A ONE CLASS OBJECT BASED SYSTEM FOR SPARSE GEOGRAPHIC FEATURE IDENTIFICATION

Christoff Fourie

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science at Stellenbosch University

Supervisor: Dr A van Niekerk
Co-supervisor: Prof L Mucina

March 2011

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature : . . . . . .

Date : . . . . . .

ABSTRACT

The automation of information extraction from earth observation imagery has become a field of active research. This is mainly due to the high volumes of remotely sensed data that remain unused and the possible benefits that the extracted information can provide to a wide range of interest groups. In this work an earth observation image processing system is presented and profiled that attempts to streamline the information extraction process, without degradation of the quality of the extracted information, for geographic object anomaly detection. The proposed system, implemented as a software application, combines recent research in automating image segment generation and automatically finding statistical classifier parameters and attribute subsets using evolutionary inspired search algorithms.

Exploratory research was conducted on the use of an edge metric as a fitness function to an evolutionary search heuristic to automate the generation of image segments for a region merging segmentation algorithm having six control parameters. The edge metric for such an application is compared with an area based metric. The use of attribute subset selection in conjunction with a free parameter tuner for a one class support vector machine (SVM) classifier, operating on high dimensional object based data, was also investigated. For common earth observation anomaly detection problems using typical segment attributes, such a combined free parameter tuning and attribute subset selection system provided superior statistically significant results compared to a free parameter tuning only process. In some extreme cases, due to the stochastic nature of the search algorithm employed, the free parameter only strategy provided slightly better results. The developed system was used in a case study to map a single class of interest on a 22.5 x 22.5km subset of a SPOT 5 image and is compared with a multiclass classification strategy. The developed system generated slightly better classification accuracies than the multiclass classifier and only required samples from the class of interest.

KEYWORDS AND PHRASES: segmentation evaluation, image classification, support vector machine, evolutionary search algorithms

OPSOMMING

Die outomatisering van die verkryging van inligting vanaf aardwaarnemingsbeelde het in sy eie reg 'n navorsingsveld geword as gevolg van die groot volumes data wat nie benut word nie, asook na aanleiding van die moontlike bydrae wat inligting wat verkry word van hierdie beelde aan verskeie belangegroepe kan bied. In hierdie tesis word 'n aardwaarneming beeldverwerkingsstelsel bekend gestel en geëvalueer. Hierdie stelsel beoog om die verkryging van inligting van aardwaarnemingsbeelde te vergemaklik deur verbruikersinteraksie te minimaliseer, sonder om die kwaliteit van die resultate te beïnvloed. Die stelsel is ontwerp vir geografiese voorwerp anomalie opsporing en is as 'n sagteware program geïmplementeer. Die program kombineer onlangse navorsing in die gebruik van evolusionêre soek-algoritmes om outomaties goeie beeldsegmente te verkry en parameters te vind, sowel as om kenmerke vir 'n statistiese klassifikasie van beeld segmente te selekteer.

Verkennende navorsing is gedoen op die benutting van 'n rand metriek as 'n passings funksie in 'n evolusionêre soek heuristiek om outomaties goeie parameters te vind vir 'n streeks kombinering beeld segmentasie algoritme met ses beheer parameters. Hierdie rand metriek word vergelyk met ‘n area metriek vir so ‘n toepassing. Die nut van atribuut substel seleksie in samewerking met 'n vrye parameter steller vir 'n een klas steun vektor masjien (SVM) klassifiseerder is ondersoek op hoë dimensionele objek georiënteerde data. Vir algemene aardwaarneming anomalie opsporings probleme met 'n tipiese segment kenmerk versameling, het so 'n stelsel beduidend beter resultate as 'n eksklusiewe vrye parameter stel stelsel gelewer in sommige uiterste gevalle. As gevolg van die stogastiese aard van die soek algoritme het die eksklusiewe vrye parameter stel strategie effens beter resultate gelewer. Die stelsel is getoets in 'n gevallestudie waar 'n enkele klas op 'n 22.5 x 22.5km substel van 'n SPOT 5 beeld geïdentifiseer word. Die voorgestelde stelsel, wat slegs monsters van die gekose klas gebruik het, het beter klassifikasie akkuraathede genereer as die multi klas klassifiseerder.

SLEUTELWOORDE EN FRASES: segmentasie evaluasie, beeld klassifikasie, steun vektor masjien, evolusionêre soekalgoritmes

ACKNOWLEDGEMENTS

I would like to thank the following people and institutions that assisted me in some way or another during the course of this research project:

- The academic staff of the Department of Geography and Environmental Studies, Stellenbosch University, for helpful comments.
- Dr. Adriaan van Niekerk, supervisor, for facilitating numerous sessions of academic discussions.
- Prof. Ladislav Mucina, co-supervisor, for allocating a National Research Fund (NRF) bursary that made this research possible and for suggesting the Cape thicket problem.
- Dr. Gaul Costa, Catholic University of Rio de Janeiro, for providing their experimental segmentation parameter tuner.
- Prof. Zhigang Liu, Beijing Normal University, for providing his work on semi supervised object based one class support vector machine classification.
- The Satellite Applications Centre (SAC), which is part of the Council for Scientific and Industrial Research (CSIR) of South Africa, for supplying the SPOT 5 satellite imagery used in this study.
- My family, for greatly appreciated support.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS AND ACRONYMS

ANN	Artificial neural network
CFR	Cape Floristic Region
DE	Differential evolution
DEM	Digital elevation model
DT	Decision tree
ECA	Evolutionary computation algorithm
FNEA	Fractal net evolution approach
GA	Genetic algorithm
Geo-ND	Geographic object novelty detector
GEOBIA	Geographic object based image analysis
GLCM	Gray level co-occurrence matrix
HRG	High resolution geometric
HSI	Hue, saturation and intensity
JDE	Self adapting differential evolution
KNN	k-nearest neighbour
LSB	Larger segments booster
NDVI	Normalised difference vegetation index
NFI	National Forest Inventory
NIR	Near infrared
ORM	Object relational model
PSO	Particle swarm optimisation
PTASS	Parameter tuning and attribute subset selection
PTO	Parameter tuning only
RBF	Radial basis function
RS	Remote sensing
RVI	Ratio vegetation index
SAC	Satellite Applications Centre
SCRM	Size constrained region merging
SPARK	Spatial reclassification kernel
SVDD	Support vector domain descriptor
SVM	Support vector machine
SWIR	Short wave infrared
VHR	Very high resolution

CHAPTER 1: INTRODUCTION

The automation of information extraction from earth observation imagery is receiving increasing attention owing to the advent of high resolution earth observation imagery and the accelerated increase of desktop computing power (Lang 2008). Traditional means of information extraction from such imagery is a labour intensive and complex task (Lück 2004). The large volume of daily captured earth observation data that remain unused and the numerous benefits the extracted information can provide to society, stresses the need for highly automated information extraction workflows (Blaschke 2010; Blaschke, Lang & Hay 2008; Lang 2008; Rogan et al. 2008). This is especially true for national forest inventory (NFI) mapping projects as traditional large scale mapping endeavours are prohibitively expensive in terms of finances and time. NFIs have been identified as an application field that can greatly benefit from more automated information extraction techniques (Louw & Scholes 2002; McRoberts & Tomppo 2007; Mucina & Geldenhuys 2002; Sanchez-Hernandez, Boyd & Foody 2007a; Tomppo et al. 2008; Wannenburg & Mabena 2002; Wulder et al. 2008). Although autonomy and efficiency are desirable properties in the processing of remotely sensed images, they should not come at the expense of degradation in the accuracy of the extracted information. Unfortunately, full autonomy of the information extraction process is still an elusive goal.

Another reason why more autonomous and efficient remote sensing (RS) workflows are desirable is that they will make this technology available to a wider scientific audience. The analysis of satellite imagery is a highly technical field, with its own specialist methods and associated literature (Newton 2007). The relatively new sub discipline of geographic information sciences, called geographic object based image analysis (GEOBIA), attempts to address this problem (Blaschke & Lang 2006; Hay & Castilla 2008). A key objective of GEOBIA is the creation of tools and methodologies to replicate human interpretation of RS imagery in automated or semi automated ways.

1.1 THE IMAGE PROCESSING WORKFLOW

The image processing workflow, in an earth observation or RS context, describes the process whereby information is extracted from remotely sensed data (Schott 2007). Image processing workflows can be as varied as the different application fields in RS. Arguably, the most

important step in such a workflow is the classification or identification of specific land cover elements. Image classification, or land cover mapping, has traditionally been done with manual interpretation of aerial photographs (Bock et al. 2005; Goetz et al. 2003; Schott 2007; Schowengerdt 2007). Modern image classification techniques use statistical methods or expert system approaches to identify land cover elements in a supervised, semi supervised or unsupervised manner (Campbell 2002). The quality of a classification depends on the fidelity of the spatial, spectral, radiometric and temporal resolution of the data source. Land cover elements are identified by their spectral, textural, contextual and temporal signatures. The trained system is then applied to classify the rest of the scene(s).

Various supervised classification techniques exists, each with different characteristics. In supervised image classification a system is usually trained on a small sample of a scene's pixels or objects. Although image classification is often considered to be the most important step in an image processing workflow, a number of other steps are also needed. Schott (2007) describes the RS image processing workflow as an "image chain", where the complete chain is only as strong as its weakest link. Additional steps in the image processing workflow may include image pre-processing, image segmentation, parameter tuning, object/pixel attribute selection, and information formatting. All these steps influence the accuracy of the final product. As with image classification, there are many approaches to carrying out these steps and each approach has varying levels of efficiency, accuracy and autonomy. These characteristics, as well as the characteristics of the land cover features to be identified, should be considered when designing or choosing an image processing workflow (Lu & Weng 2007).

The automation of steps in image processing has received much research attention in recent years. Examples include automating image pre-processing for accurate and large scale change detection applications (Leprince et al. 2007; Lück 2004; Toutin 2004); automatic attribute selection in object based classification (Nussbaum, Niemeyer & Canty 2006; Van Coillie, Verbeke & De Wulf 2007); automatically obtaining high quality image segmentation results (Espindola et al. 2006; Feitosa et al. 2006; Fredrich & Feitosa 2008); and automatic semantic rule generation for expert system applications (Tseng et al. 2008). There is a large operational gap between developments in the primary information sciences (e.g. the fields of pattern recognition, machine learning and computer vision) and applied RS. The RS discipline could benefit from adopting concepts and methods from these fields (Hay & Castilla 2008).

