THE ROBUST CLASSIFICATION OF HYPERSPECTRAL IMAGES USING ADAPTIVE WAVELET KERNEL SUPPORT VECTOR DATA DESCRIPTION

A Thesis

by

REVATHI SHARMA KOLLEGALA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2012

Major Subject: Electrical Engineering

THE ROBUST CLASSIFICATION OF HYPERSPECTRAL IMAGES USING ADAPTIVE WAVELET KERNEL SUPPORT VECTOR DATA DESCRIPTION

A Thesis

by

REVATHI SHARMA KOLLEGALA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Deepa Kundur
Committee Members,	Jim Ji
	Ulisses Braga-Neto
	Jyh. C. (Steve) Liu
Head of Department,	Costas Georghiades

May 2012

Major Subject: Electrical Engineering

ABSTRACT

The Robust Classification of Hyperspectral Images Using Adaptive Wavelet Kernel Support Vector Data Description. (May 2012)

Revathi Sharma Kollegala, B.S, National Institute of Technology Karnataka

Chair of Advisory Committee: Dr. Deepa Kundur

Detection of targets in hyperspectral images is a specific case of one-class classification. It is particularly relevant in the area of remote sensing and has received considerable interest in the past few years. The thesis proposes the use of wavelet functions as kernels with Support Vector Data Description for target detection in hyperspectral images. Specifically, it proposes the Adaptive Wavelet Kernel Support Vector Data Description (AWK-SVDD) that learns the optimal wavelet function to be used given the target signature. The performance and computational requirements of AWK-SVDD is compared with that of existing methods and other wavelet functions.

An introduction to target detection and target detection in the context of hyperspectral images is given. This thesis also includes an overview of the thesis and lists the contributions of the thesis. A brief mathematical background into one-class classification in reference to target detection is included. Also described are the existing methods and introduces essential concepts relevant to the proposed approach. The use of wavelet functions as kernels with Support Vector Data Description, the conditions for use of wavelet functions and the use of two functions in order to form the kernel are checked and analyzed. The proposed approach, AWKSVDD, is mathematically described. The details of the implementation and the results when applied to the Urban dataset of hyperspectral images with a random target signature are given. The results confirm the better performance of AWK-SVDD compared to conventional kernels, wavelet kernels and the two-function Morlet-Radial Basis Function kernel. The problems faced with convergence during the Support Vector Data Description optimization are discussed. The thesis concludes with the suggestions for future work.

DEDICATION

To my family and friends

ACKNOWLEDGMENTS

This thesis would not have been possible without the motivation given by Dr. Andrew Chan and the work done by his former student, Dr. Wesam Sakla. I am thankful to my advisor Dr. Deepa Kundur for guiding me and the rest of my committee including Dr. Jim Ji, Dr. Ulisses Braga-Neto and Dr. Steve Liu for taking their time out for the sake of this thesis. I am indebted to my family and friends for their support throughout my graduate education. I thank the department graduate officers for their enormous patience and help. I acknowledge the Texas A & M Supercomputing Facility for providing computing resources useful in conducting the research reported in this thesis. Last but not the least, a hearty thanks to anyone and everyone that I might have forgotten to mention, who helped me during my graduate education.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGMENTS vi

TABLE OF CONTENTS vii

LIST OF TABLES ix

LIST OF FIGURES x

1. INTRODUCTION 1

1.1 Overview 1
1.2 Target Detection Problem 1
1.2.1 Target Detection 1
1.2.2 Target Detection Of Hyper-spectral Images 2
1.3 The Contributions Of This Thesis 2
1.3.1 Organization Of The Thesis 2
1.3.2 The Proposed Approach: AWK-SVDD 2

2. A BRIEF MATHEMATICAL BACKGROUND 4

2.1 Hyper-Spectral Imagery 4
2.1.1 The Hyper-spectral Data Cube 4
2.1.2 Challenges In Hyper-Spectral Image Processing 5
2.2 A Brief Introduction to Wavelets 6
2.2.1 Mother Wavelet And The Concept Of Wavelet Decomposition 6
2.2.2 Common Wavelet Functions 9
2.3 Pattern Classification Using Support Vector Data Description 10
2.3.1 Support Vector Machines 12
2.3.2 The Significance Of Kernel Functions 15
2.3.3 Support Vector Data Description 17
2.4 The Need for Adaptive Wavelet Kernel SVDD (AWKSVDD) 19
2.4.1 The Advantages Of Support Vector Classification Methods . . 20
2.4.2 The Relevance Of One-class Classification 20
2.4.3 The Need For An Adaptive Wavelet Kernel 21

Page

3. ADAPTIVE WAVELET KERNEL - SUPPORT VECTOR DATA DESCRIPTION (AWK-SVDD) 23
3.1 SVDD With Wavelet Kernels 23
3.1.1 Admissibility Of Wavelet Kernels 23
3.1.2 Morlet Wavelet Kernel 24
3.1.3 Morlet-RBF Wavelet 25
3.2 Adaptive Wavelet Kernel - Support Vector Data Description 26
3.2.1 Multiple Kernel Learning 27
3.2.2 Adaptive Wavelet Kernel Learning for Support Vector Data Description 27
4. IMPLEMENTATION AND RESULTS 32
4.1 Data And Preparation 32
4.1.1 Hyper-spectral Image Data 32
4.1.2 Gaussian Model Of Target Detection 33
4.2 Comparison Of Performance With Previously Established Methods . 34
4.2.1 F-measure Value 34
4.2.2 Comparison of F-measure 35
4.2.3 Variation Of Performance With The Number Of Training Samples 35
4.3 Analysis Of Computational Time Requirements 36
4.4 Results And Discussion 36
4.4.1 Tabulated Results 37
4.4.2 Discussion And Short-comings 39
5. CONCLUSION 40
5.1 Advantages Of The Proposed Solution 40
5.2 Scope For Improvement And Future Work 40
REFERENCES 42

LIST OF TABLES

TABLE Page

2.1 Popular kernel functions 17

4.1 Average precision values 37

4.2 Average recall values 37

4.3 Average F-measure values 38

4.4 Average computational time 38

LIST OF FIGURES

FIGURE Page

2.1 The Gaussian mother wavelet function 10

2.2 Morlet wavelet function 11

3.1 The Morlet Mother Wavelet Function 25

4.1 Urban dataset hyperspectral image visualized at R = 420nm, G = 440nm and B = 2400nm 33

4.2 F-measure plotted as a function of the number of training points 35

4.3 Computational time taken by each method 36

1. INTRODUCTION

An overview of the thesis is included in Section 1.1. An introduction to the target detection problem is considered in Section 1.2 and Section 1.3 lists the contributions of this thesis.

1.1 Overview

Sensors capable of recording images over a wider spectral range were discovered only over the past decade. However, this technique of hyperspectral imagery has already been the focus of increasing interest with the potential for a wide range of applications. [1] In specific, hyperspectral imagery has proves to be a useful tool in the area of remote sensing to distinguish types of vegetation, building materials and similar spectrally similar materials that are otherwise indistinguishable using conventional methods of imagery. Many tools have been developed ([2], [3]) in order to work with hyperspectral images. Target detection is one of the main applications of hyperspectral images. This thesis proposes a more robust approach, Adaptive Wavelet Kernel Support Vector Data Description (AWKSVDD), for target detection, and compares it with conventional methods.

1.2 Target Detection Problem

1.2.1 Target Detection

Target detection, in general, involves distinguishing target material in available data. This is described as a classification problem, where each sample of data obtained needs to be classified as being either a target, or not. The rejected samples are termed outliers. Target detection has a wide range of applications in several

This thesis follows the style of *IEEE Transactions letters*.

areas including medical imaging, multimedia signal processing, remote sensing and telecommunications.

1.2.2 Target Detection Of Hyper-spectral Images

Target detection in the context of hyperspectral images refers to the spectral classification of each pixel as a target or an outlier. Example target spectral signatures are used in order to characterize the target. However, the presence of other materials in the same spatial region representing a pixel leads to spectral mixing. This is a challenge for accurate target detection. Several methods have been proposed in order to achieve better target detection. [4]. This thesis proposes a new method,Adaptive Wavelet Kernel Support Vector Data Description, for better target detection.

