To the Graduate Council:

I am submitting herewith a thesis written by Yuxin Tian entitled "Target Detection and Classification Using Seismic Signal Processing in Unattended Ground Sensor Systems." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements of the degree of Master of Science, with a major in Electrical Engineering.

Hairong Qi

Hairong Qi, Major Professor

We have read this thesis
and recommend its acceptance:

M. J. Roberts

Daniel B Koch

Accepted for the Council:

Vice Provost and Dean of Graduate Studies

Target Detection and Classification Using Seismic Signal Processing in Unattended Ground Sensor Systems

A Thesis
Presented for the
Master of Science
The University of Tennessee, Knoxville

Yuxin Tian

August, 2001

Abstract

This thesis studies the problem of target detection and classification in Unattended Ground Sensor (UGS) systems. One of the most challenging problems faced by target identification process is the design of a robust feature vector which is stable and specific to a certain type of vehicle. UGS systems have been used to detect and classify a variety of vehicles. In these systems, acoustic and seismic signals are the most popularly used resources. This thesis studies recent development of target detection and classification techniques using seismic signals. Based on these studies, a new feature extraction algorithm, Spectral Statistics and Wavelet Coefficients Characterization (SSWCC), is proposed. This algorithm obtains a robust feature vector extracted from the spectrum, the power spectral density (PSD) and the wavelet coefficients of the signals.

Shape statistics is used in both spectral and PSD analysis. These features not only describe the frequency distribution in the spectrum and PSD, but also shows the closeness of the magnitude of spectrum to the normal distribution. Furthermore, the wavelet coefficients are calculated to present the signal in the time-frequency domain. The energy and the distribution of the wavelet coefficients are used in feature extraction as well. After the features are obtained, principal component analysis (PCA) is used to reduce the dimension of the features and optimize the feature vector.

Minimum-distance classifier and k-nearest neighbor (kNN) classifier are used to carry out the classification. Experimental results show that SSWCC provides a robust feature set for target identification. The overall performance level can reach as high as 90%.

Acknowledgment

Foremost, I would like to express my deepest gratitude to my advisor, Professor Hairong Qi, for her excellent guidance, advice and endless support during my course work and research at The University of Tennessee, Knoxville.

I also wish to give my thanks to Professor Michael J. Roberts and Professor Daniel B. Koch for their guidance and suggestion in solving many practical problems.

I would like to give my special thanks to my wife, Hongliu Mo, who gives me the strongest support during my study life.

Finally, I want to thank my dearest father who has given me infinite courage to explore the scientific world, but would never have the chance to share my happiness again.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Unattended Ground Sensor (UGS) Systems

An *Unattended Ground Sensor* (UGS) is a device placed on the ground which automatically gathers and interprets sensor data from remote "targets", then transfers results back to some higher level processing center. The targets can include aircraft, moving vehicles, human beings, weapons, etc. A simple example of a UGS is a microphone sensor coupled with an automatic acoustic signal processor and a telemetry system. Such an acoustic UGS could be used to detect and report the sound of an aircraft flying overhead [42]. Without a human's interference, a UGS is always set in the vicinity of the targets to monitor the emanation from the targets. Until now, UGS systems have been used in a variety of applications ranging from industrial monitoring to military information gathering [41].

Recent advances in commercial sensors and digital signal processing (DSP) techniques have spurred the development of UGS systems to be used for real-time ground target detection, classification, localization and tracking. By using a distributed sensor network in the battlefield, UGS systems can detect the existence of a ground target, classify the type of the target and activate the alarm for the soldiers. Typically used sensors in UGS systems include acoustic and seismic sensors [39].

An acoustic ground sensor is used to detect, identify, and locate sources of acoustic emanations from targets, which propagate as compression waves with a sound speed of 340m/sec, and potentially impinge upon the acoustic transducer (such as microphone) at the ground sensor system. The acoustic transducer then outputs a voltage that is proportional to the incident dynamic pressure. Omni-directional acoustic transducers are always used as acoustic ground sensors. These transducers provide excellent sensitivity over a frequency from 20Hz to 10,000Hz. However, the background acoustic noise, especially wind noise, will reduce the performance of an acoustic sensor system significantly. Meanwhile, the influence of atmospheric and terrain variations will result in reflection, refraction, scattering and attenuation to the acoustic signals, which will also decrease the sensitivity of ground acoustic sensors [41].

Similar to acoustic sensors, a seismic sensor is used to detect, identify, and locate sources of seismic emanations which result from the vibration of the components within the target coupled to the ground. The seismic emanations from the target propagate as seismic waves, and potentially impinge upon the seismic transducer (such as geophone) at the ground sensor system; then the seismic transducer outputs a voltage that is proportional to the velocity of the ground vibrations. Both wind noise and natural noise also affect the performance of seismic ground sensors. When the wind strikes the ground near the seismic sensor, the ground motion will be transferred to the seismic sensors, thus resulting in the output background noise of the seismic sensors. Burying the transducer at significant depths can reduce the wind noise significantly [41].

1.2 Target Detection and Classification in UGS Systems

No matter what sensing modalities are used, target detection and classification usually follow the same framework as shown in Fig. 1.1: the raw sensor readouts are first processed in different processing domains, such as frequency domain and time domain. Signal features which can represent different vehicle types such as track type, engine cylinder number, the relative weight, the velocity of the vehicle, etc., are then extracted. Based on these features, pattern recognition techniques are used to classify these vehicles. One of the most challenging problems faced by the ground vehicle identification process is how to select a robust feature vector that is stable and specific to a certain type of vehicle.

Target detection and classification using seismic sensors has an advantage over acoustic systems in that propagation through the ground is less sensitive to ambient weather conditions [6][23]. Acoustic vehicle recognition will be affected by Doppler effects, by noises introduced from various moving parts of vehicles, and by atmospheric and terrain variations, such as wind effects across acoustic sensors, noises from friction between terrain and wheels or tracks, weather and environmental effects, and echoes [28][41]. On the other hand, seismic wave propagation is highly dependent on the underlying geology. Seismic propagation includes compression waves,

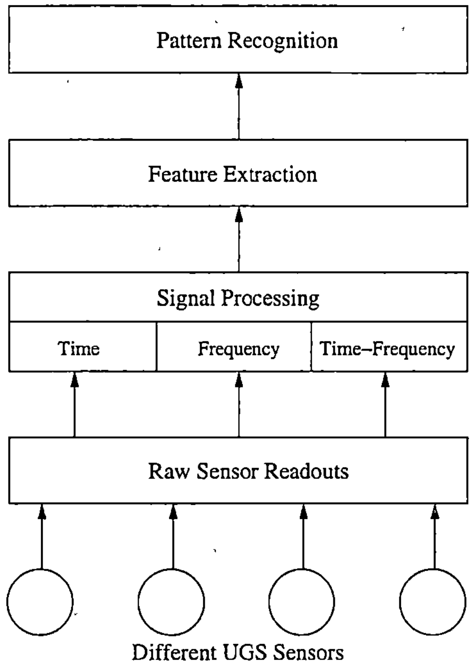


Figure 1.1: Target Detection and Classification

shear waves, surface waves, reflected waves, and refracted waves. These waves have different directions and different propagation speeds, ranging from 200m/sec (surface waves in loose, dry sediments) to over 4000m/sec (compression waves in hard rocks) [41]. Therefore, the analysis of seismic waves is more complicated. Another influence on seismic ground sensors is frequency-dependent and range-dependent propagation effects [22]. Seismic absorption, scattering and spherical spreading will result in a decrease in signal amplitude at both increasing ranges and increasing frequencies [41]. These effects are highly dependent on the geologic medium, the operating bandwidth and the required target-to-sensor range.

The focus of this thesis is to use seismic signal processing techniques in UGS systems for target detection and classification. New features are developed from seismic signals that could be used for pattern recognition. The algorithm will be implemented and tested on data collected from an experiment planned by the DARPA SensIT program, where a dense field of nodes have been developed to analyze real-world targets.

1.3 Contribution of the Research

This thesis concentrates on the research of seismic signal processing in the vehicle identification problem for UGS systems. Based on current research results, a new feature extraction algorithm, which is called spectral statistics and wavelet coefficient characterization (SSWCC), is proposed in this thesis. This new algorithm extracts the feature vector from power spectral analysis and a wavelet transform, followed by the principal component analysis (PCA) to extract the most significant features from the potential feature set. The performance of the classification is evaluated by the minimum distance classifier and k-nearest neighbors (kNN) classifier. The results of the experiments show that wavelets and spectral analysis provide a robust feature set which could be used in vehicle identification. The overall accuracy of the classification could reach as high as 90%.

1.4 Thesis Outline

The organization of this thesis is as follows:

Chapter 2 first reviews the principles and properties of seismic waves and popularly used seismometers. It then summarizes existing algorithms for seismic signal processing including feature extraction and classification.

Chapter 3 describes the proposed seismic signal feature extraction algorithm —

Spectral Statistics and Wavelet Coefficients Characterization (SSWCC). There are 41 features extracted from Fourier spectrum, power spectral density (PSD), and wavelet coefficients by SSWCC. Principal component analysis (PCA) is applied to reduce the dimension of the feature space.

Chapter 4 presents the fundamental theories of pattern classification. The minimum-distance classifier and k-nearest neighbor (kNN) classifier are chosen to classify the feature space.

Chapter 5 first describes the data collection process. It then develops several experiments to thoroughly evaluate the performance of SSWCC. Results show that the combination of features extracted by SSWCC and a simple classifier like kNN could reach a correct classification rate as high as 90%.

Chapter 6 discusses the results of the thesis and points out future work in the study of target analysis in UGS systems.

Chapter 2

Seismic Waves and Seismic Signal Processing for UGS Systems

The study of the generation, propagation, and recording of seismic waves in the Earth is called *Seismology* [32]. As an important branch of earth science and geophysics, it has been used in various applications, such as the study of earthquakes, modern exploration for oil and gas, and the study of the Earth structure. Seismology began with the study of earthquakes. The transmission of elastic waves inside the Earth was realized by Hooke in 1668 and Michell of Cambridge in 1670. In the 19th century French and later English mathematicians developed the theory of seismic waves through a solid body [19]. In 1830, Poisson was the first person to show that there are two interior (so-called body) waves, compressive (P) and shear (S), produced by a disturbance in a solid body. Surface waves were predicted by Rayleigh in 1887, and he also demonstrated one type of surface wave. Love discovered the second surface wave in 1911 [19]. The use of seismic wave detection in

UGS systems could date to 1970s, when the UGS systems are developed to detect enemy troop movement in the battlefields [1]. This chapter reviews the properties of seismic waves and the signal processing techniques used in the study of seismic waves.

2.1 Seismic Waves

Seismic waves are elastic waves propagating through the Earth. Both natural and human-made sources of deformational energy can produce seismic waves. The study of seismic waves has a long history. It started as early as the study of earthquakes. However, the theoretical analysis of seismic waves was not developed until 1800s.

In general, seismic waves can be classified into two categories: body waves and surface waves. Body waves travel at a higher speed through the interior of the Earth and propagate in three dimensions, while surface waves travel near the surface of the Earth and propagate in two dimensions.

2.1.1 Body Waves

In 1830, Poisson used the equation of motion and elastic constitutive laws to show the existence of two solutions to the motion equations for the Earth shown in Eq. 2.1 [32]:

$$\nabla[(k+\frac{4}{3}\mu)\nabla^2\phi-\rho\ddot{\phi}]+\nabla\times[\mu\nabla^2\Psi-\rho\ddot{\Psi}]=0 \tag{2.1}$$

where

$$k = \text{the incompressibility or bulk modulus of the rock,}$$
$$\mu = \text{the rigidity of the body,}$$
$$\rho = \text{the density of the body.}$$

Suppose

$$\alpha = \sqrt{\frac{k + \frac{4}{3}\mu}{\rho}} \tag{2.2}$$

and

$$\beta = \sqrt{\frac{\mu}{\rho}} \tag{2.3}$$

then Eq. 2.1 will be solved if

$$\nabla^2\phi - \frac{1}{\alpha^2}\ddot{\phi} = 0 \tag{2.4}$$

and

$$\nabla^2\Psi - \frac{1}{\beta^2}\ddot{\Psi} = 0 \tag{2.5}$$

Equations 2.4 and 2.5 describe a scalar wave equation for ϕ and a vector wave equation for Ψ respectively. α is the velocity of the wave solution of ϕ, which is called the *P-wave*. Similarly, β is the velocity corresponding to the solution of Ψ, which is called the *S-wave*. A general wave potential solution for P waves and S waves are shown in Eq. 2.6 and Eq. 2.7 respectively.

$$\phi(\mathbf{x}, t) = Ae^{\pm i(\omega t \pm k_1 x_1 \pm k_2 x_2 \pm k_3 x_3)} \tag{2.6}$$

$$\Psi(\mathbf{x}, t) = Be^{i(\omega t - \mathbf{k}_\beta \mathbf{x})} \tag{2.7}$$

where

$$\begin{aligned}
\omega &= \text{the frequency of waves,} \\
x_1, x_2, x_3 &= \text{the position of the particle,} \\
k_1, k_2, k_3 &= \text{the constants that satisfy the condition } k_1^2 + k_2^2 + k_3 = \omega^2/\alpha^2, \\
|\mathbf{k}_\beta| &= \omega/\beta.
\end{aligned}$$

Because P waves and S waves only travel through the body of interior of the Earth, people also named them as *body waves.*

P waves are *compressional waves.* They transmit the pressure changes through the Earth as a series of alternating compressive and decompressive waves. The direction of the particle motion is parallel to the direction of propagation. On the other hand, S waves, also called *shear waves* or *transverse waves*, have only shearing deformation and no volume change, which is why they cannot propagate in fluids [32]. In S waves, the direction of particle motion is perpendicular to the direction of propagation [19]. The schematic difference between P waves and S waves is shown in Fig. 2.1. A more detailed analysis would show that S waves can be decomposed into two components, one is in purely horizontal (x_2) motion, which is called *SH* component, another is in purely vertical (x_3) motion, and is called *SV* component.

These two types of S waves will be useful in the analysis of surface waves.

From Eqs. 2.2 and 2.3, we can see that α is always greater than β, and the velocity of S waves is about 60% of the P waves. Since P waves always arrive earlier than S waves, they are the most accurately measured and the most commonly used waveforms in the study of earthquake location and the reflection and refraction methods of exploration [19].

2.1.2 Surface Waves

Surface waves are defined as those produced in media with a free surface which propagate parallelly to the free surface and whose amplitudes decrease with the distance from the surface. Surface waves are generated by the incidence of body waves to the free surface [46].

In 1887, Lord Rayleigh demonstrated the existence of additional solutions of elastic equations of motion for bodies with free surface. These waves, named as *Rayleigh waves*, are formed by the interaction of incident P and SV waves with the free surface boundary condition as described in Eq. 2.8:

$$\begin{aligned} \sigma_{33} &= (k - \frac{2\mu}{3})(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_3}{\partial x_3}) + 2\mu\frac{\partial u_3}{\partial x_3} = 0 \\ \sigma_{13} &= \mu(\frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3}) = 0 \end{aligned} \tag{2.8}$$

where

μ = the rigidity of the rock,

k = the incompressibility or bulk modulus of the rock,

σ_{ij} = the stresses in direction x_j acted on plane whose normal is x_i,

u_1, u_2, u_3 = displacement of the waves in the direction of x_1,x_2 and x_3, respectively.

The displacements of Rayleigh wave are given by Eq. 2.9,

$$\begin{aligned} u_1 &= -An\sin(nx_1 - \omega t)(e^{-0.85nx_3} - 0.58e^{-0.39nx_3}) \\ u_3 &= -An\sin(nx_1 - \omega t)(0.85e^{-0.85nx_3} - 1.47e^{-0.39nx_3}) \end{aligned} \tag{2.9}$$

where

n = wave number of Rayleigh defined as $n = \omega/c$ (c is the velocity).

By 1911, a second type of surface-wave motion, which is produced in a bounded body with layered material properties, was characterized by Love and are hence called the *Love waves.* Love waves are generated by the reflection of SH waves at the free surface combined with internal layering of the Earth to trap SH reverberations near the surface. The displacements of Love-wave are described by Eq. 2.10.

$$\begin{aligned} v_1 &= Ae^{i\omega(px_1+\eta_{\beta_1}-t)} + Be^{i\omega(px_1-\eta_{\beta_1}-t)} \\ v_2 &= Ce^{i\omega(px_1+\eta_{\beta_1}-t)} \end{aligned} \tag{2.10}$$

where v_1 is the SH displacement in the layer, and v_2 is the SH displacement in the half-space [32], p is called the *horizontal slowness*, and η is called the *vertical slowness*, which are determined by the properties of the travel media.

Rayleigh and Love waves travel only on the surface of the Earth and are thus called the *surface waves.* Compared with body waves, surface waves have a lower speed and travel in two dimensions, not three dimensions. Moreover, because they are guided by the surface, their amplitudes decrease more slowly with distance and they tend to dominate recording at a large distance from a shallow earthquake. For Rayleigh waves, the ground motion is elliptical, as for a water wave. While for Love waves, because they are shear waves similar to S waves, the ground motion is in a horizontal plane. People have found that surface waves are not short wavelets as for P and S waves, but spread out in time and space, and their velocities are a function of wave period and wavelength [19].

The schematic of the particle motion of the four fundamental waves is shown in Fig. 2.1 [32]. The waves all propagate from left to right. This figure clearly shows the relationship between propagation direction and the particle motion.

The propagation paths of the four seismic waves are shown in Fig. 2.2. When the vibrations generated by the source propagate through the Earth, the wave field becomes very complicated. Rayleigh and Love waves travel along the surface, while P and S waves travel through the deeper parts of the Earth in multiple forms, one form of "multipath" occurs in the direct path (B) from the source to the sensor, other forms, such as reflections (A) and refractions (D) also arise. When seismic media reflection occurs, the mode conversion are also generated, which is shown by path C where P waves (solid line) convert to S waves (dashed line). S to P conversions (not shown) can also occur at the interface between medium 1 and medium 2 [12].

Research of seismic waves shows that if the disturbance, such as an earthquake, explosion or human interference occurs near the surface, a significant amount of

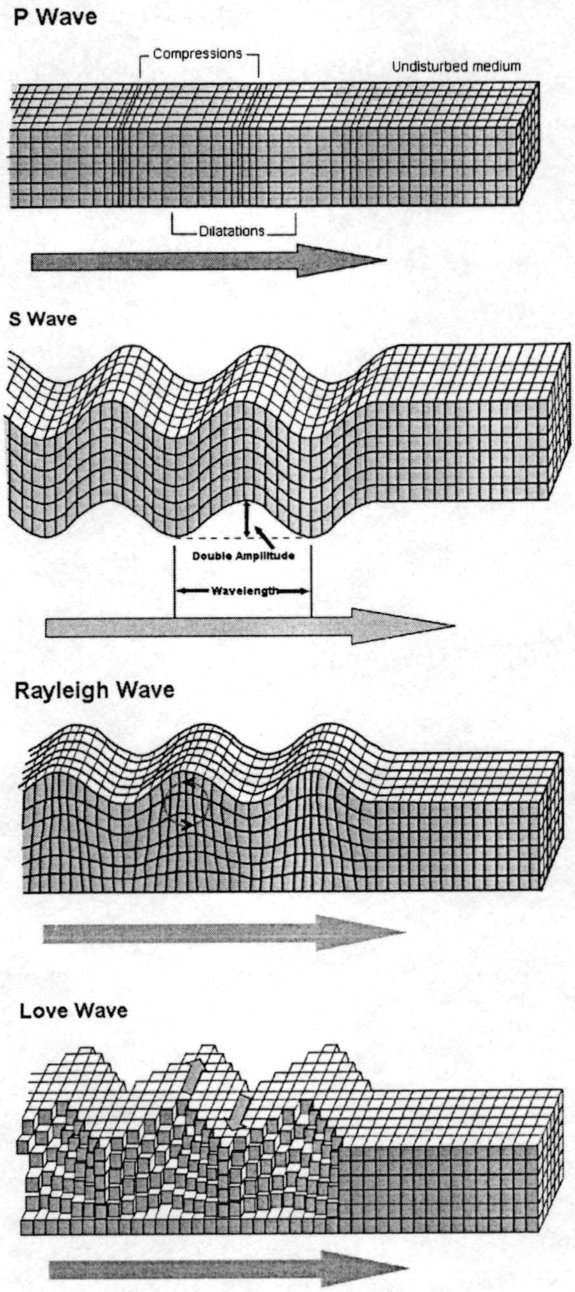


Figure 2.1: Schematic of the Four Fundamental Elastic Waves [2]

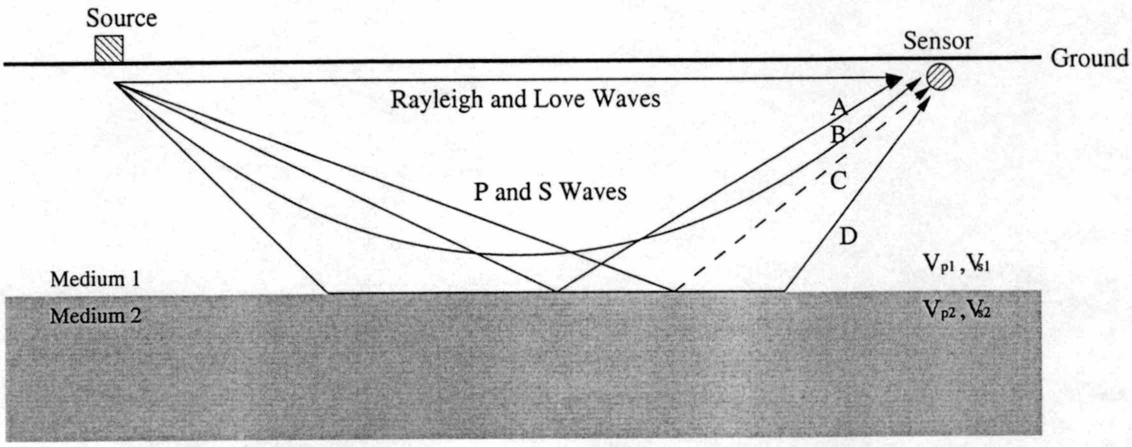


Figure 2.2: The Propagation of Four Different Seismic Waves

seismic energy is dispersed near the surface and is transmitted as surface waves (70% of the energy of the impact is distributed in the Rayleigh waves, and the remaining 30% of the energy is transmitted into the earth via body waves [44]). Therefore, in UGS systems, when we study the seismic waves generated by the ground vehicles, we pay more attention to the surface waves, especially on Rayleigh waves.