1.2 CAPE THICKET AS A MODEL REMOTE SENSING ANOMALY DETECTION PROBLEM

Cape thicket is an endangered vegetation type dominated by shrubs and small trees and it is found on steep, rocky slopes, around boulder formations and in ravines within the Fynbos biome. Not much is known about Cape thicket (it has not yet been studied ecologically or floristically) (Mucina & Rutherford 2006), but Campbell (1985) identifies two structural variations this vegetation type within the mountainous areas of the Western Cape, namely Cape thicket and Mitchell thicket. The former occupies dryer sites than Afrotemperate forest and is described as a mid-high to tall, evergreen, closed, microphyllous shrubland. Mitchell thicket has a more open canopy with less than 70% continuous cover. The gaps in cover are usually filled with large boulders and/or screes. The two vegetation types also share many floristic similarities. Figure 1.1 illustrates a Cape thicket patch found around a sandstone scree on the slopes of Bothmaskop near Stellenbosch. In the photograph the thicket is strongly contrasted by the remains of a veld fire that destroyed much of the surrounding fynbos vegetation. Cape thicket patches are highly fire resistant. (More photographs of Cape thicket and related vegetation structural types are presented in Appendix A).

Figure 1.1: Patch of Heeria argentea tallus forest (Cape thicket) around a sandstone scree on the slopes of Bothmaskop, near Stellenbosch

Owing to its sparse distribution and small patch size, occurrences of Cape thicket have not been mapped to date. A cost effective methodology to map these thickets using earth observation methods will not only contribute to scientific knowledge, but will also have practical value, particularly for their conservation and management. The geographical and environmental characteristics of these thickets will also be of interest to evolutionary biologists (e.g. vegetation paragenesis and ecological synthesis).

Cape thicket is difficult to detect and delineate using existing RS techniques. This is mainly due to the structural and spectral heterogeneity of the vegetation type and the complexity of the landscapes in which it is located. The ability to accurately map heterogeneous Mediterranean vegetation has been identified as a weakness in applied RS (Mallinis et al. 2008) and a subject of much controversy among botanists (Cowling & Holmes 1992). According to Shoshany (2000), the dynamic patterns and spatial variation of Mediterranean type vegetation are due to local differences in habitat conditions and climatic gradients. This results in the formation of various types of transition patterns that complicate the classification of vegetation types. Until recently the resolution of satellite sensors was low compared to the spatial configuration of Mediterranean landscapes and this prohibited the use of RS in effectively mapping Mediterranean land cover (Mallinis et al. 2008). Although the advent of high resolution satellite imagery offers new mapping possibilities (Johansen & Phinn 2006), Vogiatzakis, Mannion & Griffiths (2006) found that high resolution satellite imagery has limited use for monitoring Mediterranean vegetation using existing RS methods. Consequently, novel image processing techniques are needed to facilitate generalised patch delineation from high resolution satellite imagery to better understand the ecology of Mediterranean vegetation (Shoshany 2000). In particular, new methods are needed to cost effectively map and monitor Cape thicket patches using high resolution satellite imagery.

1.3 PROBLEM FORMULATION

Hay et al. (2001) and Forster & Kleinschmit (2008) have identified object based classification as a promising method to map sparsely distributed vegetation types such as Cape thicket. The extra information obtained from the objects, as opposed to just the spectral information of a per-pixel approach, hold promise for differentiating such anomalous features from their surroundings. However, according to Baatz, Hofmann & Willhauck (2008), the success of such an approach is greatly reliant on an optimal image segmentation. Selecting the appropriate segmentation algorithm and parameters can be a very time consuming and

subjective exercise (Baatz, Hofmann & Willhauck 2008; Feitosa et al. 2006). Fredrich & Feitosa (2008) present a promising technique to automate segmentation parameter tuning, as opposed to the traditional trial and error process. Unfortunately, the technique and its search fitness functions have only been tested on homogeneous land cover features such as those found in an urban setting. Other fitness functions, and possibly search techniques too, need to be investigated for segmenting features with less well defined boundaries such as Cape thicket.

A kernel based statistical anomaly detection approach such as one class support vector machines/support vector domain descriptors (SVM/SVDD), has the potential to extract sparsely located vegetation types (e.g. Cape thicket) from satellite imagery (Sanchez-Hernandez, Boyd & Foody 2007a; 2007b). The main advantage of SVM/SVDD is that it requires only training samples for the class of interest and generally requires fewer samples than traditional probability density approaches (Sanchez-Hernandez, Boyd & Foody 2007b; Tax & Duin 2004). Although an SVM/SVDD requires the user to tune the free parameters (the nature and number of parameters depend on the kernel) and to select an appropriate attribute subset, population based search techniques can be used to automatically determine these parameters (Huang & Wang 2006; Lin et al. 2008; Van Coillie, Verbeke & De Wulf 2007).

Unfortunately, the effectiveness of population based search techniques has only been demonstrated on hypothetical datasets employing binary or multiclass techniques and, compared to simpler classifier tuning strategies, did not deliver superior results over all test datasets. However, no research has been reported on applying population based search techniques to optimise a kernel based anomaly (or multiclass) detector for free parameters and attribute subsets on object based or per-pixel based data, to map sparsely located land covers (such as Cape thicket). The only published example of a population based search technique being used on remotely sensed imagery (for classifier free parameter tuning and/or attribute subset selection) is by Van Coillie, Verbeke & De Wulf (2007) who used a different classifier (artificial neural networks) and did not perform free parameter tuning (it was only used for attribute subset selection). A standard genetic algorithm was used in their study for attribute subset selection. A kernel based statistical classifier optimised with a population based search algorithm in terms of the used attributes and classifier free parameters should be evaluated

using different sets of algorithm parameters and input data for object based remote sensing image processing.

Finally, no integrated system exists that allows for image segmentation to be tailored to a specific class of interest (and one nominal scale of observation) followed by an optimised kernel based statistical anomaly detection. An investigation of the use of such a system for resolving mapping problems could contribute to the debate about effective techniques for information extraction from earth observation imagery.

1.4 AIMS AND OBJECTIVES

The primary aim of this research is to investigate the use of segmentation, classification, free parameter tuning, and attribute subset selection techniques for identifying and delineating sparsely located land cover features from moderate-to-high resolution multispectral earth observation imagery in automated or semi automated ways. The secondary aim is to demonstrate the potential of these techniques in a mapping exercise aimed at identifying Cape thicket (as a case study).

The five objectives for reaching these aims are to:

1. Conduct a literature review on image classification techniques, segmentation algorithms, segmentation evaluation methods and similar RS case studies to help formulate an effective classification workflow;
2. Collect and prepare RS data suitable for use in exploratory research and the case study;
3. Develop a user friendly software application that combines supervised, automated segmentation parameter tuning and band selection, a one class kernel based classifier, and population based search techniques;
4. Use the developed system to carry out various experiments to evaluate which of these techniques are superior for mapping sparsely located land cover features; and
5. Carry out a case study in which the most promising anomaly detection techniques are compared to a popular multi-class classification strategy.

1.5 RESEARCH AGENDA

This research is exploratory in nature as it investigates the use of various techniques for solving a specific earth observation problem (i.e. automating the identifying and delineating of sparsely located land cover features from moderate-to-high resolution multispectral earth observation imagery). The study employs empirical data (satellite images and field observations) to carry out quantitative experiments for assessing the suitability of the various techniques. A new software system is also developed to facilitate these experiments.

Figure 1.2 schematically illustrates the stages of the research. In this chapter an overview was given of the research problem, the aims and objectives. Chapter 2 provides a review of the fundamental principles of RS image classification. Key concepts and technologies used in this research are explained and supported by appropriate examples of existing implementations. In Chapter 3 the selected data source is described and the study area is introduced. All the preprocessing steps performed on the chosen data are briefly noted. Chapter 4 presents the developed image processing system while Chapter 5 elaborates on its segmentation component.

Exploratory research results with different attribute selection and parameter tuning strategies for the classification component of the system are presented and interpreted in Chapter 6. Chapter 7 describes the mapping of a single class of interest as a case study using the developed methodology and qualitatively compares the workflow efficiency of the system to a more traditional classification approach. Chapter 8 concludes the thesis with a summary of the research and suggestions for future work.

