

Research Proposal
**Target Detection in Hyperspectral Mineral Data using Wavelet
Analysis**

Student: Michael Mitchley
Supervisor: Professor Michael Sears

September 4, 2008

Abstract

In this document, research is proposed to develop and implement a multistage method for the detection of target spectral signatures in hyperspectral mineral data obtained using the Hyperspectral Core Imager device. The multistage method relies on wavelet analysis, a powerful and relatively new mathematical field used to decompose data into components of varying scales and positions. Such a method would provide the means for detecting minerals indicative of valuable natural resources such as platinum-bearing ore.

Declaration

I, Michael Mitchley, hereby declare the contents of this research proposal to be my own work. This proposal is submitted for the degree of Bachelor of Science with Honours at the University of the Witwatersrand. This work has not been submitted to any other university, or for any other degree.

Acknowledgements

I would like to thank Professor Michael Sears for presenting this problem at the MISG and allowing me to continue working on it as a research topic, as well as for his advice and guidance in the creation of this document. I would also like to thank Professor Steve Damelim for his help and advice in narrowing the research topic. Furthermore, I would like to thank AngloGold-Ashanti for the HCI data obtained thus far.

Contents

Abstract	i
Declaration	ii
Acknowledgements	iii
1 Introduction	1
2 Background	3
2.1 Introduction	3
2.2 Hyperspectral Data	3
2.2.1 Background	3
2.2.2 Traditional Usage	4
2.2.3 Problem Usage and Challenges	5
2.3 Wavelets	5
2.3.1 Background	5
2.3.2 Construction	8
2.3.3 Properties of Orthonormal Wavelets	9
2.3.4 Extension schemes	10
2.3.5 Wavelets and Hyperspectral Data	10
2.4 Denoising	11
2.4.1 Background	11
2.4.2 Denoising and Wavelets	11
2.4.3 Threshold Selection	12
2.4.4 Correlation-based Denoising	12
2.5 Feature Reduction	12
2.5.1 Background	12
2.5.2 Wavelet Histogram Variance	13
2.5.3 Wavelet Histogram Signature	13
2.5.4 ROC curves	14
2.6 Target Detection	14
2.6.1 Background	14
2.6.2 Binary Classification by Features	15
2.7 Conclusion	15

3	Research Methodology	16
3.1	Introduction	16
3.2	Research Hypothesis	16
3.3	Hyperspectral Data	17
3.3.1	Motivation	17
3.3.2	Methodology	17
3.4	Wavelet Transformation	17
3.4.1	Motivation	17
3.4.2	Methodology	17
3.4.3	Implementation and Testing	18
3.5	Denoising	18
3.5.1	Motivation	18
3.5.2	Methodology and Implementation	19
3.5.3	Testing	19
3.6	Feature Extraction	20
3.6.1	Motivation	20
3.6.2	Methodology	20
3.6.3	Implementation and Testing	20
3.7	Target Detection	21
3.7.1	Motivation	21
3.7.2	Methodology	21
3.7.3	Implementation	21
3.7.4	Testing	21
3.8	General Implementation	22
3.9	Conclusion	22
4	Research Plan	23
4.1	Introduction	23
4.2	Time Plan	23
4.3	Deliverables	24
4.3.1	Algorithms	24
4.3.2	Programs	25
4.3.3	Data and documentation	25
4.4	Potential Risks and Solutions	25
4.4.1	Denoising	25
4.4.2	Feature Extraction	26
4.4.3	Target Detection	26
4.5	Conclusion	26
5	Conclusion	28

Chapter 1

Introduction

Hyperspectral imagery is a recently developed field of remote sensing that yields far richer data than traditional colour or multispectral imagery. However, this data comes at the price of high dimensionality, causing greatly increased difficulty in classification. Further challenges are encountered when hyperspectral imagery is used on mineral data.

The Hyperspectral Core Imager (HCI) is a device owned by AngloGold-Ashanti used for obtaining hyperspectral scans of mining cores. Cylindrical mining cores are cut in half, and the flat surface scanned by the HCI to obtain a continuous spectrum ranging from visible light to the far infra-red for each pixel. A typical spectrum is shown in figure 1.1. Derivative-based feature extraction is used to reduce the dimension of the spectral vector associated with each pixel, and the reduced feature set is clustered using self-organising memory.

Two problems are then encountered: firstly, the process is slow, and it was found that pixels clustered together often had sufficiently different spectral signatures to be considered different minerals. This implies the need for a system that is both computationally efficient and accurate. The second problem leads to the idea of target detection. At present, it is not possible to automatically discover the mineral group associated with each cluster. It is necessary for a spectral geologist to examine the spectral signatures found in each cluster to identify each mineral.

Thus, a problem of target detection can be defined. Given the hyperspectral data of a core scan, and a target mineral to find within the core (under the assumption that the spectrum of the target mineral is known), is it possible to detect instances of the target in the data?

The chapters of this document are divided up as follows. In chapter 2, the background of the research topics will be presented. These topics will establish the context of the research through an examination of literature in a number of areas. A number of techniques are examined for the transformation and denoising of data using wavelet techniques, with a focus on those techniques that have been shown to work on hyperspectral data. Feature reduction methods that operate on wavelet coefficients, and target detection methods that work on features are presented. Thus, a common thread is kept throughout the chapter: the techniques presented are compatible with each other, and have been shown to work in hyperspectral data applications.

Chapter 3 extends this thread into a multistage research methodology. The context established in the previous chapter and the problem definition of this chapter are used to define a research hypothesis and justify techniques and methods to be used in the proposed research. Furthermore, this chapter defines the scope of the research. The multistage method involves a wavelet transform, denoising and feature extraction methods, and a target detection method.

In chapter 4 a research plan is presented. This plan provides a time frame to the proposed research

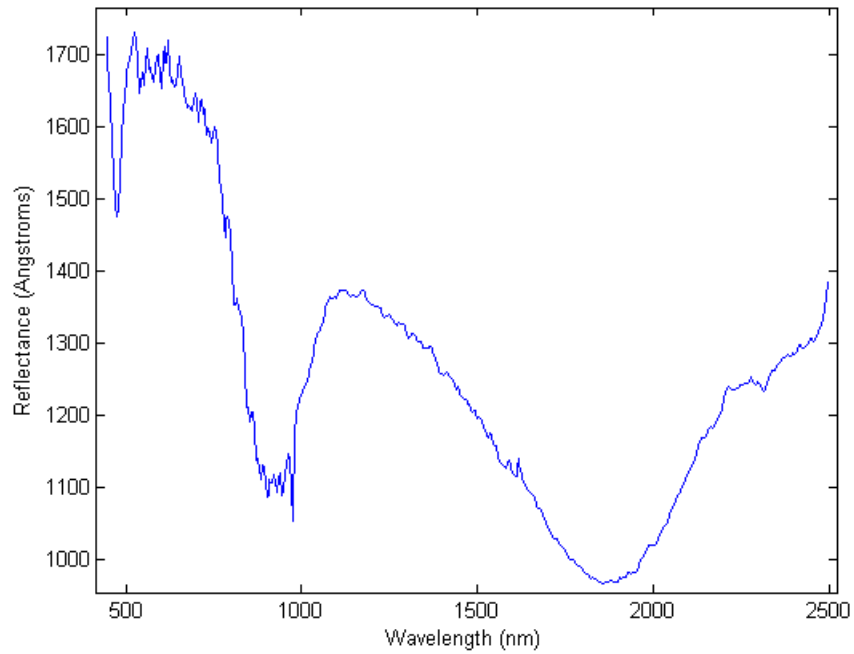


Figure 1.1: A typical spectrum obtained from the HCI, with overlapping sensor values removed

methodology of the previous chapter, and establishes the expected final products of the research in the form of deliverables. The risks involved in the steps of the method are discussed, with solutions. The combination of background, methodology and plan thus establish the context, postulate a hypothesis, propose a method, and provide a comprehensive risk analysis, time plan and expected outcome of the research.

Chapter 2

Background

2.1 Introduction

Before the research methodology is proposed, it is necessary to establish the context of the research. This chapter will focus on each component of the research in turn. In section 2.2, hyperspectral data will be introduced, together with the challenges of this relatively new data type. As the usage of hyperspectral data in this problem is quite different from traditional usage, it is recommended that the reader experienced in the field of hyperspectral imaging read section 2.2.3. However, the remainder of the chapter covers basic theory, and so the experienced reader may wish to skip ahead to chapter 3. In section 2.3, wavelet analysis is presented from the viewpoint of functional analysis. This section provides only a brief introduction into the many fascinating properties of this new field of analysis, with a particular focus on the orthonormal wavelets of Daubechies, and the applications of wavelets to traditional hyperspectral data. In section 2.4 techniques for dealing with imperfect data collection are discussed, with emphasis on denoising techniques using wavelet transforms. Section 2.5 introduces a number of methods of feature extraction for the purposes of dimensionality reduction for efficient classification. Finally, in section 2.6, a number of methods for feature-based target detection are discussed.

Through each section, a common thread will be drawn, linking it to previous sections. Each method presented here can be used in conjunction with methods discussed in other sections of this chapter. Thus, step by step, a multistage method will be suggested. The resulting method will be discussed further in chapter 3.

2.2 Hyperspectral Data

2.2.1 Background

Hyperspectral data are data recorded using hyperspectral sensors. While most digital cameras and multispectral imaging systems measure reflectance using between three and ten bands (from wavelengths corresponding to the colours red, green and blue to the infra-red spectrum), hyperspectral systems measure across hundreds of bands, evenly spaced. Typically, these bands are measured from the visible spectrum through to the far infra-red. By selecting bands that are uniformly spaced, a more continuous sampling is obtained.

Each pixel is a spatial location on the face of the hyperspectral data. This is illustrated in figure 2.1, where pixels at each coordinate (x, y) are associated with a spectral vector extending in the direction of z .