2.2 Seismometer

Seismic waves are detected and recorded by seismometers. The ideal seismometer will record the ground motion relative to a stationary port above the ground. One of the most popularly used seismometer is *land geophone*, which is a highly sensitive ground motion transducer. The principle of a geophone is shown in Fig. 2.3.

The arrival of the seismic waves sets the surface in motion. The geophone case with rigidly attached coil is coupled to the ground and thus moves harmoniously. The magnet which is suspended on springs remains effectively stationary because

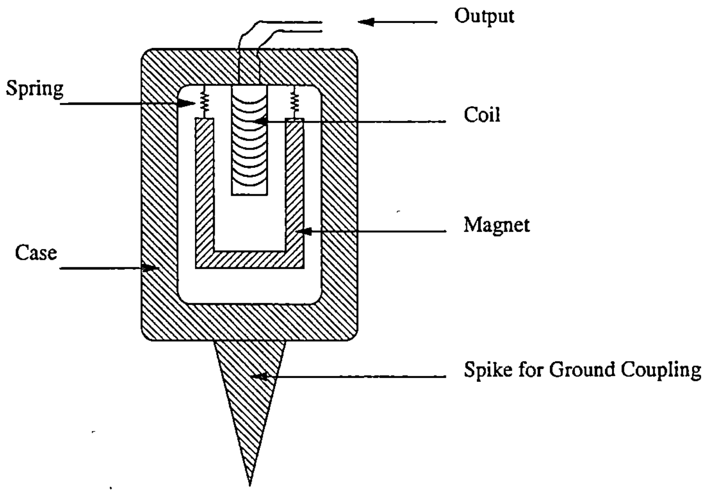


Figure 2.3: Structure of Geophone

of its inertia. Movement of the coil within the magnetic field induces an electrical voltage across the coil which is proportional to the relative velocity of the coil and the magnet [36]. Therefore, the response of a seismometer is commonly proportional to the ground velocity as it depends on the relative velocity of the magnet and the coil [19].

The photo of a typical geophone is shown in Fig. 2.4 [4], a quarter serves as reference for scale. This particular geophone is side cut out so that we can see its working parts. The coils (copper wire in this case) can be clearly seen inside the geophone. The spring connecting the geophone to the case can not be seen but is just above the mass. The silver colored case just inside the plastic external case is magnetized. The black wires coming out from either side of the case transmit the variations in voltage to the recording system. The long silver spike below the blue case is used to firmly attach the geophone to the ground. This spike is pressed into the ground by stepping on the top of the geophone until it is completely buried [4].

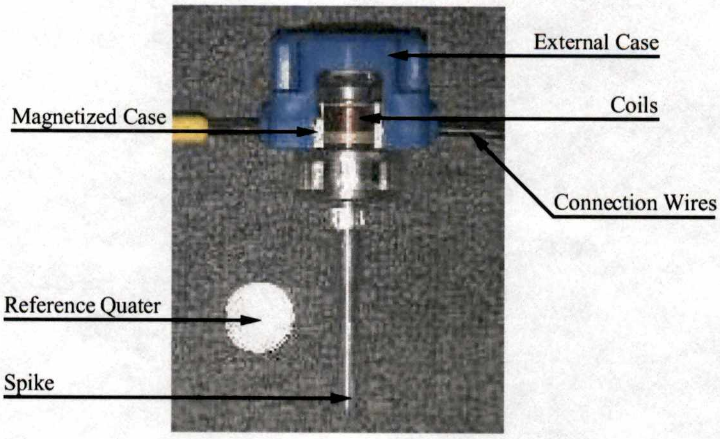


Figure 2.4: A Typical Geophone [4].

Most of the modern geophones use a dual coil system which reduces the external interference [36]. An alternative geophone design is to mount the magnet rigidly and suspend the coil around an inertial mass, which is called *moving coil geophones.* In both cases, the output is independent of frequencies above the natural frequency of the suspended element, and dependent on frequencies below the natural frequency.

The geophone must be damped to give a reasonably flat response curve over the necessary frequency. This stops the resonant vibrations which would swamp the records. Damping is usually obtained by an appropriate electrical resistance in parallel with the device which absorbs the unwanted vibration energy [19]. Typically, a geophone is only used for high frequency (10Hz - 500Hz) seismology experiments since its resolution degrades dramatically at low frequency (5mHz - 10Hz). We can see the poor frequency response at low frequency range from a typical frequency response of the geophone shown in Fig. 2.5 [3].

Geophones are used today in a variety of applications far from their original purpose as earthquake detectors. Different types of geophones for specific use have

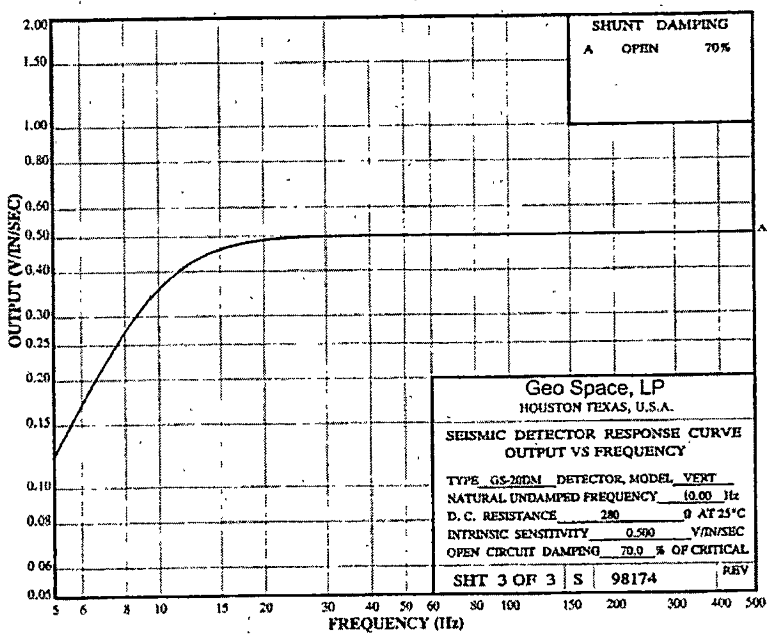


Figure 2.5: A Typical Frequency Response of Geophone [3].

also appeared: low frequency geophone, which is used to detect the machinery vibration, has a flat frequency response in low frequency range from 1Hz to 20Hz; High frequency geophone, which has a flat frequency response range from 100Hz to 2500Hz, is used to detect the high frequency vibration of the machine; Three-axis geophone, which can detect vibrations in three directions, has been used for 40-50 years in earthquake detection and moving vehicles detection. Meanwhile, high temperature geophone, high sensitive geophone, and some other types of geophones are also developed. For moving vehicle detection, three-axis geophones are most popularly used [6][15][43].

2.3 Seismic Signal Processing for UGS Systems

A traveling ground vehicle will introduce ground motion via two distinct ways. First of all, the vertical and tractive forces from the vehicle's wheels or tracks on the road surface will cause seismic waves to radiate outward through soil. Secondly, acoustic waves from the vehicle's exhaust pipe and engine will radiate through the air and will be transmitted into the ground as well [6]. In UGS systems, the seismic signal processing stage extracts feature vectors from the received seismic signal and then use pattern recognition techniques to classify and identify the vehicle from the feature vector. This section reviews these seismic signal processing techniques.

2.3.1 Feature Extraction

Based on different processing domains, feature extraction for seismic signals can be classified into three categories: in time domain, in frequency domain, and in time-frequency domain.

Time Domain Feature Extraction

Sleefe et al. [42] proposed a method based on time domain signals to locate the stationary target. First of all, the algorithm assumes the source and the sensor are located at position (x_0, y_0, z_0) and (x_s, y_s, z_s), respectively. The source emits a signal represented by time-varying waveform $u(t, x_0, y_0, z_0)$. This waveform is coupled into the local medium which can be modeled as a source-coupling function $c(t, x_0, y_0, z_0)$. Considering the frequency-dependent attenuation, reflection, refraction and scattering, the source function undergoes temporal and spatial modifications that can be represented as a medium-dependent function $m(t, x, y, z)$. Meanwhile the UGS sensor is coupled to the medium through the coupling function $s(t, x_s, y_s, z_s)$. Then the signal that sensor received can be represented as the convolution of source function with the coupling functions as Eq. 2.11:

$$\begin{aligned} r(t, x_s, y_s, z_s) &= u(t, x_0, y_0, z_0) * c(t, x_0, y_0, z_0) * m(t, x, y, z) * s(t, x_s, y_s, z_s) \\ &\quad + n(t, x_s, y_s, z_s) \end{aligned} \tag{2.11}$$

where $n(t, x_s, y_s, z_s)$ is the interfering noise sources, "$*$" stands for the convolution operation. If $u(), c(), m()$ and $s()$ are combined together to represent the *in-situ* received sensor signal, and then separate out the direct waveform arrival [37], then Eq. 2.11 may be simplified to Eq. 2.12.

$$r(t) = q_0(t - T_0) + n(t) + \sum_{i=1}^{N} q_i(t - T_i) \tag{2.12}$$

where

$$\begin{aligned} r(t) &= \text{the received sensor signal,} \\ q_0 &= \text{the received direct wave,} \\ n(t) &= \text{the additive interfering noise source,} \\ T_0 &= \text{the propagation time from the source to the UGS.} \end{aligned}$$

Suppose the direct wave propagates through a homogeneous, isotropic medium, then T_0 could be represented as Eq. 2.13.

$$T_0 = \frac{\sqrt{(x_0 - x_s)^2 + (y_0 - y_s)^2 + (z_0 - z_s)^2}}{c_d} \tag{2.13}$$

where

$c_d =$ the nominal propagation velocity of direct wave.

After the calculation of the average of the multiple observations of the emitted wave fields, the incoherent noise will be reduced. Then the received signal $r(t)$ would be processed using a matched filter based on $u(t)$ to detect the target. In general, some features of $u(t)$ should be known, from which a near optimal digital signal enhancement filter can be applied. However, the applications of correction effects for source coupling, signal propagation, and receiver coupling are also needed to achieve acceptable detection results [42].

For optimal identification, the received signal $r(t)$ would be processed using a series of matched filters followed by an M-ary classification process. *A-priori* information about the suspected target should be known so that near-optimal filters can be applied. Similar to the detection of the target, the correction for propagation and coupling effects is essential [42].

As a unique inverse problem, the location estimation of the target can be approached by two methods: Array processing using multiple UGS sensors; and/or signal feature extraction from multiple UGS. Both approaches are based on arrival time measurements and the implementation of the inverse relationship as Eq. 2.13. As the signal is stationary, multiple observations at multiple locations are dependent on the arrival times. Thus, a desirable location accuracy can be achieved through a multiplicity of spatial and temporal UGS measurements if application permits [42].

This time-domain detection algorithm has several constraints on unattended ground sensors. First of all, because the UGS relies on the measurement of emitted wave fields to characterize the target, the general location of the target should be known *a-priori*, very wide area searches using UGS are generally not possible. Secondly, *a-priori* information on the signatures of the emitted wave fields must also be available. To optimally detect and identify the target, the full time-varying source signature, source/receiver coupling characteristics and wave field propagation effects would need to be known. However, in practice, accurate estimation of the target features and the coupling/propagation effects are not so easy. Thirdly, the algorithm can only detect the stationary target, the test results [42] show that the accuracy is limited in detecting mobile targets.

Generally speaking, detection and identification of the ground vehicle only in the time-domain is always not very accurate. Considering the interfering noise, the complicated waveforms, the reflection and refraction of the seismic waves, and the variation of the terrains, the time domain signal would give us little feature information about the target. Therefore, frequency domain analysis is necessary.

Frequency Domain Feature Extraction

Succi et al. [44] discussed the problem of target tracking using surface waves. They analyze several special features of seismic measurement, including the contamination of the seismic signal by local acoustic waves, the excess non-geometric attenuation of the seismic signal, the influence of reflections from layered soil and a method of ranging using the periodic impact signature of vehicles. The algorithm they developed has been used to measure the bearing as well as the velocity of the vehicle. The measurement of the bearing is based on the analysis of Raleigh waves, which has the property that the vertical component is 90 degrees out of phase with respect to the horizontal component. The influence of the contamination of acoustic signal to seismic signal is reduced by doing the cross-correlation and the spectral analysis. The velocity is estimated based on the assumption that the vehicle velocity is in direct proportion to the seismic frequency. Meanwhile, a Kalman filter is used to track the vehicle in real time. The results show that the range is accurately predicted with bearing and velocity, and the vehicle track computed is nearly identical to the actual track.

Der et al. [15] present a higher-order spectral analysis algorithm to detect and identify the ground vehicle. Higher-order spectral analysis is an extension of the second order measurements (such as the auto correlation function and power spectrum), and it provides supplementary information to the power spectrum which is useful in detecting and characterizing nonlinear properties of signals [35]. Based on the fact that machine-generated seismic signals commonly contain a number of phase coupled harmonics related to some fundamental rotation frequency, the paper develops a multi-mode mixed seismic propagation model, then applies the higher-order spectral analysis which will suppress the effect of the seismic background noise as well as uncoupled noise spectral lines. Finally, the bispectral estimates similar to the block-averaged periodogram are computed to decompose compounds in seismic signals and discriminate different machines as shown in Eq. 2.14,

$$M(f_1, f_2) = \sum X(f_1)X(f_2)X^*(f_1 + f_2) \tag{2.14}$$

where $X(f_i)$ is the Fourier transform of the time series, "$*$" denotes complex conjugation and the summation is performed over Fourier transforms of numerous time windows. However, test results of this algorithm show that only in limited instances could bispectral analysis be effective, and if the bispectrum of two different machines are similar enough, such as the generator and the concrete compactor, the effectiveness of the discrimination was the least [15].

The HLA (harmonic line association) combined with seismic shape feature extraction algorithm, proposed in [48], takes the advantage of spectral characteristics that are dominated by narrow band spectral peaks. First of all, this algorithm selects peaks that are harmonically related to create harmonic line sets for each frame of data samples. Secondly, split-window peak-piking is performed on the PSE (power spectral estimates) feature set to find the maximum peak in the frequency domain. These peaks are assumed as some kth harmonic lines of the fundamental frequency. Thirdly, the algorithm calculates total signal strength in the HLA set. The integer value k, is assumed to be the correct harmonic line number, and the harmonic lines of this particular set are retained as a feature vector. Finally, the algorithm uses the higher order shape statistics of seismic feature sets which are defined in Eq. 2.15 to identify different kind of vehicles [29].

$$\begin{aligned}
\text{Mean:} \qquad & \mu_{shape} = \frac{1}{S}\sum_{i=1}^{N} iC(i) \\
\text{Standard deviation:} \qquad & \theta_{shape} = \sqrt{\frac{1}{S}\sum_{i=1}^{N}(i-\mu_{shape})^2 C(i)} \\
\text{Skewness:} \qquad & \gamma_{shape} = \frac{1}{S}\sum_{i=1}^{N}\left(\frac{i-\mu_{shape}}{\theta_{shape}}\right)^3 C(i)
\end{aligned}$$

$$\text{Kurtosis:} \qquad \beta_{shape} = \frac{1}{S}\sum_{i=1}^{N}(\frac{i-\mu_{shape}}{\theta_{shape}})^4 C(i) - 3 \tag{2.15}$$

where

$$\begin{aligned} S &= \sum_{i=1}^{N} C(i), \\ C(i) &= \text{the power spectral magnitude for the } i\text{th frequency bin}, \\ N &= \text{the number of the frequency bins.} \end{aligned}$$

The shape statistics parameters tell us how the variation distributed around the expected value of the signal spectrum, the symmetry of the spectral distribution, and whether data are peaked or flat relative to the normal distribution.

HLA technique has two advantages: (1) the feature vector is normalized and solely dependent on harmonic line number but not a function of frequency; (2) the peak energy is tracked frame by frame, thus producing a predominance pattern for the signal energy source. The results show that the feature space obtained from seismic shape statistic plus the HLA is far more robust than the feature space used before [48].

Time-Frequency Domain Feature Extraction

Technical reports from Applied Research Laboratories in the University of Texas, Austin [23][33][43], present a time-spectral combination analysis for ground vehicle detection. This algorithm processes data from a three-axis seismometer and estimates the vehicle bearing from the sensor as well as the direction of the vehicle. It uses both time series and spectral information to derive the features of the vehicles,

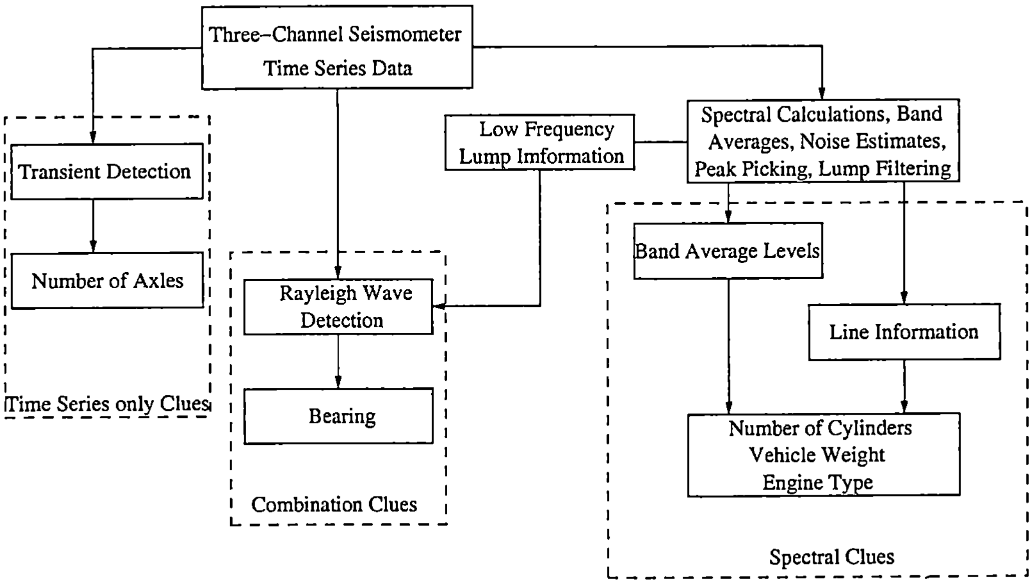


Figure 2.6: Basic Feature Hierarchy

such as the vehicle's engine, transmission, and approximate weight. The determination process of these features is shown in Fig. 2.6. First of all, the multi-channel data are decomposed into continuous data arrays for prepossessing in the spectral, transient and combination algorithm. Then single-channel time series data are used for transient analysis. Meanwhile, spectra from the three channels are calculated and used in classification algorithms. Also spectra from the three axes are averaged, forming a single averaged spectrum to improve the signal-to-noise ratio (SNR) of the received data, and this result is analyzed in many of the spectral routines as well. Finally, the three-channel data are used to infer a bearing to the vehicle, when Rayleigh wave propagation is present. From this figure, we can see that time series data are used to detect transient information and to detect the Rayleigh wave propagation related to vehicle generated energy, while spectral information is used to identify the vehicle weight, the number of cylinders, and the engine type [23].

2.3.2 Pattern Classification

Once vehicle features are identified and extracted, classification process will proceed. The goal of pattern classification is to partition the feature space into regions. Every region corresponds to one category. Generally speaking, the pattern classifier design problem can be categorized into two classes: one is based on *supervised learning* and another is based on *unsupervised learning.*

In supervised learning, the discrimination function is "learned" from examples in the training set which contains elements of paired values of input features and output classes. Many algorithms have been developed for this task including decision trees, neural networks, Hidden Markov Models, association rules and the nearest neighbor rule. In contrast to the supervised learning, the unsupervised learning problem could not obtain such a training set of examples showing what kind of desired input/output relations there should be. In another words, in unsupervised learning, there is no target specified for the learning process, and the learning process is mainly based on the input data and self-organization. The generally-used method for the unsupervised learning problem is self-organizing maps (SOM) and k-mean analysis. In both supervised and unsupervised learning problems, samples $\mathbf{x}$ are assumed to be obtained by selecting a state of nature ω_i with probability $P(\omega_i)$, and then independently selecting $\mathbf{x}$ according to the probability law $p(\mathbf{x}|\omega_i)$. The difference is that with supervised learning we know the state of nature for each sample, that is, the training samples used to design a classifier are labeled to show their category membership. However, with the unsupervised learning we do not. Therefore, the problem of unsupervised learning is more difficult than the supervised one [20].

The weight classification algorithm proposed in [33] is based on the assumption that the source-coupling strength and the corresponding signal to noise ratio (SNR) levels increase with vehicle velocity. The peak-averaged SNR levels corresponding to the horizontal seismometer axes are time-averaged over a narrow window centered on CPA (Closest Point of Approach). The time-averaged SNR and relative velocity estimates are tabulated for each vehicle detected, thereby establishing the range of reference values used to classify the relative weight categories. The accuracy of the relative weight classification depends on the range of observed vehicle velocities as well as weights and will improve with time. The results show that relative weight were most accurate for the heavy vehicles, and the algorithm successfully distinguished between heavy vehicles with large velocities and light vehicles with small velocities.