1.3 The Contributions Of This Thesis

1.3.1 Organization Of The Thesis

This thesis describes a brief mathematical background of essential concepts and existing methods in Chapter 2. Chapter 3 introduces the proposed method and alternate methods. Chapter 4 contains details of the implementation, the results and observations. The thesis concludes with Chapter 5 containing a brief summary and the scope for future work.

1.3.2 The Proposed Approach: AWK-SVDD

The contributions of this thesis and the proposed approach are as follows:

- The thesis proposes the use of wavelet functions, including the Morlet wavelet function and the Adaptive Wavelet Kernel, as kernels in the context of target detection in hyperspectral images with Support Vector Data Description.

- It compares and analyzes the performance of conventional kernel functions, such as the Radial Basis Function and the Sigmoid function, with that of wavelet kernels, including the Morlet wavelet function, in the context of one-class classification.

- The thesis introduces the use of two kernel functions in the context of once-class classification. It discusses the implementation of one such implementation composed of the Morlet wavelet function and Radial Basis Functions and compares their performance with the use of single function kernel.

- The thesis proposes the Adaptive Wavelet Kernel Support Vector Data Description (AWK-SVDD) for one-class classification. It discusses and compares the performance and computational time requirements of this method with existing kernel methods and the wavelet kernel methods mentioned previously.

2. A BRIEF MATHEMATICAL BACKGROUND

This chapter gives a brief overview of the essential theoretical concepts. Section 2.1 talks about hyper-spectral images, how they contain information in the spectral domain and the technical challenges in processing them. This is followed by a brief introduction to the wavelet transform and a description of some commonly used wavelets used in this thesis in Section 2.2. Section 2.3 describes the concepts of Support Vector learning methods and kernel functions. In Section 2.4, we explain the relevance of the proposed scheme compared to the existing methods.

2.1 Hyper-Spectral Imagery

Most of the visual data acquisition systems acquire data in the red, green and blue(RGB) bands of the visual spectrum. The human eye is also accustomed to RGB images. However, for the past few decades, high resolution remote sensing or spectral imaging has been an active focus of research [5].

Spectral imaging is the technique of acquiring and using images beyond the visible spectral range. In this technique, the spectral region is sampled at an increased number of points.

Multi-spectral imaging samples the spectrum at 10-15 points, whereas hyper-spectral images involve the sampling of the spectrum at more than 20-30 points. Acquisition of hyper-spectral images requires the use of specialized equipment, such as the NASA's AVIRIS sensor [2]

2.1.1 The Hyper-spectral Data Cube

The hyper-spectral sensors acquire several images across the spectrum, with each image containing information corresponding to the reflectance of the objects across a portion of the spectral range. Thus, each image corresponds to a "spectral band".

Thus, together, the images, can be combined to form a three-dimensional data cube across the spectral dimension [3]. This is known as the spectral cube. The spectral cube contains information about the area imaged in both the spatial and the spectral domain.

The spectral domain

In remote sensing images, each image cell or pixel in a spectral band physically represents an area covered. This is known as the spatial resolution of the image. Existing imaging systems have resolutions of upto 1 sq. m. per pixel.
However, across the spectral cube, each image cell can also be represented as a spectral signal. This is referred to as the representation of the area in the spectral domain. Thus, the spectral signal contains information about the reflectance properties of the objects in the spatial area covered. This information has been shown to be useful for data mining and classification of the image cells [3], [6], [4], [7], [8]. This thesis aims to improve upon the previous attempts on classification using the spectral domain.

2.1.2 Challenges In Hyper-Spectral Image Processing

Large spatial variability of the hyper-spectral target signature

The hyperspectral signature of the target signature can vary widely depending on the spatial characteristics.

Atmospheric effects

The hyperspectral data contains the reflectance values of the objects. The atmospheric effects may cause a huge margin of error. However, the image datasets used were pre-processed to remove the atmospheric effects.

The dimensionality of the hyperspectral information

All the spectral information contained in the hyperspectral image may not be essential for classification. This dimensionality of the data increases the complexity in data processing.

2.2 A Brief Introduction to Wavelets

Wavelets have been used extensively over the past decade to analyze signals. Wavelet analysis works by the representation of the given signal in terms of a set of basis functions. Due to the nature of the basis functions involved, wavelet analysis has been shown to better represent the given signal better and achieve greater numerical stability in reconstruction. As a result, they are easier to manipulate and hence, a useful tool in data analysis [9], [10], [11], [12] and [5].

2.2.1 Mother Wavelet And The Concept Of Wavelet Decomposition

The basis functions used in wavelet analysis are termed as wavelets. These wavelets are obtained as shifted and scaled forms of a fixed function called the mother wavelet.

The mother wavelet function

A complex-valued function, ψ, satisfying the following two conditions can be considered for use as a mother wavelet:

Finite energy condition

$$\int_{\infty}^{\infty} |\psi(t)|^2 dt < \infty, \tag{2.1}$$

Admissibility criterion

$$c_\psi = 2\pi \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty, \tag{2.2}$$

where Ψ is the Fourier Transform of ψ. The second condition implies that if $\Psi(\omega)$ is smooth, then $\Psi(0) = 0$.

The continuous wavelet transform

Given a mother wavelet, ψ, the continuous wavelet transform of a real signal $y(x)$ with respect to the wavelet function $\psi(x)$ is defined as:

$$y(b, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \tilde{\psi}\left(\frac{x - b}{a}\right) y(x)dx, \tag{2.3}$$

where $\tilde{\psi}$ denotes the complex conjugate of ψ, b $\exists$ $\mathbf{R}$, $a > 0$. Here, b corresponds to the time shift and a corresponds to the scale of the wavelet.
The translated and scaled version of the mother wavelet can be represented as:

$$\psi_{a,b}(x) = \frac{1}{\sqrt(a)}\psi\left(\frac{x - b}{a}\right). \tag{2.4}$$

Using this terminology, the equation (2.3) can be rewritten as

$$Y(b, a) = \int_{-\infty}^{\infty} \tilde{\psi}_{a,b}\ y(x)dx. \tag{2.5}$$

The inverse transform is given by:

$$y(x) = \frac{1}{c_\psi} \int_{-\infty}^{\infty} Y(b, a)\psi_{a,b}(x)\frac{db.da}{a^2}. \tag{2.6}$$

The Discrete Wavelet Transform

The discrete form of the continuous wavelet transform given in Eq. (2.5) is obtained by substituting $a = a_0^m, b = nb_0, m, n \in \mathbf{I}$. This gives us the discrete wavelet transform defined as follows:

$$Y(m,n) = \int_{-\infty}^{\infty} \tilde{\psi}_{m,n}(x) y(x) dx, \tag{2.7}$$

where

$$\psi_{m,n}(x) = a_0^{-m/2} \psi \left(\frac{x - nb_0}{a_0^m} \right). \tag{2.8}$$

Similarly, the discrete form of (2.6) gives us the inverse discrete wavelet transform as follows:

$$y(x) = k_\psi \sum_m \sum_n Y_{m,n} \psi_{m,n}(x), \tag{2.9}$$

where k_ψ is the normalization factor.

Parametrized Quadrature Mirror Filter (QMF) Wavelet Decomposition

The discrete wavelet transform (DWT) coefficients of a given signal can be computed using a Quadrature Mirror Filter (QMF) bank. A QMF consists of high-pass and low-pass filters, h and g and are related to a single mother wavelet. Analytic formulae for these filters are specific to the filter length, but are useful for the parametrization of QMF generation.

The following equations are an example of parametrization of a QM Filter bank of length, L = 4 [13]:

$$\begin{aligned} i = 0,3: \quad h[i] &= \frac{1 - \cos\alpha + (-1)^i \sin\alpha}{2\sqrt{2}} \\ i = 1,2: \quad h[i] &= \frac{1 - \cos\alpha + (-1)^{i-1} \sin\alpha}{2\sqrt{2}} \end{aligned} \tag{2.10}$$

for a given angle, $\alpha\epsilon[0, 2\pi]$.

2.2.2 Common Wavelet Functions

The proposed approach, Adaptive Wavelet Kernel Support Vector Data Descriptor (AWK-SVDD), takes advantage of the improved representation of the signal using wavelet functions. While any function that satisfies the conditions given in (2.1) and (2.2) can be used as a wavelet function, the use of particularly a few of them have been explored in this thesis. These wavelets are explained in the following sub-section.