However, this introduces numerous challenges compared to multispectral data. The dimensionality of

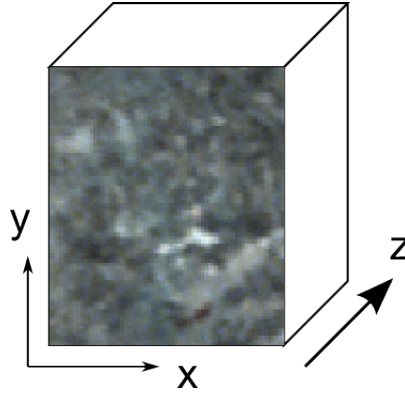


Figure 2.1: A representation of a hyperspectral datacube.

System	Type	Altitude	Spatial resolution	Spectral resolution	Spectral bands
Hyperion	Satellite	705 km	30 m	10 nm	242
AVIRIS	Airborne	20 km	20 m	10 nm	224
ASD Fieldspec 3	Handheld	1.22 m	0.25 m	3.5 nm	384
HCI	Fixed	0 m	0.5 mm	4.18 nm	500

Table 2.1: Typical spatial and spectral resolutions of a number of hyperspectral imaging systems

the data is greatly increased, as each pixel will now be associated with a vector of hundreds of values, corresponding to the spectral reflectance at the series of wavelengths. High dimensional data has caused many problems for traditional target detection and clustering techniques. For a data set of dimension n (in this case the number of spectral bands), it has been found by Kaewpijit *et al.* [2003] and Bruce *et al.* [2002], and characterised further by Serpico *et al.* [2007] that the amount of training data required for automated clustering or target detection increases linearly to exponentially with n , depending on the classifier used. For even a modest hyperspectral system, this implies several thousand training sets in the best case.

The advantages to be gained from using hyperspectral data outweigh these challenges. Some of the usages of hyperspectral data will be explored in the following section.

2.2.2 Traditional Usage

Hyperspectral data are most often used in remote sensing [Kerekes and Schott 2007]. Scenes to be imaged are naturally lit with sunlight, introducing issues of absorption of certain wavelengths of light by the atmosphere. When data are collected from a remote sensing platform, it is costly to obtain and it must often be compressed for transmission [Pickering and Ryan 2006], as transmission from satellite platforms is expensive, and storage on both satellite and airborne platforms may be limited. Hyperspectral data is often noisy. Methods of eliminating noise will be discussed in section 2.4.

Spatial resolution in remote sensing systems is often low. Spatial and spectral resolutions of a number of hyperspectral imaging systems are given in table 2.1, with data from Kerekes and Schott [2007], Bruce *et al.* [2002] and Kruse *et al.* [2003]. Figures are only approximate, and show typical usage. When the spatial resolution of a hyperspectral image is low, it is possible for multiple macroscopic objects to appear in a single pixel, each contributing to the resultant spectrum of that pixel. If we assume there is no interaction between the objects in terms of reflected light, we might assume that the resultant spectrum will be a linear

combination of the spectra of the objects. This linear unmixing model is a natural assumption, and for some applications it allows one to determine the proportion of each pixel that is comprised of each material, assuming the endmembers are known in advance. Such applications correspond to non-intimate mixtures, where the mixing scale is macroscopic.

Hyperspectral data has been used to detect the presence of pitted morning glory in soy bean crops [Bruce *et al.* 2002], [Koger *et al.* 2003], and detect ice on Mars [Schmidt *et al.* 2007]. It has also been used to classify crop types [Kaewpijit *et al.* 2003], [Jiang *et al.* 2004]. Hyperspectral data has been used to detect subtle changes, such as vegetation undergoing nitrogen stress, using spectral signatures [Kempeneers *et al.* 2004]. Another use is the detection and classification of ground-level minerals using satellite or airborne data [Kruse *et al.* 2003].

2.2.3 Problem Usage and Challenges

In this project, hyperspectral data collected by the Hyperspectral Core Imager (*HCI*) will be analysed. Before discussing the method, it is important to highlight the differences between this application and more traditional remote sensing applications. Firstly, while the area to be investigated is smaller (a mere 15cm by 2m), the spatial resolution of the imager is very high, with each pixel measuring 0.5mm to a side. The implication of this is a much higher spatial dimensionality, where the spatial dimensionality refers to the number of pixels in the image. This leads to more data than is traditionally dealt with. Processing the data thus becomes costly.

Secondly, due to the minerals being mixed on a microscopic scale, linear unmixing would be difficult (as discussed in Nascimento and Dias [2007]). The intimacy of the mixture, and its crystalline nature causes nonlinear mixing effects, shifting spectra. This implies that unmixing the hyperspectral data into component minerals and their proportions would be a nontrivial problem, even if the minerals to be found in the core are known in advance, and reference spectra are available as templates. For this reason, it will not be considered in this project.

Thirdly, the scene is not naturally illuminated. This means that atmospheric absorption of bands does not take place in any meaningful way, and the incidence angle is fixed. This simplifies preprocessing, shifting the focus towards obtaining meaningful information from the data, rather than manipulating the data to account for physical phenomena.

Fourthly, transmission and storage are not considered costly in this application. Processing costs dominate due to the high spatial and spectral dimensionality. We will thus look at compression from the viewpoint of reducing processing costs by reducing dimensionality rather than the viewpoint of Pickering and Ryan [2006] which is based on efficient transmission and storage. Obtaining the hyperspectral data is also inexpensive.

Hyperspectral data is used in this application as minerals can only be distinguished by features found in the infra-red spectrum. Due to shifting, these features might not be found at precise, fixed wavelengths. A continuous view of the spectrum is thus needed.

2.3 Wavelets

2.3.1 Background

Wavelet theory has generated a lot of interest in recent years. Wavelet decomposition separates a function or signal into responses localised in time and frequency (or in the case of hyperspectral data, localised in spectral band rather than time). Fourier analysis only gives frequency information. This allows wavelets

to better approximate nonperiodic data, where features may be highly localised [Pereyra and Mohlenkamp 2004]. Wavelets, in essence, are functions that obey certain properties, which are obtained from a mother wavelet by means of scales and shifts. If the mother wavelet is given by $\psi(x)$, the scaled and shifted daughter wavelet $\psi_{j,k}(x)$ for dyadic wavelets is given by

$$\psi_{j,k}(x) = 2^{0.5j} \psi(2^j x - k) \quad (2.1)$$

Some of the properties they may have are orthogonality, compact support or smoothness. Orthogonality means that the inner product of any two wavelets within a specific family will be zero if those wavelets have dissimilar scale and shift coefficients. If a wavelet is also orthonormal, the inner product of a wavelet with itself will be one. Wavelets with a compact support are zero everywhere outside of a region known as the support. If a support is compact, it includes its end-points and is bounded above and below. Thus, it is of a finite length. Smoothness refers to the number of times a wavelet can be differentiated with respect to its argument.

They are broadly split into two categories; continuous wavelets and discrete wavelets. Continuous wavelets have continuous shift and scale coefficients, while discrete wavelets have discrete shift and scale coefficients. While continuous wavelets have many interesting properties, we will focus on discrete wavelets only, as these can be implemented as a fast transform.

Another important function related to the mother wavelet is the scaling function. A scaling function $\phi(x)$ is related to the mother wavelet, and together they define a wavelet family. For this reason, the scaling function is occasionally referred to as the father wavelet [Aboufadel and Schlicker 1999]. The relationship between the wavelet and scaling function will be explored further in section 2.3.2.

To illustrate how wavelets are localised, we will consider the Haar wavelet family, given by

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

$$\psi(x) = \begin{cases} 1 & \text{if } 0 < x \leq 0.5 \\ -1 & \text{if } 0.5 < x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

with daughter wavelets defined as in equation 2.1. This wavelet was the first wavelet to be discovered, and is conceptually the simplest wavelet. The mother wavelet $\psi(x)$, scaling function $\phi(x)$, and the first and second generation wavelet daughters are shown in figure 2.2.

It can be seen that the amplitude of the wavelet grows as the support shrinks. These properties will be discussed further in section 2.3.3. Each wavelet is orthogonal to every other wavelet across all generations. Wavelets with a small support correspond to high frequencies, localised by the shift parameter.

We now discuss a result from approximation theory that will demonstrate how the coefficients of a wavelet transform may be found. The following is taken from Burden and Faires [2001]. Consider a function $f(x)$ and family of functions $\phi_j(x)$, with $j = 0..∞$, orthogonal for all j with respect to a weight function $w(x)$ on an interval $[a, b]$. Then, the least-squares approximation to f on the interval $[a, b]$ using the family ϕ is given by

$$P_N(x) = \sum_{i=0}^N a_i \phi_i(x) \quad (2.4)$$

where N is the order of approximation, and a_i are the coefficients of the approximation. These coefficients can be found through simple algebraic manipulation, with the result of such manipulation given in equation

