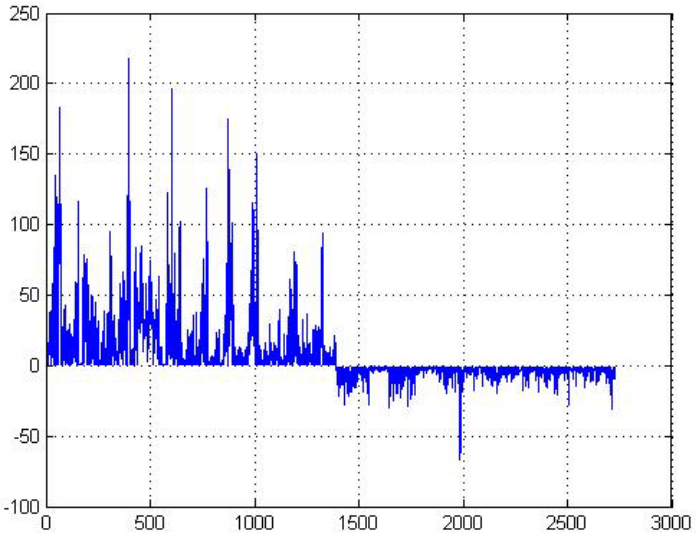


Figure 3. 6 Training result, Human running = 1and Human walking = -1

Table 3. 3 Classification result of Human running versus Human walking

Test data	*Human running*	*Human walking*
Trained distance (2,4,6m B/F, 8m B)	45/45 (100%)	44/45 (97.78%)
Total distance (D 2:2:16, H: 2:2:10m)	50/50 (100%)	48/50 (96%)

3.4.3.2 Binary tree architecture classifier

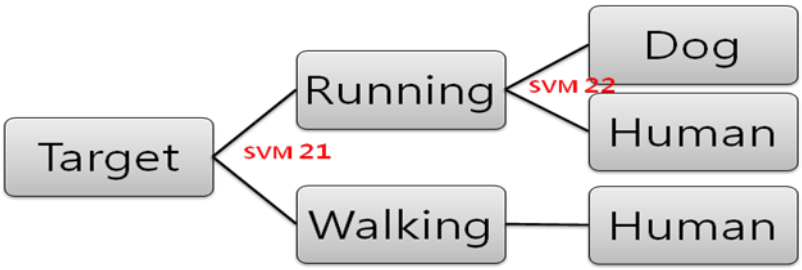
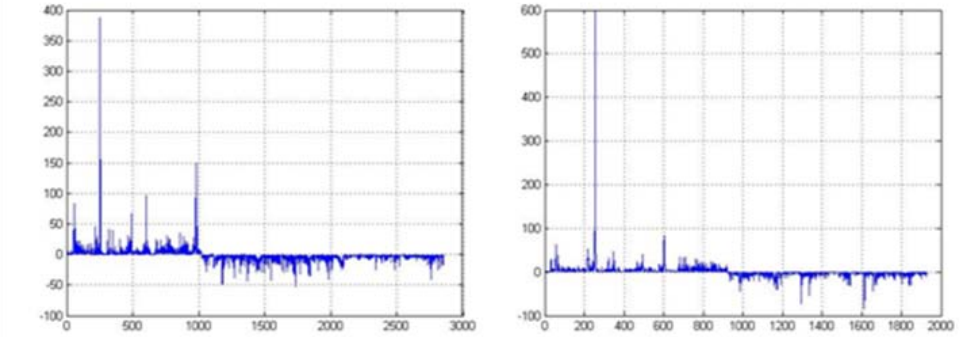


Figure 3. 7 Binary tree architecture classifier using SVM

Previous section dealt only 2 classes classifier, binary classifier. Dog and human walking are perfectly classified. Human activity such as running and walking are classified over 95% performance. But dog and human running classifier performance is not enough. For improve the performance of dog and human classifier, in this section will consider special classifier using binary tree architecture (BTA) and SVM. The classifier is in figure 3.7. First tree, SVM21, classify the activity running and walking even it is dog or human. Usually dog is classified the running class, because dog has four legs. After that, second tree, SVM22, classify human and dog. Both of SVM parameters are $C = \infty$ and $\gamma = 0.5$ and Gaussian kernel is used. The training result is shown in figure 3.8. The classification performance is in table 3.4. Accuracy is 93.33% for dog and 97.78% for human.