Wellman et al. [48] develop an ANN (Artificial Neural Network) that can provide a robust classifier as well as a measure of confidence in the classification decision. The algorithm classifies the vehicles by a back propagation neural network (BPNN) with an adaptive learning rate, which allows fine-grain adjustments during training. Further refinements to the learning rate are accomplished through an interlayer multiplier, which only affects the learning rate in the hidden neuron layer. A smoothing technique allows the control of weight adjustment based on the past values of gradient descent and can prevent the training process from terminating in local minima. Compared with the other classifier structures, such as "fuzzy" adaptive resonance theory (ART) maps, genetic algorithms, generalized potential neural networks, and the application of homogeneous and heterogeneous combination of classifiers, BPNN classifer is simpler to implement. Recent work using only seismic features has produced an average probability of correct identification as high as 86%. The neural network performance is qualified by a confusion matrix that provides the percentage of correct identification (C_{ID}) which is determined by Eq. 2.16. Meanwhile, the confidence levels for classification of each target are calculated by Eq. 2.17.

$$C_{ID} = \frac{N_{P_{CID}}}{K} \tag{2.16}$$

$$C_{level} = \frac{1}{K}\sum_{i=1}^{K}\left(\sum_{j=1}^{L}\frac{P_{CID_i} - P_{FID_j}}{2(L-1)}\right) \tag{2.17}$$

where

K = Total number of observation,

L = Total number of the output vehicles,

$N_{P_{CID}}$ = Number of correct decisions,

P_{CID_i} = Predicted value of the correct identification for vehicle class i,

P_{FID_j} = Predicted value of the false identification for vehicle class j with respect to vehicle class i,

P_{CID} = Output for the BPNN output node dedicated to a particular vehicle class,

P_{FID} = Output for the other output nodes.

The result [48] shows that the overall confidence level for the classification is 81%, and the average probability of correct identification has reached 86%.

Wolford et al. [49] generalized several existing pattern classification algorithms including the simple competitive learning network, the self organizing map (SOM), the learning vector quantization (LVQ) network, and fuzzy cluster means (FCM).

The competitive learning network contains $2M$ neurons, where M is the number of events measured. The synaptic weights are initialized randomly. The activation function defined in Eq. 2.18 maps the input vector to a unit vector. The weight is updated by Eq. 2.19 [11]. Thus, the output of a simulation event will be zero for all neurons except the "winner" whose value is 1.

$$O_i(n) = \begin{cases} 1 & \text{if } \mathrm{dist}(x, W_i(n)) \leq \mathrm{dist}(x, W_j(n)) \text{ for all } j, \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

$$W_i(n+1) = W_i(n) + \eta(x - W_i(n))O_i(n) \tag{2.19}$$

where

$$\begin{aligned} \mathrm{dist}(x, W_i(n)) &= \text{The Euclidean distance between vector } x \text{ and the weight} \\ &\quad\ \text{vector } W_i(n), \\ \eta &= \text{Learning rate.} \end{aligned}$$

Like the competitive learning network, SOM is an unsupervised technique. It usually consists a defined number of neurons in a two-layer network. The SOM is trained with competition learning: A discriminant function for all neurons is computed and the highest function value is identified, then the selected neurons and its neighbors are activated. Finally, it uses an adaptive process to strengthen the discriminant value of the activated neurons in response to the inputs. SOM has the advantage that the models are automatically organized into a meaningful two-dimensional order in which similar models are closer to each other in the grid than the more dissimilar ones [30]. Since only a portion of the neurons allocated for training remain alive, the obvious disadvantage of this algorithm is a loss of computational efficiency.

LVQ network consists of a competitive layer coupled to a linear output layer, it is a supervised version of vector quantization [8]. The competitive layer classifies the feature vectors into "subclasses" which are mapped by the linear output layer to the user's target classes. LVQ networks work well for data which are not linearly separable.

FCM performs well when the clusters overlap only slightly, it is a generalization of the k-means clustering technique. The membership function is defined as Eq. 2.20, and the degree of the membership is defined in Eq. 2.21.

$$J = \sum_{k=1}^{N}\sum_{i=1}^{M} u_{ik}^{m}||x_k - v_i||^2 \tag{2.20}$$

$$u_{ik} = \frac{1}{\sum_{j=1}^{M}\left(\frac{||x_k-v_i||}{||x_k-v_j||}\right)^{\frac{2}{m-1}}} \tag{2.21}$$

where

x_k = The feature vector,

v_i = The cluster center,

N = The number of feature vectors,

M = The number of cluster centers,

m = The fuzziness of the set membership.

Wolford et al. adapt [49] FCM algorithm provided by Bezdek [7], and it converges either to a local minimum or to a saddle point of the cost function. To be more general, FCM requires a more favorable feature space distribution.

Generally speaking, although some of the algorithms described above can reach an accuracy rate around 90% [49], considering the sensitivity of the algorithms and the requirement of the computational efficiency, additional enhancement of these algorithms is still needed to make the algorithm more robust and easier to implement.

2.4 Summary

In this chapter, we first reviewed the generation and the properties of four different seismic waves: P, S, Rayleigh and Love waves. P and S waves belong to body waves and transmit in three dimensions, while Rayleigh and Love waves are surface waves that only transmit along the surface of the Earth. Among them, the most useful waves in vehicle detection are Rayleigh waves. We then introduced a popularly used seismometer — geophone. We also discussed current techniques used in seismic signal processing for UGS systems, including feature extraction and vehicle classification. Feature extraction aims at extracting robust features from seismic waves, and pattern classification can then use pattern recognition techniques to classify feature vectors in order to identify the ground vehicles.

Chapter 3

Feature Extraction

Feature extraction is the most emphasized part of this thesis. In Chapter 2, we studied existing algorithms developed for seismic signal feature extraction. Based on the study, the thesis develops a new algorithm, spectral statistics and wavelet coefficients characterization (SSWCC), to extract a series of feature vectors from spectrum, power spectral density (PSD) and wavelet coefficients on the seismic signals acquired from the UGS systems. Compared with other algorithms, SSWCC synthesizes more information from different domains of the original signal and generates a series of more robust feature vectors. Considering the correlation existing among feature elements, principal component analysis (PCA) is applied to reduce the dimension of the features and select the most significant feature elements.

Feature extraction is the process of obtaining signal characteristics from time series data. It can be considered as a data-compression process which removes irrelevant information and preserves relevant information from the raw data [29]. Feature extraction plays an important role in the pattern recognition problem, since the performance of the classifier largely depends on the quality of feature vectors. Human beings have a strong capability to extract features from signal. Although we could not develop machine feature extraction architecture exactly the same as human beings, we can still get some instruction from the feature extraction process human beings use. Generally speaking, in classification problems, good features should process the following properties [29]:

1. Large interclass mean distance and small inner class variance;
2. Insensitivity to extraneous variables (little SNR dependency);
3. Computationally inexpensive to measure;
4. Uncorrelated with other features;
5. Mathematically definable;
6. Explainable in physical terms.

SSWCC extracts features from frequency and wavelet domains. The block diagram of the algorithm is shown in Fig. 3.1.

The first step of feature extraction is signal preprocessing. In this step, the DC component of the original signal is first eliminated, and wavelet denoising is applied to remove the noise in the original signal. The next step is feature extraction. The spectrum, the PSD and the wavelet transforms of the preprocessed signals are derived and potential features are extracted. These features include the shape statistics and amplitude statistics of the signal spectrum, the shape statistics of PSD, the largest three peak positions in the PSD, the average of wavelet coefficients, and the energy of the wavelet coefficients. To eliminate the scaling effects among different features, all the potential features are normalized. The final step is principal

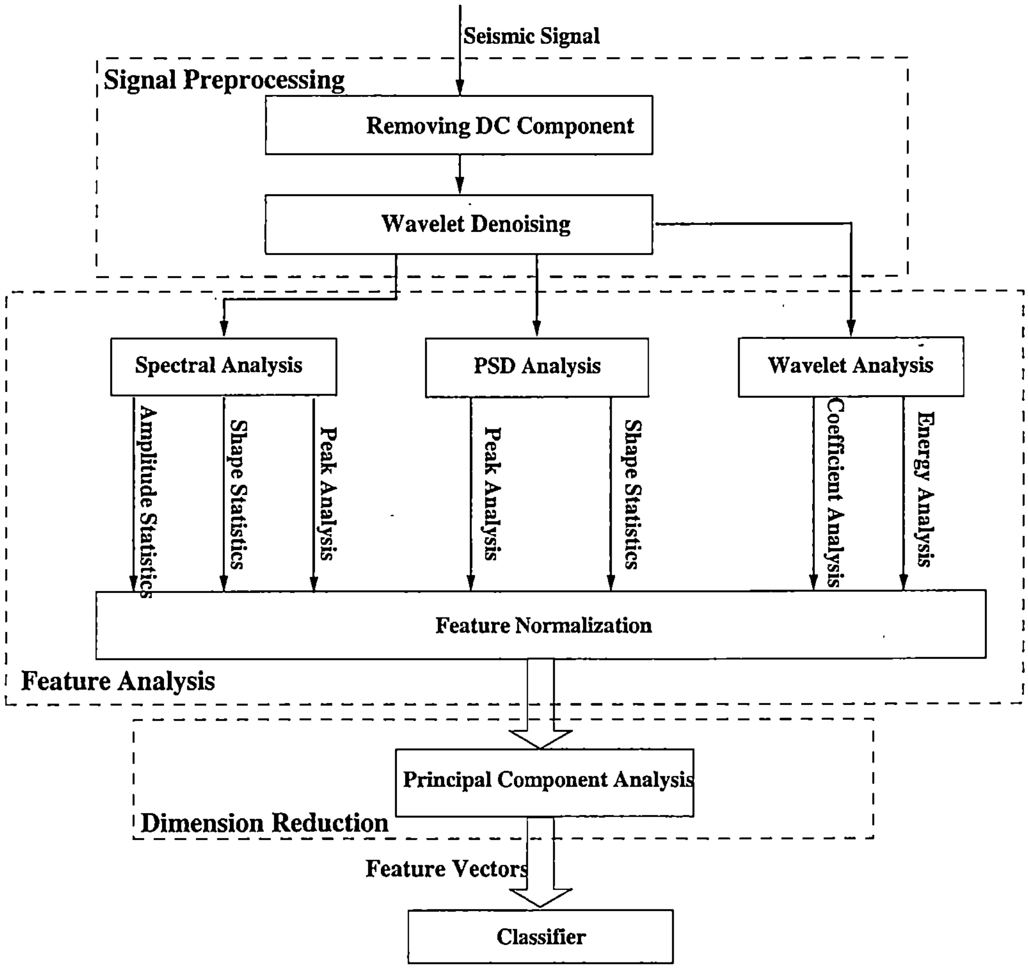


Figure 3.1: Block Diagram of Feature Extraction

component analysis (PCA), which helps optimize the feature vector, reduce the dimension of the feature space, and select the most significant feature components. The output of PCA will be used as the input to the classifier.

3.1 Signal Preprocessing

Because of the electronic and the environmental noise, the raw data captured from the seismic sensors are always noisy, which could affect the performance of feature extraction and classification. Therefore, we need to preprocess the signal first before we further process it. There are many ways to denoise a signal, such as averaging, low pass filtering, wavelet denoising, etc. A comparison of the effects of different denoising methods is shown in Fig. 3.2. The advantages of wavelet denoising is obvious as shown in Fig. 3.2 (D), where the resulting signal is much clearer than the ones in Fig. 3.2 (B) and Fig. 3.2 (C). The noise levels of the signals processed by the first two methods are still high.

3.1.1 Removing the DC Component

The raw data we obtained from sensors contain the DC component which we are not interested in. To avoid its effect on the frequency domain analysis, we need to eliminate it first. Therefore, the first step in signal preprocessing is to eliminate the DC component by Eq. 3.1

$$y(i) = x(i) - \mu_x, i = 1, 2, ...N \tag{3.1}$$

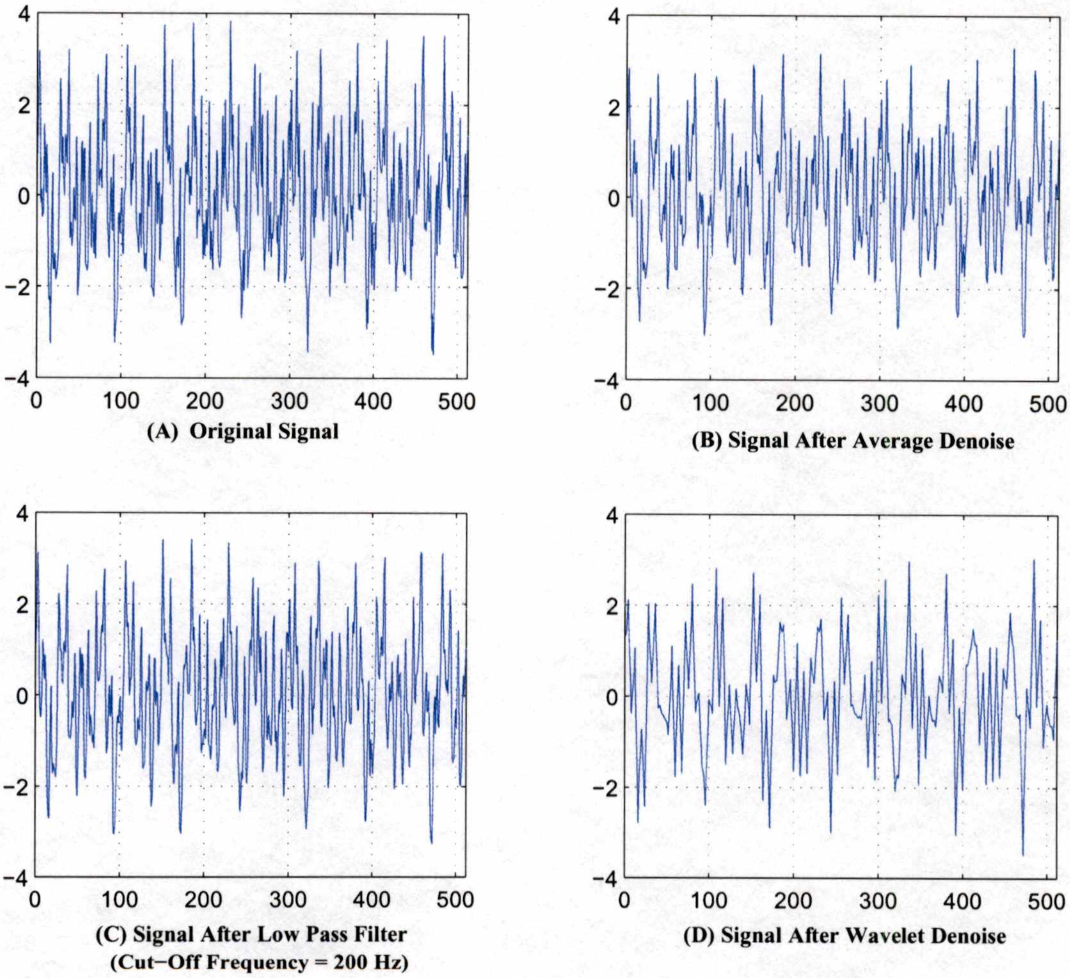


Figure 3.2: Comparison of Three Different Denoising Methods

where $x(i)$ is a sample of the original signal, $y(i)$ is a sample of the signal with zero DC component, μ_x is the average value of the original signal. The signal has N samples in total. After removing the DC component, the signal x with average value μ_x is converted to signal y whose average value is zero, and contains no DC component. Signal y will then undergo denoising process. We choose to use wavelet denoising.

3.1.2 Introduction to Wavelets

The basic idea of wavelet denoising is to ignore the lower values of the wavelet transform coefficients, and then by inverse wavelet transform, reconstruct the original signal with lower noise level. The wavelet transform is designed to analyze non-stationary signals. The original idea of wavelets comes from the Fourier transforms, which represent the signal in frequency domain. Unlike the complex sine and cosine functions of the Fourier transform, the basis functions of wavelet transform are localized in both space and frequency domain [45]. The wavelet transform aims at representing the time function in terms of simple, fixed models, which are called *wavelets.* Wavelets are mathematical functions to analyze the signal according to scale. They are a series of functions that satisfy the particular requirements and are used in representing data or other functions [24]. These wavelets are derived from a single generating function which is called *mother wavelet.* The mother wavelet should meet the condition in Eq. 3.2. The translation and scaling of the mother function will generate a family of functions described in Eq. 3.3 [9]. An important property of mother wavelet is that it is orthogonal to all the family wavelets.

$$\int \psi(t)dt = 0 \tag{3.2}$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{3.3}$$

where

$$\begin{aligned} a &= \text{Scaling factor,} \\ b &= \text{Translation factor.} \end{aligned}$$

A typically-used wavelet is $\psi = (1-t^2)e^(-t^2/2)$, which is called the *Mexican hat function*. The change of a will make the wavelets cover different frequency ranges, the greater $|a|$ is, the smaller the frequency it represents. The change of b corresponds to the change of the time center location. In other words, the wavelet transform uses narrow windows when the signal frequencies are high and wide windows when the signal frequencies are low. This representation allows two wavelet transform to "enlarge" every high-frequency component, such as the transients in signals. This is the main advantage of the wavelet transform over the Fourier transform.

The wavelet transform could be categorized into continuous wavelet transform (CWT) and discrete wavelet transform (DWT). The definitions of CWT and DWT are shown in Eq. 3.4 [14].

$$\begin{aligned} CWT_f(a,b) &= |a|^{-1/2}\int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt \\ DWT_f(m,n) &= a_0^{-m/2}\int f(t)\psi\left(a_0^{-m}t - nb_0\right)dt \end{aligned} \tag{3.4}$$

where

$$a = a_0^m \quad , \quad b = nb_0 a_0^m$$
$$a_0 > 1 \quad , \quad b_0 > 0$$
$$m, n \in \mathcal{Z}$$

In the DWT, the family of wavelets is given as Eq. 3.5

$$\psi_{m,n}(t) = a_0^{-m/2} \psi \left(a_0^{-m} t - nb_0 \right) \tag{3.5}$$

Considering the computation complexity, in sampled signals of finite length, we often use the DWT.

Some examples of the mother wavelets are shown in Fig. 3.3, including Daubchies wavelet, Coiflet wavelet, Harr wavelet, Symmlet wavelet, Meyer wavelet and Battle-Lemarie wavelet. Among these wavelets, the Daubechies wavelet, named after a famous mathematician in Princeton University, is popularly used in the engineering field because it uses a smooth finite-length kernel function to describe the details between scales of approximations, and thus suitable for solving problems in signal denoising, image compression, etc. The parameters of the corresponding Daubechies wavelets are given by sequences $(h_0, h_1, ...h_N)$ which satisfy the conditions described in Eq. 3.6.

$$\begin{aligned} \sum_{n=0}^{N} h_n &= \sqrt{2} \\ \sum_{n=0}^{N} (-1)^n n^k h_n &= 0, \text{for } k = 0, 1, 2, ...(N-1)/2 \end{aligned}$$

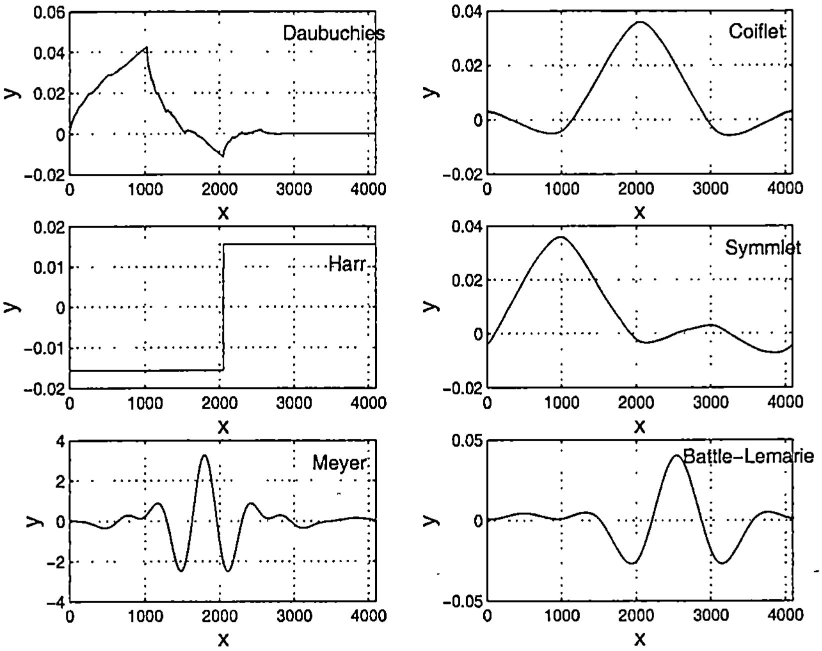


Figure 3.3: Six Popularly Used Mother Wavelets

$$\sum_{n=0}^{N-2k} h_n h_{n+2k} \;=\; 0, \text{for } k = 1, 2, 3, ...(N-1)/2 \tag{3.6}$$

In practice, the forward and inverse wavelet transform could be implemented using a set of sampling functions called *digital filter banks* as shown in Fig. 3.4. In the forward transform, the output of the filter banks are the wavelet coefficients at a particular resolution. The inverse transform is processed in the opposite direction.

First of all, the sampled data $f(t)$ will be input in parallel to a low-pass filter (H) and a high-pass filter (G). The outputs of the two filters are downsampled and kept exactly one half of the size of the input signals. After the first stage, the output of the high-pass filter becomes the wavelet coefficients $C4$ at level N (In

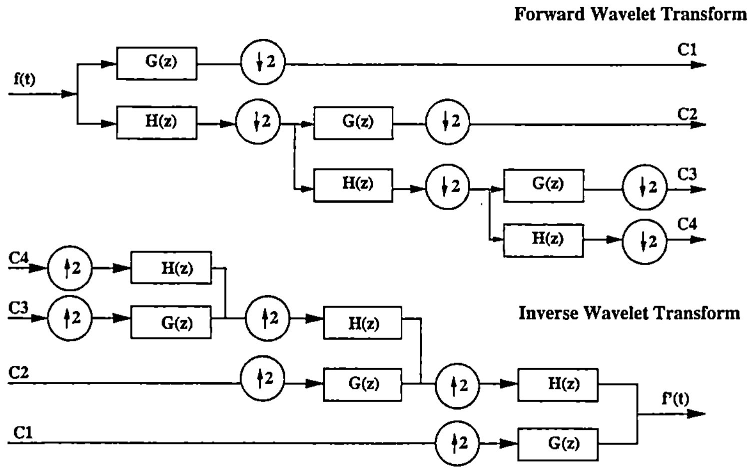


Figure 3.4: The Implementation of Wavelet Transform

Fig. 3.4, N=3). These coefficients represent the highest frequency wavelet level of the transform. The second stage uses the output of the low-pass filter of the previous stage as the input and the output of the high-pass filter of this stage becomes the wavelets coefficients $C3$ at level N-1. The same process will continue until all the stages end.