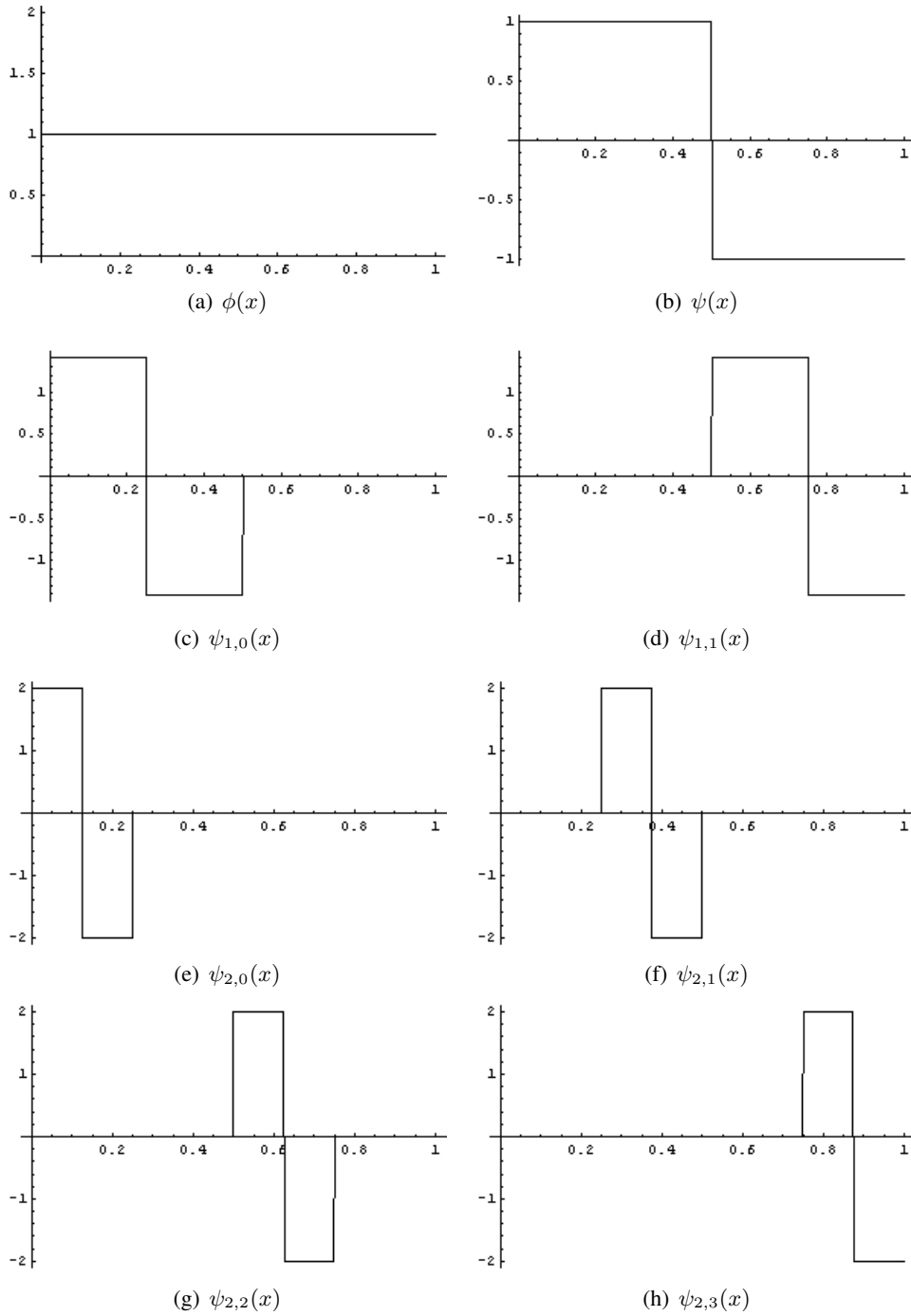


Figure 2.2: The Haar mother and father wavelets, with the first and second generation daughter wavelets.

2.5.

$$a_i = \frac{\int_a^b (f(x)\phi_i(x)w(x)dx)}{\int_a^b (\phi_i(x)\phi_i(x)w(x)dx)} \quad (2.5)$$

Since P_N is a least-squares approximation, it minimises $\|f(x) - P_N(x)\|_2^2$, where $\|\cdot\|_2$ is the $L_2[a, b]$ norm.

We now consider wavelets as our family of orthonormal functions. The results in this portion are taken from Aboufadel and Schlicker [1999]. The weight function $w(x)$ for wavelets is simply 1, and is defined on the compact support of the wavelet (or over all the real numbers, if no such support exists). Suppose this compact support exists, and is the interval $[a, b]$. Clearly, we could then obtain a similar expression to equation 2.4 for a projection onto a wavelet basis that minimises the square of the error. We thus obtain

$$P_N(x) = a_0\phi(x) + \sum_{j=0}^N \sum_{k=0}^{2^j-1} a_{j,k}\psi_{j,k}(x) \quad (2.6)$$

where $\phi(x)$ is the scaling function and $\psi(x)$, $\psi_{j,k}(x)$ the wavelet mother and daughter functions defined in the usual way. Note the bounds on the shift variable k . As before, we obtain an expression for the coefficients as follows

$$a_{j,k} = \int_a^b f(x)\psi_{j,k}(x)dx$$

$$a_0 = \int_a^b f(x)\phi(x)dx \quad (2.7)$$

Thus, we have demonstrated a method for finding the coefficients of a wavelet approximation $P_N(x)$ to a function f that minimises the square of the error of approximation. We now turn to the construction of wavelet families.

2.3.2 Construction

The construction of wavelets usually relies on multiresolution [DeVore and Lucier 1997]. Multiresolution starts with a function $\phi(x) \in L_2(\mathbb{R})$ that has the desired properties of our wavelet basis. This function will obey the refinement equation

$$\phi(x) = \sum_{j \in \mathbb{Z}} a_j \phi(2x - j) \quad (2.8)$$

with a_j refinement coefficients provided ϕ satisfies certain properties. If $\phi(x)$ is orthonormal to its integer shifts $\phi(x - j)$, $j \in \mathbb{Z}$ then the wavelet related to the scaling function is given by

$$\psi(x) = \sum_{j \in \mathbb{Z}} (-1)^j \overline{a_{1-j}} \phi(2x - j) \quad (2.9)$$

where a_j are the refinement coefficients of equation 2.8. Daubechies [1992] demonstrated that orthogonal wavelets of compact support could be generated in this way provided an orthonormal scaling function could be found with compact support. However, these functions do not have a closed-form. Instead, the refinement coefficients are used to define filters.

As a vast number of wavelet families have been defined, with many properties to choose from, it is recommended by Pereyra and Mohlenkamp [2004] that a wavelet from the literature is used, rather than constructing a new one. To this end, we will forego further details and rigorous analysis of the construction of wavelets, and discuss the properties of orthonormal wavelets, with a focus on the orthonormal wavelets with compact support of Daubechies.

2.3.3 Properties of Orthonormal Wavelets

Mathematically, orthonormal wavelets have the following properties. Let f denote an arbitrary function in L_2 , and let $\sum_{j,k} a_{j,k} \psi_{j,k}(x)$ be its wavelet transform.

$$\|f(x)\|_2^2 = \int_{-\infty}^{\infty} f^2 dx = \sum_{j,k} (a_{j,k})^2 \quad (2.10)$$

$$\int_{-\infty}^{\infty} \psi_{j,k}(x) \psi_{n,m}(x) dx = \delta_j^n \delta_k^m \quad (2.11)$$

where δ is the Kronecker delta function defined by

$$\delta_j^k = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases} \quad (2.12)$$

Equation 2.10 shows that the wavelet transform preserves energy. An important consequence of this equation is that distance measures can be performed directly on wavelet coefficients, rather than applying an inverse transform. This property will be used in clustering. Equation 2.11 demonstrates the orthonormality property. A wavelet is orthogonal to any other wavelet with scale or shift not identical to its own. In addition, wavelets have an L_2 norm of one.

If the wavelet has a compact support, we can consider integration over that support only. Clearly, we can only apply compactly supported wavelet transforms to functions with the same compact support. An important class of wavelets are those constructed by Daubechies [1992], which are orthonormal, smooth (although not infinitely smooth), and have a compact support.

The Daubechies wavelets are denoted D_{2N} , where $N \in \mathbb{N}$ is a parameter that can be varied to obtain different wavelet families. These wavelets have N vanishing moments, a filter of length $2N$, and a compact support of $2N - 1$. Equation 2.13 shows the mathematical meaning of a function $\psi(x)$ having M vanishing moments. Note that this implies such a function is orthogonal to all polynomials of order $M - 1$ or less. This gives some insight into why wavelets approximate non-periodic signals well [Pereyra and Mohlenkamp 2004].

$$\int_{-\infty}^{\infty} \psi(x) x^m dx, \quad m = 0 \dots M - 1 \quad (2.13)$$

Daubechies wavelets are also differentiable. If a Daubechies wavelet is in C^k (that is, differentiable k times) then that wavelet has at least k vanishing moments. However, the converse is not true. D_6 has 6 vanishing moments but is only in C^1 . For any finite k , a Daubechies wavelet can be found in C^k . However, as the support length of the Daubechies wavelets grows with the number of vanishing moments, there is no orthonormal wavelet of compact support that is in C^∞ . That is, no Daubechies wavelet is infinitely smooth.

An important property of the Daubechies wavelets is that they can be implemented in a fast transform. This transform is known as the discrete wavelet transform (DWT) and is in $O(n)$ for a full decomposition, making it faster than the fast Fourier transform (FFT). It relies on implementing the wavelet transform as a filter bank in a pyramid structure.

A filter bank consists of a high pass filter (H) and a low pass filter (L). The high pass filter will give the wavelet coefficients at the highest frequency for that signal length by a convolution of the signal. The low pass filter will extract the low frequency information that corresponds to the remaining detail of the signal at that level. By recursively applying both filters to the output of the low pass filter, we can fully decompose

a signal into wavelet coefficients. The proof of complexity is straightforward. The convolution of a signal of length n using the wavelet filter is linear. At each stage, the signal is downsampled by two, halving its length. The complexity in total is then

$$O\left(n \sum_{i=0}^{\log_2 n} \frac{1}{2^i}\right) = O(n) \quad (2.14)$$

2.3.4 Extension schemes

When dyadic wavelets are used to transform a signal, it is necessary for the signal to be of length 2^m , with $m \in \mathbb{Z}$. If a signal is not the right length, we must extend it artificially, as trimming it to the right length would result in an unquantifiable loss of information. Three methods of signal extension are reviewed by Kaewpijit *et al.* [2003]. Consider a signal of dimension $n < 2^m$, $n, m \in \mathbb{Z}$. We wish to extend this signal to

$$\tilde{n} = 2^{\lceil \log_2 n \rceil}$$

The easiest, and most conceptually attractive method of doing this would be to pad the signal with zeros. This method is also used by Shafri and Mather [2005]. However, this results in a 'discontinuity' in the signal, in the sense that we expect hyperspectral data to be smooth. This introduces pseudo-Gibbs phenomena, manifested as ripples around the ends of the signal where the padding took place.