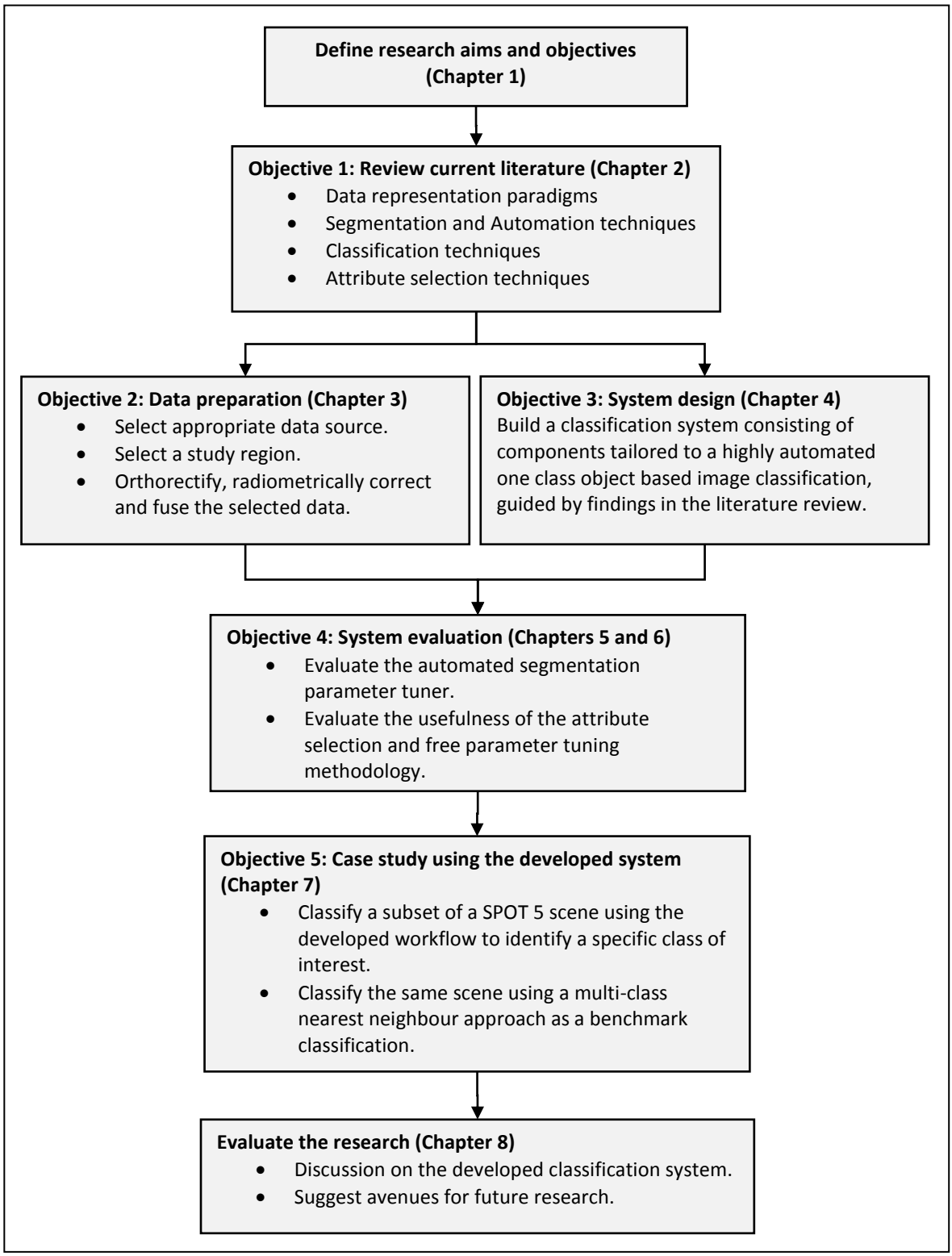


Figure 1.2: Research progression

CHAPTER 2: BACKGROUND AND RELATED WORK

This chapter gives an account of the relevant theories and concepts, which underlie the presentation and evaluation of the proposed anomaly detection system. Key concepts concerning the basics of RS, object based data representation, image segmentation, supervised image classification, population based search techniques and attribute selection are explained. Throughout, related extant research is referred to and acknowledged.

2.1 VEGETATION MAPPING FROM REMOTELY SENSED IMAGERY

The science of RS comprises the analysis and interpretation of measurements of electromagnetic radiation that is reflected from or emitted by features on the earth's surface. The sensors used for capturing the data are typically space borne (satellites) or airborne. A feature on the earth's surface can be defined as a spatial visual entity having some significance to the observer. Features on the earth's surface reflect or emit different amounts of electromagnetic radiation depending on the physical composition of the features. The amount of electromagnetic radiation emitted or reflected at different wavelengths by a feature, is known as its spectral signature. In the classification of remotely sensed images it is commonly attempted to exploit these spectral signatures by using them as the prime means of differentiating among features on a terrestrial surface. The spectral signatures can be used in expert knowledge systems, sample based statistical classification or unsupervised statistical classification (De Jong, Van der Meer & Clevers 2005; Mather 2006; Schowengerdt 2007).

Although leaves are the dominant feature in a vegetation plot and should therefore contribute greatest to its spectral signature, a vegetation plot seldom represents the spectral signature of its dominant leaf features. For example, a broad leaf forest plot consists of the combined spectral signatures of mesophyll leaves, tree bark, soil, undergrowth and dead leaves. These components are present in varying quantities, thus contributing different amounts of electromagnetic radiation to the overall spectral signature of the plot. Another factor that influences the overall spectral signature of a feature is its structure or physical form. Different textures and surface angles reflect or scatter radiation in different ways. Vegetation factors known to influence the spectral reflectance of vegetation canopies include the overall life form of the vegetation, leaf type, vegetation height or tree size, the fractional cover of vegetation, the shadows cast by vegetation structure, and the health and water content of the leaves. In addition, the soil colour and wetness contribute to the spectral response at any given location in the image (Leckie et al. 2005; Liang 2004; Lucas et al. 2008; Woodcock,

Macomber & Kumar 2002). Andrew & Ustin (2008) elaborate on the effects that these environmental factors have in vegetation classification. As an example, the spectral signatures of four structurally different vegetated earth surface features are shown in Figure 2.1. An Afrotemperate forest patch, a Cape thicket patch, a Fynbos plot and a sedimentary rock outcropping with limited herbaceous growth were arbitrarily sampled from pre-processed SPOT 5 data in the Jonkershoek valley near Stellenbosch, South Africa. The x axes in the graphs indicate the SPOT 5 bands while the y axes denote the percentage of reflected light. As vegetation cover decreases so does the reflectance in the near infrared portion of the spectrum (Band 3) due to the reduction in foliage. Less vegetation also allows more mid infrared light (point four on the x axes) through the canopy reflected by the underlying soil or rock.

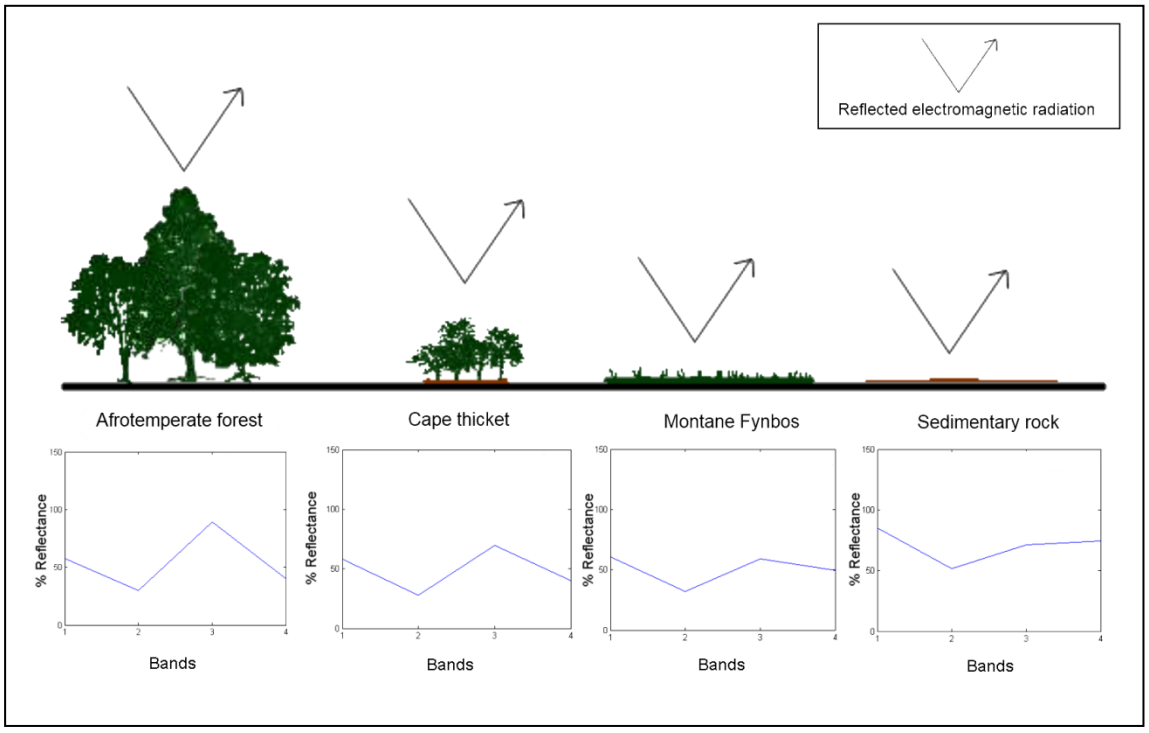


Figure 2.1: Interaction of electromagnetic radiation with different vegetation structural types

Multispectral remotely sensed vegetation mapping exercises typically exploit the spectral signatures of plots for structure or physiognomy based classifications (Campbell 2002). This is commonly the case when employing moderate-to-high resolution multispectral imagery, as the spectral fidelity of these types of sensors is insufficient to conduct genus or floristic classifications (Govender et al. 2008; Johansen & Phinn 2006; Key et al. 2001; Lucas et al. 2008; Mallinis et al. 2008). Horning (2004) overviews decreasing classification accuracy

with increasing class complexity with respect to vegetation when using moderate resolution multispectral sensor data.

2.2 TWO PARADIGMS TO REMOTE SENSING IMAGE DATA REPRESENTATION

The two prominent image data representation techniques, namely per-pixel and object based are explained and contrasted in this section. The selection of an appropriate data representation strategy for the system presented in this thesis is also justified.

2.2.1 Per-pixel data representation

Pixel based image data representation is the most used and classic technique for remotely sensed image data representation and classification. Each pixel in a scene is considered as a land cover feature or part of a feature. Pixels in the scene are classified into a predefined sample based cluster when employing supervised classification, or a self adopted cluster when employing unsupervised classification. The spectral signatures of the pixels or features are used as the numerical basis for the categorisation.

Pixel based classification is well suited to images with a resolution similar to the land cover features of interest (Wang, Sousa & Gong 2004). As the spatial resolution of a sensor increases, so does the within class spectral heterogeneity, as the resolution is finer than the features of interest, resulting in lower classification accuracies. When pixels are much larger than the features of interest, it is common for multiple features to occupy a single pixel, also resulting in a decrease in classification accuracy (Ozdarici & Turker 2006; Shaban & Dikshit 2001). This results in the so called mixed pixel effect. The scale dependency of landscape features and the problem it poses in per-pixel classifications has been widely noted in the literature (Atkinson & Aplin 2004; Chen, Stow & Gong 2004; Fassnacht, Cohen & Spies 2006; Ju, Gopal & Kolaczyk 2005). Per-pixel classifications may also lead to results that display homogeneous regions as heterogeneous and scattered (the so called salt and pepper effect). A majority filter is often applied to reduce this noise (Lu & Weng 2007).

Per-pixel texture measures can be used in a per-pixel classification approach. An important issue pertaining to their use is the kernel window size and orientation to be used in determining per-pixel texture measures, usually done with the aid of semi-variogram analysis (Franklin, Wulder & Lavigne 1996; Woodcock, Strahler & Jupp 1988). The texture of land

cover units is distinguishable at different spatial scales and also dependent on the selection of input bands (Johansen et al. 2007).