Gaussian wavelet

A Gaussian wavelet has the form:

$$\psi(x) = \exp\left(-\frac{x^2}{2} + \jmath\omega x\right), \tag{2.11}$$

where usually, $\omega > 5$, so that the DC component of the Fourier Transform of the mother wavelet is negligible [9]. Fig. 2.1 shows the Gaussian mother wavelet.

Morlet wavelet

The Morlet wavelet is one of the popular complex wavelets in practice. The mother wavelet, as used in this thesis, for the Morlet wavelet is defined as [11]:

$$\psi(x) = \cos(1.75x)\exp\left(\frac{x^2}{2}\right) \tag{2.12}$$

Fig. 2.2 shows the Morlet wavelet function.

Fig. 2.1. The Gaussian mother wavelet function

2.3 Pattern Classification Using Support Vector Data Description

The classification of the individual pixels of hyper-spectral images into the target and the background is a specific case of pattern recognition.

Pattern recognition, in a broad sense, can be described as the problem of approximating a mapping from the n-dimensional data, x_i to the corresponding class labels, y_i. In other words, it aims to estimate the function f, such that:

$$(x_i, y_i), \ldots (x_n, y_n) \in \mathbf{R}^N \rightarrow \{\pm 1\} \tag{2.13}$$

Many methods have been traditionally used for pattern recognition. Some of them assume the knowledge of the structure of the classes involved. These are known to be

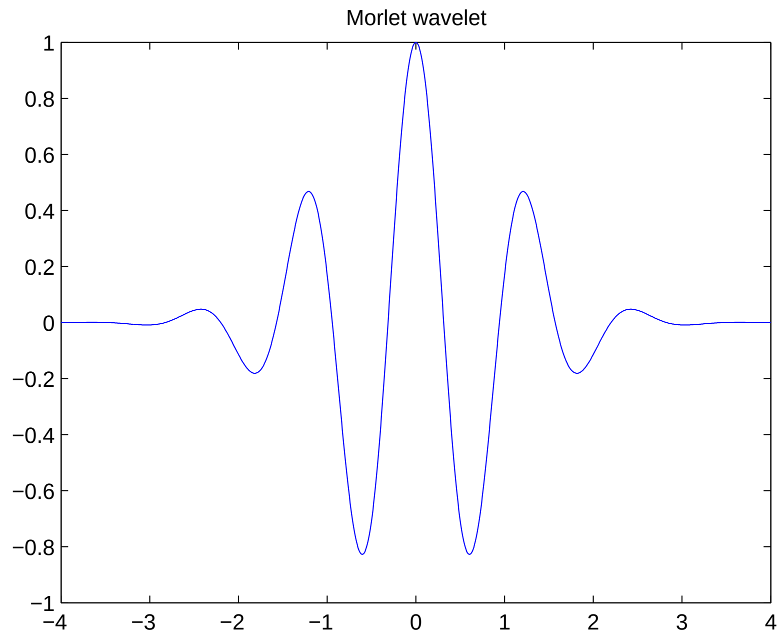

Fig. 2.2. Morlet wavelet function

parametric methods. The Linear Discriminant Analysis method is one such method and has been applied in a variety of classification problems [14]. The methods that do not assume any prior knowledge are known as the non-parametrized methods. Most of the pattern recognition methods such as the Neural Networks employ Empirical Risk minimization. This implies that the methods attempt to minimize the error of misclassification. However, Support Vector Machines (SVM) and Support Vector Data Descriptions (SVDD) employ structural risk minimization - They attempt to minimize the probability of misclassifying a randomly drawn, unseen, data point from a fixed but unknown probability distribution. This has made these methods popular in several applications. AWK-SVDD, proposed in this thesis, is derived from the SVDD (Support Vector Data Description) method.

2.3.1 Support Vector Machines

The basic formulation of Support Vector Machines is used for two-class classification. Support Vector Data Description is derived from a Support Vector Machine, specifically when the focus is on classifying one class from the rest.

Support Vector Machines - Linear decision surface

The Support Vector Machine method aims at finding the Optimal Separating Hyperplane between the two classes of data points. [14]
Given that $x_i,\ i\ \in\ [1,n]$ are the linearly separable data points given with the corresponding labels, $y_i\ \in\ \{-1,+1\}$, a linearly separating hyperplane is characterized by the parameters, $(w,b),\ w\ \in\ \mathbf{R}^n,\ b\ \in\ \mathbf{R}$, satisfying the condition:

$$y_i\left(\langle w \cdot x_i\rangle\ +\ b\right)\ \geq 1 \tag{2.14}$$

The equation of the hyperplane is given by:

$$\langle w \cdot x\rangle\ +\ b\ =\ 0 \tag{2.15}$$

It can be shown that $\frac{1}{||w||}$ is the lower bound on the distance between x_i and the hyperplane, defined by (w,b).
The optimal separating hyperplane is defined as the hyperplane for which the distance between the data points and the hyperplane is maximum. In other words, this boils down to maximizing $\frac{1}{||w||}$, subject to the constraint, (2.14). This is formulated as

$$\min\ \frac{1}{2}\langle w \cdot w\rangle \tag{2.16}$$

$$\text{subject to } y_i\left(w \cdot xi\ +\ b\right) \geq 1,\ i\ =\ 1,2,\ldots,n \tag{2.17}$$

The above problem, (2.16) can be solved by the method of Lagrangian multipliers. The corresponding Lagrangian function is given by:

$$L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_{i=1}^{n} \alpha_i \{ y_i (w \cdot x_i + b) - 1 \} \tag{2.18}$$

where the Lagrangian multipliers are given by $\alpha = \{\alpha_1, \alpha_2, \ldots \alpha_i \ldots, \alpha_n\}$, $\alpha_i > 0, i = 1, 2, \ldots, n$. At the saddle point, $L(w, b, \alpha)$ reaches a minimum for $w = \hat{w}, b = \hat{b}$ and a maximum for $\alpha = \hat{\alpha}$. Thus,

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} y_i \alpha_i = 0 \tag{2.19}$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \tag{2.20}$$

given that

$$\frac{\partial L}{\partial w} = \left(\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \ldots \frac{\partial L}{\partial w_n} \right) \tag{2.21}$$

Substituting (2.19) and (2.20) into (2.18), the Lagrange function to be maximized can be written as:

$$L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1} n \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle \tag{2.22}$$

subject to the constraint, (2.19) and given that, $\alpha \geq 0$. This is known as the dual formulation of (2.18) and can be written as:

$$\min \quad -\frac{1}{2} a^T D a + \sum_{i=1}^{n} \alpha_i \tag{2.23}$$

$$\text{subject to} \sum_{i=1} n y_i \alpha_i = 0, \alpha \geq 0 \tag{2.24}$$

where D is an $N \times N$ matrix such that

$$D_{ij} = y_i y_j \langle x_i \cdot x_j \rangle \tag{2.25}$$

Also, from (2.20),

$$\hat{w} = \sum_{i=1}^{n} \hat{\alpha}_i y_i x_i \tag{2.26}$$

And from the Kharush Kuhn-Tucker conditions of optimization theory [15],

$$\hat{\alpha}_i \left(y_i \left(\hat{w} \cdot x_i + \hat{b} \right) \right) = 0, \ i = 1, 2, \ldots, n \tag{2.27}$$

Most of the $\hat{\alpha}_i$ are usually null. As a result, the vector $\hat{w}$ is a linear combination of only some of the points, x_i, known as support vectors. Only these points are needed to determine the Optimal Separating Hyperplane. The parameter,$\hat{b}$ is obtained from the Kuhn Tucker conditions. And hence, the classification of a new data point is determined by the sign of:

$$\langle \hat{w} \cdot x \rangle + \hat{b} \tag{2.28}$$

Support vector machines - The non-linearly separable case

If the set of support vector data points are non-linearly separable, the previous analysis fails. N slack variables are introduced for a penalty in case of misclassification and analysis is conducted as above.

Support Vector Machines - The kernel trick

The previous section, 2.3.1, describes the solution for a linear decision surface using Support Vector Machines. However, the strength in the Support Vector learning methods is in their applications to general decision surfaces. The following section elaborates on the use of kernel functions.