The inverse discrete wavelet transform (IDWT) works in the reverse direction. Coefficients $C4$ and $C3$ are upsampled and then passed through the low-pass filter H and high-pass filter G respectively. The sum of the outputs is then calculated and served as the input of the next stage. This process will continue until the all the wavelets are upsampled and filtered. The final result will be the reconstructed signal $f'(t)$.

In the SSWCC algorithm, I choose the functions from Daubechies wavelet family as the low-pass filter H and high pass filter G, which are shown in Fig. 3.5.

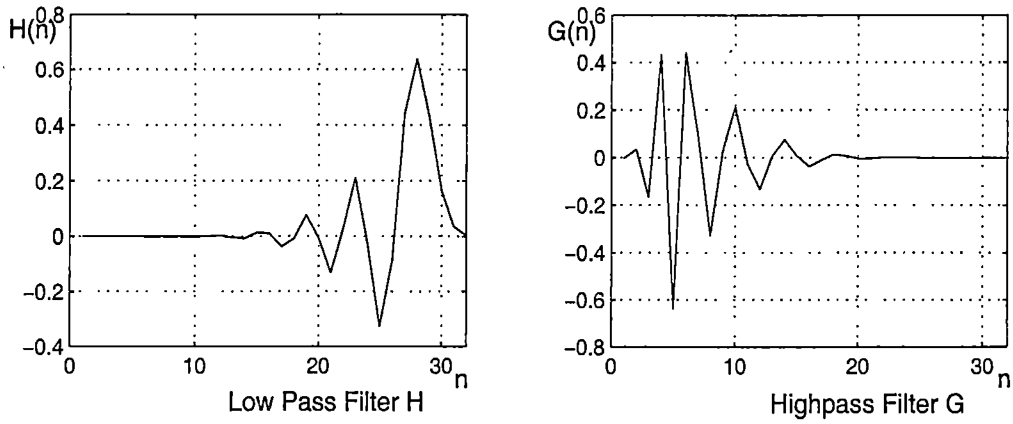


Figure 3.5: Wavelet Transform

3.1.3 Wavelet Denoising

The raw data we received from the sensor include unavoidable noise, which will affect the accuracy of feature extraction. Therefore, a denoising process is needed. The classical technique of removing noise is filtering. In the wavelet transform, the filter results are represented by wavelet coefficients. So by processing the wavelet coefficients, the noise effect to the signal could be ignored.

Donoho et al. [16][17][18] states the details of the use of wavelet transform in denoising. Suppose signal $x(t)$ is corrupted by white noise $n(t)$ with variance σ^2, which results in the noisy signal $y(t)$, then we can describe the noisy signal as Eq. 3.7.

$$y(t) = x(t) + n(t) \tag{3.7}$$

Equation 3.8 indicates the wavelet transform of the noisy signal $y(t)$.

$$W_\Psi(y(t)) = W_\Psi(x(t) + n(t)) = W_\Psi(x(t)) + W_\Psi(n(t)), \tag{3.8}$$

where W_Ψ denotes the wavelet transform. Since the basis functions of the wavelet transform are orthogonal, the wavelet transform of the white noise is still white noise $w(t)$ with the same variance [13]. Therefore, we can derive the original signal as Eq. 3.9.

$$x(t) = W_\Psi^{-1}\left(W_\Psi(y(t)) - w(t)\right) \tag{3.9}$$

where W_Ψ^{-1} represents the inverse wavelet transform. Thus, we can estimate the denoised signal by Eq. 3.10.

$$\hat{x}(t) = W_\Psi^{-1}\left(W_\Psi(y(t)) - \lambda\right) \tag{3.10}$$

This formula shows us that if we can estimate the noise contribution λ in wavelet coefficients, we can remove the noise effect from signal $y(t)$ by applying the inverse wavelet transform to the adjusted wavelet coefficients $W_\Psi(y(t)) - \lambda$.

The value of λ is called the *threshold* in the wavelet denoising problem. It is an important parameter since it affects the performance of denoising directly. The estimation of λ is a complicated problem. Many algorithms have been developed for this purpose [26][27][38]. We choose the *universal threshold* developed by Donoho and Johnstone [16][17], which is shown in Eq. 3.11.

$$\lambda = \sigma\sqrt{2log(N)} \tag{3.11}$$

where N is the number of samples and σ is the standard deviation of the noise level. Since we cannot calculate σ directly, we should estimate it from the wavelet coefficients. A common estimator is shown in Eq. 3.12.

$$\sigma = \frac{1}{\frac{N}{2} - 1} \sum_{i=1}^{N/2} (d_i - \mu)^2 \tag{3.12}$$

where d is the wavelet coefficients in the first level, which is shown in Fig. 3.4 as $C1$, and μ is the mean of the wavelet coefficients. Another estimation to σ is:

$$\sigma = \frac{MAD}{0.6745} \tag{3.13}$$

where MAD is the median value of the wavelet coefficients in the first level.

From the above analysis, we can draw the following conclusion: when wavelet coefficients are small enough, they can be ignored without substantially affecting the main features of the original data set.

Equation 3.14 represents the thresholding method we used to denoise the signal. Since we apply the same threshold to all the wavelet coefficients, it is called a *hard-threshold selection method* [10].

$$W_i = \begin{cases} W_i, & \text{if } |W_i| \geq \sigma\sqrt{2log(N)} \\ 0, & \text{otherwise} \end{cases} \tag{3.14}$$

Figure 3.6 illustrates the effect of wavelet denoising to a segment of seismic signal. Figure 3.6 (A) represents the original time series signal which contains some sharp peaks as noise. After three-stage wavelet transform, the wavelet coefficients are shown in Fig. 3.6 (B). A threshold is then calculated as the two red lines depicted in Fig. 3.6 (B), indicating the positive threshold line and negative threshold line respectively. All the coefficients whose values are between these two lines are

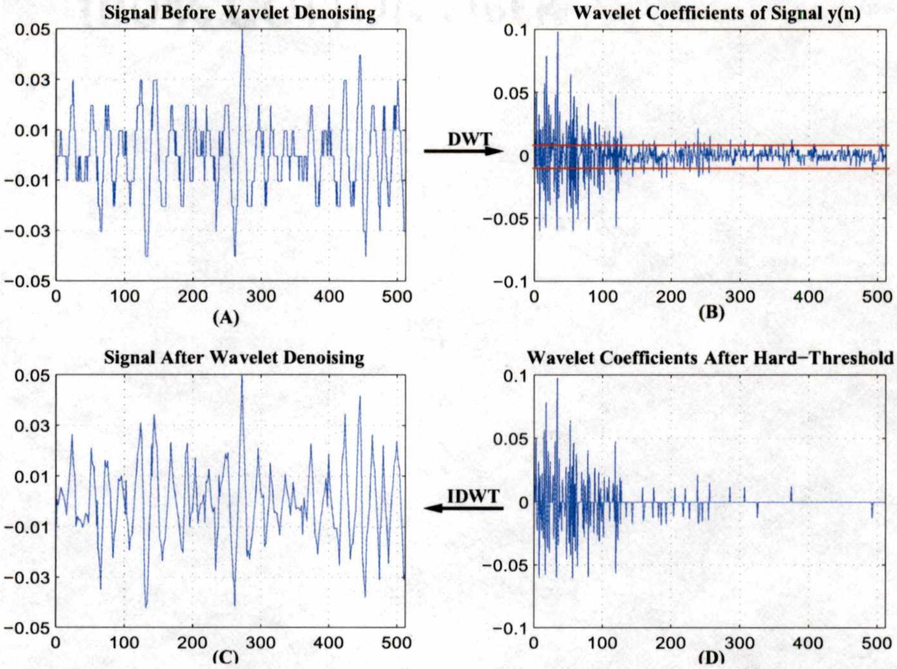


Figure 3.6: The Signals Before and After Wavelet Denoising

adjusted to zeros, others are kept unchanged. The new coefficients after thresholding are shown in Fig. 3.6 (D). Finally, IDWT is applied to these coefficients, and we obtain the denoised signal as shown in Fig. 3.6 (C).

Compared to the original signal, the denoised signal in Fig. 3.6 (C) is much smoother, and some of the sharp peaks are removed. This result demonstrates the efficiency of wavelet denoising. After denoising, the signal in Fig. 3.6 (C) serves as the input signal of the feature extraction step.

3.2 Feature Analysis

After the original signal is preprocessed, signal features will be extracted from the spectral, the PSD and the wavelet analysis. Feature normalization is then applied to normalize the features to the same scale. This section is the most important part for feature extraction. Our goal is to obtain a set of robust features from different analysis domains.

3.2.1 Spectral Analysis

Spectral analysis is one of the most popularly used methods in signal processing. The idea of spectral analysis comes from the Fourier transform, which describes the relationship between the signal in time domain and the signal in frequency domain. Fourier transform is defined as Eq. 3.15.

$$F[s(t)] = S(f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft}dt \tag{3.15}$$

where f is the frequency, and t is the time. With the Fourier transform, it is convenient for us to analyze the signal in the frequency domain. In signal processing problem, the study of signal in the frequency domain is named as *spectral analysis.*

First of all, Discrete Fourier transform (DFT) is used to convert the signal from the time domain to the frequency domain. Then features of the signal are extracted from the spectra of the signals.

As discussed in Chapter 2, seismic signals from ground vehicles are generated from the vibration of vehicle engines and their strike on the ground. Thus, the seismic signal frequency of the signal should be identical for a certain engine. Spectral analysis is able to reveal the frequency information of a vehicle, and serve as one of the major features of the vehicle. Given a series of sampled data $x(k)$ from a certain

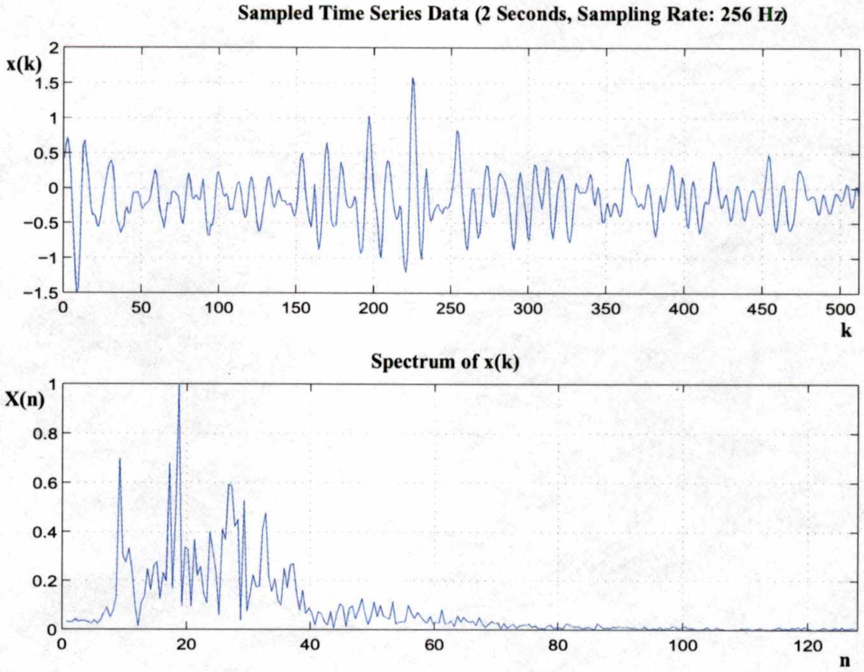


Figure 3.7: A Typical Spectrum of a Segment of Seismic Signal

signal, Equation 3.16 calculates the DFT of this signal.

$$X(n) = \sum_{k=0}^{N-1} x(k) e^{(-2\pi j)\frac{nk}{N}} \tag{3.16}$$

where

$$\begin{aligned} N &= \text{The number of samples,} \\ x(k) &= \text{The time series data,} \\ X(n) &= \text{The DFT series} \end{aligned}$$

Figure 3.7 shows a typical spectrum of a segment of seismic signal.

From the spectrum of the signal we extract the following features:

1. Shape Statistics: The shape statistics provides a higher order statistics in terms of mean, standard deviation, skewness and kurtosis. They are defined in Eq. 3.17 [29].

$$\begin{aligned}
\text{Mean:} \quad & \mu_{shape} = \frac{1}{S}\sum_{i=1}^{N} iC(i) \\
\text{Standard deviation:} \quad & \theta_{shape} = \sqrt{\frac{1}{S}\sum_{i=1}^{N}(i-\mu_{shape})^2 C(i)} \\
\text{Skewness:} \quad & \gamma_{shape} = \frac{1}{S}\sum_{i=1}^{N}(\frac{i-\mu_{shape}}{\theta_{shape}})^3 C(i) \\
\text{Kurtosis:} \quad & \beta_{shape} = \frac{1}{S}\sum_{i=1}^{N}(\frac{i-\mu_{shape}}{\theta_{shape}})^4 C(i) - 3
\end{aligned} \tag{3.17}$$

where

$$\begin{aligned}
S &= \sum_{i=1}^{N} C(i), \\
C(i) &= \text{the power spectral magnitude for the } i\text{th frequency bin,} \\
N &= \text{the number of the frequency bins.}
\end{aligned}$$

2. Amplitude Statistics: The amplitude statistics are described in Eq. 3.18. They provide a statistical measurement of local spectral energy content over the signal band [29].

$$\begin{aligned}
\text{Amplitude:} \quad & \mu_{amp} = \frac{1}{N}\sum_{i=1}^{N} C(i) \\
\text{Standard deviation:} \quad & \sigma_{amp} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(C(i)-\mu_{amp})^2}
\end{aligned}$$

$$\text{Skewness:} \qquad \gamma_{amp} = \frac{1}{N}\sum_{i=1}^{N}(\frac{C(i) - \mu_{amp}}{\sigma_{amp}})^3$$

$$\text{Kurtosis:} \qquad \beta_{amp} = \frac{1}{N}\sum_{i=1}^{N}(\frac{C(i) - \mu_{amp}}{\sigma_{amp}})^4 - 3 \tag{3.18}$$

3. The location of the largest three peaks from the spectrum. The peak location in the spectrum tells us the location of the dominant frequency. In other words, the frequencies corresponding to the peaks in the spectrum represent the dominant frequencies of the vibration of vehicles.

Until now, we have extracted 11 features from the spectrum. However, in the above analysis, when we sample the infinite signal, we take a signal segment over a certain time period T and consider the signal to be a periodic signal of period T. The sampling process multiplies the signal with a unit rectangular function of width T. Since the rectangular function turns "on" and "off" very rapidly, it will bring the high frequency components that cause contamination of the original signal. This effect is called *leakage*. Although the rectangular window generates the maximal sharpness, it brings the frequency leakage to the spectrum simultaneously. To avoid this effect, we should consider to use other window functions.

Several popularly used window functions include Hanning window, Hamming window, and Blackman window. They are shown in Fig. 3.8. Figure 3.9 illustrates the spectra derived by using different window functions. We can observe that Hamming window and Hanning window have very similar results, but Blackman window reduces the ripple ratio of the spectrum further and makes the spectrum smoother. Therefore, in SSWCC, we adapt Blackman window as the window function. Black-

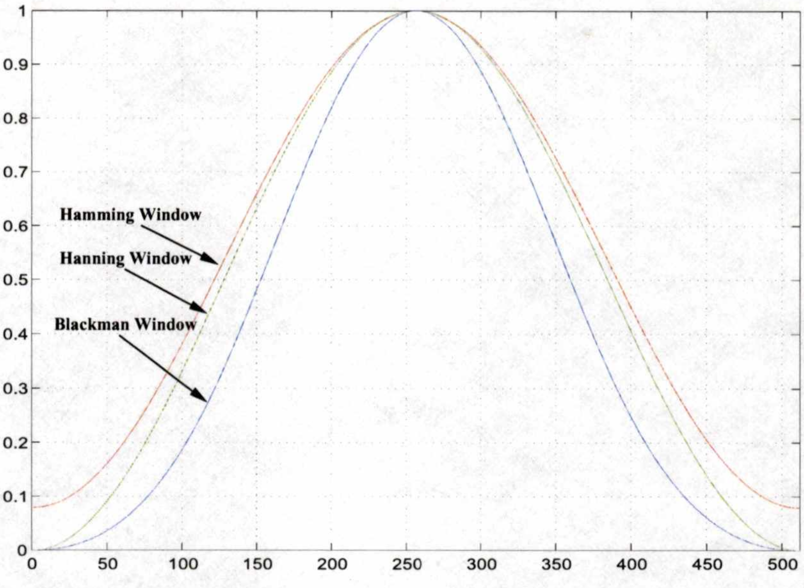


Figure 3.8: Three Different Window Functions

man window is defined as Eq. 3.19 [25].

$$\omega_M(n+1) = 0.42 - 0.5cos(\frac{2\pi n}{N_1 - 1}) + 0.08cos(\frac{4\pi n}{N_1 - 1}),\ n = 0, 1, 2, ...N_1 - 1 \quad (3.19)$$

The result of using Blackman window is shown in Fig. 3.10. It is obvious that the noise level in Fig. 3.10 (D) is much less than that in Fig. 3.10 (B), and the dominant frequency peaks are more distinctive than those in Fig. 3.10(B) as well. We also extract the same 11 features from the Blackman windowed spectrum as we did for the spectrum without using the Blackman window. Thus, from spectral analysis, we have derived 22 features. They are obtained from two groups, one is the spectrum without using window function, the other is the spectrum by using the Blackman window function. Performance evaluation is conducted in Chapter 6 between the two groups of features.

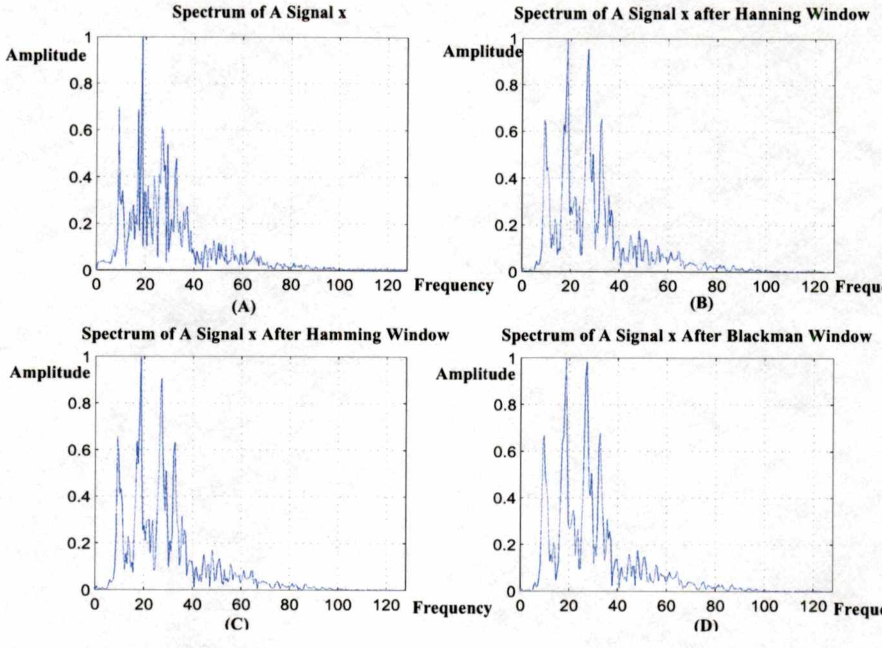


Figure 3.9: The Effects of Window Functions on Spectrum

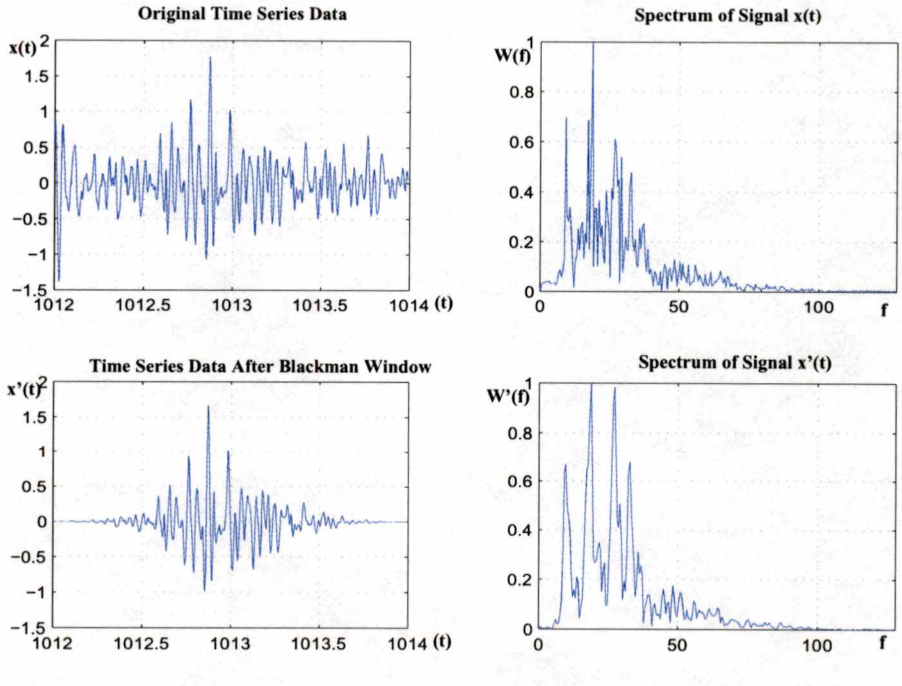


Figure 3.10: Effect of the Blackman Window

3.2.2 Wavelet Analysis

In Sec. 3.1.2, we have reviewed the basic ideas about the wavelet transform, and have used it to denoise the signal. In this section, wavelet transforms are used again to extract useful feature vectors.