(a) (b)

Figure 3. 8 Training result.

(a)SVM21(Run=1, Walk=-1), (b)SVM22(Dog=1, Human=-1)

Table 3. 4 Classification result of Figure 3.7

	Predicted class			
		Dog	Human	% correct
Actual class	Dog	84	6	93.33 %
	Human	4	176	97.78 %

3.5 Target classification using complex-valued SVM

In PDR applications the used signals are complex-valued. In order to process them, it is not a good idea to use two separate processors for real and imaginary parts, since there is cross-information that cannot be ignored. In this section complex-valued SVM classifier which process the complex-valued signals measured by PDR to identify moving targets from the background [23].

3.5.1 complex-valued SVM

SVM is widely applied in the field of pattern recognition, but features which used to classify are almost real valued data. Complex-valued SVM can classify the moving target using real valued data, imaginary valued data, and cross-information data. To design complex-valued SVM, consider slack variables of real and complex axis, and use the Karush-Kuhn-Tucker (KKT) conditions for complex data. Also, apply radial basis function (RBF) as a kernel function which use a distance of complex values. This section presents a summarized derivation of the complex-valued SVM classifier, but similar derivations hold true for other SVM approached. Also, using the kernel function, the result is directly applicable to nonlinear machines. Figure 3.9 illustrates the basic concept of complex-valued SVM with two slack variables ζ and ξ.

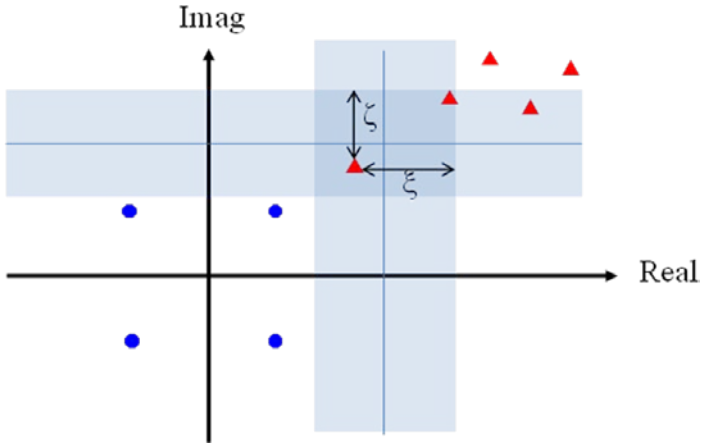


Figure 3. 9 Complex-valued support vector machine

Derivation of the complex-valued SVM is similar with real-valued SVM. Maximizing the margin requires minimizing the follow objective function

$$L_p = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i + C\sum_{i=1}^{n}\zeta_i \qquad 3.24$$

subject to the following constraints

$$\mathrm{Re}[y_i(\mathbf{w}^T\mathbf{x}_i + b)] \le 1-\xi_i, \qquad \xi_i \ge 0 \qquad 3.25$$

$$\mathrm{Im}[y_i(\mathbf{w}^T\mathbf{x}_i + b)] \le 1-\zeta_i, \qquad \zeta_i \ge 0 \qquad 3.26$$

where ξ_i is the slack variable in the real part of the real part analogously for ζ_i in the imaginary part.

The prime-dual Lagrange functional can be written with Lagrange multipliers $\alpha_n, \beta_n, \lambda_n, \eta_n \ge 0$.

$$\begin{aligned} L_{pd} = {} & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i + C\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\alpha_i(\mathrm{Re}[y_i(\mathbf{w}^T\mathbf{x}_i + b)] - 1 + \xi_i) - \sum_{i=1}^{n}\lambda_i\xi_i \\ & - \sum_{i=1}^{n}\beta_i(\mathrm{Im}[y_i(\mathbf{w}^T\mathbf{x}_i + b)] - j + j\zeta_i) - \sum_{i=1}^{n}\eta_i\zeta_i \end{aligned} \qquad 3.27$$