2.3.2 The Significance Of Kernel Functions

Definition of a kernel

For a general decision surface, the set of input data vectors, $\{x_1, x_2, \ldots, x_n\}$ are transformed to a higher dimensional feature space, using a mapping, $\Phi(x_i) :\rightarrow z_i$, such that the feature vectors obtained after mapping, $\{z_1, z_2, \ldots, zn\}$, are linearly separable. It can be observed that the solution of this problem only requires the computation of the inner product of the feature vectors, $\langle \Phi\left(x_i\right), \Phi\left(x_j\right) \rangle$ in the higher-dimensional space. Thus, the accuracy can be significantly improved by the choice of a suitable function, K, such that

$$\langle \Phi(x_i), \Phi(x_j) \rangle = K(x_i, x_j) \tag{2.29}$$

This function is known as a Kernel function.

Mercer's theorem

Mercer's theorem gives the conditions to be met by a function in order to be considered for use as a kernel.
Let $K(x_i, x_j)$ be a continuous symmetric kernel defined in the closed interval, $a \leq x_i, xj \leq b$. Then,

1. The kernel expansion is given by

$$K(x_i, x_j) = \sum_{i=1}^{\infty} \lambda_i \Phi(x_i) \Phi(x_j') \tag{2.30}$$

   with positive coefficients, $\lambda_i > 0, \forall i$

2. A necessary and sufficient condition for the above expansion, (2.30), is, $\forall g(x)$, given that, $\int_a^b g^2(x)dx < \infty$,

$$\int_a^b \int_a^b K(x_i, x_j)g(x_i)g(x_j)dx_i dx_j \geq 0 \tag{2.31}$$

The numbers λ_i are known as the eigenvalues of the expansion and the functions, $g_i(x)$ are called the eigenvalues of the expansion. All the eigenvalues are positive and hence, a kernel function, is positive definite.

Commonly used kernels

Using a suitable kernel function to classify non-linear data using Support vector learning methods is known as the kernel trick. The kernel chosen highly influences the result of the classification.
Several kernel functions are very commonly used. Some of these functions are listed as follows:

Polynomial kernel

$$K(x_i, y_j) = (x_i \cdot x_j \ + \ 1)^d \tag{2.32}$$

Here, the order of the polynomial, d is a kernel parameter.

Radial Basis Function (RBF)

$$K(x_i, y_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) \tag{2.33}$$

where $\sigma \in \mathbf{R}$, is a kernel parameter.

Table 2.1
Popular kernel functions

No	Kernel	Formula ($K(x,y) =$)
1	Polynomial	$(x_i \cdot x_j \ + \ 1)^d$
2	Radial Basis Function	$\exp\left(-\frac{\|\|x_i - x_j\|\|^2}{2\sigma^2}\right)$
3	Exponential	$\exp\left(-\left(\|\|x_i - x_j\|\|\right)/\beta\right)$

Exponential kernel

$$K\left(x_i, x_j\right) = \exp\left(\frac{-\left(||x_i - x_j||\right)}{\beta}\right) \tag{2.34}$$

where β is a kernel parameter.

Table 2.1 summarizes the kernels described above.

2.3.3 Support Vector Data Description

The Support Vector Data Description(SVDD) is a solution to the data domain classification problem. ([16]) The data domain classification aims to distinguish a desired class of data as given in the training data and reject all other possible data points. The desired class is known as "the target class" and the rejected data belong to "the outlier class".

SVDD for a spherical distribution

It is assumed that a description is required of a data set containing N data points, given by, $\{x_1, x_2, x_3, \ldots, x_N\}$. The idea is to obtain a sphere of minimum volume containing the data objects. Since the volume of the sphere is extremely sensitive to the most outlying object, slack variables, ξ_i, are introduced to allow for data points to lie outside the sphere.

Given that the sphere is described by the center a, and radius R, the problem is to minimize the radius of the resultant sphere, given by,

$$F\left(R, a, \xi_i\right) = R^2 + C\sum_i \xi_i \tag{2.35}$$

where C gives the trade-off between the volume of the sphere and the number of target data points allowed to be rejected. This represents the trade-off between simplicity and the number of errors. The constraints for this minimization are given by:

$$\left(x-a\right)^T\left(x-a\right) \leq R^2 + \xi_i \ \forall i, \xi_i \geq 0 \tag{2.36}$$

Incorporating the constraints, (2.36), into (2.35), the following Lagrangian is constructed:

$$L(R, a, \alpha_i, \xi_i) = R^2 + C\sum_i \xi_I - \sum_i \alpha_i\{R^2 + \xi_i - \left(x_i^2 - 2ax_i + a^2\right)\} - \sum_i \gamma_i\xi_i \tag{2.37}$$

with the Lagrange multipliers, $\alpha_i, \gamma_i \geq 0$. Finding the saddle point of the Lagrangian and setting the partial derivatives to zero, we get:

$$\sum_i \alpha_i = 1 \tag{2.38}$$

$$a = \frac{\sum_i \alpha_i x_i}{\sum_i \alpha_i} = \sum_i \alpha_i x_i \tag{2.39}$$

$$C - \alpha_i - \gamma_i = 0 \ , \forall i \tag{2.40}$$

$$\tag{2.41}$$

Rewriting eq.(2.37) using the equations in (2.40) to maximize w.r.t α_i:

$$L = \sum_i \alpha_i\left(x_i \cdot x_i\right) - \sum_{i,j} \alpha_i\alpha_j\left(x_i \cdot x_j\right) \tag{2.42}$$

subject to the constraints, $0 \leq \alpha_i \leq C, \sum_i \alpha_i = 1$. As in the case of the support vector machine, the parameters of the sphere only depends on the support vectors. A test point, $\bar{x}$, is accepted when the distance to the center of the sphere determined as above is smaller than the radius. Expressing this in terms of the support vectors, the test point is accepted when:

$$(\bar{x} \cdot \bar{x}) - 2\sum_i \alpha_i (\bar{x} \cdot x_i) + \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2 \tag{2.43}$$

SVDD for the general case

The equation,(2.43), obtained in the previous section, assumes a spherical distribution of data. However, SVDD can be used in the general case too, by applying the kernel trick, i.e, replacing all inner products $\langle x_i, x_j \rangle$ by the value of the kernel, $K(x_i, x_j)$. In this case, the problem (2.42) is rewritten as:

$$L = \sum_i \alpha_i K(x_i, x_j) - \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \tag{2.44}$$

subject to the constraints, $0 \leq \alpha_i \leq C, \sum_i \alpha_i = 1$. Similarly, as in (2.43), a test point, $\bar{x}$, is accepted when

$$K(\bar{x}_i, \bar{x}_j) - 2\sum_i \alpha_i K(\bar{x}_i, \bar{x}_j) + \sum_{i,j} \alpha_i \alpha_j K(\bar{x}_i, \bar{x}_j) \leq R^2 \tag{2.45}$$

As in the case of an SVM, the choice of the kernel highly influences the classification results.

2.4 The Need for Adaptive Wavelet Kernel SVDD (AWKSVDD)

This section highlights the advantages of using Support Vector learning methods for pattern recognition and goes on to describe why one-class classification is more relevant for the hyper-spectral target classification problem. The section also explains why the proposed approach AWK-SVDD is required and offers an advantage over the existing methods.

2.4.1 The Advantages Of Support Vector Classification Methods

Support vector-based classification methods have been shown to be more relevant for the classification of hyper-spectral data ([6]) due to the following reasons:

- Support Vector methods have been shown to be efficient at handling large input spaces. [17], [18]
- Previous research has demonstrated that Support Vector methods are quite robust in the presence of noisy samples. [19]
- The resulting solution depends only on the support vectors and hence, the solution requires computations only on a subset of the training set. This produces sparse solutions that are easier to compute. [18], [20]
- The flexibility offered due to the use of a kernel supports the use of the methods for different data distributions, subject to the application of a suitable kernel function. [20]
- Compared to traditional methods, with a suitable kernel, the support vector learning methods have been shown to work well in the absence of a-priori information about the data distribution. [18]

2.4.2 The Relevance Of One-class Classification

One-class classification has been used mainly to recognize a particular class, independent of the characteristics of the other classes. Support Vector Data Description has been successfully used for the one-class classification of several types of data. [21], [4], [7], [8]

- SVDD inherits all the advantages of using an SVM for binary-class classification as described in the previous section. These include, efficient output for large input spaces, robustness w.r.t noisy samples, sparse structure of the solution and the flexibility due to the kernel trick.