As described in Sec. 3.1.2, the decomposition of signals into wavelets begins from the data series $x(n)$. After the decomposition, we will obtain a series of wavelet coefficients in the same length as $x(n)$, as shown in Fig. 3.4. Similar to the spectral representation of the signal, we can measure the original signal $x(n)$ by wavelet coefficients as Eq. 3.20 [40],

$$x(n) = \sum_{k=-\infty}^{\infty} [h_{n-2k}a_{1k} + (-1)^n h_{2k+1-n}d_{1k}] \tag{3.20}$$

where h is the Daubechies wavelets that satisfy the conditions described in Eq. 3.6, a and d are wavelet coefficients.

Another important concept in wavelet analysis is the energy of the wavelets, defined as:

$$||p||^2 = \sum_{n=-\infty}^{\infty} p_n^2 \tag{3.21}$$

where p_n is a data series. The wavelet theory has proven [40] that the energy of the signal is equal to summation of the energies of the wavelet coefficients sequence. That is,

$$||x_n||^2 = ||a_M||^2 + \sum_{m=1}^{M} ||d_m||^2 \tag{3.22}$$

where M is the maximum level of the wavelet transforms. and d_m and a_M are the wavelet coefficients from different levels.

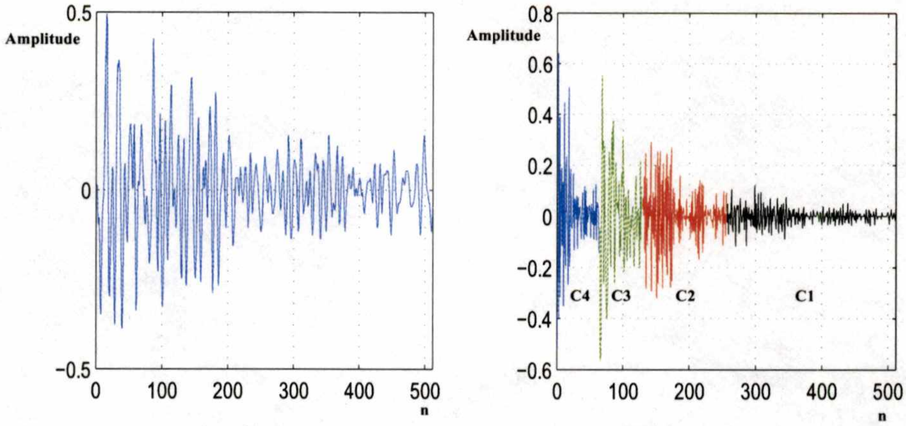


Figure 3.11: Different Levels of Wavelet Coefficients

From the above description, we see that the coefficients of the wavelets can represent the original signal in two different ways: the decomposition described in Eq. 3.20, and the energy. Since these two aspects of wavelets are related to the type of the signal as well, we use them to derive another series of features.

From Fig. 3.4, we already know that for a three-band wavelet transform, we can derive four series of coefficients, $C1$ to $C4$. Figure 3.11 shows a segment of seismic signal and its wavelet transform. Different colors correspond to different levels of coefficients. Since the time series data we analyzed are 2-second intervals with a sampling rate of 256 Hz, the wavelet coefficients also have 512 samples. The first 64 points correspond to $C4$ in Fig. 3.4, which belong to the third scale residue of the wavelet transform. The second 64 points correspond to $C3$ that belong to the third scale high pass band. The next 128 points correspond to $C2$ that belong to the second scale high pass band. The last 256 points are $C1$ that belong to the first scale high pass band.

Existing methods use the wavelet coefficients themselves to serve as a feature set.

However, in our case, we have 512 coefficients, which is hard for classifiers to use the whole coefficient set as features. On the other hand, it can make features from other spaces be ignored because the amount of features from the wavelet coefficients are far more than those from other domains. Therefore, we choose to calculate only three features from each level of the transformation: the average, the variance, and the energy. Since we use three-band wavelet transform, we get 4 levels of coefficients for each group. Therefore, we have $4 \times 3 = 12$ features from the wavelet transform.

The use of the wavelet transform in pattern recognition has been a hot research field, many new algorithms are used to develop robust feature vectors from the wavelet transform [31] [34]. These algorithms require more complex computation and are not as efficient as the algorithm we proposed. Since our algorithm aims at real-time signal processing, these simplified feature extraction will save a lot of computation cost.

3.2.3 Power Spectral Density Analysis

Power spectral density (PSD) describes the energy distribution of a signal in the frequency domain. Mathematically, PSD is defined as the Fourier transform of the autocorrelation of the time series signal, which is shown in Eq. 3.23

$$P_{xx}(n) = F[R_{xx}(k)] = \sum_{k=0}^{N-1} R_{xx}(k) e^{-j2\pi \frac{nk}{N}} \tag{3.23}$$

where $R_{xx}(k)$ is the autocorrelation function of $x(m)$, defined in Eq. 3.24

$$R_{xx}(k) = \frac{1}{N-k} \sum_{k=1}^{N-k} x(m)x(m+k) \tag{3.24}$$

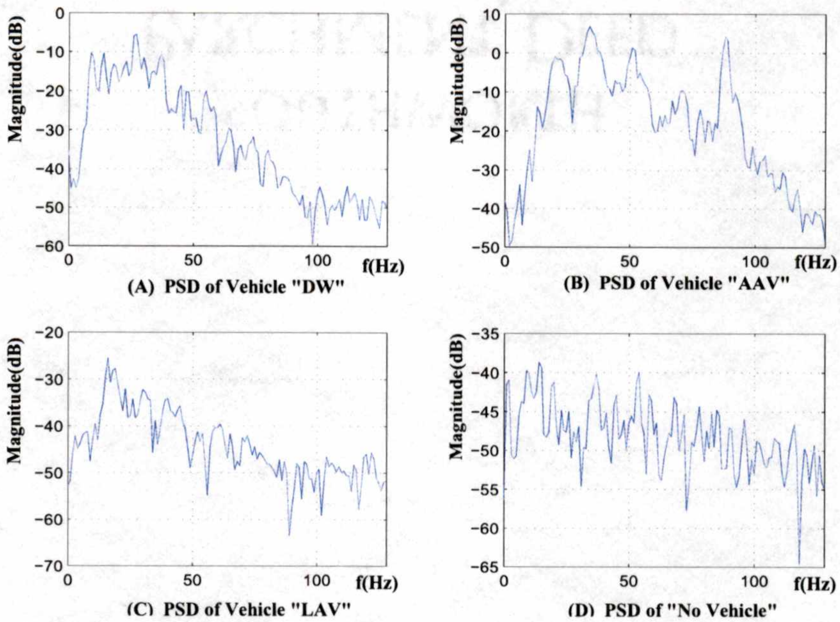


Figure 3.12: PSDs of Different Vehicles

Although PSD can be calculated directly, we always use Welch's averaged, modified periodogram method for engineering use. Detailed description of this algorithm can be found in [47].

At a certain noise level, the peaks in the PSD shows the most powerful frequency components in the signal, while the lower ones show the weak components in the signal. Therefore, the shape of the PSD also represents the characteristics of the signal. Figure 3.12 shows four different PSD shapes, among which the first three (Fig. 3.12 (A) to (C)) indicate the PSDs of seismic signals generated from three different vehicles, and the fourth one in Fig. 3.12 (D) is the PSD of the seismic signal when no vehicle is present. The difference is obvious in the peak locations and the energy distribution pattern.

The conventional method of analyzing the PSD is to calculate the averaged PSD in a series of time blocks to represent the signature of the signal. However, in our problem, signals are required to be processed in real time, which means information from other time blocks is not available. Therefore, we select the locations of the largest three peaks from each PSD as the features of the signal. To describe the shape and the distribution of the signal, we use the shape statistics described in Eq. 3.17 again. Thus we have 7 features from the PSD analysis.

In summary, a total of 41 features are obtained from the spectral, the PSD and the wavelet analysis. A summary of these features are shown in Table. 3.1

3.3 Feature Normalization

Although we have extracted all the potential features, the scales of these features are not the same. For example, feature 1 is the shape of the spectrum, which is always around 30, but feature 5 is the amplitude of the normalized spectrum, which is always between 0 and 1. If we use these features directly as the inputs to the classifier, feature 5 will be ignored since it is too small. Therefore, all the features should be normalized to the same scale to be compared.

The normalization process is expressed as Eq. 3.25 [29]:

$$\hat{x}(i,j) = \frac{x(i,j) - \mu_j}{\sigma_j} \tag{3.25}$$

where

$$\begin{aligned} i &= 1, 2, 3, ...M, \\ j &= 1, 2, 3, ...N, \end{aligned}$$

Table 3.1: The Feature Vector

Feature Number	Description
1	Mean of Spectrum
2	Standard Deviation of Spectrum
3	Skewness of Spectrum
4	Kurtosis of Spectrum
5	Amplitude of the Normalized Spectrum
6	Standard Deviation of Normalized Spectrum
7	Skewness of the Normalized Spectrum
8	Kurtosis of the Normalized Spectrum
9	Mean of Balckman Windowed Spectrum
10	Standard Deviation of Blackman Windowed Spectrum
11	Skewness of Blackman Windowed Spectrum
12	Kurtosis of Blackman Windowed Spectrum
13	Amplitude of the Blackman Windowed Normalized Spectrum
14	Standard Deviation of Blackman Windowed Normalized Spectrum
15	Skewness of the Blackman Windowed Normalized Spectrum
16	Kurtosis of the Blackman Windowed Normalized Spectrum
17	The First Peak Location in the Spectrum
18	The Second Peak Location in the Spectrum
19	The Third Peak Location in the Spectrum
20	The First Peak Location in the Blackman Windowed Spectrum
21	The Second Peak Location in the Blackman Windowed Spectrum
22	The Third Peak Location in the Blackman Windowed Spectrum
23	The First Peak Location in PSD
24	The Second Peak Location in PSD
25	The Third Peak Location in PSD
26	Shape of PSD
27	Standard Deviation of PSD
28	Skewness of PSD
29	Kuritosis of PSD
30	Average of the wavelet coefficient series 1
31	Variance of the wavelet coefficient series 1
32	Average of the wavelet coefficient series 2
33	Variance of the wavelet coefficient series 2
34	Average of the wavelet coefficient series 3
35	Variance of the wavelet coefficient series 3
36	Average of the wavelet coefficient series 4
37	Variance of the wavelet coefficient series 4
38	Energy of the wavelet coefficient series 1
39	Energy of the wavelet coefficient series 2
40	Energy of the wavelet coefficients series 3
41	Energy of the wavelet coefficients series 4

$$
\begin{aligned}
\mu_j &= \frac{1}{M}\sum_{i=1}^{M} x(i,j), \\
\sigma_j &= \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(x(i,j)-\mu_j)^2} \\
M &= \text{the number of samples in the whole set} \\
N &= \text{the number of elements in one feature vector} \\
\hat{x}(i,j) &= \text{normalized feature element,} \\
x(i,j) &= \text{original feature element,}
\end{aligned}
$$

The effect of this normalization is to make the feature vector have zero mean and unit variance.

3.4 Principal Component Analysis

The elements of the feature vectors can be strongly correlated with each other. Therefore, after we extract as many features as we can from the signal, it is necessary to reduce the feature set to a minimum but sufficient one. The principal component analysis (PCA) technique is one of the most popularly used algorithms in dimensionality reduction.

PCA is also known as the Karhunen-Loeve transformation. It can be used to reduce the dimension of the features while keeping the most important information untouched in the feature sets by a linear transformation matrix. The transformation matrix is made up of the eigenvectors of the feature vector covariance matrix.

Suppose after the features are normalized, we obtain an $M \times N$ feature matrix X, where M is the number of observations, and N is the dimension of the feature vector. The goal of the PCA transformation is to find an $N \times K$ matrix R, on which the original feature set can be transformed to get a lower-dimension feature set. The procedure can be represented in Eq. 3.26.

$$Y_{M\times K} = X_{M\times N} R_{N\times K} \tag{3.26}$$

where K is the feature vector sets after the transformation, and K is less than N.

The process to find matrix R is as follows: first of all, we calculate the covariance matrix S of the feature vector matrix, which is defined in Eq. 3.27.

$$S = E(\mathbf{x}-\mu)^T(\mathbf{x}-\mu) = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x_i}-\mu)^T(\mathbf{x_i}-\mu) \tag{3.27}$$

where $\mathbf{x_i}$ is the feature vector, and μ is the mean vector defined as Eq. 3.28.

$$\mu = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x_i} \tag{3.28}$$

Since after normalization, the mean of the feature vectors in the training set is zero, Equation 3.27 can be simplified as:

$$S_{N\times N} = \frac{1}{M}X^TX = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x_i}^T\mathbf{x_i} \tag{3.29}$$

The next step is to calculate the eigenvalues (λ) and the eigenvectors (v) of the covariance matrix S. We choose the K largest λ's and the corresponding eigenvectors. Suppose $\lambda_1 \geq \lambda_2 \geq ...... \geq \lambda_K$, then the eigenvectors will form the transformation matrix R as Eq. 3.30.

$$R_{N\times K} = [v(1)v(2)v(3)......v(K)] \tag{3.30}$$

The choice of K is determined by Eq. 3.31

$$\frac{\sum_{i=1}^{K}\lambda_i}{\sum_{i=1}^{M}\lambda_i} \geq 1-\eta \tag{3.31}$$

where η is the loss of the energy.

After K is selected, the transformation matrix R in Eq. 3.30 is formed. By applying this equation to the training set and the test set independently, we can then derive a set of the most dominant features.

3.5 Summary

This chapter describes the feature extraction procedure in the proposed algorithm — SSWCC. The algorithm extracts 41 features from the spectral, the PSD and the wavelet analysis. The feature set are then normalized and optimized by principal component analysis (PCA). The results of the feature extraction serve as the input to the classifier described in the next chapter to obtain the final identification results.

Chapter 4

Classifier Design

After the feature extraction and the dimension reduction, we need to do the classification. Classifers are actually many-to-one mapping functions which assign a class to each input sample. For a two-class problem, the mapping functions can be described as Eq. 4.1 [29].

$$y = f(x) \Longrightarrow \begin{cases} x \in \text{class 1 if} & y \geq \gamma \\ x \in \text{class 2 if} & y < \gamma \end{cases} \tag{4.1}$$

where x is the input feature vector, y is the output of the classification function. γ is the threshold used to determine which class the input feature vector belongs to. The classifier design is to construct function $f(\cdot)$ from the training set and then apply it to the test set for classification.

In this chapter, Bayesian decision theory, the fundamental theory of pattern classification problem, is reviewed. We then focus the discussion on the design of two classifiers: the *minimum-distance classifier* which assumes the probability density function (pdf) of the sample variable is Gaussian, and the *k-nearest-neighbor classifier* which does not make any assumption to the pdf of the sample variable.

4.1 Bayesian Decision Theory

Bayesian decision theory is a statistical approach to pattern classification problems. This approach is based on the assumption that the decision problem is posed in probabilistic terms, and all of the relevant probabilities are known [20]. Suppose $\mathbf{x}$ is a measurement vector in d dimension, n is the number of classes. The decision statistics is to compare the *posterior* probabilities so that:

$$\text{if } p(\omega_i|\mathbf{x}) > p(\omega_j|\mathbf{x}), (1 \leq i, j \leq n) \Longrightarrow \mathbf{x} \in \omega_i \tag{4.2}$$

The *posterior* probabilities $p(\omega_i|\mathbf{x})$ can be calculated using Bayes formula:

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{Z} \tag{4.3}$$

where Z is a normalization factor.

Thus, the decision statistics can be written as

$$\text{if } p(\mathbf{x}|\omega_i)P(\omega_i) > p(\mathbf{x}|\omega_j)P(\omega_j), (1 \leq i, j \leq n) \Longrightarrow \mathbf{x} \in \omega_i \tag{4.4}$$

We define *discriminant function* as a function to discriminate or differentiate among multiple categories. In the above example, the discriminant function can be defined

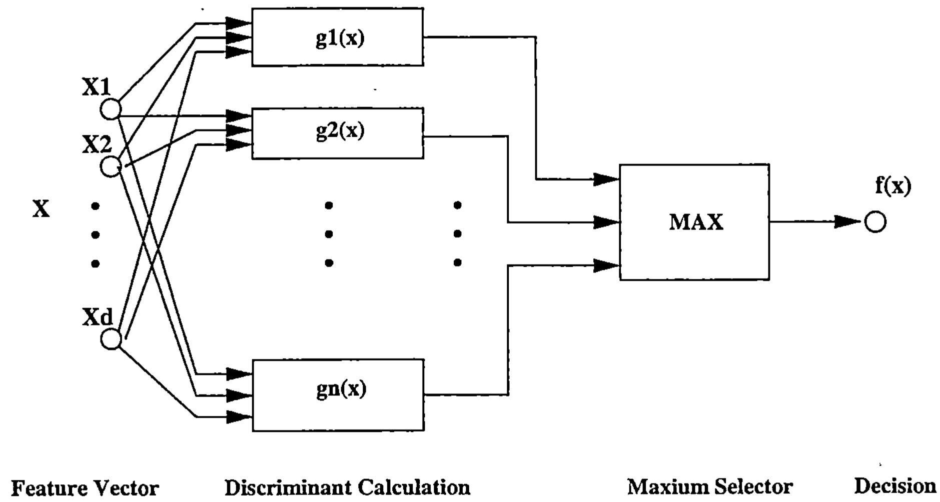


Figure 4.1: A Typical Pattern Classifier Structure

as:

$$g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i) \tag{4.5}$$

From the description above, we find another way to represent pattern classifier in terms of a set of discriminant functions $g_i(\mathbf{x})$. The feature vector $\mathbf{x}$ is classified to class ω_i if $g_i(\mathbf{x}) > g_j(\mathbf{x})$ for any $i \neq j$.

Thus the classifier can be viewed as a machine that computes n discriminant functions and selects the category corresponding to the largest discrimination function. This procedure is illustrated in Fig. 4.1. All classifiers can be viewed as a discriminant calculator or a function mapping from the feature space to the decision space [20].

4.2 Minimum-Distance Classifier

From Eq. 4.5, we know that in Bayesian classifier, the discriminant function $g(\mathbf{x})$ is determined by the conditional density $p(\mathbf{x}|\omega_i)$ and the prior probability $P(\omega_i)$. Among the density functions we known, the multivariate normal or Gaussian density is the most widely used density function. The minimum-distance classifier is designed by assuming that the conditional density is a Gaussian distribution. The general multivariate normal density in d dimensions can be written as Eq. 4.6 [21]:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\left[-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right] \tag{4.6}$$

where

$$\begin{aligned} \mathbf{x} &= d\text{-component column vector,} \\ \mu &= d\text{-component mean vector,} \\ \Sigma &= d \text{ by } d \text{ covariance matrix,} \end{aligned}$$

To further simplify the computation, we apply a natural logarithm on both sides of the discriminant function, that is, replace every $g_i(\mathbf{x})$ with function $f(g_i(\mathbf{x}))$, where $f(\cdot)$ is the natural logarithm. Since $f(\cdot)$ is a monotonic function, we do not change the same classification results. Thus the discriminant function in Eq. 4.5 can be simplified to Eq. 4.7

$$f(g_i(\mathbf{x})) = \ln(p(\mathbf{x}|\omega_i)) + \ln(P(\omega_i)) \tag{4.7}$$

If we assume the conditional densities $p(x|\omega_i)$ are multivariate normal functions —$N(\mu_{\mathrm{i}}, \sigma_{\mathrm{i}})$, and substitute Eq. 4.6 into Eq. 4.7, then we can derive a simplified discriminant function shown in Eq. 4.8

$$g_i(\mathrm{x}) = -\frac{1}{2}(\mathrm{x} - \mu_{\mathrm{i}})^T \Sigma_{\mathrm{i}}^{-1}(\mathrm{x} - \mu_{\mathrm{i}}) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_{\mathrm{i}}| + \ln P(\omega_i) \tag{4.8}$$

If we further assume that all features are statistically independent and each feature has the same variance σ^2, we can write the covariance matrix as: $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$, and the inverse covariance matrix as: $\mathbf{\Sigma_i}^{-1} = (1/\sigma^2)\mathbf{I}$. The final version of the discriminant function is shown in Eq. 4.9 [21].

$$g_i(\mathrm{x}) = -\frac{||\mathrm{x} - \mu_{\mathrm{i}}||^2}{2\sigma^2} + \ln P(\omega_i) \tag{4.9}$$

where $||\cdot||$ denotes the *Euclidean distance* and $||\mathbf{x} - \mu_{\mathbf{i}}|| = (\mathbf{x} - \mu_{\mathbf{i}})^T(\mathbf{x} - \mu_{\mathbf{i}})$.

If the prior probabilities $P(\omega_i)$ are the same for all classes, then $\ln P(\omega_i)$ can be ignored in the discriminant function. Therefore, we can use the first term in Eq. 4.9 as the discriminant function, that is, the Euclidean distance $||\mathbf{x} - \mu_{\mathbf{i}}||$ as the discriminant function.

By measuring the Euclidean distance between each feature vector and the mean vector of each class, we can assign $\mathbf{x}$ to the class of the nearest mean. This classifier is named the *minimum-distance classifier.* We apply it to classify the feature vectors in Chapter 5.

4.3 k-Nearest Neighbor Classifier

It is well known that Bayesian classifier is the optimal classifier since it is solved as an optimization problem. However, in practice, we have to make several approximations in order to carry out this classifier. For example, in practice, we use finite number of samples to estimate probabilities, and assume the probability density function (pdf) to have a special form, like Gaussian, as we did in the derivation of the minimum-distance classifier. Therefore, in practice, Bayesian classifier can never achieve its optimum performance. Another type of classifiers do not assume a particular pdf form. We named this procedure as *nonparametric* estimation.

There are several types of nonparametric methods for pattern recognition problems. The k-nearest neighbor (kNN) rule is one of the popularly used methods. This method is an extension to the nearest neighbor rule. (This rule classifies $\mathbf{x}$ by assigning it to the label that most frequently represented among the k nearest neighbors.)