Besides, the KKT conditions are equation 3.28~3.31.

$$\frac{\partial L_{pd}}{\partial \mathbf{w}} = 0: \quad \frac{\partial L_{pd}}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i - \sum_{i=1}^{n} j\beta_i y_i \mathbf{x}_i = 0 \qquad 3.28$$

$$\frac{\partial L_{pd}}{\partial b} = 0: \quad \frac{\partial L_{pd}}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i - \sum_{i=1}^{n} j\beta_i y_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} (\alpha_i + \beta_i) y_i = 0 \qquad 3.29$$

$$\frac{\partial L_{pd}}{\partial \xi_i} = 0: \quad C - \lambda_i - \alpha_i = 0 \quad \Rightarrow \quad C = \lambda_i + \alpha_i \qquad 3.30$$

$$\frac{\partial L_{pd}}{\partial \zeta_i} = 0: \quad C - \eta_i - \beta_i = 0 \quad \Rightarrow \quad C = \eta_i + j\beta_i \qquad 3.31$$

$$\alpha_i, \beta_i, \lambda_i, \eta_i \geq 0 \qquad 3.32$$

$$\alpha_i y_i (\mathrm{Re}[y_i (\mathbf{w}^T \mathbf{x}_i + b)] - 1 + \xi_i) = 0 \qquad 3.33$$

$$\beta_i y_i (\mathrm{Im}[y_i (\mathbf{w}^T \mathbf{x}_i + b)] - 1 + \zeta_i) = 0 \qquad 3.34$$

$$\lambda_i \xi_i = 0, \quad \eta_i \zeta_i = 0 \qquad 3.35$$

Applying equation 3.28~3.35 to equation 3.XX, we obtain an optimal solution for the complex-valued SVM classifier weights

$$\mathbf{w} = \sum_{i=1}^{n} \psi_i^* y_i^* \mathbf{x}_i^* \qquad 3.36$$

where $\psi_i = \alpha_i + \beta_i$. This result is analogous to the one for the real-valued SVM classifier problem, except that now Largrange multipliers α_i and β_i for both the real and the imaginary components have been considered.

The norm of the complex-valued coefficients can be written as

$$\|\mathbf{w}\|^2 = \sum_{i=1}^{n} \sum_{j=n_0}^{n} \psi_i^* \psi_j y_i^* y_j \mathbf{x}_i^H \mathbf{x}_j = \Psi^H Y^H R Y \Psi \qquad 3.37$$

where $R_{ij} = \mathbf{x}_i^{\,H}\mathbf{x}_j$.

The functional for complex-valued SVM classifier model has exactly the same form as real-valued SVM. For complex-valued SVM classifier, the expression for the functional is simply

$$\textit{Maximize}: \quad L_{pd} = -\frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{n}(\alpha_i + j\beta_i) = -\frac{1}{2}\Psi^H Y^H R Y \Psi + \Psi$$

$$\textit{Subject to}: \quad \sum_{i=1}^{n}\psi_i y_i = 0, \quad 0 \le \psi_i \le C(1+j)$$

$$\Downarrow$$

$$\textit{Minimize}: \quad \frac{1}{2}\Psi^H Y^H R Y \Psi - \Psi \qquad \textit{Where} \quad \psi_i = \alpha_i + j\beta_i, \quad R_{ij} = \mathbf{x}_i^{\,H}\mathbf{x}_j$$

$$\textit{Subject to}: \quad \sum_{i=1}^{n}\psi_i y_i = 0, \quad 0 \le \psi_i \le C(1+j).$$

Figure 3.9 illustrates the basic concept of complex-valued SVM. SVM is widely applied in the field of pattern recognition, but features which used to classify are almost real valued data. Complex-valued SVM can classify the moving target using real valued data, imaginary valued data, and cross-information data. To design complex-valued SVM, consider slack variables of real and complex axis, and use the Karush-Kuhn-Tucker (KKT) conditions for complex data. Also, apply radial basis function (RBF) as a kernel function which use a distance of complex values.