The second method is to pretend the signal is periodic, and repeat as much of it as is necessary to reach the required length. This, however, introduces the same discontinuity problem as zero-padding. Kaewpijit *et al.* [2003] demonstrates that this method gives accuracy comparable to zero padding. The third method is to mirror part of the signal until the required length is reached. This method avoids introducing discontinuities, and consistently gives the lowest classification error of the three methods.

2.3.5 Wavelets and Hyperspectral Data

Wavelets have been used to compress hyperspectral data for transmission by both Fowler and Rucker [2007] and Tang and Pearlman [2006]. Similar approaches were used: the hyperspectral data was subjected to a three dimensional lossless wavelet transform, and encoding techniques were used to compress the resultant data. Wavelets were investigated in Bruce *et al.* [2001], using both continuous and discrete wavelets to detect subpixel targets in hyperspectral data using a variety of features. Features will be discussed further in section 2.5. The techniques that were found to be most effective were then applied to target detection in vegetation data gathered using an ASD handheld spectrometer (the details of which are given in table 2.1) by Bruce *et al.* [2002] and Koger *et al.* [2003]. Kempeneers *et al.* [2004] investigated another target detection problem, using wavelet-based features to differentiate between subtly different vegetation spectral signatures. Wavelets have been used in noise reduction (which will be discussed in more detail in section 2.4) of hyperspectral data in Shafri and Mather [2005]. Dimension reduction using wavelet transforms on hyperspectral data has been investigated by Kaewpijit *et al.* [2003] and Schmidt *et al.* [2007]. Kaewpijit *et al.* [2003] followed a technique of discarding high-frequency information, reasoning that the remaining information best approximates the data in a least-squares sense. Schmidt *et al.* [2007] followed a different approach, discarding low-frequency information, reasoning that the high frequency coefficients of the wavelet transform contain the discriminating features.

2.4 Denoising

2.4.1 Background

Denoising is the process of recovering a signal from an input that is assumed to be noisy. This noise is most commonly thought of as Gaussian white noise, but other types of noise also exist. Noise can be introduced at a number of points in the collection of hyperspectral data. The method of introduction, and the type of noise introduced will be discussed in this section, using material from Young *et al.* [1998].

Photonic noise forms the limiting case for noise removal. As photons are a statistical phenomenon governed by quantum physics, it is impossible to predict the exact number of photons that will be recorded in a scene, no matter how tightly controlled that scene is. Photonic noise is not Gaussian, is dependant on the signal and is not additive. Regions of low reflectance can be dominated by photon noise.

Noise is introduced by the scanning equipment at several stages. Thermal noise (or dark current) is caused by thermal energy residing in the sensors. This is especially problematic for infra-red sensors (which are more sensitive to thermal energy), and is usually combated by cooling the sensor. Other sources of noise stem from the electronics: namely, the measurement, sampling, amplification and quantisation of signals. This noise is Gaussian and additive, and is independent of the signal.

Noise introduced by malfunctioning sensors manifests as bands where the values are either zero (for dead sensors) or at a maximum (for malfunctioning sensors). This type of noise is independent of the signal, and usually occurs uniformly over all the pixels. Correcting for this noise is usually part of the preprocessing of the data, where the average of the neighbouring sensors will be substituted for the erroneous value.

2.4.2 Denoising and Wavelets

Noise resides in higher frequencies [Pereyra and Mohlenkamp 2004], and so a simple way to denoise a signal is to discard the high-frequency information of the wavelet transform. This is the approach applied to hyperspectral data by Kaewpijit *et al.* [2003] (although the principle aim in that paper is to reduce dimensionality). However, this naive approach has drawbacks. By discarding all high frequency information, we lose not only noise, but also high frequency features. For the purposes of denoising, a more sophisticated technique should be used. Denoising can also be performed using wavelet coefficients via a process called thresholding. Two types of thresholds are commonly used. In a hard thresholding scheme, all wavelet coefficients $a_{j,k}$ smaller than some threshold t are reduced to zero. That is, the denoised coefficients $\hat{a}_{j,k}$ can be defined as

$$\hat{a}_{j,k} = \begin{cases} a_{j,k} & \text{if } |a_{j,k}| \geq t \\ 0 & \text{otherwise} \end{cases} \quad (2.15)$$

Another scheme is soft thresholding, where all wavelet coefficients are attenuated by the threshold t . Thus, as above, the denoised coefficients $\hat{a}_{j,k}$ can be defined as

$$\hat{a}_{j,k} = \begin{cases} a_{j,k} - t & \text{if } |a_{j,k}| \geq t \\ 0 & \text{otherwise} \end{cases} \quad (2.16)$$

Empirical results of Shafri and Mather [2005] indicate that hard thresholding is more accurate than soft thresholding for certain applications. This agrees with Pan *et al.* [1999]. However, the results of Shafri and Mather [2005] exhibited a ripple effect. Shafri and Mather [2005] argues that this is due to the discontinuities introduced by the zero padding, while Pan *et al.* [1999] suggest such artefacts are caused by the shift variance of the orthogonal wavelet transform.

2.4.3 Threshold Selection

Choosing a threshold t automatically is far from trivial. A commonly used threshold is the Universal Threshold (UT) introduced by Donoho and Johnstone [1995] which is based off Stein's unbiased risk estimate (SURE). In theory when then is used with a soft threshold, it minimises the error introduced by noise. However, researchers have found that hard thresholding gives better results in practical applications [Shafri and Mather 2005], [Pan *et al.* 1999]. This threshold is given by

$$t = \sqrt{2 \log N} \sigma \quad (2.17)$$

where σ is the noise deviation and n the length of the signal. Shafri and Mather [2005] estimated this noise variance as a weighted median of absolute deviations (MAD), but found that the resulting threshold severely underestimated the noise. Other thresholding methods are similar to the UT, but introduce different estimations of the noise deviation, and different weights. It was suggested by Pickering and Ryan [2006] that error estimation using mean absolute deviation is inflexible, and as such percentage mean absolute deviation (PMAD) was suggested as a replacement. A survey of techniques was undertaken by Kaur *et al.* [2002], although with a focus on two-dimensional image thresholding.

2.4.4 Correlation-based Denoising

A more sophisticated technique could use the correlation between scales in a wavelet transform. That is, significant features are expected to be represented by a number of different coefficients, while noise will have an effect on high-frequency coefficients only. This insight forms the basis of the method of Xu *et al.* [1994], which is known as spatially selective noise filtration (SSNF). Pan *et al.* [1999] expands on the SSNF, demonstrating a technique to choose a noise power reference σ by using the two wavelet scales of highest frequency. The SSNF technique is outlined briefly: Let $Corr_2(j, k)$ be the correlation between wavelet scales at position k given by

$$Corr_2(j, k) = \prod_{i=0}^{l-1} a_{j+i, k} \quad (2.18)$$

where l is a parameter (chosen to be 2 by Pan *et al.* [1999]). We use this correlation between scales to characterise wavelet coefficients as being produced by noise, or features. Only those coefficients that are produced by features are accepted into a new set of coefficients. The process stops once the wavelet coefficients left are nearly equal to σ .

2.5 Feature Reduction

2.5.1 Background

Features are distinguishing elements in data comparison and analysis. A feature might be a sharp change in a signal, the variance in a histogram, the coefficients of a Fourier transform or the data itself. Feature extraction is the problem of finding these features (which is clearly trivial if the features are the data elements, and potentially complex if the data must be transformed), although feature reduction is often performed as part of feature extraction. Feature reduction is the problem of selecting a subset of the 'best' features from the entire feature set. Often a cost function is used to determine the best features. In general classification problems, we wish to select features that maximise interclass variance, and minimise intraclass variance. In

the problem of target detection, there are only two classes, so the 'best' features are those that most readily distinguish the target.

The principle problem of hyperspectral data is its high spectral and spatial dimensionality. The wavelet transform discussed in section 2.3 will in fact increase the dimensionality. To reduce the dimensions of the data, we must extract features, and select the best ones. In this application we wish to concentrate on feature reduction methods that use the wavelet transform coefficients (or data obtained from the coefficients) as features. This will keep the operations in the coefficient domain, reducing computational expense. Feature reduction is also important outside of wavelet analysis, and was suggested as a data compaction procedure by Ghassemian and Landgrebe [1988] for the transmission of data from satellite platforms. Three different feature extraction methods that operate on wavelet coefficients will be discussed.

2.5.2 Wavelet Histogram Variance

This method was introduced in Bruce *et al.* [2001] as wavelet energy levels, and elaborated upon by Kempeneers *et al.* [2004] as the variance of the wavelet coefficient histogram for each scale. It works as follows: for each scale j , the energy feature E_j is computed as the root mean square of the wavelet coefficients $a_{j,k}$ at scale j , as shown in equation 2.19.

$$E_j = \sqrt{\frac{1}{2^j} \sum_{k=0}^{2^j-1} (a_{j,k})^2} \quad (2.19)$$

The original n wavelet coefficients are reduced to $\log_2 n$ features, one for each scale. From equation 2.10, this can be seen as the mean norm of the signal at each frequency level.