Suppose we have a training set A, which contains N samples belong to class $\omega_1, \omega_2, ..., \omega_n$, also suppose in these N samples, the number of samples from ω_i is N_i, where i is from 1 to n. Then the problem is to classify an unknown sample $\mathbf{x}$ to a certain class ω_x. The principle of the kNN is to estimate the *posterior* probability directly, and then draw the decision rule. To estimate this probability, we need a center cell around $\mathbf{x}$ and suppose it is in the shape of a hypersphere cell $V(\mathbf{x})$ which encloses just k points in its volume. If in these points, k_i points belong to class ω_i, where $i = 1, 2, 3, ..., n$, then we could estimate the joint density $p(\mathbf{x}, \omega_i)$ by Eq. 4.10. Following Bayes rule in Eq. 4.11, we can estimate the *posterior* probability density as Eq. 4.12. Substitute Eq. 4.10 into Eq. 4.12, we can get the simplified *posterior* probability density in Eq. 4.13

$$
\begin{aligned}
p(x, \omega_i) &= \frac{k_i/n}{V(\mathbf{x})} &(4.10)\\
p(x, \omega_i) &= p(\omega_i|\mathbf{x})p(\mathbf{x}) &(4.11)\\
p(\omega_i|\mathbf{x}) &= \frac{p(\mathbf{x}, \omega_i)}{\sum p(\mathbf{x}, \omega_i)} &(4.12)\\
p(\omega_i|\mathbf{x}) &= \frac{\frac{k_i/n}{V(\mathbf{x})}}{\sum \frac{k_i/n}{V(\mathbf{x})}} = \frac{k_i}{k} &(4.13)
\end{aligned}
$$

From the above analysis, we can define the discriminant function as Eq. 4.14 and the decision rule as Eq. 4.15 directly.

$$g_i = k_i, i = 1, 2, 3, ..., n \tag{4.14}$$

$$g_i(x) = max(k_i) \Longrightarrow x \in \omega_i \tag{4.15}$$

As described above, how to measure the distance between the samples is the most important step in this algorithm. Suppose $d(\mathbf{x}, \mathbf{y})$ is a distance measure between sample $\mathbf{x}$ and $\mathbf{y}$, it should satisfy the following three conditions:

$$
\begin{aligned}
d(\mathbf{x}, \mathbf{x}) &= 0,\\
d(\mathbf{x}, \mathbf{y}) &= d(\mathbf{y}, \mathbf{x}),\\
d(\mathbf{x}, \mathbf{y}) &\leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{x}, \mathbf{z}), \text{for any } \mathbf{z} \text{ in the sample space}
\end{aligned}
$$

The distance measure we often used is the Euclidean distance. It is defined as

Eq. 4.16.

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\left(\sum_{i=1}^{n} |x_i - y_i|^2\right)} \tag{4.16}$$

The performance of the kNN is also dependent on the choice of k. Experimentally, k is choosen to $\sqrt{n}$ where n is the number of samples in the training set. We show the performance difference by using different k in Chapter 5.

4.4 Summary

This chapter reviews the fundamental theory of pattern classification, and then derives the minimum-distance and the kNN classifiers. The minimum-distance classifier assumes the pdf is Gaussian which simplifies the discriminant function to calculating the Euclidean distance between the feature vector and the mean vector, then classifies $\mathbf{x}$ to the nearest mean vector. kNN is a nonparametric estimation algorithm based on its nearest neighbor classifier. It estimates the density functions from the nearest k neighbors of the vector $\mathbf{x}$, and uses this estimation to derive the discriminant function. The performance of these two classifers will be evaluated in the next chapter.

Chapter 5

Performance Evaluation

In Chapter 3 and Chapter 4, we have developed the proposed feature extraction algorithm — SSWCC, and reviewed the minimum-distance and the kNN classifiers for pattern classification. This chapter first briefly explains the training set and the test set used, and then applies SSWCC to extract the features. The minimum-distance and the kNN classifiers are used to do the classification. The performance of the feature extraction algorithm will be evaluated from several aspects.

5.1 SensIT Experiment and SITEX00

SensIT stands for Sensor Information Technology, it is a program developed by DARPA. SITEX00 is the nickname of a data set collected from a field experiment carried out in August 2000. The goal of SITEX00 is to collect the time series data from UGS systems which include acoustic, seismic and Infrared (IR) sensors to support the laboratory experiments in ground vehicle identification and tracking. These data are generated from ground targets of interest for the battlefield surveillance problem [5].

5.1.1 SensIT Multi-Node System

SensIT multi-nodes system include thirty-seven testing nodes. These nodes are organized into three groups, which are named as A, B and C group respectively. Each node collects the sensor data, processes the data and transfers the data back to the user via a tactical RF network. The auxiliary Ethernet link provides the high-reliable collection of time series and detection event data, remote node control and logging of diagnostic information [5].

Most of these 37 sensors include three sensors (acoustic, seismic and IR) located within 1m of the node. Seismic sensors used in the experiment are geophones on channel #1. They are installed vertically to collect seismic signals propagated in the vertical direction. Acoustic sensors are omni-directional microphones assigned to channel #2. IR sensors are on channel #3. All these sensors are working simultaneously at a sample rate of 256Hz/channel. The frequency response and the detailed technical description of the geophone is shown in Fig. 5.1 and Fig. 5.2 respectively.

Figure 5.3 and Fig. 5.4 illustrate the distribution of the sensor nodes in a dense field. There are 25 nodes in group A, scattered at the intersection of the roads; 5 nodes in group B, clustered at the north checkpoint; and 7 nodes in group C, placed along the west side of the road from group A to group B.

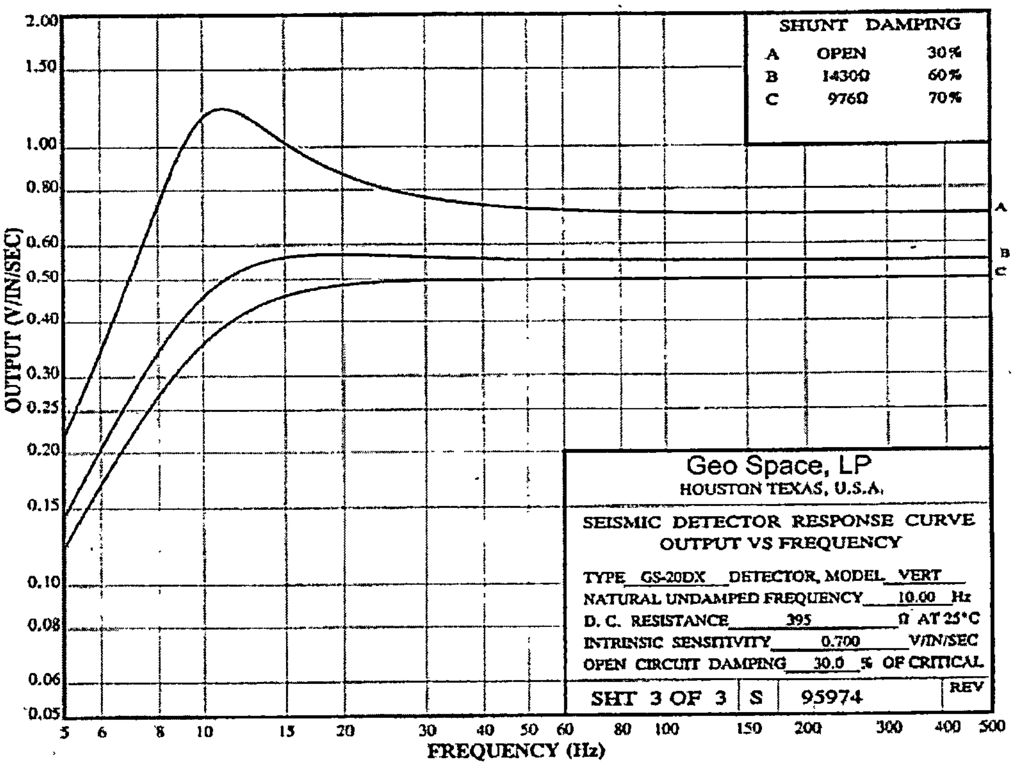


Figure 5.1: The Frequency Response of Geophone GS-20DX [3]

Geo Space GS-20DX Specifications			
(At 25° C)			
Natural Frequency (Fn)	8 Hz	10 Hz	14 Hz
Frequency Tolerance	± 0.5 Hz	± 5 %	± 5 %
Maintains Fn Specifications to Tilt Angle of	15°	20°	20°
Typical Spurious Frequency	>200 Hz	>250 Hz	>300 Hz
Harmonic Distortion with Driving Velocity of 0.7 in/sec (1.8 cm/sec)P-P	<.2%	<.2%	<.2%
Distortion Measured at	12 Hz	12 Hz	14 Hz
Open Circuit Damping (Bo) (for all coil resistances)	0.38 ± 10%	0.30 ± 10%	0.22 ± 10%
Standard Coil Resistance (Rc) ± 5%	———395 Ohms———		
Intrinsic Voltage Sensitivity ± 10%			
V/in/sec	———.70———		
V/cm/sec	———.28———		
Optional Coil Resistances (Rc) ± 5%	———155 Ohms / 630 Ohms———		
Intrinsic Voltage Sensitivity ± 10%			
V/in/sec	———.44 / .88———		
V/cm/sec	———.173 / .346———		
Damping Constant at Fn (Rt Bc Fn)	5500 for 395 Ohms, 2170 for 155 Ohms, 8760 for 630 Ohms		
Normalized Transduction Constant	.035 (sq.root Rc)V/in/sec (.0138 (sq.root Rc)V/cm/sec))		
Moving Mass ± .5g	11g	11g	11g
Case to Coil Motion P-P			
Min	——.8 mm (0.32 in)——		
Max	——1.5 mm (0.60 in)——		
Operating Temperature Range (°C)	——45° to +100° ——		
Dimensions (less terminals*)			
Height	——3.30 cm (1.30 in) ——		
Diameter	——2.54 cm (1.00) ——		
Weight	——87.3 g (3.08 oz.) ——		
*terminal height is .16 inches			

Figure 5.2: The Detailed Technical Data of Geophone GS-20DX [3]

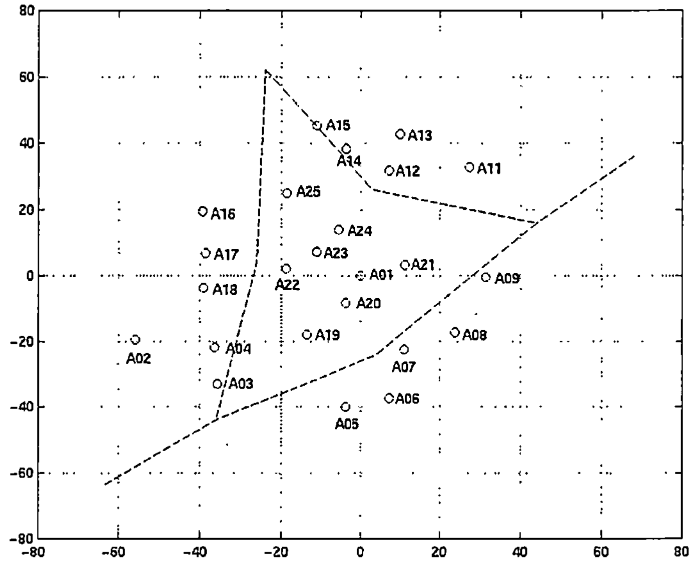


Figure 5.3: Detailed Location of Sensor Group A

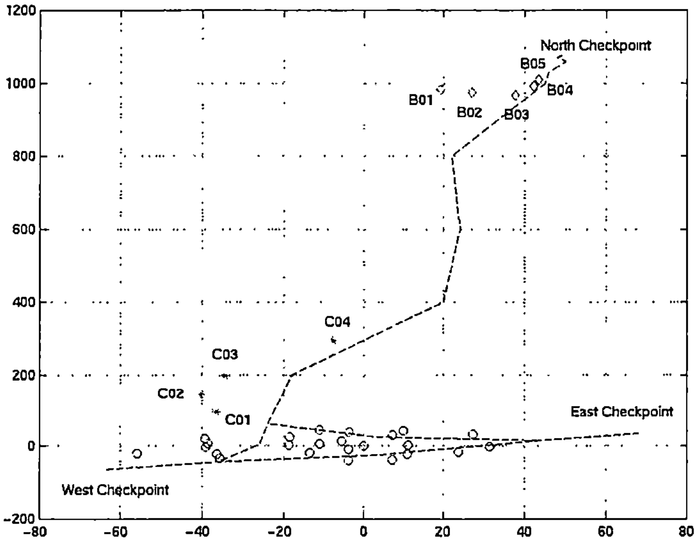


Figure 5.4: Detailed Location of Sensor Group B and C

5.1.2 SITEX00 Data Set

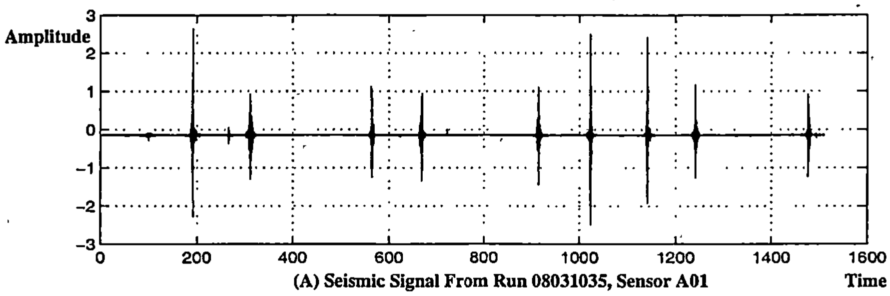
As described in the previous section, the experimental data are collected from the sensors grouped in A, B and C. A portable wideband data recording system based on eight channel digitizing system stores data directly to a file on the host computer disk. An appendix module provides anti-alias filtering, gain, and simultaneous sampling for all the channels. The data are stored in binary multiplexed format to save space and each data sample represents a 16-bit word. Vehicle operations in the experiments are monitored and recorded by a Global Positioning System (GPS) receiver [5].

The experiments are repeated about 32 times, with different vehicles combination and different operating time interval. Each experiment is named as one "Run". Every run has a complete raw data set from the sensors and an event log. The event log records the operation of the vehicles and a local time stamp, providing the information of the vehicle traffic during a data run. A sample event log for Run "08020830" is shown in Table 5.1. In this run, two vehicles are operated, one is "Dragon Wagon", the other is "HUMMVV". The event record begins from 8:32:15, and ends at 8:56:53, with every operation of the vehicles recorded.

Every node has one data set to record the time series data from the three sensors. A typically received signal from the acoustic sensor and the seismic sensor at one node is shown in Fig. 5.5. compared with the time index to the run eventlog, it is obvious that when the vehicle is moving, the peaks in the seismic data are significant. In the following section, the time series data from channel #1, which is the seismic data, is analyzed.

Table 5.1: A Sample Event Log

Time	Event
8/3/00 10:44:01	COMEX
8/3/00 10:44:21	Synchronized Shot
8/3/00 10:48:03	AAV arriving East Checkpoint
8/3/00 10:48:13	Helicopter CH53 to the west
8/3/00 10:48:25	AAV leaving East Checkpoint
8/3/00 10:51:27	AAV arriving North Checkpoint
8/3/00 10:51:50	AAV leaving North Checkpoint
8/3/00 10:54:03	AAV arriving West Checkpoint
8/3/00 10:54:31	AAV leaving West Checkpoint
8/3/00 10:57:18	AAV arriving North Checkpoint
8/3/00 10:57:39	AAV leaving North Checkpoint
8/3/00 11:00:03	AAV arriving East Checkpoint
8/3/00 11:00:22	AAV leaving East Checkpoint
8/3/00 11:01:42	AAV arriving West Checkpoint
8/3/00 11:02:26	AAV leaving West Checkpoint
8/3/00 11:03:43	AAV arriving East Checkpoint
8/3/00 11:04:01	AAV leaving East Checkpoint
8/3/00 11:06:47	AAV arriving North Checkpoint
8/3/00 11:07:06	AAV leaving North Checkpoint
8/3/00 11:09:14	AAV arriving West Checkpoint



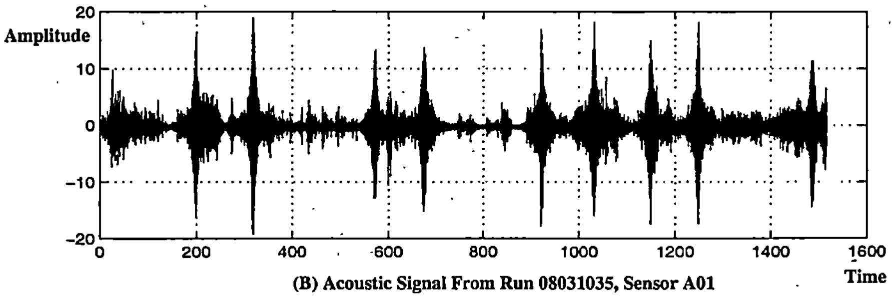


Figure 5.5: Seismic Signals and Acoustic Signals

5.2 Training Set and Test Set

Since we use the supervised learning algorithms, a training set and a test set need to be generated for the classification. The training set refers to a set used to derive the classifier, that is, to determine the parameters of the classifier. The classifier is then used to classify the data in the test set, and to make the final decision to each sample in the test set.

As described in Sec. 5.1, the raw data come from the seismic sensors, and a typical seismic signal is shown in Fig. 5.5 (A). The obvious peaks in the figure represent the vehicle-generated seismic signals received by geophone. Since we have a series "RUN"s from the SITEX00 data set, each "RUN" represents different combinations of vehicle operations, we choose some typical "RUN"s as the source of the training set, and use others as the test set. In this thesis work, we focus on single target

Table 5.2: The Selection of the Training Set and the Test Set

Set Type	Class	Run Name	Potential Sensors
Training Set	"DW"	08020830	A01, A03, A04, A07, A08, A09, A11, A24, A25, B04, B05, C01
Training Set	"AAV"	08031035	A01, A03, A04, A07, A08, A11, B02, B05, C03
Training Set	"No vehicle"	08031035 & 08020830	A01, A03, A04, A07, A08, A11, A08, A09
Test Set	"DW" & "No Vehicle"	08040645	A01, A03, A04, A07, A08, A09, A11, A23, A24, A25, B03,B04,C01
Test Set	"AAV" & "No Vehicle"	08021030	A01, A03, A04, A07, A08, A11, B02, B05, C03

detection and classification. We only take some "RUN"s with one vehicle operation in the experiment. We select two types of vehicles, "DW" and "AAV" to test the feature extraction algorithm we developed. "DW" is the acronym for "Dragon Wagon". Both "DW" and "AAV" are military vehicles. Thus, our classifier needs to distinguish three classes: "DW", "AAV", and "No Vehicle". The data source of training set and the test set are shown in Table 5.2.

Among the selected sensor nodes above, only the first channel of each sensor node, which is sampled at 256Hz, is used to extract the training and the test set. To simulate the real time requirement of the identification system, we use a window in size of 2 seconds to analyze the data. That is, we use 512 points selected from the raw data to begin our analysis.

5.3 Effect of Principal Component Analysis (PCA)

We apply SSWCC to the training set. The performance of some of the features from the training set before the PCA is shown in Fig. 5.6. Figure 5.6 (A) illustrates the distributions of three features — mean, standard deviation, and skewness of the spectrum without using the Blackman window. Figure 5.6 (B) illustrates another three features — mean, standard deviation, and skewness of the Blackman windowed spectrum. Figure 5.6 (C) shows the shape features from the PSD analysis — mean, standard deviation, and skewness of the PSD. Figure. 5.6 (D) draws the energy of the wavelet coefficients from level 1 to level 3. Each point in the figure corresponds to a certain sampled data series from the training set. We can see from Fig. 5.6 that the "No Vehicle" class is very well separated from the "Vehicle" classes. However, the "DW" class and the "AAV" class are mixed together and it is difficult to draw a clear decision boundary to separate these two classes.

After applying the PCA, we evaluate feature performance again. Some of the results are shown in Fig. 5.7. Figure 5.7 illustrates the distribution of three feature elements projected onto the three eigenvectors, which correspond to the three largest eigenvalues. We can see from the figure that PCA does a very good job in generating the most significant features based on the largest eigenvalues. It is much easier to draw a clear boundary among these three classes than it is from the feature space before applying the PCA.