3.5.2 Feature extraction

Figure 3.10 shows inphase component and quatradure component from the PDR when human move across the PDR 4m distance. Spectrogram using inphase or quadrature component only is shown in figure 3.11(a). Figure 3.11(b) shows the spectrogram using both of inphase and quadrature component as a complex value. There is no negative frequency in spectrogram of real value, so figure 3.11 (a), the spectrogram is symmetric long the frequency axis. Compare two of figure 3.11 (a) and (b), spectrogram of complex value has more information than real value. Previous section ignored the complex information.

In this section, 16th order LPC coefficients used as feature vector, same reason as previous section but the spectrogram is different. Section 3.4 used real value of LPC coefficients, in that case cross information in real and complex LPC coefficient is ignored. Section 3.5 used complex value of LPC coefficients

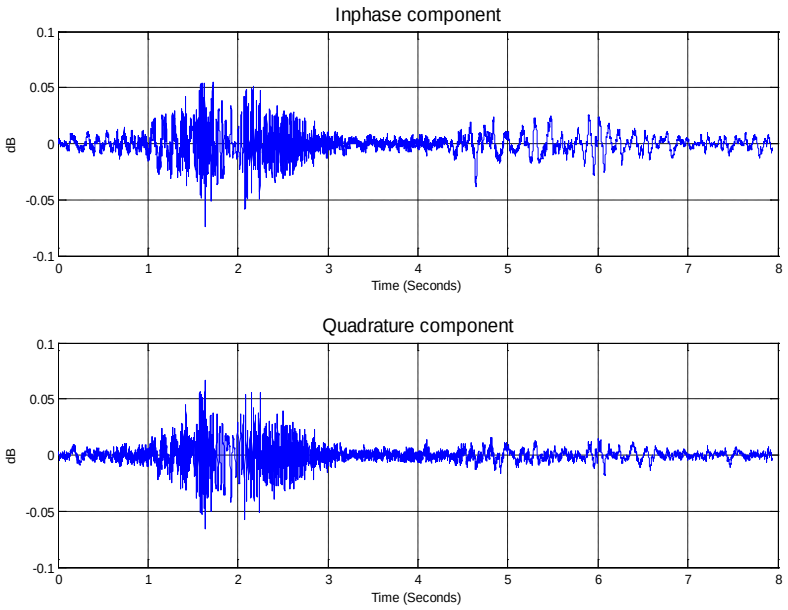
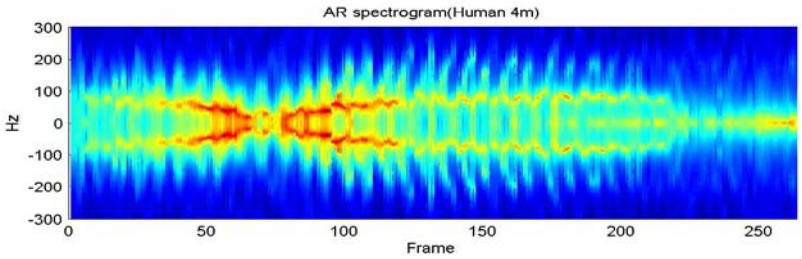
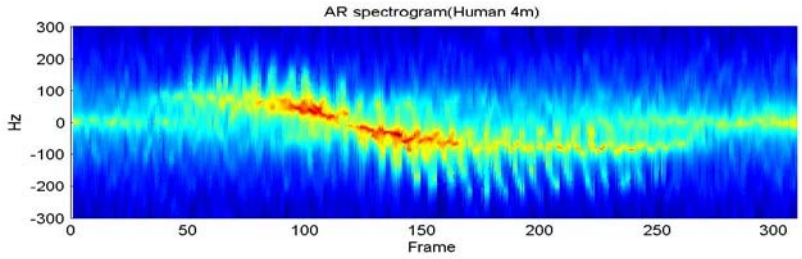


Figure 3. 10 Data from the PDR



(a)



(b)

Figure 3. 11 Spectrogram of (a)real value, (b)complex value

3.5.3 Classifier performance

To evaluate the performance of the complex-valued SVM, complex valued data from PDR were classified using real-valued SVM and complex-valued SVM. Data is collected 10 times each distance(2, 4, 6, 8, 10m) between the target and PDR. Target limited human and animal(dog). Movement path of the targets is across the PDR. Each target have 50 data, 7% of them used for training and last of them used for the performance of the classifier. The radial basis function (RBF) was chosen for the support vector machine kernel. The RBF kernel is well-behaved mathematically, uses only a single free parameter γ, and performs well on many classification tasks. When using the RBF kernel, the SVM designer must provide appropriate values for the cost parameter *C* and the RBF kernel parameter γ. SVM parameters are $C = 2048$ and $\gamma = 0.5$ for the real-valued SVM, $C = 1448$ and $\gamma = 0.5$ for the complex-valued SVM.