A per-pixel classifier can also use adjacent pixel information in the classification process. The spatial reclassification kernel (SPARK) is a good example of such an approach. SPARK examines the spatial and spectral properties in a kernel and classifies a pixel on pre-defined decision rules. Sluiter (2005) reports on the use of SPARK for Mediterranean vegetation classification, noting an increase in classification accuracy for a difficult heterogeneous vegetation classification problem compared to a maximum likelihood per-pixel classification.

Recently, new approaches to per-pixel classification incorporating the spatial domain and image context have been proposed. Camps-Valls et al. (2007) and Gómez-Chova et al. (2008) put forward a pixel based classification framework incorporating spectral graph theory, in addition to its standard manifold learning and kernel based algorithm components. The method is a semi-supervised support vector machine (SVM) classifier, employing composite kernels to handle context (domain knowledge). Each pixel in the scene spreads its label information to adjacent pixels (unknown classes are modelled in a semi-supervised manner to make this methodology more robust) until a stable classification is achieved over the whole scene (graph component). This methodology delivers superior results when classifying urban areas and identifying clouds, compared to standard SVM and transductive SVM classifiers using both linear and RBF kernels. This methodology also has the ability to mitigate the salt and pepper effect common to per-pixel classification (Camps-Valls 2008).

2.2.2 Object based data representation

This section introduces the object based data representation paradigm, discusses scale of representation and describes recent theories on object based data representation and classification.

2.2.2.1 Basics

In RS image processing, an image object can be defined as a grouping of pixels of similar spectral and/or textural properties. In the object based data representation paradigm, these objects (also referred to as object primitives or segments) form the basis unit for image interpretation and classification. An object has, compared to a pixel, other attributes,

including various measures of its geometry, spectral characteristics, texture and morphology that can be used in image analysis (Navulur 2007). The paradigm of object based data representation and classification was introduced and demonstrated in the 1970s but general uptake lagged due to computing power constraints (De Kok, Schneider & Ammer 1999).

For most applications, modern satellite sensors exhibit the H-resolution phenomenon – meaning that most ground features are composed of multiple pixels (Hay et al. 2005). With these sensors, individual pixels rarely correspond to true geographical objects. Homogeneous land cover classes could consist of heterogeneous pixel groupings (Atkinson 2004; Hay et al. 2005). Object based data representation and classification was designed to overcome these limitations of pixel based data representation and classification. Hay et al (2005) describe additional motives for the development of object based image analysis approaches. In object based image classification, object primitives are generated through the process of image segmentation.

Image segmentation is a process whereby pixels in the scene are grouped into regions on the basis of varying criteria related to pixel and neighbourhood properties. Usually, segmentation algorithms group regions in an image having homogeneous spectral or textural properties. Numerous image segmentation approaches exist, all with varying characteristics. The result of an image segmentation process is heavily dependent on the free parameters needed as input to this process. Image segmentation decreases the level of detail in a scene, reduces image complexity, and makes image content more graspable (Lang 2008). Creating image segments that correctly delineate features of interest is a popular research topic in GEOBIA (Baatz, Hofmann & Willhauck 2008; Blaschke 2010; Castilla 2003; Castilla & Hay 2008; Feitosa et al. 2006; Kim & Madden 2006; Lang 2008; Osman, Inglada & Christophe 2009; Wulder et al. 2008).

2.2.2.2 Scale considerations

A widely acknowledged research focus of image segmentation is the determination of the appropriate scale of observation or the size of the object primitives obtained from image segmentation (Addink, De Jong & Pebesma 2007; Benz et al. 2004; Hay & Marceau 2004; Kim & Madden 2006; Wu 2004). Different land cover features are observable or detectable at different scales. In object based image analysis the size of the generated image segments is defined as the scale of observation (Benz et al. 2004). Many image segmentation algorithms

have a free parameter that controls the relative scale of the resulting segments. A good segmentation algorithm attempts to segment an image so that object primitives correspond to the appropriate scale of observation of the land cover feature(s) of interest. Kim & Madden (2006) demonstrate the effect that the scale of observation has on classification accuracy in an object based classification approach.

A solution to the problem of defining the optimal scale of observation has manifested in a popular paradigm in object based image analysis, namely multiscale image segmentation and analysis (Baatz & Schäpe 1999; Burnett & Blaschke 2003). Multiscale image segmentation and analysis is inspired by the theory of hierarchical patch dynamics developed in the field of landscape ecology (Wu & David 2002; Wu & Loucks 1995). The central idea of this working paradigm is that land cover features are composed of different elements observable at different scales. It is argued that RS data is a particular case of the modifiable areal unit problem (Hall, Hay & Marceau 2004; Hay et al. 2003; Wu 2004). A scene is segmented at different scales representing these different elements, termed multiscale segmentation. A hierarchy of image segments is created. At each level of segmentation, an object can be identified by its spectral, spatial, context and textural properties. Higher level objects can be identified by their composition of lower level objects, in addition to the standard object properties. This identification is done using semantic rules defined by the analyst and is called object relational model building (ORM). Figure 2.2 illustrates this principle.

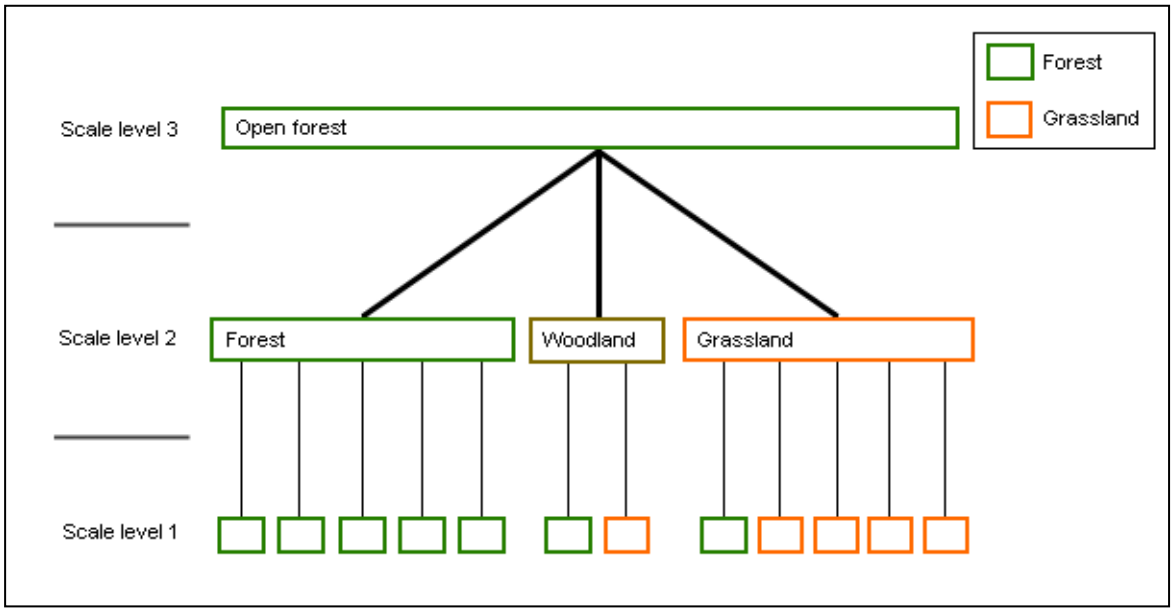


Figure 2.2: Hierarchical object relational model illustrating how higher level objects are classified through their composition of lower level objects

The objects at each scale level are classified according to their composition of lower level objects. As an example, a semantic rule could label an object at scale level two as "woodland" if its composition at scale level one is roughly equal parts forest and grassland.

Schiewe, Tufte & Ehlers (2001) and later Benz et al. (2004) extented this basic workflow by incorporating ORM in a fuzzy classifier framework. A holistic modelling strategy is advocated that incorporates expert knowledge, ORM, image data and ancillary data in a fuzzy classification framework (Lang, Albrecht & Blaschke 2006) termed the fractal net evolution approach (FNEA) or spatial semantic network. Such a modelling framework is implemented in *eCognition*, a commercial object based earth observation image analysis software. This strategy, using *eCognition* as the implementation tool, has been successfully applied to a wide range of fields (Blaschke 2010; Chen et al. 2009; Laliberte et al. 2004; Lamonaca, Corona & Barbati 2008; Matinfar et al. 2007). A common criticism of this strategy is that it is a fully manual and often time consuming process, relying heavily on the knowledge and expectations of the analyst. Another major disadvantage of FNEA is that segmentation scales are usually determined by trial and error (see above references for examples) guided by analysts' knowledge of the image.

It has been argued that multi-scale segmentation and analysis should be guided by the intrinsic scale of the dominant landscape objects in an image (Hay et al. 2003; Hay et al. 2001; Hay, Niemann & Goodenough 1997). A methodology, entitled multi-scale object specific analysis/segmentation, influenced by scale space theory (Lindenberg 1994), has been proposed that implements this idea as a core heuristic. In this technique, the variance, area and mean are calculated for pixels through a spectrum of kernel sizes (object specific analysis). A weighting scheme is employed using the area values of pixels to upscale a scene to a desired resolution. This scale space analysis approach can automatically extract the dominant landscape features at desired scales using watershed segmentation algorithms. Few case studies exist that use this technique (Hay et al. 2003). A general set of scaled segmentations can be generated automatically; guided by the intrinsic scale space nature of the image under consideration. The scaled segmentations hold promise for multiscale ecological studies (Castilla, Hay & Ruiz 2008). Hay et al. (2003) contrasts this approach with FNEA.

The theory of multiscale analysis in RS is not uncontested as the most viable means of information extraction. Corcoran & Winstanley (2008) argue that the creation of image objects at multiple scales does not conform to modern theories of visual perception as visual perception is influenced by a priori knowledge of the scale of the features of interest. Corcoran & Winstanley (2008) present an innovative segmentation algorithm that operates on both spectral data and texture measures and strives to represent land cover features at one nominal scale. The assumption is made that land cover features exhibit areas of uniform texture and/or intensity. Texture boundaries are suppressed to derive more robust segment boundaries (it is difficult to localise texture gradient boundaries). This methodology has been tested in an urban environment only. Lang & Langanke (2006) propose the use of a scale variant segmentation algorithm for a one level representation for landscape elements.