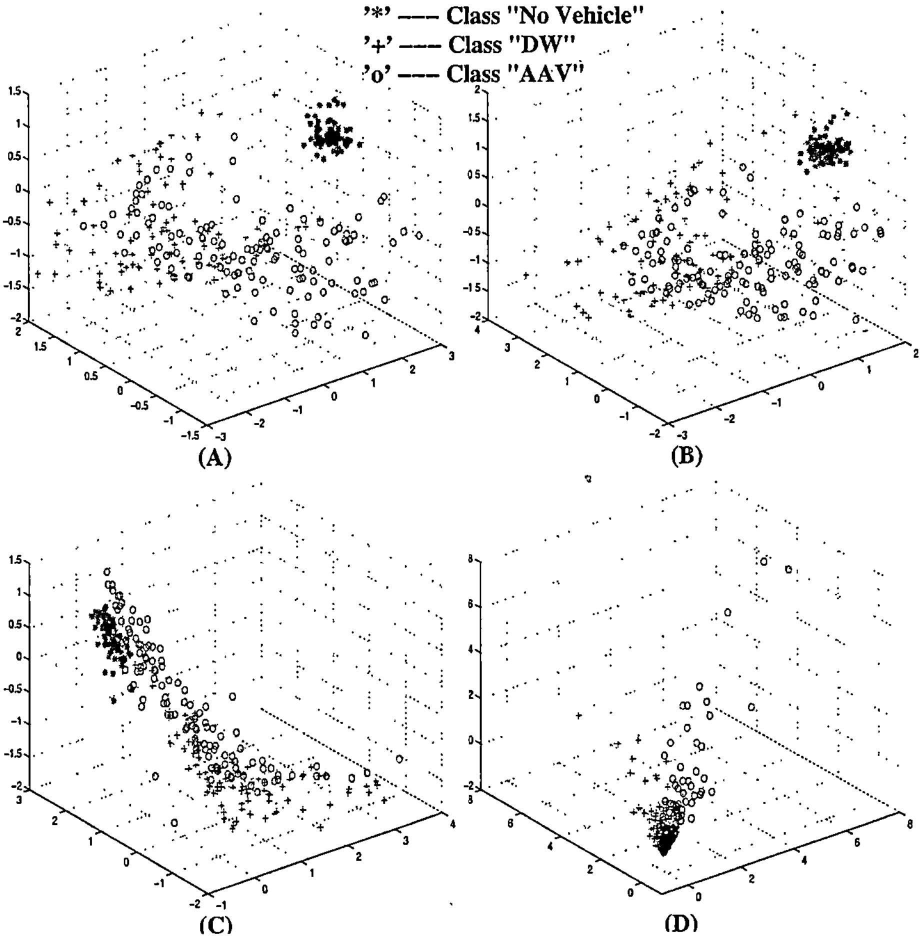


Figure 5.6: Selected Feature Elements Before Using PCA

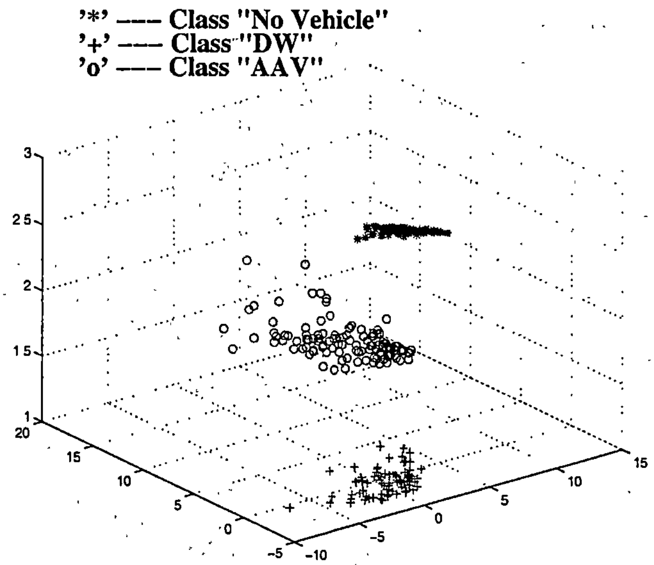


Figure 5.7: The Three Most Significant Feature Elements After Using PCA

5.4 Effect of Parameter k in kNN and Energy Loss η

As we have introduced before, the training set and the test set are selected from different "RUN"s. We have built a training set including 320 samples and a test set including 500 samples. These samples belong to 3 classes: "DW", "AAV" and "No Vehicle". We use *Performance Level* (PL) to measure the quality of the classification. PL is defined in Eq. 5.1.

$$PL = \frac{\text{Number of Correct Classification}}{\text{Number of Total Observation}} \tag{5.1}$$

Since the energy loss, η, selected in principal component analysis determines the number of feature vectors used in the classifier, and the number of the nearest neighbors included in the classification, k, also affects the classifier performance, we design two experiments to illustrate the effects of η and k on the performance level.

1. Fix η, and change k to evaluate the classifier;

2. Fix k at the maximum performance level and change η to evaluate the classifier.

Figure 5.8 shows the performance level with a changing k but a fixed η. We can see that in most cases, the performance is above 75% when k varies from 1 to 180. The highest performance level can reach 87% when k is within the interval [130, 170].

Next, we fix k from 121 to 181, and change η from 0 to 0.5, the performance is shown in Fig. 5.9. The results show the maximum peak always appears when the energy loss η is around 0.2 — 0.3, and the maximum PL can reach 87%.

From the above experiments, we can see that when k changes from 120 to 180, the performance level is stable, in most cases, around 85%. On the other hand, when the energy loss η is set between 0.2 and 0.3, the classifier has the best performance, and the performance level can reach 87%.

5.5 Effect of Blackman Window

In Chapter 3, we obtain 11 features from the spectrum and another 11 features from the Blackman windowed spectrum. In fact, this will bring some correlation between the features, and will defect the performance of the classifier. To see the effect of the Blackman window on the performance of the classifier, we eliminate the 11 features from the spectrum and only keep the features we obtained from the Blackman windowed spectrum. That is, we use $41 - 11 = 30$ features to do another

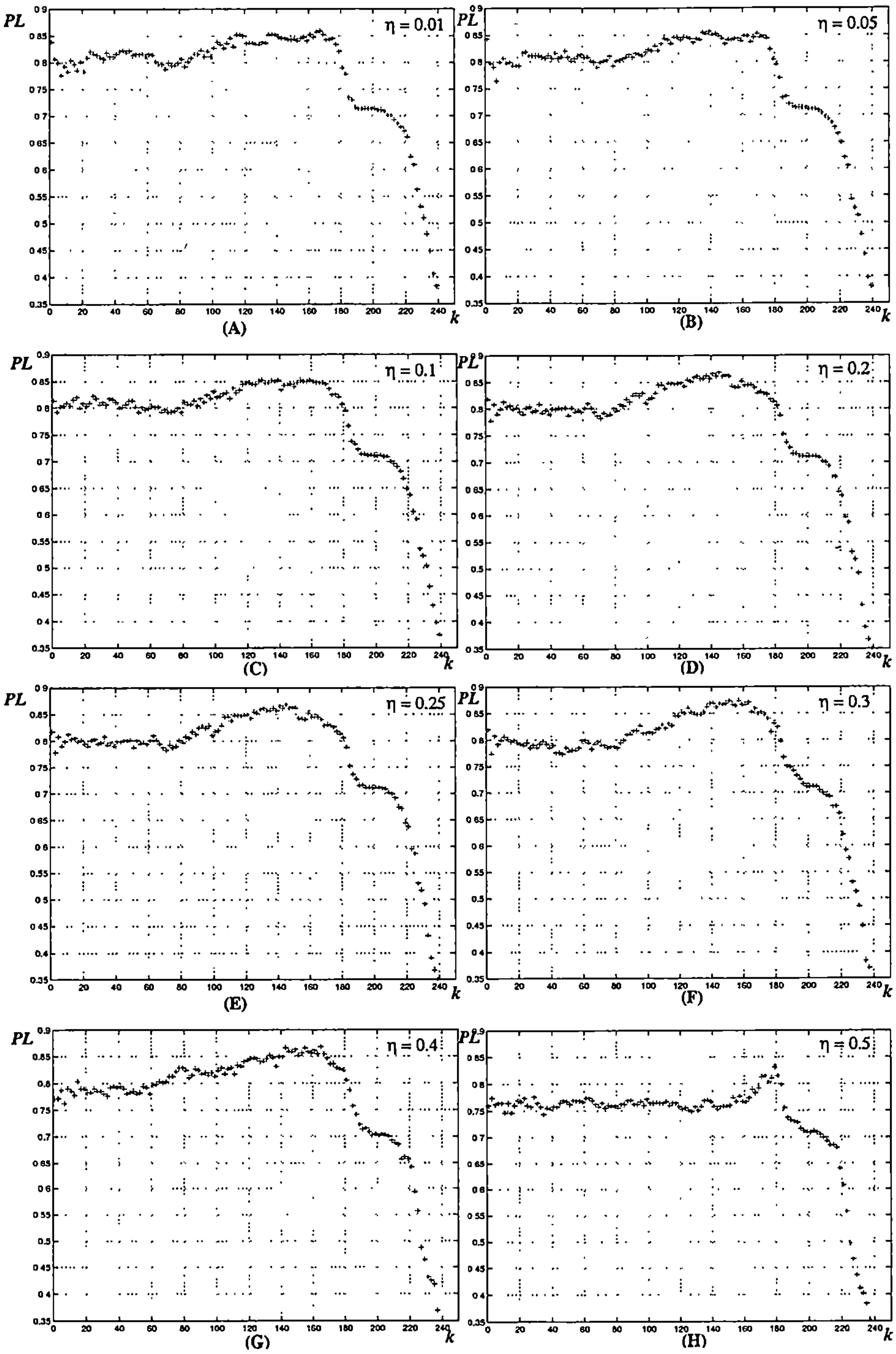


Figure 5.8: Performance Level of kNN at Different η, where k Ranges from 0 to 240

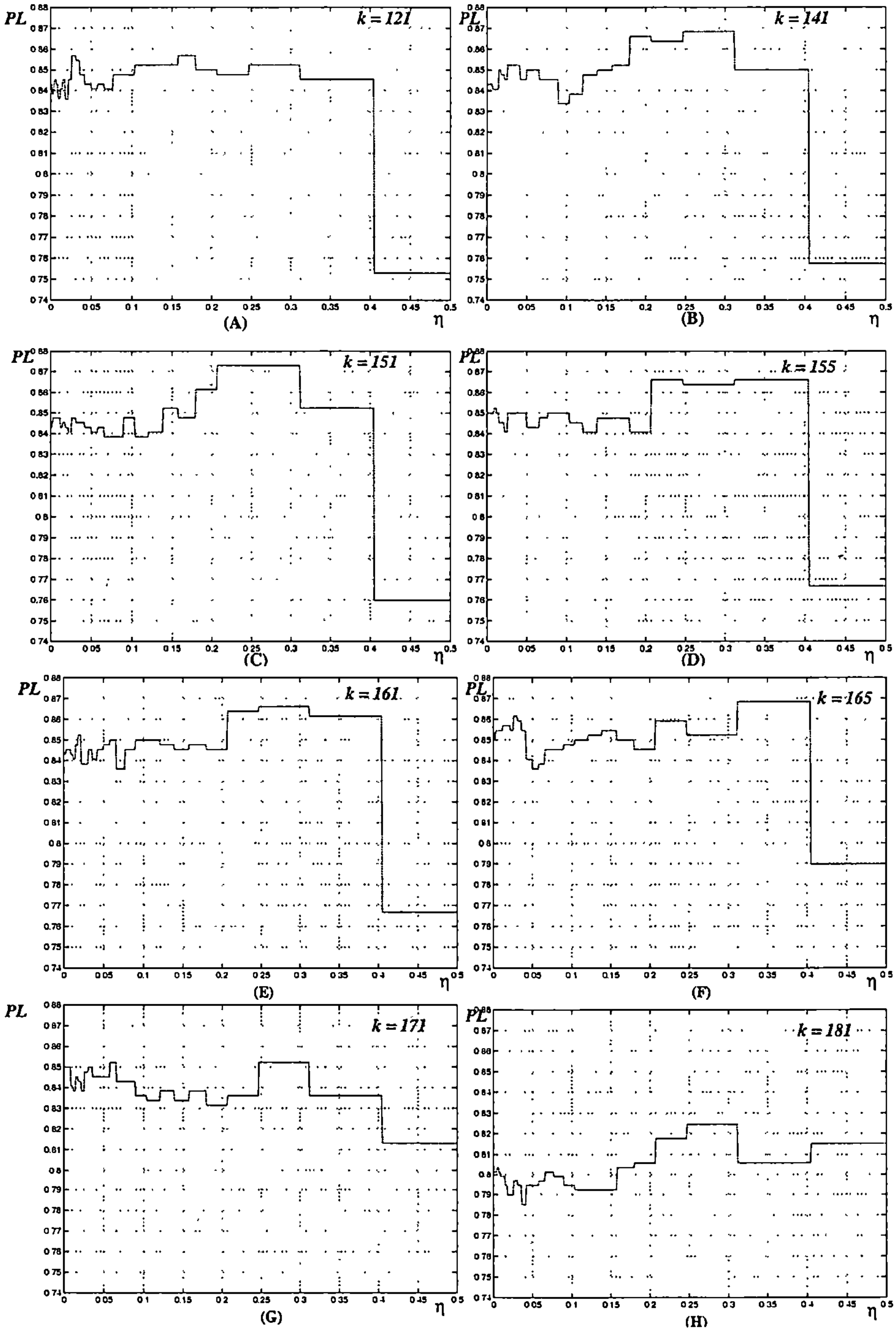


Figure 5.9: Performance Level of kNN with Different Value of k, Where the Energy Loss η Ranges from 0 to 0.5

set of performance evaluation.

The procedure of the evaluation is the same as what we did in Sec. 5.4. First of all, we fix the energy loss η, and then observe the effect of k on the performance of classifier. Figure 5.10 shows the performance level using different η. Compared with Fig. 5.8 and Fig. 5.9, the overall performance is increased by about 5%. In most cases, the performance can reach 85%.

Then we fix k from 131 to 171, and change η from 0 to 0.5, the performance is shown in Fig. 5.11. We can see that the best performance can still reach 87%, and in most cases, above 85%. The maximum peak always appears when the energy loss η is at the range of 0.2 — 0.4.

5.6 Effect of Selection of Training Set and Test Set

No matter how good the feature extraction algorithm is or how well the classification algorithm performs, if the training set is not representatively chosen, the performance level can still be very low. In the previous experiments, we select different "RUN"s to generate the training set and the test set. Since different "RUN"s might have different environmental conditions, classifiers derived from one "RUN" might not behave well when applied to the test set derived from another "RUN". In this experiment, we try to change the content of the training set and the test set in order to test the algorithms on a more reasonable basis. Our approach is to combine the training set and the test set to form a new set which contains about 800 samples. Then we separate this set into two parts randomly and treat one set

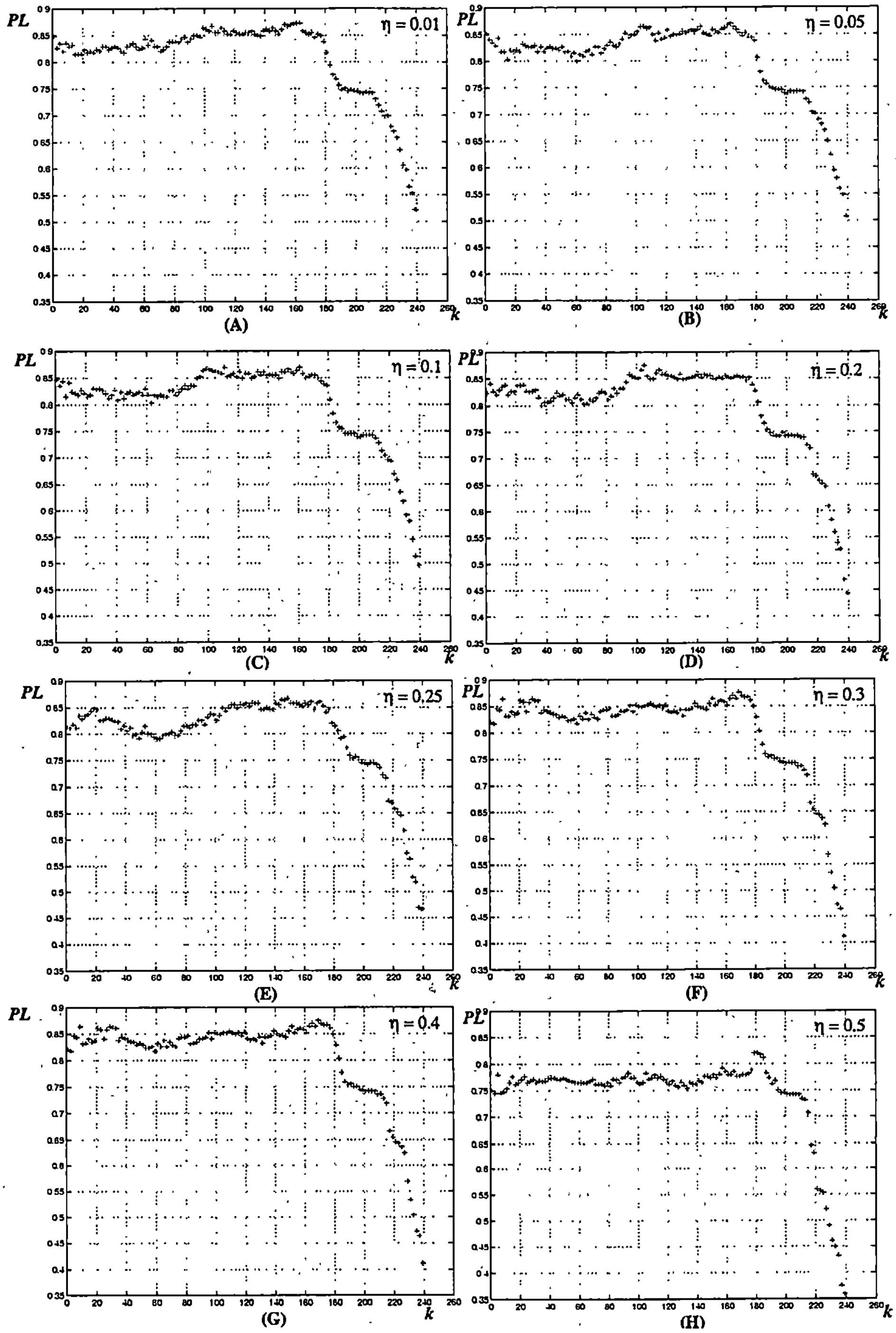


Figure 5.10: Performance Level of kNN When η is Fixed, and the Feature Elements Number is 30

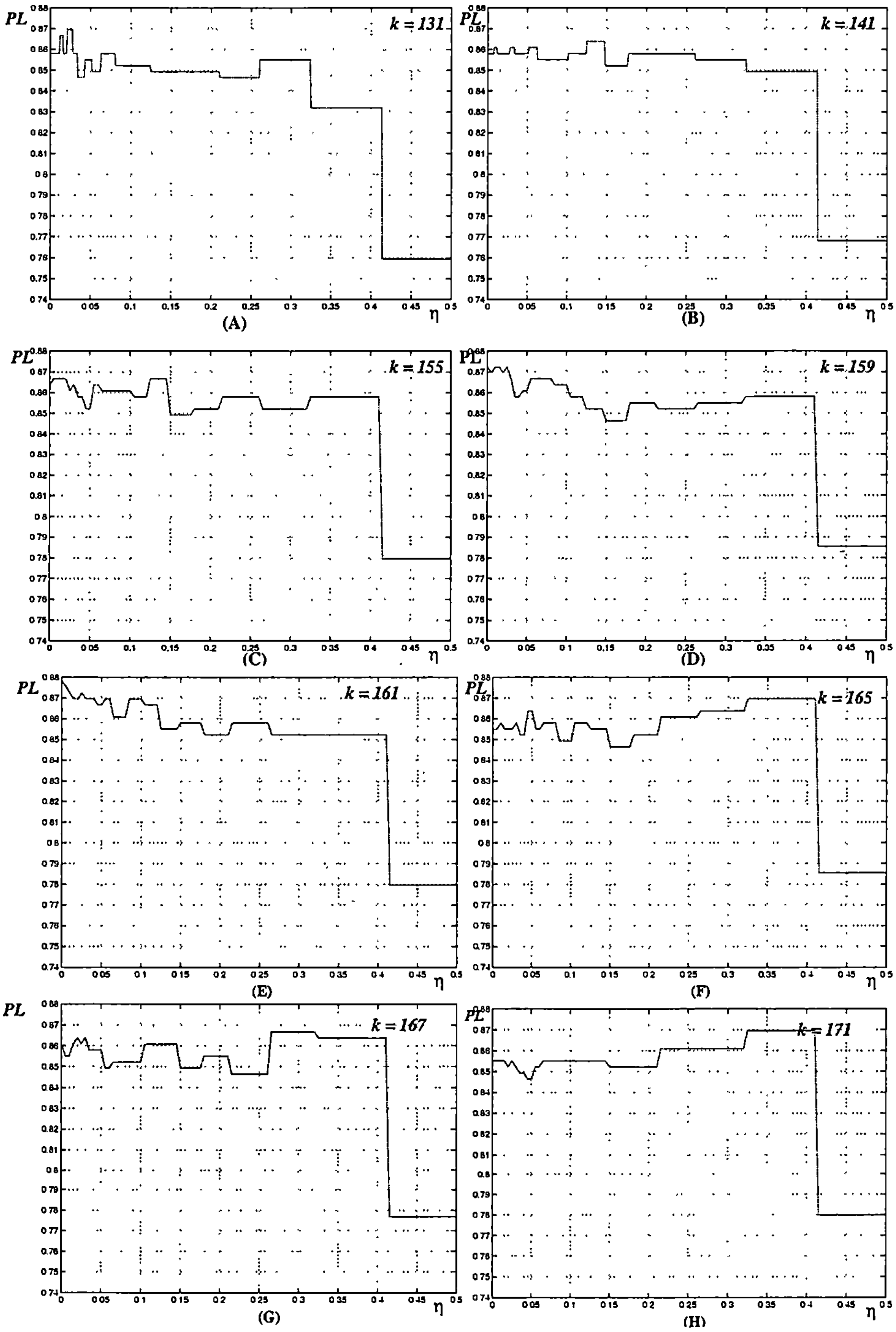


Figure 5.11: Performance Level of kNN Classifier When k is Fixed, and the Feature Elements Number is 30

as the training set and the other one as the test set.

We apply the kNN to the newly formed training set and the test set. The performance level is shown in Fig. 5.12. For each energy loss η, we randomly divide the data set into the training set and the test set for 450 times. For the first 50 times, we choose $k = 5$, for the next 50 times, k is increased by 20. The performance level profile in Fig. 5.12 is obviously better than the performance level we obtained from the last two experiments, which shows the importance of the generation of the training set. We can achieve a better performance in this experiment. The highest performance level can reach 97%. When η is from 0.15 to 0.3, the performance level is stable and more insensitive to the selection of k.

5.7 Performance of Different Classifiers

In this experiment, the minimum-distance classifier is used to classify the feature vector. Similar to the previous experiments, we evaluate the classifier by two tests. We use 30 feature vectors for these experiments, that is, we eliminate the 11 features derived from the rectangular windowed spectrum. In the first test, we use the original training set and the test set to do the classification. By changing the energy loss rate η, we obtain different performance level shown in Fig. 5.13. In the second test, we combine the test set and the training set, select the training set and test set randomly and repeat the experiment for 300 times. The result is shown in Fig. 5.14.