- Previous research has shown that SVDD can be successfully used to determine the boundary of a target data set in the case of availability of a very small sample set. [22]

- SVDD, with the optimal choice of decision threshold and tuning of false alarm rate, has been applied successfully when pure target data is difficult to obtain. [21]

- Data description has also been successfully used in the case of highly imbalanced datasets, where the samples of each of the classes are not equally distributed. [23]

2.4.3 The Need For An Adaptive Wavelet Kernel

The use of SVDD for the classification of hyperspectral images is a relatively new area. Particularly, the use of wavelet kernels with SVDD has not been studied in detail yet. However, several attempts have been made to use the discrete wavelet transform followed by support vector machines [12] or support vector machines with wavelet kernels [10] in pattern recognition problems. An attempt has also been made to use SVDD for target detection in hyper-spectral images, however, these approaches suffer from several discrepancies and, as a result, the proposed approach with an adaptive wavelet kernel, presents itself as an intuitive solution.

- It has been shown that the wavelet transform is an admissible kernel using Mercer's conditions and, under optimization, it models the data better than the normal Gaussian kernel. [10]

- Hence, several wavelets including the ones mentioned in this chapter satisfy the Mercer's conditions and can be used as kernels. The resultant efficiency of classification for each wavelet, however, varies with the data.

- Hyper-spectral images usually have a large number of spectral bands and due to the high correlation between these bands, feature selection is essential. This requires that the optimal wavelet coefficients be chosen during the computation of the kernel.

- A kernel learning approach to decide the type of wavelet would solve the problem of the choice of the mother wavelet function. Yger, Rakotomamonjy *et. al.* [24] proposed the Wavelet Kernel Learning method (WKL) that learns the shape of the mother wavelet, selects the best wavelet coefficients and learns a large-margin classifier, when used with a support vector machine. Thus, it achieves optimal kernel learning and feature selection.

- Adaptive Wavelet Kernel - Support Vector Data Description (AWK-SVDD), while retaining the advantages of WKL, uses kernel learning with SVDD, making it suitable for one-class classification. As explained previously, this is particularly relevant in the context of target detection in hyper-spectral images.

3. ADAPTIVE WAVELET KERNEL - SUPPORT VECTOR DATA DESCRIPTION (AWK-SVDD)

This chapter explains the proposed approach Adaptive Wavelet Kernel -Support Vector Data Description in Section 3.2. It also discusses mathematical background in Section 3.1.1 necessary for the validity of AWKSVDD. The Morlet wavelet kernel and the Morlet-RBF wavelet kernel, analyzed novelly in the context of one-class classification, are discussed in Section 3.1.2 and Section 3.1.3 respectively.

3.1 SVDD With Wavelet Kernels

In this section ,we theoretically analyze the use of wavelet kernels with Support Vector Data Description. In particular, Section 3.1.1 discusses the general conditions for admissibility of a wavelet function as a kernel and specifically, the Morlet and Morlet-RBF wavelet functions are described.

3.1.1 Admissibility Of Wavelet Kernels

In the previous chapter, it has been mentioned that the conditions, Eq.2.31 and Eq.2.30, are required for a function to be used as a kernel. The formula for a wavelet transform in terms of the mother wavelet, Eq.2.10 has also been seen. Wavelet functions, being translation invariant functions, can be shown to satisfy the Mercer's conditions, and hence be used as kernels. ([10])

Translation-invariant kernels

A translation-invariant kernel is one that satisfies the following property:

$$K\left(x_i, x_j\right) = K\left(x_i - x_j\right) \tag{3.1}$$

It has been shown that the necessary and sufficient condition for translation-invariant kernel to satisfy the Mercer's conditions is given by:

$$F[K](\omega) = (2\pi)^{-N/2} \int_{R^N} \exp\left(-\jmath\left(\omega \cdot x\right)\right) K(x)dx \geq 0 \tag{3.2}$$

where $F[K](\omega)$ gives the Fourier Transform of $K(x)$.

Wavelet functions as translation-invariant kernels

From the previous chapter, the translated and scaled version of a mother wavelet,ψ, is given by

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

where $b \in R$ gives the shift in time and $a > 0$ gives the scale of the wavelet.
The continuous wavelet transform (Eq.(2.5)) of a function $y(x)$ can also be written as:

$$Y\left(b, a\right) = \langle y(x) \cdot \psi_{a,b}(x)\rangle \tag{3.3}$$

where $\langle \ \cdot \ \rangle$ denotes an inner product in the L_2-normed space. This thesis aims to analyze the performance of wavelet functions as kernel functions. Specifically, the performance of Morlet wavelet kernel, discussed in Section 3.1.2, the Morlet-RBF wavelet kernel, discussed in Section 3.1.3 and the proposed Adaptive Wavelet Kernel discussed in Section 3.2 are analyzed in the context of one-class classification.

3.1.2 Morlet Wavelet Kernel

From the previous chapter, the mother wavelet for a Morlet wavelet is given by Eq.(2.12) as:

$$\psi(x) = \cos(1.75x)\exp(\frac{x^2}{2})$$

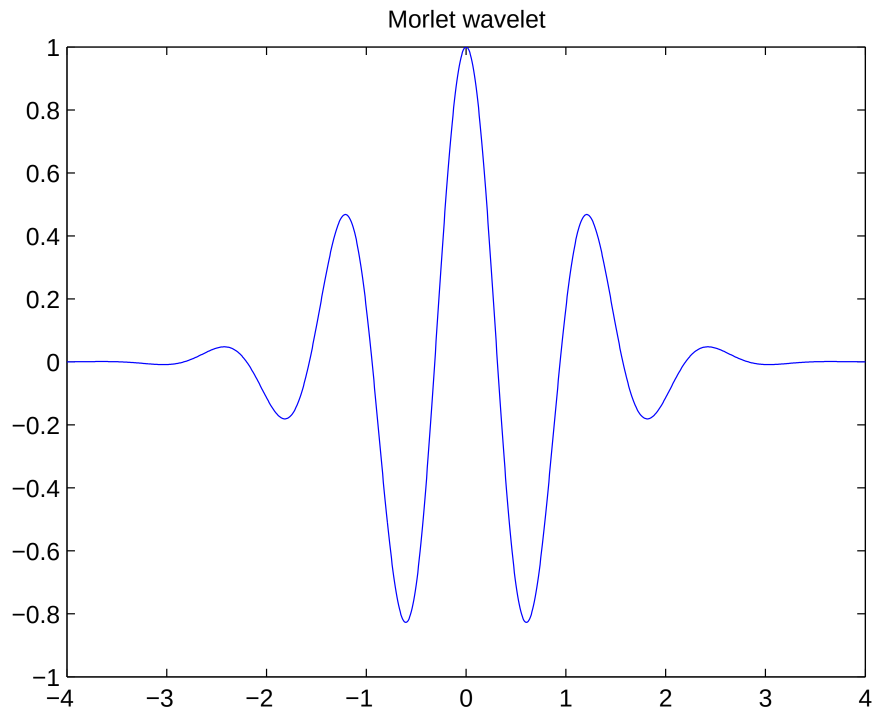

Fig. 3.1. The Morlet Mother Wavelet Function

Substituting the above equation in the formula for a translation-invariant kernel(Eq. 3.1), we get the wavelet kernel for the Morlet wavelet as:

$$K(x_i, x_j) = \prod_{k=1}^{N} \left(\cos\left(1.75\left(\frac{x_i^k - x_j^k}{a}\right)\right) \exp\left(-\frac{||x_i^k - x_j^k||^2}{2a^2}\right)\right) \tag{3.4}$$

where x_i^k and x_j^k denote the k^{th} element of the data vector x_i and x_j respectively. a is a kernel parameter. Fig. 3.1 shows a Morlet mother wavelet function.

3.1.3 Morlet-RBF Wavelet

The use of two kernels, and the Morlet-RBF kernel in particular, was shown by Jiang. et al. [25] to perform better than the existing kernels in medical image classification with Support Vector Machines. This thesis attempts to use the Morlet-RBF kernel in the context of one-class classification and analyze the performance.

The use of two kernels

The use of Morlet-RBF kernel is based on the assumption that, in instances where one time mapping of the data is not sufficient to make data linearly separable, twice mapping can be used. However, the two kernels need to be similar in order to have a consistent transform process. The Morlet and Gaussian kernels are similar in distribution and hence, are a suitable choice.