The choice of a multiscale versus single scale representation in the framework of the object based image processing paradigm should be guided by the nature of the application; and the extent to which the features of interest can be modelled with semantics in the scale space. Weinke, Lang & Preiner (2008) contrast the FNEA approach with a one level representation approach (Lang & Langanke 2006). Although no quantitative measure of accuracy between these approaches was given, both approaches delivered satisfactory results when used to characterise alpine forest habitats.

2.2.2.3 Advanced object based techniques and theories

Recently, some authors (Baatz, Hofmann & Willhauck 2008; Castilla & Hay 2008; De Kok 2006; De Kok & Wezyk 2008; Lang 2008) have debated about the greater importance image objects should have in the information extraction workflow.

Lang (2008) suggests that the object based image analysis process is cyclic and not linear in nature (e.g. segmentation followed by classification). Lang (2008) contends that an effective image segmentation and classification task is done iteratively with finer results achieved at each iteration. Baatz, Hofmann & Willhauck (2008) refine this notion by suggesting an advanced workflow where the image objects are the focus of the analysis. It is held that the segmentation process (not considering data capabilities) has the greatest influence on the accuracy of the obtainable results. The general aim of the workflow is to create accurate geographic objects from the object primitives obtained from segmentation in manageable iterative steps. Baatz, Hofmann & Willhauck (2008) suggest that results from classification

and semantic knowledge should drive the segmentation process. As the classification process proceeds, more tailored tools, segmentation algorithms and expert knowledge are used to refine the results. Baatz, Hofmann & Willhauck (2008) suggest that the structure of the final thematic product should be modular, that is, each land cover class is modelled or treated individually because classes might not have the same discriminative characteristics through the scale space.

De Kok (2006) and De Kok & Wezyk (2008) suggest an iterative classification and segmentation process having more refined results in later stages. They recommend that edge information should play a greater role in land cover feature modelling. The use of pure statistical techniques to classify land cover features has also been criticised, particularly the ability of such techniques to classify spectrally inhomogeneous objects in very high resolution imagery.

These techniques and theories all stress the importance of knowledge driven image interpretation by an analyst focussing on individual land cover features. Considering these notions, a procedure is advanced in Chapters 4 and 5 which gives considerable attention to supervised segment generation under the assumptions that a single class is of interest and that the features belonging to said class are discernable at a single scale of observation.

2.2.3 Per-pixel versus object based classification

The major criticism of per-pixel classification is that a pixel rarely corresponds to a real world object; that is it is a nondescript spatial unit. Homogeneous regions could consist of a collection of heterogeneous pixels (Hay et al. 2005; Lang 2008; Schiewe, Tufte & Ehlers 2001), although new techniques incorporating spatial information, such as those presented by Camps-Valls (2008), have been shown to mitigate this effect. As a final output, the object based classification approach attempts to represent real world objects through meaningful image segments. Wang, Sousa & Gong (2004) caution that although the spectral generalisation ability of object based classification effectively separates spectrally mixed classes, it also risks merging distinct classes with less well defined spectral boundaries, thus greatly decreasing classification accuracy. Consequently, the results of the segmentation process heavily influence the final classification results. Segmentation results are very much dependent on the chosen segmentation algorithm, segmentation parameters, the choice of

appropriate scale of observation and the analyst's knowledge of the scene or features of interest.

Comparative studies of these two classification methodologies abound in the literature. The results of these studies vary across application domains, the statistical classifiers used and the sensor characteristics. For example Barlow, Martin & Franklin (2003) report a substantial increase in classification accuracy when employing the object based paradigm in landslide detection on moderate resolution imagery. Generally, in land cover applications the discrepancy is more moderate. Oruc, Marangoz & Buyuksalih (2004), Matinfar et al. (2007) and Perveen, Nagasawa & Husnain (2008) report moderate increases in classification accuracy when using the object based paradigm, although they note that an increase in accuracy is not achieved for all land cover classes. Stuckens, Coppin & Bauer (2000) report a slight increase in accuracy when using objects instead of pixels in general classification problems.

In vegetation specific studies the results of comparisons also vary. For example, Yu et al. (2006) conducted a study in characterising 48 vegetation types from very high resolution airborne imagery using different statistical classifiers on both object based and per-pixel data. They reported that the use of objects instead of pixels significantly improved the results for most of the vegetation types. In contrast, Dorren, Maier & Seijmonsbergen (2003) found that using objects instead of pixels in mapping heterogeneous forest stands in a mountainous area resulted in lower accuracies. Dorren, Maier & Seijmonsbergen (2003) gave the spatially discontinuous edges of the forest stands, and the inability of the segmentation algorithm to detect these edges, as reasons for lower achieved accuracies. Wang et al. (2004) proposed an integrated classification scheme combining per-pixel and object based approaches and achieved higher classification accuracies using this novel combined approach.

Although no data representation and classification strategy has been shown to be generally superior to the other, most authors acknowledge the usefulness of the greater information content presented by the object based approach when exploited by an expert systems approach or an advanced statistical classifier (see Hay et al. 2005; Lang, Albrecht & Blaschke 2006). For example, in characterising heterogeneous Mediterranean vegetation using the object based classification paradigm, Mallinis et al. (2008) report significantly improving

classification accuracy (53% versus 79% overall accuracy) when using more complex per object texture measures and an advanced statistical classifier.

Considering the recommendations made in the literature, the object based paradigm is the preferred choice for this work as the basis form of data representation. A one level representation strategy will be followed. Due consideration has to be given to the choice of segmentation algorithm, scale of observation, segmentation parameter tuning and the pre-processing of data as input to the segmentation process, in addition to the standard statistical classification considerations.

2.3 IMAGE SEGMENTATION

This section overviews the image segmentation algorithms commonly employed to process earth observation images and describes the metrics used in quantitative segmentation evaluation. Recently proposed techniques to automate image segmentation in supervised and unsupervised ways are reviewed and the often overlooked problem of optimal band selection in RS image segmentation is also considered.

2.3.1 Segmentation algorithms

Image segmentation is the process of partitioning an image into regions. The characteristics of these regions (size, shape and locality) are controlled by various criteria, parameters and the underlying structure of the segmentation algorithm itself. The image segmentation process typically attempts to delineate areas in an image that have some meaning for the observer, or to generate segments that will aid more accurate information extraction via classification or modelling (Baatz, Hofmann & Willhauck 2008).

Pavlidis & Horowitz (1974) (in Lang, Albrecht & Blaschke 2006) give a mathematical definition of a segmented image entailing four governing rules. In simplified format these rules are:

- a union set of regions comprises the image;
- regions do not overlap;
- some homogeneity criterion applies to the characteristics of the regions; and
- the homogeneity criteria of neighbouring regions may differ.

RS image segmentation as a viable image analysis technique became prevalent due to the problems posed by VHR imagery (H-res problem) captured by sensors such as IKONOS, QuickBird and OrbView-3 (Carleer, Debeir & Wolff 2005). The unique characteristics of satellite imagery gave rise to a new collection of domain specific segmentation algorithms.

Segmentation algorithms commonly employed in RS image analysis can be grouped into two broad categories, namely region based and boundary based algorithms (Bins et al. 1996; Lang, Albrecht & Blaschke 2006; Zhang 1997). Region based segmentation contains the sub categories region merging, region growing and region splitting techniques. In boundary based techniques, segmentation is guided by finding edges to delineate possible segments or through gradient based watershed flooding. The choice of an appropriate segmentation algorithm is highly task dependent (Neubert, Herold & Meinel 2008). Common region based, boundary based and hybrid segmentation approaches are reviewed below and two segmentation algorithms are discussed.

2.3.1.1 Region based segmentation approaches

Seeded region growing image segmentation is an iterative process involving merging adjacent pixels with similar spectral properties. An initial random or supervised selection of seed pixels is designated, forming the basis of the seed sets. In this iterative algorithm, one additional pixel is added to a seed set based on some homogeneity criteria of the region before and after the merge. The new centre of the seed set is calculated after a successful merge (Bins et al. 1996; Freixenet et al. 2002). Pronouncements on the region growing approach are that segmentation results are dependent on the order in which pixels are merged and on the designation of seed points. Region growing algorithms typically suffer from over segmentation and under segmentation (over segmentation and under segmentation are discussed in Section 2.3.2) with non optimal parameter settings (Sonka, Hlavac & Boyle 2007). New pre-processing steps, applying edge detection and filtering techniques have been shown to increase the robustness of automatic seed point generation (Fan et al. 2005). Initially, Ketting & Landgrebe (1976) proposed a variation of this technique as an alternative to pixel based classification on coarse resolution satellite imagery.

In split and merge segmentation an image is recursively split into smaller regions if a region does not fulfil a certain homogeneity criterion. Commonly a region is split into four equal sub regions, a process referred to as quad tree segmentation. These sub regions are

subsequently compared to their neighbours. Neighbours are merged if they satisfy a certain homogeneity criterion (Lang, Albrecht & Blaschke 2006). The homogeneity criteria are typically spectral values, although variations exist that use edge detection filters as derived input to the homogeneity function (Freixenet et al. 2002). The basic variants of quad tree segmentation algorithms have varying sized and unequally distributed segments as output, which is problematic in semantic based modelling approaches. Wuest & Zhang (2009) prescribe an extension of this approach by introducing agglomerative merging and boundary refinement as post segmentation steps (based on earlier work), greatly increasing the utility of this approach for earth observation applications.

In region merging image segmentation, segments are merged with neighbouring segments according to some homogeneity criteria. In this context a segment can imply a pixel or a group of pixels. In the initial stages of the algorithm, segments consist of single pixels. As the segmentation process continues, larger segments are generated. The segmentation process ends when no further mergers in the entire image will satisfy the homogeneity criteria (Sonka, Hlavac & Boyle 2007). Many variations of this basic design exist with various aspects that need consideration, for example the (Castilla 2003):

- way in which the initial regions are chosen;
- similarity measure(s) used to merge regions;
- merging procedure (threshold and merging order); and
- stop criterion.