From these experiments, we can see that the performance level of the minimum-distance classifier can only reach around 70% when the training set is fixed, and

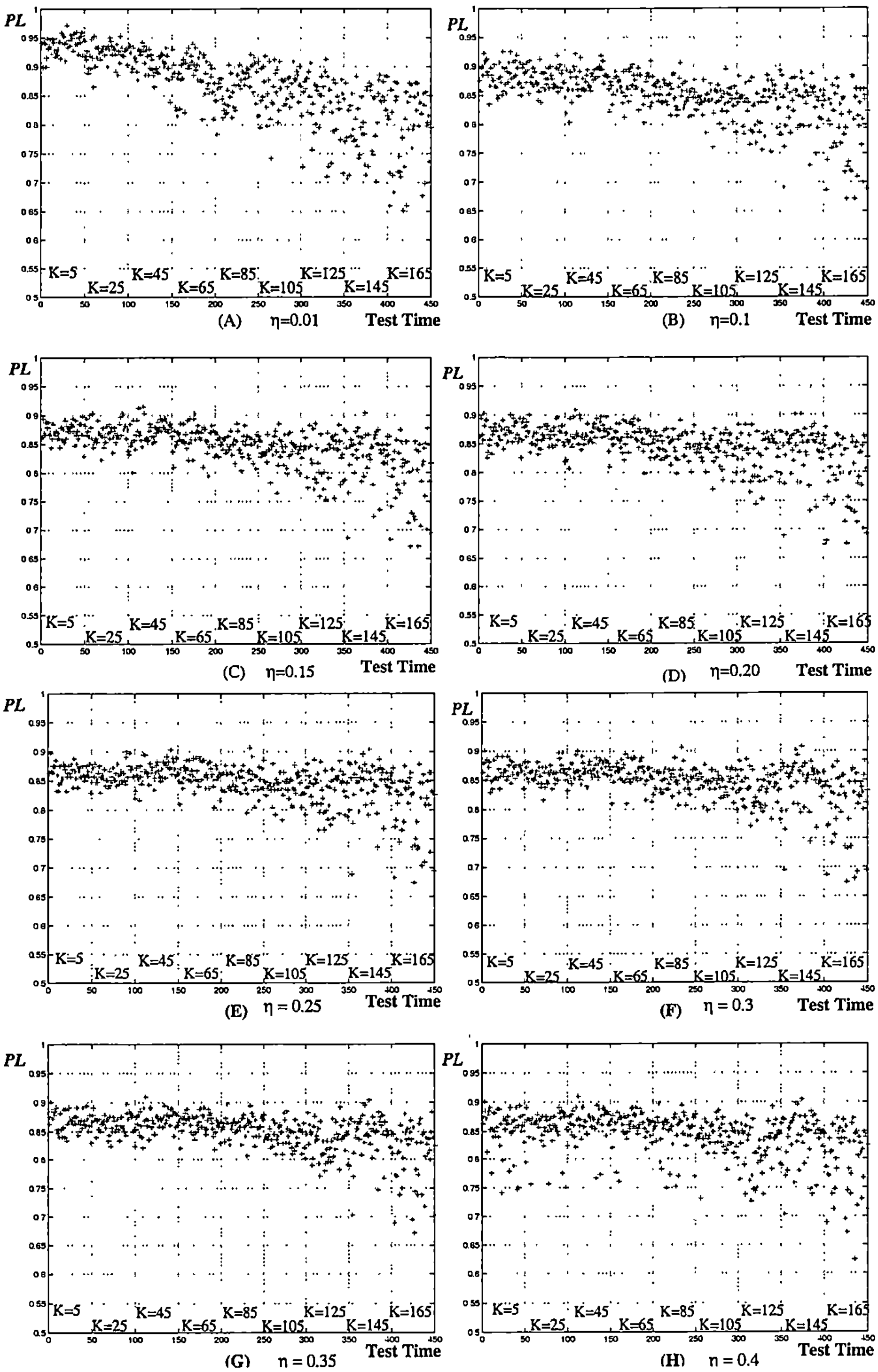


Figure 5.12: The Performance of the kNN Classifier When the Training Set and Test Set Combined

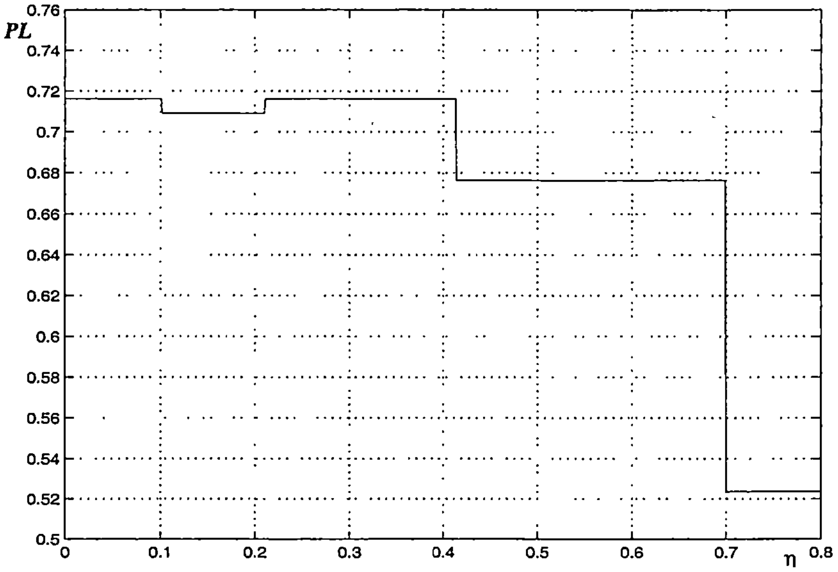


Figure 5.13: The Performance of the Minimum-Distance Classifier vs. Energy Loss η

around 82% when the training set is randomly chosen. The minimum-distance classifier always performs worse than the kNN, whihc is due to the assumption the minimum-distance classifier made about the pdf of the sample variables.

5.8 Summary

In this chapter, we evaluate both the performance of feature vector derived from SSWCC and the different classifiers as well. We showed the effect of the PCA, the Blackman windowed spectrum, the different energy loss rate η, the different k in the kNN classifier, and the difference between the kNN and the minimum-distance classifier.

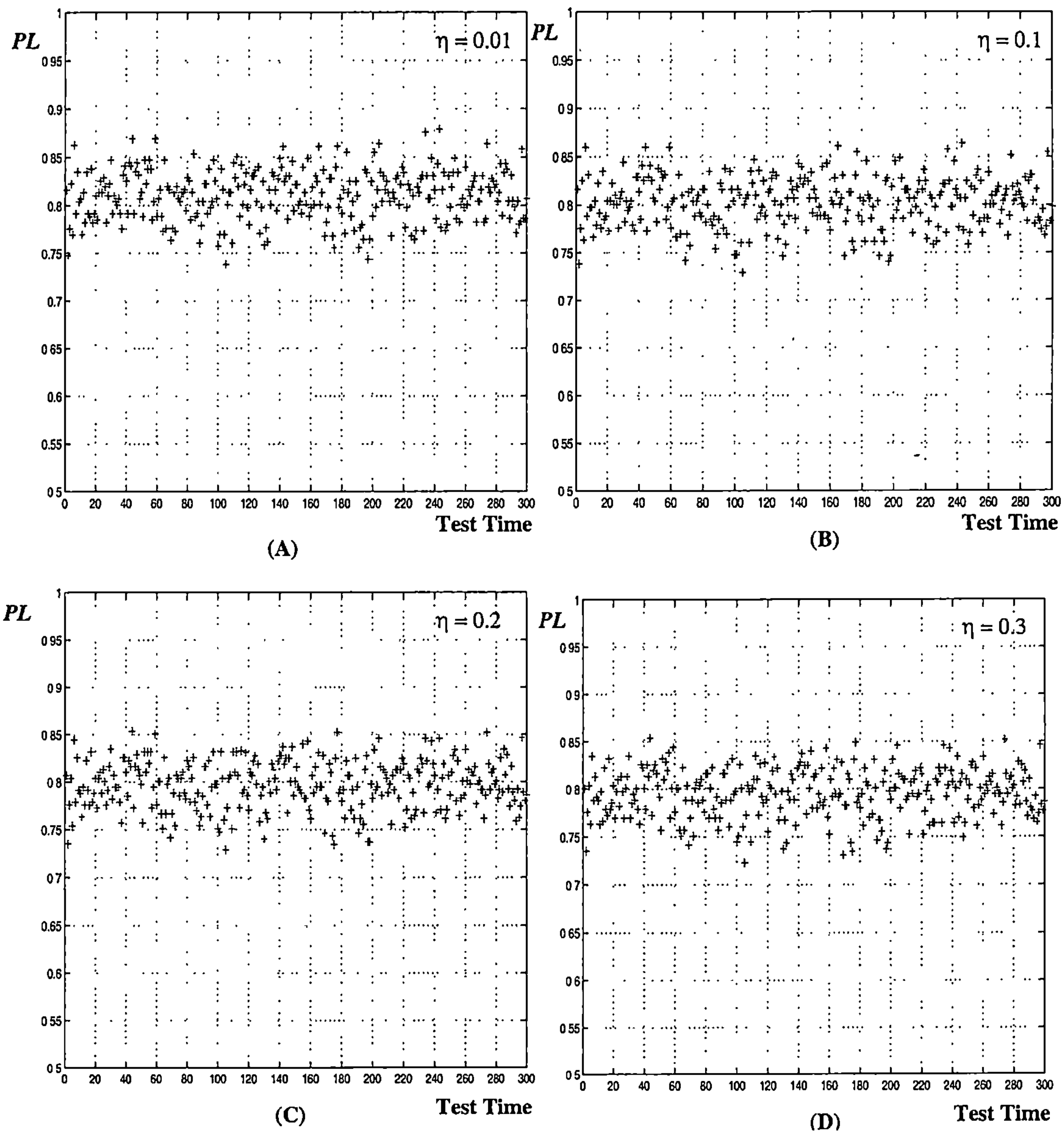


Figure 5.14: The Performance of the Minimum-Distance Classifier When the Training Set is Randomly Chosen

Chapter 6

Conclusions and Future Work

Target detection and classification has been widely used in military vehicle identification systems, where the most challenging problem is the selection of a robust feature vector which can represent a specific type of vehicle. Unattended Ground Sensor (UGS) systems have been used to detect and classify a variety of vehicles. Among the signals collected from UGS systems, acoustic and seismic signals are the most popularly used ones to identify the ground vehicles. This thesis concentrates on the study of the vehicle detection and classification problems using seismic signals. Based on recent research results, a new feature extraction algorithm, Spectral Statistics and Wavelet Coefficients Characterization (SSWCC), is proposed. This algorithm obtains shape statistic features from Fourier spectrum, PSD as well as the energy and distribution of the wavelet coefficients. The minimum-distance classifier and kNN classifier are used to do the classification. The performance level of the classification could reach as high as 90%. The following conclusions can be drawn from the experiments:

1. Principal component analysis (PCA) plays an important role in feature extraction. It reduces the dimension of the potential feature components and keeps the most significant ones as well.

2. The performance analysis shows that kNN performs better than a minimum-distance classifier in this problem. This is because the minimum-distance classifier assumes the distribution of the feature vectors to be Gaussian, which is not always true, especially when the sample set is not big enough.

3. The use of the Blackman window function on the raw signal will smooth the spectrum as well as enhance the classifier performance. The experiments show the overall performance could improve about 5%.

4. The energy loss rate η can also affect the performance level of the classification. In the kNN classifier, the best performance appears when η is around [0.2, 0.3]. In this range, the classifier is stable and insensitive to the selection of k.

5. The performance of kNN is also affected by the selection of k. When the training set and the test set are fixed, the classifier performs more stable and could reach the best performance level when k is between 120 and 180. When the training set and the test set are selected randomly, the best performance of the classifier will appear when k is between 5 and 145. In this case, the performance is not so sensitive to the selection of k.

6. The selection of the training set and the test set will affect the performance level as well. When choosing the training set and the test set representatively, the overall performance of the classifier could increase about 5%. And the overall performance in this case can reach as high as 90%.

Future work is still needed in following aspects:

1. Additional analysis of seismic signals is still needed. In this thesis work, the original signal is collected from one channel, which only records the vertical vibration of the object. That is, a one-axis geophone is used. However, seismic vibrations also exist in other directions, like the horizontal direction. A three-axis geophone could be used to capture these vibrations. Additional analysis of these waves will also be helpful in solving the ground vehicle localization problem.

2. The estimation of speed and location of the target by seismic signals. As we know, seismic waves are very complex waveforms, the speed and the direction of the waves varies in a significant range. By the use of a 3-axis geophone and multiple sensor readouts from a distributed sensor network, target speed and location can be estimated.

3. More targets need to be classified. In this thesis work, we only applied SSWCC to extract features to classify two targets: "DW" and "AAV". More targets need to be analyzed to further show the efficiency of SSWCC.

4. The recognition of multiple targets in real time. Multi-target classification and tracking has always been a challenging problem. This work needs to separate the signals generated by different vehicles. It also requires a more robust feature vector and more advanced classifier.

Bibliography

Bibliography

[1] `http://www.fas.org/`.

[2] `http://www.geo.mtu.edu/`.

[3] `http://www.geospacelp.com/`.

[4] `http://www.mines.edu/`.

[5] SensIT August'00 Experiment Plan. Technical report, BBNT, July 2000.

[6] D. G. Albert. The effect of snow on vehicle-generated seismic signatures. *Journal of the Acoustical Society of America*, 81(4):881–887, 1987.

[7] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.

[8] F. Camastra and A. Vinciarelli. Cursive character recognition by learning vector quantization. *Pattern Recognition Letters*, 22(6-7):625–629, May 2001.

[9] B. H. Chen, X. Z. Wang, S. H. Yang, and C. McGreavy. Application of wavelets and neural networks to diagnostic system. *Computers & Chemical Engineering*, 23:899–906, 1999.

[10] P. C. Ching, H. C. So, and S. Q. Wu. On wavelet denoising and its application to time delay estimation. *IEEE Transaction on Signal Processing*, 47(10):2879–2882, December 1999.

[11] C. Chou, B. Ho, and J. T. Sheu. Material characterization by ultrasonics using unsupervised competitive learning. *Pattern Recognition Letter*, 16(7):769–777, July 1995.

[12] J. Claassen, M. Ladd, and G. Elbring. The feasibility of monitoring continuous wave sources with seismic arrays. *Proceedings of SPIE*, 3713:22–32, 1999.

[13] R. R. Coifman and D. L. Donoho. Translation-invariant denoising. Technical report, Yale University and Stanford University, 1995.

[14] I. Daubechies. *Ten Lectures on Wavelets.* The Society of Industrial and Applied Mathematics, 1992.

[15] Z. A. Der and D. R. Baumgardt. Applications of higher-order spectral analyses to detection and identification of seismic and acoustic signals generated by machinery. In G. Yonas, editor, *Proceedings of SPIE: Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors*, volume 3081, pages 99–109. SPIE, July 1997.

[16] D. L. Donoho and I. M. Johnstone. Ideal denoising in an orthonormal basis chosen from a library of bases. Technical report, Department of Statistics, Stanford University, 1994.

[17] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[18] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, December 1995.

[19] H. Doyle. *Seismology*. Wiley, 1995.

[20] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1 edition, 1973.

[21] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2nd edition, 2001.

[22] I. A. Erteza. Case study: Unattended ground sensor phenomenology and signal processing. In G. Yonas, editor, *Proceedings of SPIE: Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors*, volume 3081, pages 88–98. SPIE, July 1997.

[23] R. A. Gramann, M. B. Bennett, and T. D. O'Brien. Vehicle and personnel detection using seismic sensors. In Carapezza, Edward M., and D. B. Law, editors, *Proceedings of SPIE: Sensors, C3I, Information, and Training Technologies for Law Enforcement*, volume 3577, pages 74–85. SPIE, January 1999.

[24] A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2), Summer 1995.

[25] L. B. Jackson. *Digital Filters and Signal Processing*. Kluwer Academic Publishers, 3 edition, 1995.

[26] M. Jansen and D. Roose. Bayesian correction of wavelet threshold procedures for image denoising. In *Joint Statistical Meetings*, Anaheim, CA, USA, August 1997.

[27] I. Johnstone and B. M. Silverman. Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society B*, 59:319–351, 1996.

[28] R. E. Karlsen, Grant R. Gerhert, and T. Meitzler. Wavelet-based ground vehicle recognition using acoustic signal. In *Proceedings of SPIE: Wavelet Applications III*, volume 2762, pages 434–445. SPIE, March 1996.

[29] D. H. Kil and F. B. Shin. *Pattern Recognition and Prediction with Application to Signal Characterization.* American Institute of Physics, 1996.

[30] T. Kohonen. The self-organizing map. *Neurocomputing*, 21:1–6, November 1998.

[31] A. P. Kozaitis and M. A. Getbehead. Optical wavelet feature extraction using a multiple-input phase-only encoded joint-transform correlator. *Optics Communications*, 151:15–20, May 1998.

[32] T. Lay and T. C. Wallace. *Modern Global Seismology.* Academic Press, 1995.

[33] J. LeMond. Vehicle weight estimates using a buried three-axis seismometer. In Carapezza, Edward M., and D. B. Law, editors, *Proceedings of SPIE: Sensors, C3I, Information, and Training Technologies for Law Enforcement*, volume 3577, pages 106–116. SPIE, January 1999.

[34] Y. Mallet, D. Coomans, J. Kautsky, and O. D. Vel. Classification using adaptive wavelets for feature extraction. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(10):1058–1066, October 1997.

[35] S. McLaughlin, A. Stogioglou, and J. Fackrell. Introducing higher order statistics for the detection of nonlinearities. *UK Nonlinear News*, (2), Sept. 1995.

[36] R. McQuillin, M. Bacon, and W. Barclay. *An Introduction to Seismic Interpretation.* Gulf Publishing Company, 1984.

[37] D. Middleton. A statistical theory of reverberation and similar first-order scattered fields - part i: Waveforms and the general process. *IEEE Transactions on Information Theory*, IT-13(3):372–392, July 1967.

[38] T. Ogden and M. Hilton. Data analytic wavelet threshold selection in 2d signal denoising. *IEEE Transactions in Signal Processing*, 45:496–500, 1996.

[39] J. M. Ortolf. Research supporting the unattended ground sensor mission. In G. Yonas, editor, *Proceedings of SPIE: Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors*, volume 3081, pages 42–49. SPIE, July 1997.

[40] S. Pittner. Feature extraction from wavelet coefficients of pattern recognition tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(1), January 1999.

[41] G. E. Sleefe, M. D. Ladd, T. S. McDonald, and G.J.Elbring. Acoustic and seismic modalities for unattended ground sensors. In *Proceedings of SPIE: Unattended Ground Sensor Technologies and Applications.*

[42] G. E. Sleefe, S. Peglow, and R. Hamrick. The application of unattended ground sensors to stationary targets. In G. Yonas, editor, *Proceedings of SPIE: Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors*, volume 3081, pages 21–29. SPIE, July 1997.

[43] S. A. Stotts, R. A. Gramann, and M. B. Bennett. Vehicle bearing determination from a tri-axial seismometer using rayleigh wave propagation. In Carapezza, Edward M., and D. B. Law, editors, *Proceedings of SPIE: Sensors, C3I, Information, and Training Technologies for Law Enforcement*, volume 3577, pages 168–179. SPIE, January 1999.

[44] G. Succi, G. Prado, R. Gampert, T. Pedersen, and H. Dhaliwal. Problems in seismic detection and tracking. In Carapezza, Edward M., and T. M. Hintz, editors, *Proceedings of SPIE: Unattended Ground Sensor Technologies and Applications II*, volume 4040, pages 165–173, July 2000.

[45] Y. Tao, E. CM. Lam, and Y. Y. Tang. Feature extraction using wavelet and fractal. *Pattern Recognition Letters*, 22:271–287, 2001.

[46] A. Udias. *Principles of Seismology.* Cambridge University Press, 1st edition, 1999.

[47] P. D. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transaction on Audio Electroacoustics*, AU-15:70–73, June 1967.

[48] M. C. Wellman, N. Srour, and D.B. Hillis. Feature extraction and fusion of acoustic and seismic sensors for target identification. In G. Yonas, editor, *Pro-*

ceedings of SPIE: Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors, volume 3081, pages 139–145. SPIE, July 1997.

[49] J. K. Wolford. Event identification from seismic/magnetic feature vectors-a comparative study. In G. Yonas, editor, *Proceedings of SPIE: Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors*, volume 3081, pages 131–138. SPIE, July 1997.

Vita

Tian, Yuxin was born in Shannxi, P. R. China in 1975. His major interests and research areas are signal processing and digital system design. He graduated from Tsinghua University, P. R. China in 1997 with a B. S. Degree in Electrical Engineering. During his five years' undergraduate study, he designed the Intelligent Electric Energy Consumption Management System (IEECM System), which has been granted with National Invention Patent and widely used in Northeastern of China. He continued his graduate study in Tsinghua University, P. R. China since 1997, where he obtained his M. S. degree in Electrical Engineering in January, 2000. During that period, he developed the Optically Coupled Current Measurement System, which has significantly reduced the cost of the traditional measuring devices in power systems. After graduation, he worked at Unisun Co., Ltd. Beijing Branch as a hardware engineer. In summer 2000, Yuxin came to The University of Tennessee, Knoxville, U.S.A. as a graduate student in Electrical & Computer Engineering Department. He worked in the Lab of Dr. Hairong Qi with a major research topic on Target Detection and Classification Using Seismic Signal Processing in Unattended Ground Sensor Systems. He proposed a new feature extraction algorithm named Spectral Statistics and Wavelet Coefficients Characterization (SSWCC), which has been successfully implemented in unattended ground sensor systems. Yuxin's publications include: "Continuous Signals and Systems with MATLAB – Solution Manual" (CRC Press, ISBN: 0-8493-1124-1) , "Study of Optical Current Transformer in Power System," published in *Automation of Electric Power Systems*, January, 2000.