Mathematical formula for the Morlet-RBF kernel

If $K_1(x_1, x_2)$ is used to represent the Morlet kernel between the set of feature vectors, x_1 and x_2 and $K_2(x_1, x_2)$ is used to represent the Gaussian wavelet, the Morlet-RBF kernel, $K(x_1, x_2)$ can be represented as follows:

$$
\begin{aligned}
K(x_1, x_2) &= \psi(\varphi(x_1), \varphi(x_1)) \\
&= K_2(\varphi(x_1), \varphi(x_1)) \\
&= \exp(-\gamma \|\varphi(x_1) - \varphi(x_2)\|^2) \\
&= \exp(-\gamma[\varphi(x_1) - \varphi(x_2) - 2\varphi(x_1) \cdot \varphi(x_2) + \varphi(x_1) \cdot \varphi(x_2)]) \\
&= \exp = \gamma[2 - 2K_1(x_1, x_2)]
\end{aligned}
\tag{3.5}
$$

3.2 Adaptive Wavelet Kernel - Support Vector Data Description

Adaptive Wavelet Kernel Support Vector Data Description (AWK-SVDD) was motivated by the wavelet kernel learning method, proposed by F.Yger and A.Rakotomamonjy [24].

3.2.1 Multiple Kernel Learning

Recent research has shown that using multiple kernels instead of a single kernel improves classification performance. [26]. One of the commonly used approaches is to consider the resultant kernel $K(x_1, x_2)$ as a convex combination of basis kernels as follows:

$$K(x_1, x_2) = \sum_{m=1}^{M} d_m K_m(x_1, x_2) \tag{3.6}$$

$$\text{given} \sum_{m=1}^{M} d_m = 1,\ d_m \geq 0 \tag{3.7}$$

where each $K_m(x_1, x_2),\ m = 1, \ldots, M$ is a positive definite kernel belonging to the reproducing kernel Hilbert Space (RKHS), $\mathbf{H}$, and d_m are the corresponding weights. Thus, the problem of finding an accurate kernel representation is reduced to finding the optimal weights by solving the above formulation. One such solution is the Simple Multiple Kernel Learning (Simple MKL) was proposed by Rakotomamonjy et. al. ([26]). Wavelet Kernel Learning is another solution based on Simple MKL and is discussed in the following section. [24]

3.2.2 Adaptive Wavelet Kernel Learning for Support Vector Data Description

Wavelet Kernel Learning assumes that each of the kernels, $K_m(,)$ in Eq.3.6 is formed by the coefficients given by a wavelet decomposition, as described in Eq.2.10. The method uses an iterative approach to find the classification boundary and the kernel weights using a Support Vector Machine as the classifier.

Wavelet Kernel using a Quadratic Mirror Filter

A multiple kernel learning approach is adopted to determine the optimal wavelet features and the optimal mother wavelet to be used. Suppose the orthogonal mother wavelet is given by ϕ_α obtained by parametrization of the wavelet using QMF banks as in Eq. 2.10. Let this wavelet be dilated at scale s and translation t, giving $\phi_{\alpha,s,t}$. Then, given the signals, x, y $\epsilon\mathbb{R}^d$, the linear wavelet kernel is given as follows:

$$K_{\alpha,s,t}(x, y) =< \phi_{\alpha,s,t}x > \cdot < \phi_{\alpha,s,t}y > . \tag{3.8}$$

where $K_{\alpha,s,t}$ is the product of the wavelet coefficients obtained at scale s and translation t.

Formulation of wavelet kernel learning

Consider a training set $(x_i, y_i)_{i=1}^n$, where $x_i \in \mathbb{R}^d$, and the labels are given by $y_i = \{1, -1\}$. Using wavelet kernels $K_m(,)$ parametrized with Quadratic Mirror Filters as defined in the previous section, it has been shown that the decision function for the Support Vector Machine, $f(x)$ can be obtained from the solution of the Multiple Kernel Learning problem with the following primal formulation: [26]

$$\begin{aligned} \min_d J(d) = & \begin{cases} \min_{\{f_m\},b,\xi} & \frac{1}{2}\sum_m d_m \|f_m\|^2_{H_m} + C\sum_i \xi_i \\ \text{s.t.} & y_i \sum_m f_m(x_i) + y_i b \geq 1 - \xi_i,\ \xi_i \geq 0,\ \forall i \end{cases} \\ \text{s.t.} & \sum_m d_m = 1,\ d_m \geq 0,\ \forall m \end{aligned} \tag{3.9}$$

The kernel weights d_m obtained from the above solution and the dual variables α^* give the following decision function:

$$f(x) = \sum_i \alpha_i^* y_i (\sum_m d_m K_m(x_i, x_i)) + b \tag{3.10}$$

AWK-SVDD considers a similar formulation that can be applied to one-class classification as follows:

$$\min_{d} J(d) = \begin{cases} \min_{\alpha} & J(d) = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \sum_{m} d_m K_m(x_i, x_j) - \sum_{i=1}^{N} \alpha_i \sum_{m} d_m K_m(x_i, x_i) \\ \text{s.t.} & \sum_{i=1}^{N} \alpha_i = 1, 0 \leq \alpha_i \leq C, \forall i \end{cases}$$

$$\text{s.t.} \quad \sum_{m} d_m = 1,\ d_m \geq 0,\ \forall m \tag{3.11}$$

An infinite number of kernels needs to be dealt with for the solution to the above multiple kernel learning framework. Yger et. al. proposed an iterative approach in order to solve this with a reduced computational complexity and a finite kernel space. [24] This is described in the following section.

Iterative solution to the wavelet kernel weight learning problem

The wavelets used in the adaptive wavelet kernel problem(Eq. 3.11) are parametrized using the parameter θ as described previously in Eq. 2.10. An iterative approach similar to the one suggested by Yger. et. al [24] is used to solve this. This section describes the implementation of this approach. A finite set of kernels, described by $\Theta = \theta$ is sampled from the otherwise infinite kernel space, described by $[0, 2\pi]^{M-1}$. This leads to an exponential number of kernels. However the iterative approach proposed by Yger. et. al. [24] provides a reasonably efficient solution.

The problem can be rewritten as the following non-linear convex and differentiable simplex formulation:

$$\begin{aligned} \underset{d}{\text{minimize}} \quad & J(d) \\ \text{subject to} \quad & \sum_{m} d_m = 1 \\ & d_m \geq 0,\ \forall m. \end{aligned} \tag{3.12}$$

Using Kharush Kuhn Tucker conditions, at optimality, it follows that:

$$\begin{aligned} \frac{\partial J(d)}{\partial d_m} &= -\lambda \quad \text{if } d_m > 0 \\ \frac{\partial J(d)}{\partial d_m} &\geq -\lambda \quad \text{if } d_m = 0 \end{aligned} \tag{3.13}$$

where λ is the corresponding Lagrangian multiplier. It can also be shown that:

$$\sum_{m:d_m>0} d_m \frac{\partial J(d)}{\partial d_m} = -\lambda \tag{3.14}$$

It follows that, at optimality, all kernels with non-zero d_m have equal gradients. The kernels with $d_m > 0$ are termed as Active Kernels and non-active otherwise. The Wavelet Kernel Learning method uses this property to check on each iteration if optimality is reached using Eq. 3.13, and update the set of kernels if not. Three update strategies are used as proposed by Yger. et. al.:

Exhaustive search

In this update strategy, the kernel which maximally violates the contraint is chosen first.

Stochastic search

It is assumed that $\theta \in \Theta$, where Θ represents an infinite set of kernels. In order to find a constraint violating kernel, a QMF is randomly generated by randomly sampling on Θ. The first violating kernel is selected by visiting all wavelet coefficients in ascending scale order. If a violating kernel is not found after a certain pre-set number of samplings (20), the solution is considered optimal.

Full stochastic search

This search strategy involves random sampling on the wavelet scale and translation alongwith random sampling on the infinite set $Theta$. The first violating kernel through randomly generated QMF, scale and translation until a pre-decided number of sampling is reached.

The iterative method of update has been proved to be robust compared to the existing methods. In addition, multiple kernel learning has been shown to have greater accuracy compared to conventional kernels [24]. However, they have been applied only in the case of binary or multi-class classification. The proposed Adaptive Wavelet Kernel Support Vector Data Description implements these features in the context of one-class classification, better suitable for many applications, including the target detection in hyperspectral images.