Baatz & Schäpe (2000) proposed a scale dependent region merging algorithm that has seen widespread implementation in earth observation applications. This algorithm is implemented in the commercial software *eCognition*. Baatz and Schäpe (2000) consider several heuristics related to the region merging process and its general effect on the creation of meaningful image segments in earth observation imagery.

A few strategies exist to decide when a segment pair should be merged. Typically, a neighbour segment of the segment of interest can be chosen arbitrarily and merged if the homogeneity criterion is fulfilled. Alternatively, the neighbour segment that best fulfils the homogeneity criteria can be chosen. In mutual best fitting, both the neighbour segment and the segment of interest are individually evaluated for neighbours that best fulfil the homogeneity criteria of a possible merge. If the neighbour segment and the segment of

interest have identified each other as best fits, the merge is performed. Baatz & Schäpe (2000) advocate the use of mutual best fitting with a distributed treatment order over the entire image, as opposed to a global mutual best fitting strategy.

Baatz & Schäpe (2000) define the degree of fitting or homogeneity criterion as:

$$h = \sqrt{\sum_d (f_{1d} - f_{2d})^2} \quad (2.1)$$

where f_{1d} and f_{2d} represent the values of two adjacent objects having *d* dimensions. In earth observation images it is typical for *d* to denote one dimensional spectral values. The degree of fitting is adjusted by adding the standard deviation (σ) over all segments (over all input dimensions) to the equation, given by:

$$h = \sqrt{\sum_d \left(\frac{f_{1d} - f_{2d}}{\sigma_{fd}}\right)^2}. \quad (2.2)$$

When two segments are evaluated for a possible merge, the change in the degree of fitting is measured as:

$$h_{diff} = h_m - \frac{h_1 + h_2}{2} \quad (2.3)$$

where h_1 and h_2 are the degree of fitting before the virtual merge and h_m the fitting after the virtual merge. The change of degree of fitting measure is extended by incorporating object size, defined by *n*, given by:

$$h_{diff} = h_m - \frac{h_1 n_1 + h_2 n_2}{n_1 + n_2}. \quad (2.4)$$

Baatz & Schäpe (2000) recommend that object size should be used as a weight to the heterogeneity measure, resulting in:

$$h_{diff} = (n_1 + n_2) h_m - (n_1 h_1 + n_2 h_2). \quad (2.5)$$

This measure of heterogeneity is extended to operate on an arbitrary number of input channels, each contributing a certain weight to the heterogeneity measure, written as:

$$h_{diff} = \sum_c w_c (n_1 (h_{mc} - h_{1c}) + n_2 (h_{mc} - h_{2c})). \quad (2.6)$$

Baatz & Schäpe (2000) broaden the homogeneity criterion by suggesting the integration of two additional heterogeneity measures based on segment form. The compactness measure is

defined as the relationship between the boundary length (l) of the segment and the square root of the number of pixels (n) in the segment, expressed as:

$$h_{compact} = \frac{l}{\sqrt{n}}. \quad (2.7)$$

The smoothness measure is defined as the relationship between the boundary length (l) of the segment and the perimeter of the bounding box (b) of the object, given by:

$$h_{smooth} = \frac{l}{b}. \quad (2.8)$$

These three measures of heterogeneity contribute different weights in the segmentation process, set by the analyst according to algorithm parameters. This segmentation algorithm of Baatz & Schäpe (2000) has three free parameters that need to be set by a user, namely:

1. The scale parameter, a unit-less measure governing the relative segment sizes defined by h_{diff} (equation 2.6).
2. The shape/colour parameter, depicting the weights these different measures should have in guiding the segment merging: h_{diff} opposed to a combined $h_{compact}$ (equation 2.7) and h_{smooth} (equation 2.8), normalised to shape/colour (range [0…1]).
3. The compactness/smoothness parameter, depicting the weights the different shape measures should have in guiding the segment merging ($h_{compact}$ opposed to h_{smooth}), normalised to compactness/smoothness (range [0…1]).

In addition to these parameters, the weights of the input bands require caution.

Figure 2.3 illustrates an arbitrary segmentation using the region merging segmentation algorithm proposed by Baatz & Shäpe (2000). A scale parameter of 40, colour/shape parameter of 0.5 and a compactness/smoothness parameter of 0.5 were arbitrarily chosen. The input band weights were evenly distributed. The yellow box in Figure 2.3 illustrates a patch of Cape thicket. A criticism of this specific region merging approach to image segmentation is the numerous free parameters that the user needs to set, usually by trial and error. Castilla, Hay & Ruiz (2008) censure this approach for incorporating two incommensurable features, namely colour and shape, in a heterogeneity measure, stating that the form criteria only lead to more visually pleasing results. The scale parameter is criticised for its unit-less and image dependent nature.

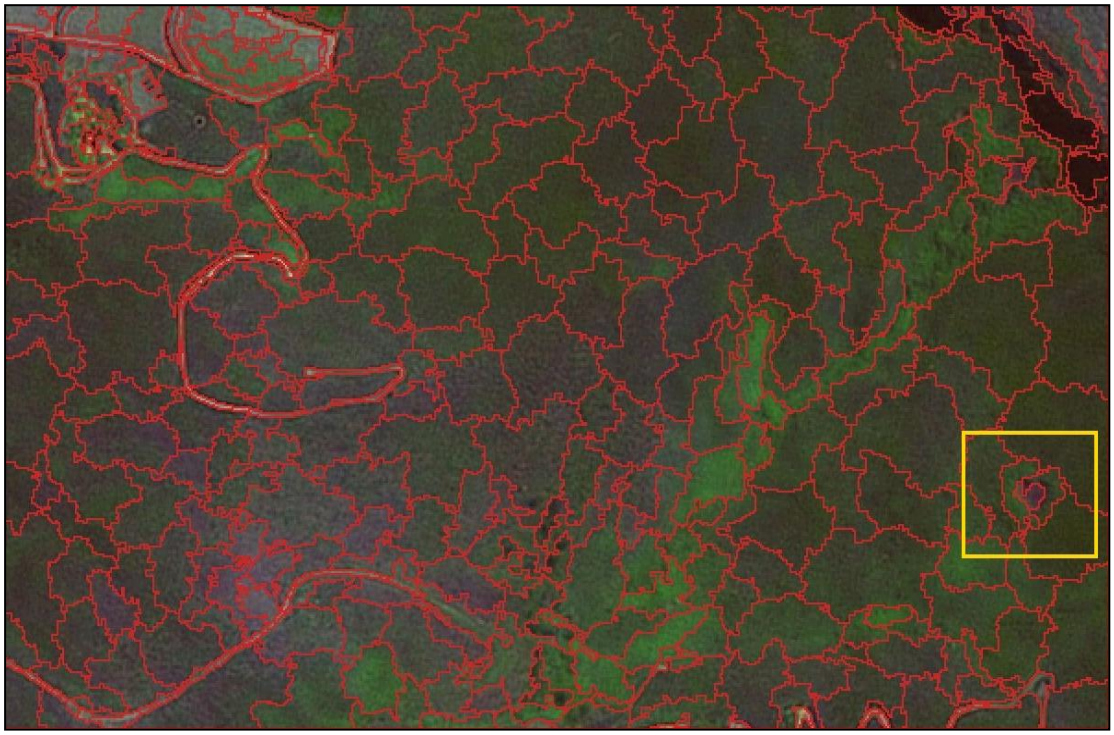


Figure 2.3: Example of an arbitrary Baatz & Schäpe region merging image segmentation of a SPOT 5 subset depicting the slopes of Bothmaskop, near Stellenbosch.

This segmentation algorithm is typically employed as a multi-scale analysis tool and has enjoyed widespread use. Over fifty percent of peer reviewed articles on applied RS employing object based image analysis use this algorithm (Blaschke 2010). Parameters are typically selected by trial and error (Feitosa et al. 2006). Tzotsos & Argialas (2006) amplify this basic region merging approach by adding global heterogeneity heuristics, edge compensation heuristics and advanced texture heuristics (Tzotsos, Losifidis & Argialas 2008).

2.3.1.2 Boundary based segmentation approaches

Edge based image segmentation relies on finding borders of homogeneous regions by means of edge detection filters, such as the Sobel, Laplace and Compass. Segments or regions are defined by post filter processes to detect edge pixels and close discontinuities (Sonka, Hlavac & Boyle 2007).

Watershed segmentation contains the properties of both edge based and region based techniques. Gradients are computed on the image (or input bands) to determine the edges of the watersheds or drainage divides (similar to a DEM). Typically, a filter is applied on the gradient image to remove irrelevant local minima pixels. Regions are subsequently grown or flooded from the local minima pixels. In the growing/flooding process, if a pixel is identified

as having two or more neighbours with different labels (e.g. four connected neighbourhood) it is considered a watershed pixel (Li & Xiao 2007; Sonka, Hlavac & Boyle 2007). In earth observation image segmentation, watershed algorithms are rarely used individually due to problems of over segmentation and they are typically employed as part of hybrid approaches (e.g. Castilla, Hay & Ruiz 2008). Li & Xiao (2007) propose a heavily modified watershed based segmentation algorithm applied to RS imagery and they cite promising results.

2.3.1.3 Hybrid segmentation approaches

Some authors (Freixenet et al. 2002; Kermad & Chehdi 2002; Munoz et al. 2003) submit that the broad segmentation types - region based and edge based - have complementary advantages and so propose the use of hybrid segmentation algorithms that have the beneficial properties of both algorithm types. Freixenet et al. (2002) overview early attempts at segmentation algorithm hybridisation.

The baseline method (Castilla 2003) is a notable example within the earth observation domain of a hybrid segmentation algorithm. An initial input image is subjected to a gradient smoothing process (gradient inverse weighted edge preserving smoothing). This filtering process creates a piecewise constant image, with each uniform region being the area of influence of a gradient minimum. Next, the image is subjected to a gradient filter and the local minima are located (standard watershed approach). A watershed partition is applied. The regions generated through the watershed partition are subjected to an innovative region merging process entitled size constrained region merging (SCRM), with the minimum mapping unit as the only homogeneity criterion. SCRM uses the original image data when deciding on a merge and not the smoothed image. This workflow is illustrated in Figure 2.4 and Figure 2.5 shows an arbitrary segmentation using this algorithm.