2.5.3 Wavelet Histogram Signature

Kempeneers *et al.* [2004] introduces a refinement on the histogram variance method. Coefficient histograms of natural images can be modelled by

$$h(u) = K e^{-\left(\frac{|u|}{\alpha}\right)^\beta} \quad (2.20)$$

where K is a normalisation constant given by

$$K = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} \quad (2.21)$$

and α, β estimated using

$$m_1 = \int |u| h(u) du \quad (2.22)$$

$$m_2 = \int |u|^2 h(u) du$$

$$\alpha = m_1 \frac{\Gamma(1/\beta)}{2/\beta} \quad (2.23)$$

$$\beta = F^{-1} \left(\frac{m_1^2}{m_2} \right) \quad (2.24)$$

where functions F and Γ are defined by

$$F(x) = \frac{\Gamma^2(2/x)}{\Gamma(3/x)\Gamma(1/x)}$$

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$$

The original n wavelet coefficients are reduced to $2 \log_2 n$ features. The variance of the histogram is modelled by α , while β describes the rate of decrease of the histogram for each wavelet coefficient scale.

2.5.4 ROC curves

This method of feature reduction was introduced in the context of hyperspectral data and wavelet transforms by Bruce *et al.* [2001], and further researched in Bruce *et al.* [2002] and Koger *et al.* [2003]. Receiver operating characteristic (ROC) curves are used in binary classification problems as a performance measure, and the area under the ROC curve can be used as a measure of separability between classes. Let Se be sensitivity and Sp specificity. Let FP , FN , TP and TN stand for false positives, false negatives, true positives and true negatives respectively. Then sensitivity and specificity can be quantified as

$$Se = \frac{TP}{TP + FN} \quad (2.25)$$

$$Sp = \frac{TN}{TN + FP} \quad (2.26)$$

Sensitivity and specificity together provide a more detailed view of classification accuracy. A ROC curve is obtained by varying a classification parameter (normally the classification sensitivity) and plotting the sensitivity against (1 - specificity). In the context of target detection and feature extraction, we use the area under the ROC curve (AUC) to characterise the contribution of each wavelet coefficient. Coefficients that contribute the most in terms of separability between classes are used as the reduced feature set, provided adding them to the reduced feature set increases the AUC of the set. A problem with this method is that it requires well-defined classes. In the case of target detection of hyperspectral mineral data, the target is well known, but the second class ('not the target') is poorly defined, as it consists of any possible spectral signature that does not match that of the target, for some suitable measure. A possible method for extending ROC curves to work with target detection without binary classes would be to use a training set of spectral signatures (or rather, the wavelet transforms of spectral signatures), and find the AUC of each wavelet coefficient with respect to each member of the training set. It would then be necessary to find the subset of wavelet coefficients that best differentiate the target from the members of the training set.

2.6 Target Detection

2.6.1 Background

Target detection in hyperspectral data is the problem of finding the set of pixels that match the specified target. In keeping with the nomenclature of the previous section, a distance metric is defined on the reduced feature set. If the features of the pixel being considered are sufficiently close to the features obtained from the target spectral signature, that pixel is considered to match the target. We will consider binary rather than

quantitative target detection: that is, either the target is present in a pixel at some predefined level, or it is not.

Target detection techniques were the focus of Bruce *et al.* [2001], Bruce *et al.* [2002], Koger *et al.* [2003], Chang and Ren [2000] and Kempeneers *et al.* [2004]. However, within the context of the HCI, target detection remains an open problem. The classification methods of Bruce *et al.* [2002] and Kempeneers *et al.* [2004] are focussed on in the following subsection.

2.6.2 Binary Classification by Features

Clustering algorithms are used to partition a data set into discrete (or potentially fuzzy), exhaustive clusters, where the variance within each cluster is minimised, and the variance between clusters is maximised. The simplest possible classifier is one based on a Euclidean distance metric, where if the distance between the features of the pixel and the features of the target is below some predefined threshold, that pixel matches the target. However, this requires a priori knowledge of the tolerance threshold. Another unsupervised classifier is k-means. This could be used as a binary classifier if $k = 2$. However, we would not be able to guarantee the selection of the target as a centroid (consider the case where the target is not present in the data set).

Kempeneers *et al.* [2004] demonstrated that the wavelet histogram signature feature set could be classified using a kNN classifier. This is a supervised classifier that requires training data. However, it permits a Euclidean distance metric. Briefly, the kNN system is given a training set of classified features. When given an unclassified feature set A , it will find the k nearest neighbours to A within the training set. A will be placed in the same class as the majority of its nearest neighbours. The parameter k must be chosen with care to prevent ties. For binary classification, any odd k will force a majority.

Another trained classifier is the minimum distance (MD) classifier. A mean feature vector is kept for each class. An unknown pixel is assigned to the class with the minimum distance between its feature vector and the mean vector. This classifier is used by Koger *et al.* [2003] and Kaewpijit *et al.* [2003].

2.7 Conclusion

In this chapter, the context of the research was established. This was broken up into several sections. In section 2.2, hyperspectral data was discussed. The differences between the traditional usage of hyperspectral data in remote sensing, and the usage in the HCI were highlighted. Primarily, HCI data require less preprocessing, but the dimensionality is higher, and the mixing is nonlinear. In section 2.3, wavelet theory was briefly introduced. While dyadic orthonormal wavelets were the focus of this section, it is worthwhile to note many different wavelet types exist. The applications of wavelets to hyperspectral data were covered, showing precedent.

Denosing techniques that operate on wavelet coefficients were presented in section 2.4. In particular, the application of denosing techniques to hyperspectral data was discussed, and some of the dangers highlighted. In section 2.5 feature reduction was discussed as a method of reducing dimensionality by selecting a subset of the available features. This feature subset forms the basis of the target detection technique, as discussed in section 2.6.

The links between each section and the overlapping contexts suggest a multistage method using all of the techniques outlined in this chapter. In the following chapter, this multistage method will be discussed in greater detail as the methodology of the research is defined.

Chapter 3

Research Methodology

3.1 Introduction

The challenges of the hyperspectral data generated by the HCI were described in the introduction and in section 2.2.3. In order to obtain accurate information from the data, these challenges must be met. The desired information is a binary mapping of target membership in the hyperspectral data set. In sections 2.3 to 2.6 numerous techniques were introduced to address the challenges of hyperspectral mineral data.

In this chapter, a research methodology will be outlined. The proposed research is a multistaged process. In order to expand on the research hypothesis given in section 3.2, the process will be divided into its logical stages, with each stage discussed and justified in relation to previous stages. In section 3.3, the use of hyperspectral data in this project will be justified and expanded upon. The hyperspectral data will be subjected to a wavelet transform described in section 3.4. The choice of wavelet basis and a justification for using wavelets at all will be presented. In section 3.5, a denoising method will be presented that uses wavelet coefficients. This links it to the output of the wavelet transformation. The resultant coefficient set will be considered as a feature set, and the method of feature reduction will be described in section 3.6. Binary clustering using the reduced feature set will be described in section 3.7. Finally, in section 3.8 general implementation details will be discussed, with a specific focus on how each method will combine with the others.

3.2 Research Hypothesis

In sections 2.2 to 2.6 a multistage process was suggested, with each stage relying on the output of the previous stage. The following research hypothesis is thus obtained.

The use of a multistage wavelet analysis technique consisting of a wavelet transform, wavelet-domain denoising and feature extraction methods and a classification step will provide an accurate and computationally tractable technique for the detection of target spectral signatures in hyperspectral mineral data.

Having a fast and accurate target detection method will lay the grounds for further research into automated classification and identification of hyperspectral mineral data. More immediately, such a technique could be used to identify minerals indicative of platinum or other rare metals within the core. This would lead to more efficient mining and resource exploitation. The stages of the proposed technique will be outlined in the remainder of this chapter, with motivation and a description of how each step could be tested, where appropriate. The verification of all the methods will thus verify the hypothesis.

3.3 Hyperspectral Data

3.3.1 Motivation

Hyperspectral data are used in this application as it is in the infra-red spectrum that minerals can be distinguished. However, the features that lead to differentiation of minerals may not occur at the expected bands, due to nonlinear mixing effects. Thus, a more continuous view of the spectrum is needed. Hyperspectral data provide such a view.

3.3.2 Methodology

The details of the hyperspectral data obtained from the HCI are given in table 2.1. An example data set has been obtained, courtesy of AngloGold-Ashanti. This data set corresponds to 2m of mining core, and will be used as a basic data set around which to develop the methods to follow. The data sets obtained are all in a format readable by the hyperspectral imaging software suite *ENVI*. *ENVI* will be used to visualise and manipulate the hyperspectral data. It will be assumed that such data has been preprocessed as normal by AngloGold-Ashanti. Further data will be obtained when meetings with AngloGold-Ashanti representatives can be arranged.

3.4 Wavelet Transformation

3.4.1 Motivation

Wavelet transforms characterise signals with functions localised in frequency and band. It has been shown that principle component analysis (PCA) performs poorly for hyperspectral data [Bruce *et al.* 2001], [Bruce *et al.* 2002], [Koger *et al.* 2003], [Kaewpijit *et al.* 2003]. The reason for this is that interclass differences may be very subtle. Principle component analysis is computationally expensive when compared to a wavelet transform, and in this application is not suitable. Other standard signal analysis techniques, such as the fast Fourier transform (FFT) and the discrete cosine transform (DCT) have also been found to perform poorly on hyperspectral data [Bruce *et al.* 2002]. This is due to spectral data having no periodic signals, but rather features that are localised to specific bands.