4. IMPLEMENTATION AND RESULTS

Hyper-spectral data from the Urban dataset was used with varying signal to noise ratio. This chapter talks about the method used for simulating the target signature in Section 4.1, the results of the comparison between the proposed AWKSVDD approach and other wavelet kernels in Section 4.2, their computational requirements in Section 4.3 and discusses these results in Section 4.4.

4.1 Data And Preparation

The dataset and assumptions taken into account for target detection are explained in this section.

4.1.1 Hyper-spectral Image Data

An image from the HYDICE (Hyperspectral Digital Imagery Collection Experiment) sensor was chosen to test the method. HYDICE is an airborne pushbroom imager. HYDICE images have 210 spectral bands covering wavelengths of 0.4 - 2.5 microns in roughly 10nm bandwidths. The image of the urban scenery is shown in Fig 4.1. For the ease of visualization, it is displayed such that the colors red, green and blue correspond roughly to the bands 420nm, 440nm and 2400nm respectively. It is practically difficult to obtain the ground truth image in order to verify the accuracy of classification. For this reason, the targets were simulated using a Gaussian model and inserted at 300 randomly chosen pixels, roughly about half percent of the total pixels, in order to simulate the low-probability target detection arrangement. [7].

Fig. 4.1. Urban dataset hyperspectral image visualized at R = 420nm, G = 440nm and B = 2400nm

4.1.2 Gaussian Model Of Target Detection

A random target signature was created and a simple Gaussian noise model was used to obtain the corrupted samples of this signature. These corrupted samples were inserted at the 300 random locationsThe corrupted target signature, t_c is obtained from the pure target signature, t by adding noise n as follows:

$$t_c = t + n \tag{4.1}$$

where n $N_k[t, \sigma^2 I]$. In other words, n belongs to a K-dimensional Gaussian distribution with mean t and per-band variance σ^2. The variance σ was adjusted so that the resulting Signal to Noise Ratio (SNR) was 6 dB. (Usually classified as a highly corrupt signal). The comparison of the results obtained on this data using AWK-SVDD and known one-class classification methods are presented in the rest of this chapter.

4.2 Comparison Of Performance With Previously Established Methods

AWK-SVDD was used to classify the data obtained as described in Section 4.1 and its performance was compared with that of known classification methods. The F-measure was chosen as the measure of performance. This section explains the significance of using F-measure and the results obtained.

4.2.1 F-measure Value

F-measure is a statistical measure used to determine the accuracy of classification. F-measure (F) is defined as the harmonic mean of the statistical parameters, precision (P) and recall (R) as follows:

$$F = \frac{2PR}{P+R} \tag{4.2}$$

Precision and recall are defined as below:

$$P = \frac{TP}{TP+FP} \tag{4.3}$$

$$P = \frac{TP}{TP+FN} \tag{4.4}$$

where TP is the rate of true positives (the probability that a given sample classified as a target is a true target), FP is the rate of false positives (the probability that a given sample classified as a target is not a target) and FN is the rate of false negative (the probability that a target sample is rejected). Thus, the value of the F-measure can vary from 0 to 1, with 0 being very poor classification accuracy and 1 signifying the highest classification accuracy. The F-measure is widely used as a performance measure when precision and recall are equally valued.

4.2.2 Comparison of F-measure

AWK-SVDD and Support Vector Data Description with conventional kernel methods were used to classify the hyperspectral image data. Cross-validation and repeated iterations were used for a better estimation of predictive measure. The average F-measure obtained was noted. The plot of corresponding F-measures is given in Fig 4.2. In the figure, the value of F-measure for AWK-SVDD is the average of the F-measure values for all update strategies. It can be seen that the average F-measure

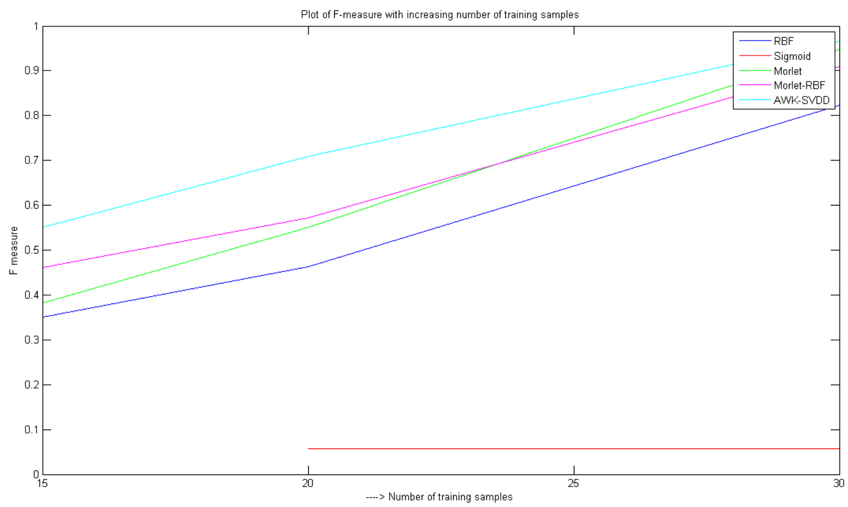

Fig. 4.2. F-measure plotted as a function of the number of training points

is the highest for the proposed AWK-SVDD method.

4.2.3 Variation Of Performance With The Number Of Training Samples

The Fig 4.2 shows the average F-measure for various kernels plotted against the number of training samples. It can be seen that, as expected, the F-measure increases with the number of training samples. For the Sigmoid kernel, the Support Vector Data Description optimization refused to converge for lesser number of training samples. It can be seen that AWK-SVDD performs well even with decreased number of training samples.

4.3 Analysis Of Computational Time Requirements

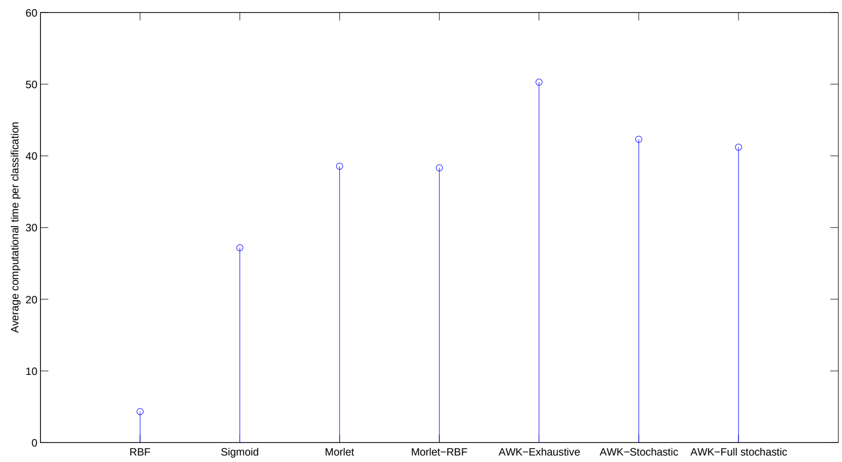

Fig. 4.3. Computational time taken by each method

The plot of average computational times for each of the methods per classification for $N = 30$ training samples is given in Fig 4.3. AWK-SVDD, being a multiple kernel learning method, requires more computational time as expected. However, it was seen that the time taken using the stochastic update strategy and the full stochastic update strategies are comparable with the time required by the Morlet kernel and the Morlet - RBF kernel. However, it should be noted that the value of the computational time is highly dependent on the implementation of the algorithm.

4.4 Results And Discussion

The results obtained are tabulated in Section 4.4 and the shortcomings of the approach are discussed in Section 4.4.2.

Table 4.1
Average precision values

METHOD	N = 15	N = 20	N = 30
RBF	1	1	1
Sigmoid	NaN	0.0291	0.0291
Morlet	1	1	1
Morlet-RBF	1	1	1
AWK-SVDD	1	1	1

Table 4.2
Average recall values

METHOD	N = 15	N = 20	N = 30
RBF	0.2267	0	0.7
Sigmoid	NaN	1	1
Morlet	0.25	0.3983	0.9
Morlet-RBF	0.31	0.41	0.8333
AWK-SVDD	0.3983	0.55	0.9333

4.4.1 Tabulated Results

Table 4.1 lists the average precision values obtained according to the Eq. 4.3for each of the methods with various number of training samples. It should be noted that the NaN values were obtained when the solution to the quadratic optimization formulation in the Support Vector Data Description did not converge.