Table 3.5 shows the classification result of the real-valued SVM and the complex-valued SVM is shown in Table 3.6. Complex-valued SVM classification was improved compared to real-valued SVM for dog and human, respectively 8%, 10% have been improved.

Table 3. 5 Performance of the real-valued SVM

		Predicted class		
		Dog	Human	% correct
Actual class	Dog	35	15	70 %
	Human	16	34	68 %

Table 3. 6 Performance of the complex-valued SVM

		Predicted class		
		Dog	Human	% correct
Actual class	Dog	39	11	78 %
	Human	11	39	78 %

Chapter 4. Conclusion

4.1 Conclusion

This thesis research achieved its primary scientific objective to perform robust, automatic detection and classification of moving targets using seismic sensor and pulse doppler radar on a stationary platform.

A major contribution of this thesis was the collection, processing, and analysis of a diverse seismic and doppler signature database. The database included datasets with targets moving at numerous ranges, different of targets and activities. The target classes included were human running human walking and animal. Evaluation of theory on realistic experimental data is vital to the advancement of knowledge. The data can be used to rigorously evaluate new classification, detection, and feature selection algorithms. Computer simulation analysis also plays a crucial role in theoretical development. The experimental data collected by this thesis research can be utilized to improve the accuracy of computer models.

This thesis contributed a novel set of high-performance seismic and doppler based features. The Fisher score used for selecting a feature set. In addition, the feature set included both statistical and the linear predictive coding (LPC) residual energy feature. The selected feature set was shown to perform well on the seismic sensor and doppler radar based target classification problem.

The design and detailed analysis of target classification algorithms based on support vector machine (SVM) and binary tree architecture (BTA) classifiers were designed to accomplish high-performance target classification. In order to process the complex-valued signals from pulse doppler radar, complex-valued SVM classifier was derived. Complex-valued SVM classifier processes the complex-valued signals measured by PDR to identify moving targets from the background.

4.3 Future work

Although the scientific objectives of this thesis were met, more research can be performed to further improve our knowledge of seismic sensor and pulse doppler radar based classification. One of the first priorities for continued research would be fusion multi-sensors data and analysis them. One more is collecting more datasets. Acquiring more data is important for several reasons. First, collecting a wider variety of target types improves our understanding of seismic and doppler signal. especially, target doppler responses can be used to improve radar simulation models and to study various electromagnetic scattering phenomena. In addition to increasing our knowledge of doppler, expanding the micro-Doppler signature database is vital for improving classification results.

The results of this thesis support the argument that feature selection is often more critical than classifier selection. Classification performance is fundamentally limited by the inherent separability of the feature set. The Fisher score initialized sequential backward selection algorithm utilized in this dissertation performed well. However, other feature selection algorithms such as those incorporating genetic algorithms (GA) may perform better. In addition to feature selection, improving feature extraction is also vital for obtaining better performances. Perhaps the most important feature to add would be in the one pulse of signal. Signals from seismic sensor consist of several footsteps, each footstep has one pulse of seismic.

References

[1] K. M. Houston and D. P. McGaffigan,"Spectrum Analysis Techniques for Personnel Detection Using Seismic Sensors", *Proc. of SPIE*, Vol. 5090, pp. 162-173, 2003.

[2] Raju Damarla and David Ufford, "Personnel detection using ground sensors", *Proc. of SPIE*, Vol. 6562, 656205, 2007.

[3] Thyagaraju Damarla, James Sabatier, and Alex Ekimov, "Personnel detection at a border crossing", Proc. of Military Sensing Symposium National, July 2010, Las Vegas.