Castilla, Hay & Ruiz (2008) qualitatively compare this approach with other common segmentation algorithms employed in earth observation image processing. The greatest criticism levelled at the baseline segmentation algorithm is its inability to detect edges of very low contrast, leading to under segmentation in such cases (Castilla, Hay & Ruiz 2008). The yellow rectangle in Figure 2.5 depicts a Cape thicket patch and illustrates an arbitrary example of this general criticism of the baseline segmentation algorithm.

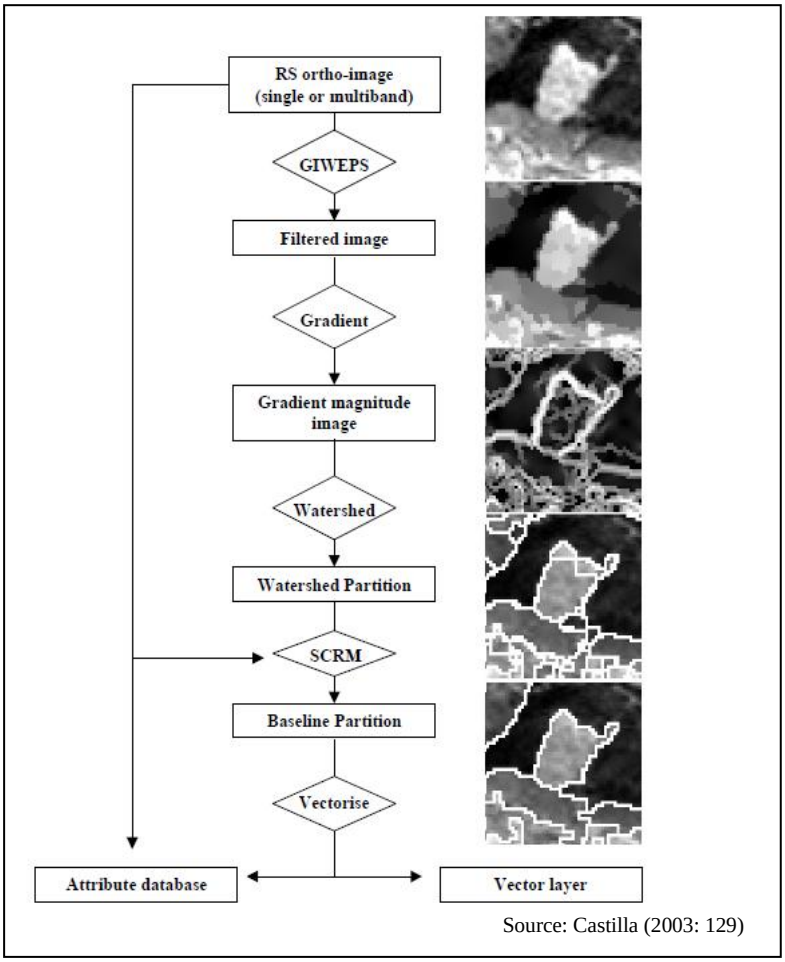


Figure 2.4: Schematic overview of the baseline segmentation algorithm

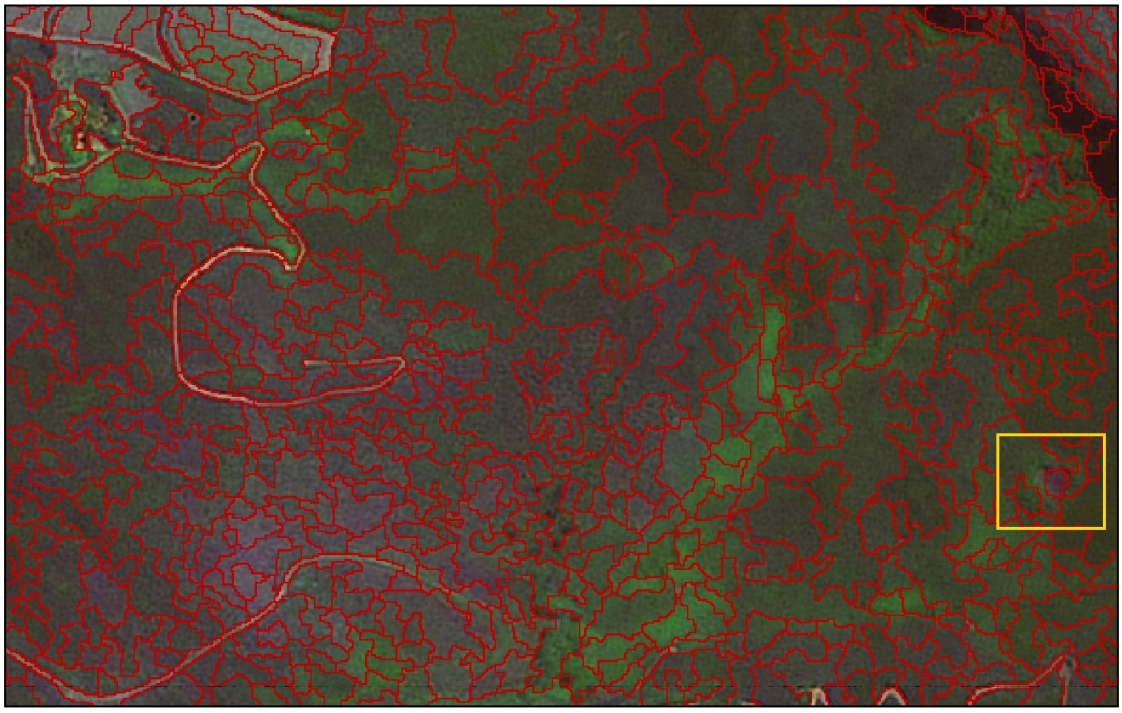


Figure 2.5: Arbitrary example of the baseline segmentation algorithm

Irrespective of the algorithm employed, a user is still confronted with numerous segmentation parameters that directly influence the results of the subsequent classification. In the next section segmentation quality measures are discussed, followed by an overview of automatic segmentation parameter tuning techniques that employ these quality measures.

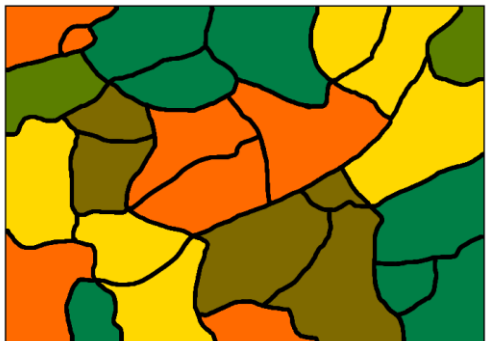
2.3.2 Segmentation evaluation techniques

An image segmentation evaluation technique, or metric, measures the quality of a segmented image in terms of supervised or unsupervised quality criteria. The traditional method of evaluating image segments, through the use of the mean opinion score, is a subjective and time consuming process (Wang, Bovic & Lu 2002). Routinely, metrics are employed to evaluate segment results, as a benchmarking tool for comparing algorithms for a specific application and for automatically tuning algorithm parameters (see Section 2.3.3) (Wang, Bovic & Lu 2002).

Measures for evaluating the quality of image segments have their origin in the field of computer vision. Many of these techniques were directly imported for use in applied RS, although a few new techniques have recently been proposed to address the unique characteristics of satellite imagery (see below). The effect of the quality of image segments on the final classification accuracy in a RS task is well documented in the literature (Addink, De Jong & Pebesma 2007; Carleer, Debeir & Wolff 2005; Dorren, Maier & Seijmonsbergen 2003; Gao et al. 2007; Kim & Madden 2006; Meinel & Neubert 2004; Wang, Sousa & Gong 2004), suggesting that adequate attention be given to the segmentation process. Despite the advice of the above authors, numerous case studies in applied RS still select segmentation algorithms and set the segmentation algorithm parameters by trial and error (Gao et al. 2007).

Frequently, a segmented image may suffer from over or under segmentation. Figure 2.6 illustrates an abstract image where the differently coloured segments correspond to the features of interest to the analyst.

Features of interest are identified by different colours. The black lines depict the edges of an ideal segmentation.



Example of over segmentation: Features are subdivided into smaller segments by the algorithm.

Example of under segmentation: Features are incorrectly merged with neighbouring features to form larger segments.

Figure 2.6: Illustrations of the concept of over- and under segmentation

If an algorithm generates too many segments with respect to the hypothetical optimal segmentation, it is referred to as over segmentation. The features of interest are subdivided into smaller segments by the algorithm. If an algorithm generates too few segments, it is referred to as under segmentation. Features are merged with neighbouring features to form larger segments. Under segmentation constitutes a bigger problem to the overall classification accuracy than over segmentation (Neubert, Herold & Meinel 2008).

This simple model of segmentation quality evaluation does not generalise well to real world problems as it does not observe segment edge offsets. The criteria for good segmentation results are usually scene and application dependent and more complex measures of segmentation quality are typically needed (Weidner 2008; Zhang, Fritts & Goldman 2008).

Zhang's (1996; 2001) overviews of the possible approaches to quantitative image segmentation evaluation are summarised in Figure 2.7.