Table 4.2 lists the average recall values for each of the methods obtained according to the Eq. 4.4 with various number of training samples. The resulting F-measure values obtained from the precision and recall values calculated as in Eq. 4.2 are listed in Table 4.3.

Table 4.4 summarizes the average computational time taken by each method with $N = 30$ training samples, plotted in the Fig. 4.3

Table 4.3
Average F-measure values

METHOD	N = 15	N = 20	N = 30
RBF	0.3503	NaN	0.8235
Sigmoid	NaN	0.0566	0.0566
Morlet	0.3826	0.551	0.9474
Morlet-RBF	0.4618	0.5714	0.9091
AWK-SVDD	0.551	0.7097	0.9655

Table 4.4
Average computational time

METHOD	Average time per classification
RBF	4.3212
Sigmoid	27.1754
Morlet	38.5478
Morlet-RBF	38.3138
AWK-Ex	50.2837
AWK-Stoc	42.3154
AWK-Ex	41.2143

4.4.2 Discussion And Short-comings

Computational complexity

As evident from table 4.4 and fig. 4.3, the average computational time taken varies widely with the update strategy. In general, the computational time varied widely between the iterations with the same dataset depending on the random seed used for the iterations. However, the average computational time is comparable with that of the regular Morlet wavelet kernel with increased performance.

Convergence of the method

During the classification tests, it was observed that the quadratic optimization formulation associated with the Support Vector Data Description did not converge at times. Checking for this and updating the kernel with updated weights resulted in a converging formulation. This might be due to an ill-conditioned kernel matrix and occurred irrespective of the kernel used. However, this needs to be checked with a wider variety of kernels and data in order to check the convergence of the method.

5. CONCLUSION

The proposed method Adaptive Wavelet Kernel Support Vector Data Description (AWK-SVDD) for the one-class classification of hyperspectral images in the context of target detection was discussed and analyzed. The Morlet and the Morlet-RBF wavelet kernels were also used in the context of one-class classification for the first time. The advantages of the proposed solution and possible improvements are discussed here in Section 5.1 and Section 5.2 respectively.

5.1 Advantages Of The Proposed Solution

- The method conveniently combines feature selection with kernel optimization.
- AWK-SVDD is based on one-class classification and hence, suits most practical applications of targeted object identification in hyperspectral images and many other classification problems.
- From the results, AWK-SVDD is quite robust even with lesser number of training samples.
- The computational time taken by AWK-SVDD can be significantly reduced using a suitable strategy as observed in the Fig. 4.3.

5.2 Scope For Improvement And Future Work

- The efficiency of classification was emphasized during the implementation of AWK-SVDD. However, computational efficiency can be further improved.
- AWK-SVDD focuses on using the spectral characteristics of the pixels for target detection. This does not take into account the inter-spatial correlation between the pixels. A spatio-spectral focus on classification would better exploit the information that hyper-spectral images have to offer.

- Several attempts have been made on combining one-class classifiers for multi-class classification. The results on combining AWK-SVDD for multi-class classification can be observed and analyzed.

- Batch processing can be implemented and the memory used in classification can be analyzed and optimized.

REFERENCES

[1] Peg Shipprt. Why use hyperspectral imagery? *American Society for Photogrammetry and Remote Sensing, University of Montreal*, 70:377 – 380, 2011.

[2] Robert O. Green, Michael L. Eastwood, Charles M. Sarture, Thomas G. Chrien, Mikael Aronsson, Bruce J. Chippendale, Jessica A. Faust, Betina E. Pavri, Christopher J. Chovit, Manuel Solis, Martin R. Olah, and Orlesa Williams. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote Sensing of Environment*, 65(3):227 – 248, 1998.

[3] D. Landgrebe. Hyperspectral image data analysis. *Signal Processing Magazine, IEEE*, 19(1):17 –28, Jan 2002.

[4] W. A. Sakla, A. A. Sakla, and A. Chan. A support vector data description approach to target detection in hyperspectral imagery. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7335 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, May 2009.

[5] T Cocks, R Jenssen, A Stewart, T Wilson, and T Shields. Wavelet-aided svm tool for impulse fault identification in transformers. *Proc. of the 1st EARSeL Workshop on Imaging Spectrometry*, pages 37–42, 1998.

[6] G Camps-Valls and L Bruzzone. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43:1351–1362, 2005.

[7] Wesam.A.Sakla and Andrew Chan. A support vector data description scheme for hyperspectral target detection using first-order markov modeling. In *Automatic Target Recognition*, Apr 2010.

[8] Wesam.A.Sakla and Andrew Chan. A decision level fusion scheme using the support vector data description for target detection in hyperspectral imagery. In *Automatic Target Recognition*, Apr 2010.

[9] F.B Tuteur. Wavelet transformations in signal detection. In *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, volume 3, pages 1435 –1438, Apr 1988.

[10] Li Zhang, Weida Zhou, and Licheng Jiao. Wavelet support vector machine. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 34(1):34–39, Feb. 2004.

[11] Arvind Tolambiya and Prem K. Kalra. WSVM with Morlet wavelet kernel for image compression. In *2007 IEEE International Conference On System Of Systems Engineering, Vols 1 And 2*, pages 558–562, 2007. IEEE International Conference on System of Systems Engineering, San Antonio, TX, Apr 16-18, 2007.

[12] C. Koley, P.Purkait, and S. Chakravorti. Wavelet-aided svm tool for impulse fault identification in transformers. *Power Delivery, IEEE Transactions on*, 21(3):1283–1290, July 2006.

[13] M Lucas, A Gaufriau, S Pascual, C Doncarli, and D Farina. Multi-channel surface emg classification using support vector machines and signal-based wavelet optimization. *Biomedical Signal Processing and Control*, 3(2):169 – 174, 2008.

[14] Richard O.Duda, Peter.E.Hart, and David.G.Stork. *Pattern Classification*, volume 1. Wiley Inter-science, 2nd edition, 2000.

[15] M. Hanson and B. Mond. Necessary and sufficient conditions in constrained optimization. *Mathematical Programming*, 37:51–58, 1987.

[16] David M. J. Tax and Robert P. W. Duin. Support vector data description. *Mach. Learn.*, 54:45–66, Jan. 2004.

[17] M. Pontil and A. Verri. Support vector machines for 3d object recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(6):637 –646, Jun 1998.

[18] E. Osuna, R. Freund, and F. Girosit. Training support vector machines: an application to face detection. In *Computer Vision and Pattern Recognition, 1997. Proc., 1997 IEEE Computer Society Conference on*, pages 130 –136, Jun 1997.

[19] M. Thulasidas, Cuntai Guan, and Jiankang Wu. Robust classification of eeg signal for brain-computer interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 14(1):24 –29, March 2006.

[20] B. Scholkopf, Kah-Kay Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik. Comparing support vector machines with gaussian kernels to radial basis function classifiers. *Signal Processing, IEEE Transactions on*, 45(11):2758 –2765, Nov 1997.

[21] Giorgio Giacinto, Roberto Perdisci, Mauro del Rio, and Fabio Roli. Intrusion detection in computer networks by a modular ensemble of one-class classifiers. *Information Fusion*, 9(1):69–82, 2005.

[22] David M.J. Tax and Robert P.W. Duin. Data description in subspaces. *Pattern Recognition, International Conference on*, 2:2672, 2000.

[23] Kiyoung Lee, Dae-Won Kim, DoKyun Na, Kwang. H Lee, and Doheon Lee. Plpd: reliable protein localization prediction from imbalanced and overlapped datasets. *Nucleic Acids Research*, 34(17):4655–4666, 2006.

[24] F. Yger and A. Rakotomamonjy. Wavelet kernel learning. *Pattern Recogn.*, 44:2614–2629, Oct. 2011.

[25] Huiyan Jiang, Xiangying Liu, Lingbo Zhou, Hiroshi Fujita, and Xiangrong Zhou. Morlet-rbf svm model for medical images classification. In *Proc. of the 8th International Conference on Advances in Neural Networks - Volume Part II*, ISNN'11, pages 121–129, 2011.

[26] Alain Rakotomamonjy, Francis Bach, Stéphane Canu, and Yves Grandvalet. Simplemkl. *Journal of Machine Learning Research*, 9, 11 2008.