[4] G. Succi, D. Clapp and R. Gambert, "Footstep, Detection and Tracking", *Proc. of SPIE*, Vol. 4393, pp. 22-29, 2001.

[5] A. Sunderesan, A. Subramanian, P. K. Varshney and T. Damarla, "A copula based semi-parametric approach for footstep detection using seismic sensor networks", *Proc. of SPIE*, Vol. 7710, 77100C, 2010.

[6] A. Mehmood, J. M. Sabatier, M. Bradley and A. Ekimov, "Extraction of the velocity of walking human's body segments using ultrasonic Doppler", *The Journal of the Acoustical Society of America*, Vol. 128, no.5, pp. EL316-EL322, Oct. 2010.

[7] Alexander Ekimov and James M. Sabatier, "Passive ultrasonic method for human footstep detection", *Proc. of SPIE*, Vol. 6562, 656203, 2007.

[8] James M. Sabatier and Alexander Ekimov, "Range limitation for seismic footstep detection", *Proc. of SPIE*, Vol. 6963, 69630V-1, 2008.

[9] G. L. Goodman, “Detection and classification for unattended ground sensors.,” in *Proceedings of Information Decision and Control 99*, pp. 419–424, 1999.

[10] Y. Tian and H. Qi, “Target detection and classification using seismic signal processing in unattended ground sensor systems,” in *International Conference on Acoustics Speech and Signal Processing*, 2002.

[11] J. Altmann, “Acoustic and seismic signals of heavy military vehicles for co-operative verification,” *Journal of Sound and Vibration*, Vol. 273, no. 4-5, pp. 713 – 740, 2004.

[12] Kay, Steven M. *Fundamentals of Statistical signal processing, Volume 2: Detection theory*. Prentice Hall PTR, pp140-141, 1998.

[13] Swets, John A.; *Signal detection theory and ROC analysis in psychology and diagnostics : collected papers*, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[14] A.Ridge, "Scalar parameter extraction from FFT data for transient association", Internal report, Sound Concepts Department, QinetiQ, Winfrith, UK, 2003.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2000.

[16] J. Geisheimer, W. Marshall and E. Greneker, “A continuous –wave (CW) radar for gait analysis,” Conf. Signals, Systems and Computers, Vol. 1, pp. 834 - 838, Nov. 2001.

[17] J. L. Geisheimer, E. F. Greneker and W. S. Marshall, “High-resolution Doppler model of the human gait,” *SPIE Proc. Radar Sensor Technology and Data Visualization*, Vol. 4744, pp. 8-18, July 2002.

[18] J. Lee, C. Lee, J. Bae, J. Kwon, "Target detection algorithm based on seismic sensor for adaptation of background noise," *Journal of The Institute of Electronics Engineers of Korea*, *50*(7), pp.258-266, 2013.

[19] Y. L. Chang, J. P. Fang, J. N. Liu, H. Ren, and W. Y. Liang, "A simulated annealing feature extraction approach for hyperspectral images," in *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pp. 3190-3193, 2007.

[20] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2000.

[21] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognition Letters*, Vol. 10, pp. 335-347, Nov 1989.

[22] J. S. Lee, "Hybrid genetic algorithms for feature selection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 26, pp. 1424-1438, Nov 2004.

[23] Y. Kang, J. Lee, J. Bae, C. Lee, "Target classification algorithm using complex-valued support vector machine," *Journal of The Institute of Electronics Engineers of Korea*, *50*(4), 182-188, 2013.

[24] Martínez-Ramón, Manel, and Christos Christodoulou. "Support vector machines for antenna array processing and electromagnetics," *Synthesis Lectures on Computational Electromagnetics* 1.1, pp. 1-120, 2005.

[25] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[26] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, Vol. 2, pp. 121-167, June 1998.

[27] B. Schölkopf, C. J. C. Burges, and A. J. Smola, *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press, 1999.

[28] Peck, L., Lacombe, J., Anderson, T. S., Gagnon, J., Fisk, D., Williams, C., & Wesson, K., "Seismic Detection of Personnel: Field Trials and Signatures Database," In *Human, Light Vehicle and Tunnel Detection Workshop, University of Mississippi, NCPA.*, 2009.