3.4.2 Methodology

In this project, a wavelet transform will be applied to the data along each spectral vector. No spatial transforms will be applied, in order to keep the transformation strictly based on local information, as the spatial dimensionality is very high. A Daubechies orthonormal wavelet will be used. Daubechies wavelets provide a number of desired properties, such as compact support, some degree of regularity, orthonormality, and the ability to be implemented as a fast transform [Daubechies 1992]. There is dissent in the literature as to which Daubechies wavelet is best for hyperspectral data. Schmidt *et al.* [2007] advises using a wavelet with a filter length to match the spectral sampling (that is, if the spectrum is undersampled, use a shorter filter). Shafri and Mather [2005] advises using a long filter, while Bruce *et al.* [2002] advises the use of the shortest possible filter (which is in fact the Haar wavelet). Both approaches are supported by empirical data.

From the recommendations in literature, a wavelet with a medium-length filter will be used. As can be seen from table 2.1, the spectral resolution of the HCI is high when compared to remote sensing systems. However, since the number of pixels in a hyperspectral data set is potentially very high, the three dimensional

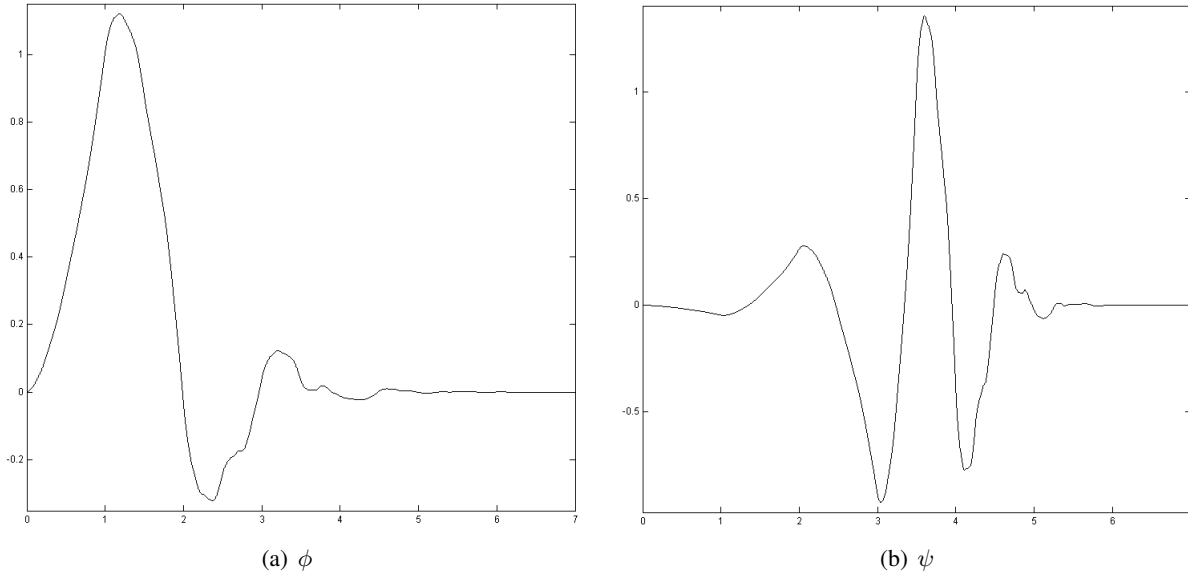


Figure 3.1: The Daubechies D8 scaling function and wavelet

transforms used by Fowler and Rucker [2007] and Tang and Pearlman [2006] will not be considered, as they require transformation of the global data set.

More specifically, the D8 wavelet will be selected. This wavelet has a filter length of 8, and is at least in C^1 (that is, it is at least once-differentiable). As argued above, this wavelet has a medium-length filter. More intuitively and somewhat informally, this wavelet is smoother than the shorter D4 wavelet, and in some sense will provide a 'better fit' to the hyperspectral data, which will aide in the accurate identification of noise. The D8 scaling function and wavelet are illustrated in figure 3.1, using graphs from Matlab's wavelet toolbox.

3.4.3 Implementation and Testing

Fast wavelet transforms are included in Matlab's wavelet toolbox. These built-in optimised functions will be used in conjunction with ENVI to produce the wavelet transform of the hyperspectral data. Following the results of Kaewpijit *et al.* [2003], the extension scheme selected will be signal mirroring. Should the wavelet toolbox not be available, the wavelet transform can be implemented as a recursive filter bank using filter coefficients given by Daubechies [1992]. Testing the output of the transform will be done to ensure correct implementation, although this testing will not be part of the final product. By applying the inverse transform to the wavelet coefficients, and cutting the extended segment of the signal, we should obtain the original signal. This is due to the wavelet transform being a true transform: all information is preserved.

3.5 Denoising

3.5.1 Motivation

As the data gathered by the HCI is noisy, more meaningful information will be obtained from it if the noise level is reduced. Noise is a low-amplitude, high-frequency phenomenon [Pereyra and Mohlenkamp 2004]. Since the wavelet transform is localised in frequency, with higher frequencies having larger amplitudes,

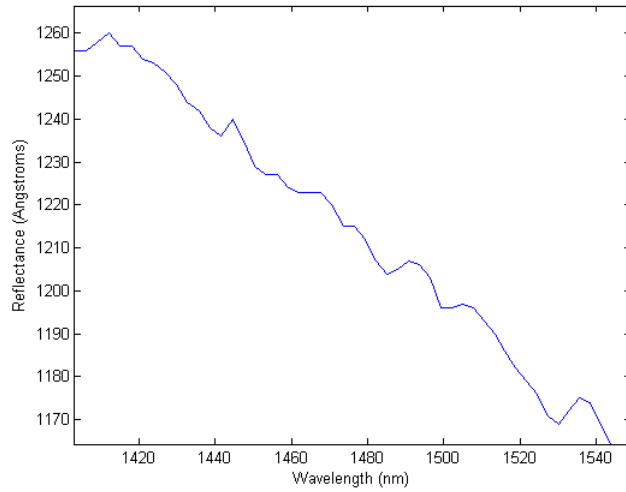


Figure 3.2: Noise in the HCI data

noise is poorly approximated by wavelets. This implies that denoising can be accomplished using wavelet coefficients. Figure 3.2 is a closer view of figure 1.1, which shows variation that is caused by noise in the HCI data, as the spectrum should ideally be smooth.

3.5.2 Methodology and Implementation

Several schemes for the reduction of noise using wavelet coefficients were presented in section 2.4. Most notably, a technique developed by Xu *et al.* [1994] and presented in further detail by Pan *et al.* [1999], called spatially selective noise filtration (SSNF) was briefly described. This scheme will be used as the denoising method. The output of the wavelet transform (an ordered vector of coefficients) will be taken as the input of this method. The output will be an ordered vector of coefficients of the same dimension as the input. It is expected that this output will be the wavelet transform of a 'cleaner' signal. That is, the coefficients corresponding to noise will be zeroed out. The SSNF algorithm will be investigated further, and implemented in Matlab. Asymptotic complexity analysis of the resulting algorithm will form part of the research.

3.5.3 Testing

The denoising method is the first major testing milestone. Artificial data will be used to measure the effect of the denoising, as in the method of Shafri and Mather [2005]. Smooth spectra that represent ideal mineral readings will be developed, and subjected to a 'noising' method that represents the addition of real noise in hyperspectral data, as described in section 2.4. The denoising method will be tested against the known smooth (or 'clean') spectra.

However, while Shafri and Mather [2005] uses an L_1 norm to measure the effect of the denoising, three different metrics will be used in this project. Both an L_1 and an L_2 norm will be used, as well as measuring the signal-to-noise ratio of the denoised signal measured against the known clean signal. The L_2 norm is traditionally used as an error metric as it reduces small errors and exaggerates large errors, but it is stated in DeVore and Lucier [1997] that the L_1 norm forms a better model for human perception of error.

The development of artificial data and noising techniques, and the interpretation of the results of the three metrics will form part of the research.

3.6 Feature Extraction

3.6.1 Motivation

'The curse of dimensionality', described by Bruce *et al.* [2002], is the relation between the amount of training data needed for accurate classification, and the dimensionality of the data set. Depending on the classifier used, this relationship can be anything from linear to exponential [Serpico *et al.* 2007]. The wavelet transform increases dimensionality, by extending the data set to the nearest power of two, and mapping the extended data set to a set of coefficients of the same dimension. Feature reduction is necessary to find a feature set of low dimensionality that in some sense allows for accurate classification of the data. Feature extraction involves finding and then reducing a feature set.

The use of feature extraction or reduction on hyperspectral data has been investigated by Ghassemian and Landgrebe [1988], Bruce *et al.* [2001], Bruce *et al.* [2002], Koger *et al.* [2003], Jiang *et al.* [2004], Schmidt *et al.* [2007] and Serpico *et al.* [2007], and is a necessary step to deal with the high dimensionality of hyperspectral data.

3.6.2 Methodology

In this project, two different feature extraction methods will be implemented. Firstly, the wavelet histogram signature method outlined by Kempeneers *et al.* [2004] will be used. This will entail the development of suitable numerical methods for estimating the features α and β for each scale. Secondly, the extension of the ROC curve method suggested in section 2.5.4 will be implemented. Finding a suitable training set and a method for selecting the best subset of coefficients given the added complication of using multiple classes will be part of the research.

Both methods will be applied to the output of the denoising method. However, only one method will be used in the final product. The two methods will not necessarily be formally compared, but rather a subjective decision will be taken based on the ease of implementation, asymptotic complexity and accuracy of the two methods.

3.6.3 Implementation and Testing

This step of the process will have the most complex implementation. In order to estimate the features α and β for the histogram signature method, a number of numerical issues must be addressed. The wavelet coefficient histogram can be used as an initial approximation to $h(u)$, which will start the method. Integration to find m_1 and m_2 can be replaced with numerical integration, although this will have to be performed carefully as the histogram will not generally be evenly sampled, unless the histogram is itself approximated. Stirling's approximation can be used in place of the Gamma function, but the inversion of the function F to find β is a challenge that will have to be part of the research.

The ROC method will be implemented in Matlab using some suitable training set of spectral signatures that do not match the target. Further implementation details and the development of the ROC algorithm will form part of the research, as a method for combining the coefficients of greatest discrimination across the various training elements will have to be developed.

The correctness of the histogram signature method could be tested by supplying coefficients constructed to have a known histogram signature. The correctness of the ROC method could be tested with carefully constructed training data or smaller examples where the coefficients of maximum discrimination are obvious. However, no tests can be performed on real wavelet data at this stage, since feature extraction accuracy metrics rely on classification.

3.7 Target Detection

3.7.1 Motivation

Clustering is the process of generating a mapping of similar clusters, where by some distance metric the variation within each cluster is minimised, and variance between different clusters is maximised. In binary target detection, we reduce the hyperspectral data to two clusters. Each pixel either matches the target, or it does not.

3.7.2 Methodology

As the focus of this project is on wavelet methods, a simple linear trained classification algorithm will be used, namely kNN. The kNN algorithm was described in section 2.6. The selection of appropriate training data will form part of the research. The final output of the multistage method will be a target detection map that describes the locations of the target in the hyperspectral image.

3.7.3 Implementation

The kNN algorithm will be implemented in Matlab, and will be made to output a target detection map in a form readable by ENVI. The kNN algorithm will keep a library of training data. This training data will have undergone the same wavelet transform and feature extraction technique as the input data, but will have additional information concerning whether it matches the target or not. The Euclidean distance between the feature vector of the pixel and each training feature vector will be found, and the pixel will be assigned to the same class as the majority of the k training elements with the smallest distance. The pixel's feature vector and class will not be added to the training set, as this could lead to increasing inaccuracy as more feature vectors are added.

3.7.4 Testing

The test of the resultant target detection map is the second major testing milestone of the research. We will use both artificial data and clustered data to test the output of this method through a measure suggested by Fowler and Rucker [2007] and Pickering and Ryan [2006], namely preservation of classification. Preservation of classification measures the percentage of pixels that have not changed classes between two classifications. Furthermore, in the case of artificial data, the true locations of the target will be known. For this reason, we can also test the target detection map using the metrics of specificity and sensitivity, as defined in section 2.5.4. The clustering method will be tested against a priori knowledge of whether the spectra match the target or not. The development of this method, as well as test cases, will form part of the research.

As the HCI has been used extensively by AngloGold-Ashanti, it will be possible to obtain processed data where clusters of minerals are known and identified. If a target is defined as the mineral mixture in a

processed cluster, we could expect the multistage method to generate that cluster as a target detection map. In the case of processed data, it will be necessary to interpret the results carefully, as it may not be known which is more accurate.

3.8 General Implementation

All the methods described above will be implemented in Matlab. A wrapper program will be developed to coordinate the flow of data between each method, converting and marshalling data where necessary. Matlab is chosen in this application as it has a large library of built-in optimised mathematical functions. The relation between Matlab's processing and the viewing abilities of ENVI will be explored when the ENVI software is available, with the desired outcome being the use of both programs to process and display the data.

3.9 Conclusion

In this chapter, the methodology of the proposed research was discussed. A multistage method will be developed to verify the hypothesis of section 3.2, consisting of a wavelet transform using the D8 wavelet, a denoising step using the SSNF method, a feature extraction step consisting of either the ROC curve method or the wavelet histogram signature method, and feature-based classification using the kNN classifier. The testing of each method was discussed, and the research scope clearly delineated.

The combination of context and methodology shows where the research draws from, and what it will entail. A plan is also needed, to show how and when the research will be performed, and what is to be expected. This plan will be the focus of the next chapter.

Chapter 4

Research Plan

4.1 Introduction

The research methodology of the previous chapter outlines the scope and method of the proposed research. Further information on the research is given in this chapter, in the form of a time plan, deliverables, and risk analysis. The time plan of the research is presented in table 4.1 and discussed in section 4.2. This time plan accounts for the allocated research hours per week, with the projected order and time division of each task. In section 4.3 the deliverables of the project are defined in three categories, namely algorithms, programs and data & documentation. Finally, in section 4.4, risks associated with each stage of the project and their solutions are given.

4.2 Time Plan

The time plan of the proposed research is presented in table 4.1. It has been broken into six phases. The design phase will be spent designing the two feature extraction methods and artificial data for testing. Much of this time will be spent designing numerical methods for the histogram signature feature extraction method, of which eight hours will be spent on the β parameter alone, as this is the most complex of the two parameters, as can be seen from equation 2.24. Four hours have been allocated to designing the noising method that will be used to create realistically noisy artificial data. Sixteen hours will be spent attempting to design an extension of the ROC curve method to the target detection problem.

The implementation phase will be spent implementing the designed algorithms in Matlab. Implementation is fairly straightforward, and includes four hours allocated to creating a data marshalling program to handle the multistage process. This time will also be spent learning to use the ENVI software. In the testing phase, the implemented algorithms will be tested for correctness. At this stage, it is estimated that a feature extraction method can be chosen from the two methods developed. This phase overlaps somewhat with the computation phase, where artificial and clustered data will be used to test the performance of the combined methods. The computation phase will also be used to run the combined multistage method on real data.

Once results are obtained from the testing and computation phases, they will be interpreted. The interpretation of results will be by a thorough examination of the measurements obtained from the three noise metrics and the three classification metrics. This then leads to the writeup phase, where the final research report and presentation will be created. Finally, the research report will be presented in the week of the 24th of November.

Week	Activity	Hours	Total
08-09	Design wavelet Histogram method	8	8
	Design β estimate	8	16
15-09	Design ROC curve method	16	32
22-09	Design artificial data	12	44
	Design noising method	4	48
29-09	Implement histogram signature method	8	56
	Implement ROC curve method	8	64
06-10	Implement denoising method	4	68
	Implement kNN method	8	76
	Implement marshalling method	4	80
13-10	Test denoising method	5	85
	Test histogram signature method	5	90
	Test ROC curve method	5	95
	Choose feature extraction method	5	100
	Test clustering method	5	105
20-10	Compute with artificial data	10	115
	Compute with clustered data	10	125
	Compute with real data	5	130
27-10	Interpretation of artificial data	15	145
	Interpretation of other data	10	155
10-11	Writeup	40	195
17-11		40	235
24-11	Presentation week	30	265

Table 4.1: Research breakdown per week

4.3 Deliverables

4.3.1 Algorithms

A number of algorithms will be deliverable at the end of the design phase. These algorithms will be analysed for asymptotic complexity. By method, they are as follows:

- Algorithm for creating noisy artificial spectral data
- Wavelet histogram signature feature extraction algorithm
- ROC curve algorithm

The generation of representative spectral data with the noisy characteristics of data obtained from the HCI is critical to an objective measurement of the performance of the denoising method, and the multistage method as a whole. The first deliverable of the project will be an algorithm for creating smooth (ideal) spectral data, and noising it in a realistic manner. Numerical techniques for the estimation of wavelet histogram signature parameters α and β must be designed, with careful consideration for practical and polynomially bounded performance. A method for extending the ROC curve method to deal with multiple

classes (but a single target) will be attempted. If such a method can be created, it will be delivered as an algorithm.

4.3.2 Programs

As the proposed multistage method is made up of several discrete stages, the resulting multistage program will encompass several independent programs. These are given below:

- Denoising program using SSNF
- Feature extraction program using wavelet histogram signatures
- Feature extraction program using ROC curves
- Clustering program using kNN
- Artificial data program for the generation of noisy hyperspectral data
- Marshalling program to unite the above with the wavelet transform of Matlab

With the exception of the marshalling program, these programs are expected to work independently. This will lead to a more robust overall program through modularity.

4.3.3 Data and documentation

Data are deliverable through the creation of artificial test data, as well as the output of the multistage program. In addition to this data, a research report on the proposed research will be submitted. The data and documentation to be delivered are as follows:

- Artificial data
- Target detection map using artificial data
- Target detection map using real data
- Research report

Should the hypothesis of section 3.2 prove valid, the output target detection maps will be of use in identifying mineral clusters in hyperspectral data. The artificial data generated will be of use in other applications concerning hyperspectral mineral data, regardless of the outcome of the research.

4.4 Potential Risks and Solutions

4.4.1 Denoising

A potential risk of the denoising method is the one encountered by Shafri and Mather [2005]. That is, the denoising method could underestimate the noise. While this would not damage the signal, it may render the denoising method useless. A solution to this would be to experiment with the noise power parameter σ using artificial data until a sufficiently high parameter is found that minimises the error in the signal. The reciprocal problem is a denoising method that overestimates the noise. This would be somewhat more

serious, as it may lead to an overly smoothed signal where features are removed. While the SSNF method chosen attempts to preserve features by looking for interscale correlation, it is possible that the denoising will be too aggressive. Again, a careful choice of σ derived using artificial data may solve this problem.

The asymptotic complexity of the SSNF method is as yet unknown. It is possible that the method will not be polynomially bound, or will otherwise be computationally prohibitive. In this case, the simpler $O(n)$ hard thresholding scheme will be implemented, with some suitable parameter σ . Complementary to the denoising method is the creation of realistic noisy artificial data. This noising must encompass all the noise introduction points discussed in section 2.4. Care must be taken not to make any false assumptions regarding the noise.

4.4.2 Feature Extraction

The biggest risks of the project are within the feature extraction method, as it is on this method that the project hangs. Without the feature extraction step, the dimensionality of the data will not be reduced, leading to inaccurate and slow clustering. The two methods that will be considered are wavelet histogram signatures, and ROC curves. The histogram method requires the use of many numerical methods to estimate the parameters. These methods might prove to be inefficient, converging on the correct parameters slowly (or worse, not converging at all). The ROC curve method, however, may prove to be too difficult to design, or impossible under the problem assumptions. Both methods may give poor results, or prove too difficult to implement.

For these reasons, the designs of the feature reduction methods are performed first in the timeplan of table 4.1. Should one method prove impossible, inaccurate or inefficient, the other method will be focussed on. If both methods fail to meet the project requirements, a derivative-based feature extraction method will be used. As mentioned in section 3.4, the wavelet selected is at least once-differentiable, allowing computation of the wavelet transform of the first derivative of the signal. This can be used to select those coefficients corresponding to sharp changes in the signal at various scales.

4.4.3 Target Detection

Possible risks in the target detection method can be broken up into two types: problems with the method itself, and problems with its output. Both may be characterised by poor performance and inaccurate output. Problems with the method may arise from poorly selected training data that does not adequately represent the spectral data of the core. If this issue arises, with no apparent solution, it may be necessary to then use an untrained classifier of greater sophistication than kNN. Poor output may also be caused by a negative result. This would be harder to detect, as a negative result may arise from an impossible hypothesis, or from an error in the process. Should a negative result arise, the multistage method will be examined in detail to find possible points of failure.

4.5 Conclusion

In this chapter, the time plan, deliverables and risk analysis of the proposed research were presented. The available time is divided into six discrete phases, namely: design, implementation, testing, computation, interpretation and writeup. In each phase, time is allocated to the separate methods present in the proposed research. The feature extraction methods are given the most time and designed first, due to their complexity and the large number of risks associated with them, which were discussed in section 4.4. The deliverables

of the proposed research were discussed in section 4.3. These deliverables are the final products of the research.

Chapter 5

Conclusion

Hyperspectral data present many challenges due to the high dimensionality in both the spatial and spectral dimensions. Techniques for dealing with this data must take into account the computational expense of dealing with global information, and be robust enough to cope with noise. Traditionally, a major issue in hyperspectral data is storage and transmission, leading to a large number of encoding-based compression techniques [Fowler and Rucker 2007], [Kerekes and Schott 2007].

The use of wavelet techniques in hyperspectral data has been shown to provide better results than the traditional techniques of PCA, FFT and DCT [Bruce *et al.* 2002], [Koger *et al.* 2003]. This is due to the non-periodic, non-linear nature of spectral data, where features of interest occur at specific scales and locations. Wavelet techniques are localised to some extent in both time (or band in the case of spectral data) and frequency. This allows the decomposition of a spectral signal into localised frequency components. In this project, the use of a D8 orthonormal Daubechies wavelet was proposed, as this can be implemented as a fast transform, and is at least once-differentiable [Daubechies 1992].

This decomposition can be used to denoise signals, since noise resides in high frequencies, and smoother wavelets approximate noise poorly [Pereyra and Mohlenkamp 2004]. Denoising is necessary for HCI data. The SSNF method of Xu *et al.* [1994] will be used to denoise the wavelet transform data. The high dimensionality of hyperspectral data (and especially the data obtained from the HCI) must be reduced, and so it will be necessary to perform feature reduction. This is the process of finding and choosing a subset of features that will lead to the most separability between classes. The feature reduction methods that will be examined are the wavelet histogram signature method of Kempeneers *et al.* [2004], and the ROC curve method of Bruce *et al.* [2002]. These will be clustered using a standard kNN binary classifier.

The final practical products of this research are algorithms and programs to detect a target spectral signature in noisy hyperspectral mineral data of high dimensionality in an accurate and efficient manner. This will lay the foundation for further research into the automatic identification of mineral data, as well as immediately providing a practical tool for the exploitation of hyperspectral data for the mining community.

References

- [Aboufadel and Schlicker 1999] E. Aboufadel and S. Schlicker. *Discovering Wavelets*. Wiley-Interscience, Hoboken, New Jersey, 1st edition, 1999.
- [Bruce *et al.* 2001] Lori M. Bruce, Cliff Morgan, and Sara Larsen. Automated detection of subpixel hyperspectral targets with continuous and discrete wavelet transforms. *International Journal of Remote Sensing*, 39:2217–2226, 2001.
- [Bruce *et al.* 2002] L.M. Bruce, C.H. Koger, and Li Jiang. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40:2331–2338, October 2002.
- [Burden and Faires 2001] R. Burden and J. Faires. *Numerical Analysis*. Brooks/Cole, Pacific Grove, California, 7th edition, 2001.
- [Chang and Ren 2000] Chein-I Chang and Hsuan Ren. An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38:1044–1063, 2000.
- [Daubechies 1992] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.
- [DeVore and Lucier 1997] Ronald A. DeVore and Bradley J. Lucier. *Wavelets*. www.math.purdue.edu/~lucier/692/wavelet.pdf, 1997. Last accessed 5 August 2008.
- [Donoho and Johnstone 1995] David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1225, 1995.
- [Fowler and Rucker 2007] James E. Fowler and Justin T. Rucker. Three-dimensional wavelet-based compression of hyperspectral imagery. In Chein-I Chang, editor, *Hyperspectral Data Exploitation: Theory and Applications*, pages 379–407. Wiley-Interscience, 2007.
- [Ghassemian and Landgrebe 1988] Hassan Ghassemian and David A. Landgrebe. Object-oriented feature extraction method for image data compaction. *IEEE Control Systems Magazine*, 8:42–48, 1988.
- [Jiang *et al.* 2004] Xiaoguang Jiang, Lingli Tang, Changyao Wang, and Cheng Wang. Spectral characteristics and feature selection of hyperspectral remote sensing data. *International Journal of Remote Sensing*, 25:51–59, 2004.
- [Kaewpijit *et al.* 2003] Sinthop Kaewpijit, Jacqueline Le Moigne, and Tarek El-Ghazawi. Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 41:863–871, April 2003.

- [Kaur *et al.* 2002] Lakhwinder Kaur, Savita Gupta, and R. C. Chauhan. R.c.chauhan, image denoising using wavelet thresholding. In *Indian Conference on computer Vision, Graphics and Image Processing, Ahmedabad, 2002*.
- [Kempeneers *et al.* 2004] Pieter Kempeneers, Steve De Backer, Walter Debruyn, and Paul Scheunders. Wavelet based feature extraction for hyperspectral vegetation monitoring. In *Proceedings of the Image and Signal Processing for Remote Sensing IX*, volume 5238, pages 297–305, 2004.
- [Kerekes and Schott 2007] John P. Kerekes and John R. Schott. Hyperspectral imaging systems. In Chein-I Chang, editor, *Hyperspectral Data Exploitation: Theory and Applications*, pages 19–45. Wiley-Interscience, 2007.
- [Koger *et al.* 2003] Cliff H. Koger, Lori M. Bruce, David R. Shaw, and Krishna N. Reddy. Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (*ipomoea lacunosa*) in soybean (*glycine max*). *Remote sensing of environment*, 86:108–119, 2003.
- [Kruse *et al.* 2003] Fred A. Kruse, Joseph W. Boardman, and Jonathan F. Huntington. Comparison of airborne hyperspectral data and eo-1 hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 41:1388–1400, 2003.
- [Nascimento and Dias 2007] Jose M. P. Nascimento and Jose M. B. Dias. Unmixing hyperspectral data: independent and dependent component analysis. In Chein-I Chang, editor, *Hyperspectral Data Exploitation: Theory and Applications*, pages 149–177. Wiley-Interscience, 2007.
- [Pan *et al.* 1999] Quan Pan, Lei Zhang, Guanzhong Dai, and Hongcai Zhang. Two denoising methods by wavelet transform. *IEEE Transactions on Signal Processing*, 47:3401–3406, 1999.
- [Pereyra and Mohlenkamp 2004] Mara Cristina Pereyra and Martin J. Mohlenkamp. *Wavelets, Their Friends, and What They Can Do for You*. <http://www.sci.sdsu.edu/compsciwork/IIIPASI/DOCUMENTS/COURSENOTES/waveletPASI-IIICPereyra.pdf>, 2004. Last accessed 5 August 2008.
- [Pickering and Ryan 2006] Mark R. Pickering and Michael J. Ryan. An architecture for the compression of hyperspectral imagery. In Giovanni Motta, Francesco Rizzo, and James A. Storer, editors, *Hyperspectral Data Compression*, pages 1–34. Springer, 2006.
- [Schmidt *et al.* 2007] Frederic Schmidt, Sylvain Doute, and Bernard Schmitt. Wavanglet: An efficient supervised classifier for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 45:1374–1385, may 2007.
- [Serpico *et al.* 2007] Sebastiano B. Serpico, Gabriele Moser, and Andrea F. Cattoni. Feature reduction for classification purposes. In Chein-I Chang, editor, *Hyperspectral Data Exploitation: Theory and Applications*, pages 245–274. Wiley-Interscience, 2007.
- [Shafri and Mather 2005] Helmi Z. M. Shafri and Paul M. Mather. Wavelet shrinkage in noise removal of hyperspectral remote sensing data. *American Journal of Applied Sciences*, 2:1169–1173, 2005.
- [Tang and Pearlman 2006] Xiaoli Tang and William A. Pearlman. Three-dimensional wavelet-based compression of hyperspectral images. In Giovanni Motta, Francesco Rizzo, and James A. Storer, editors, *Hyperspectral Data Compression*, pages 273–308. Springer, 2006.

- [Xu *et al.* 1994] Yansun Xu, J.B. Weaver, D.M. Healy, and Jian Lu. Wavelet transform domain filters: a spatially selective noise filtration technique. *IEEE Transactions on Image Processing*, 3:747–758, 1994.
- [Young *et al.* 1998] Ian T. Young, Jan J. Gerbrands, and Lucas J. van Vliet. *Fundamentals of Image Processing*. http://www2.units.it/~carrato/didatt/elab_elettr_imm_2/doc/image_processing/FIP.pdf, 1998. Last accessed 13 June 2008.