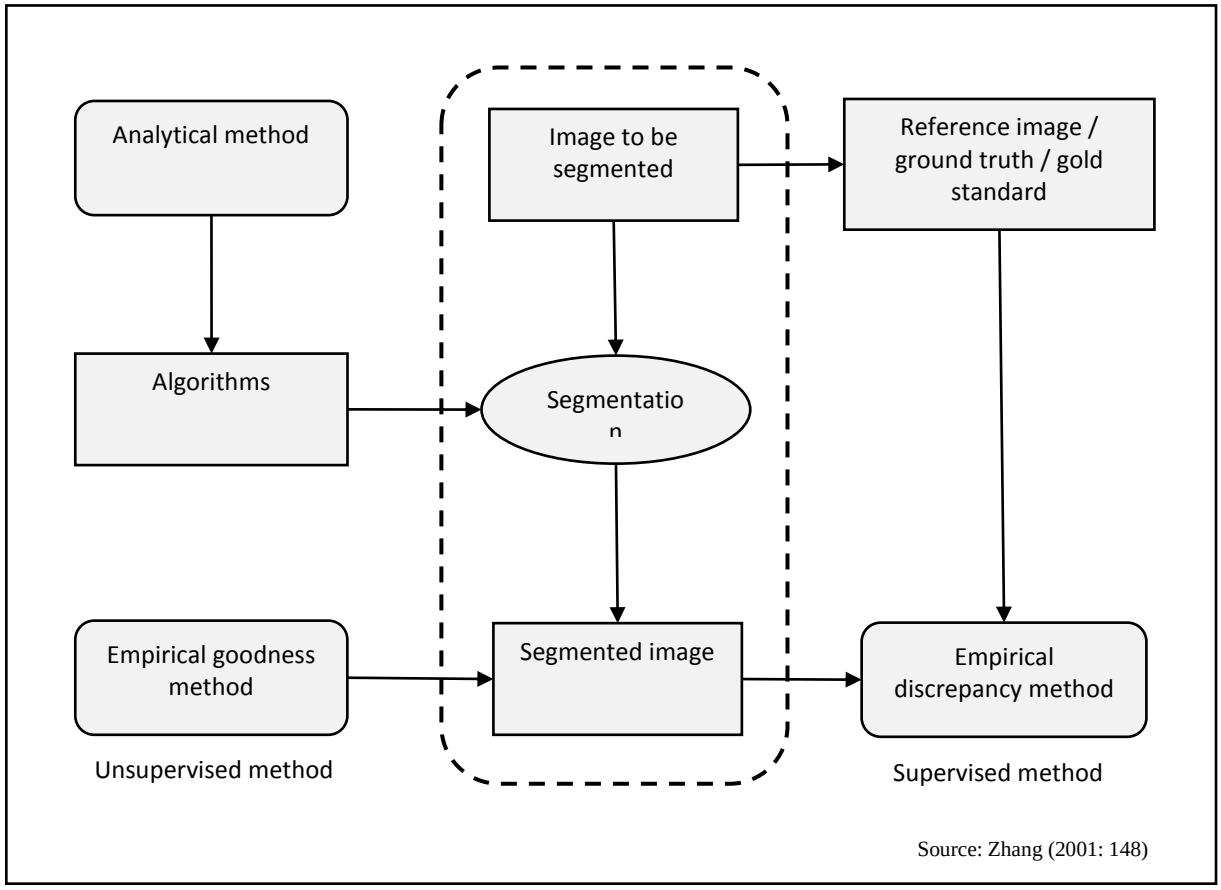


Figure 2.7: Different types of methods applied in segmentation evaluation

Three broad categories of image segmentation evaluation techniques are defined, namely analytical methods, empirical goodness methods and empirical discrepancy methods. Analytical methods for segmentation evaluation do not evaluate the results of the segmentation process, rather the intrinsic nature and behaviour of the segmentation algorithm itself. This method of evaluation is seldom used separately for algorithm and segmentation evaluation (Zhang 2001).

The empirical goodness methods evaluate the results of the segmentation process in an unsupervised manner. Notions of segment quality (given by human intuition, e.g. spectral homogeneity) are captured a priori in these types of measures that evaluate the segmented image without any analyst intervention. Usually unsupervised quality measures operate on the:

- intrasegment uniformity, based on spectral or textural properties;
- intersegment contrast, based on spectral or textural properties; and
- shape of the generated segments.

It should be noted that these criteria are also employed in the actual design of segmentation algorithms (e.g. the Baatz & Shäpe region merging algorithm) (Zhang 2001).

Empirical discrepancy methods compare the generated segments with a ground truth image; also called reference image or gold standard image. The difference between the ground truth and segmented images is quantified using some discrepancy criteria.

Typical discrepancy criteria include the (Zhang 2001):

- number of incorrectly segmented pixels, or area based metrics;
- physical position of the incorrectly segmented pixels or edge based metrics;
- number of segments in the image; and
- attributes of the segmented objects.

Zhang (1997) and Zhang, Fritts & Goldman (2008) suggest that the use of empirical discrepancy methods give more accurate segment results due to their finer resolution of evaluation as opposed to empirical goodness methods. Carleer, Debeir & Wolff (2005) criticise goodness methods on the basis that the desirable properties of the objects are chosen subjectively and the use of the same or similar measures of homogeneity as in the actual segmentation algorithm result in biased assessments. Other authors (Cardoso & Corte-Real 2005; Chabrier et al. 2006; Corcoran & Winstanley 2007; Goldman et al. 2008; Wang, Bovic & Lu 2002; Zhang, Fritts & Goldman 2008) find fault with the empirical discrepancy methods for requiring user delineated input. They contend that results are highly dependent on the quality of the user input as well as on the application, making such an evaluation method subjective. The unique ability of unsupervised methods to self tune is praised by these authors (see Section 2.3.3).

Zhang (1996) has overviewed early segmentation evaluation techniques employed in the domain of computer vision. Earth observation imagery has distinctively different characteristics to the imagery typically processed in computer vision, resulting in the development of specialised quality measures. A summary of recent empirical goodness and

empirical discrepancy methods in the earth observation image processing domain is given next.

2.3.2.1 Empirical discrepancy methods

Empirical discrepancy methods can be subdivided into two groups according to the way they evaluate segments. Typically, discrepancy methods evaluate segment edge offsets or total segment area.

Marcal & Rodrigues (2008) present a novel empirical discrepancy framework to evaluate an algorithm and accommodating parameter set for a specific earth observation application. The methodology involves generating synthetic imagery from a small set of user defined features. The synthetic image consists of a mosaic of rectangular shaped objects of different sizes having the spectral properties of the user delineated features. The analysis moves beyond the scope of the limited user input by automatically generating more training samples (semi-supervised sample generation) thus mitigating the criticism levelled against empirical discrepancy measures. This process is illustrated in Figure 2.8.

Source: Marcal & Rodriques (2008: 3)

Figure 2.8: Synthetic image generation by the framework proposed by Marcal & Rodrigues (2008)

Marcal & Rodrigues (2008) recommend the use of three discrepancy metrics in their framework, namely the Hammound distance, Rand and Jaccard coefficients. The Rand and Jaccard coefficients are measures of global categorisation accuracy operating on individual segments. The Hammound metric (H) is an area discrepancy measure indicating the percentage of correctly classified pixels. This metric measures the similarity between two segments (ground truth X and generated segments Y) by subtracting the number of non-shared pixels from the number of shared pixels, normalised by the number of shared pixels, as in:

$$H = \frac{\#(X \cup Y) - \#(X \cap Y)}{\#(X \cup Y)}, \in [0 \ldots 1]. \quad (2.9)$$

The # symbols signify the number of pixels in the following brackets. This metric can be modified to measure the percentage of incorrectly classified pixels (Carleer, Debeir & Wolff 2005), given by:

$$H = \frac{\#(X \cap Y)}{\#(X \cup Y)}, \in [0 \ldots 1]. \quad (2.10)$$

Marcal & Rodrigues (2008) propose the use of the Hammound metric for evaluating the ability of an algorithm for segmenting a specific feature type. An observed shortcoming of this technique is that numerous border types are generated in the synthetic image, many probably never occurring in the real imagery. This would over generalise the evaluation if the real world features of interest have a limited range of border types.

Weidner (2008) extends the basic area metric (equation 2.10) by incorporating a measure of geometry. The distances to the reference segment boundaries are incorporated in the metric with the total matching area by using weighting functions. Weidner (2008) advises that the choice of metric should be guided by the application as the proposed hybrid measure does not generalise well for all feature types. Möller, Lymburner & Volk (2007) propose a similar approach that combines notions of image area overlap and geometric centres.

Another area metric employed in the RS domain that combines the notions of correctly classified pixels and incorrectly classified pixels is the area fitness rate, given by:

$$Area\ fitness\ rate = \frac{X \cap Y}{X} \times \frac{X \cap Y}{Y}, \in [0 \ldots 1]. \quad (2.11)$$

Lucieer (2004) used this measure in modelling uncertainty in the visualisation of geographic data. Lübker & Schaab (2009) propose an iterative segmentation evaluation process, involving an empirical discrepancy method (equation 2.11) for broad level segment evaluation followed by the goodness method proposed by Espindola et al. (2006) to fine tune segmentation algorithm parameters with a lesser influence on the overall results (see Section 2.3.2.2).

Trias-Sanz, Stamon & Louchet (2008) offer an edge based empirical discrepancy method that operates in a multi-scale segmentation framework. This framework addresses the notion that

under segmentation is a greater problem than over segmentation. Different penalisation weights at different scales of observation are assigned for instances of under and over segmentation.

Prieto & Allen (2003) present a pixel based similarity measure (unapplied in an earth observation image processing context). Pixels in the reference and generated images are compared on a per-pixel level. A bipartite graph matching algorithm pairs the pixels in the reference and generated images. The metric computes the distances between matched pixels and the number of unmatched pixels. Jiang et al. (2006) criticise edge metrics and especially the approach by Prieto & Allen (2003) for their inability to correctly quantify region based segmentation results. Jiang et al. (2006) note that a small error in the metric can correspond to a critical edge not being segmented – possibly resulting in a large under segmentation error. More detail on this pixel based similarity measure is given in Chapter 5 where it is employed as part of a parameter tuning system.

Neubert, Herold & Meinel (2008) overview the performance of many segmentation algorithms employed in satellite RS. Various simple empirical discrepancy measures are used by these authors to evaluate segmentation algorithms for standard land cover classification tasks. They advise that careful consideration should be given to the choice of segmentation algorithms as the performance of said algorithms very much depends on the nature of the segmentation task. They conclude that no single evaluation method can be considered superior.

2.3.2.2 Empirical goodness methods

Espindola et al. (2006) suggest the use of an empirical goodness method that employs measures of intrasegment homogeneity (variance) and intersegment heterogeneity (spatial autocorrelation). Their proposed methodology operates on region merging segmentation algorithms and can be considered a post-segmentation global evaluation pass. They provide a case study using the region merging algorithm proposed by Baatz & Shäpe (2000). A similar approach proposed by Gao et al. (2007) operates on region merging algorithms. Chabrier et al. (2006) and Corcoran & Winstanley (2007) commend the use of texture measures in unsupervised segmentation evaluation.

Radoux & Defourney (2006) use the Bhattacharyya distance (Bhattacharyya 1943) as an unsupervised measure of segment quality evaluation prior to classification. A similar approach is followed by Wang et al. (2004) to derive the optimal scale parameter for mapping mangrove forest. They concluded that the use of the Bhattacharyya distance to select the critical scale parameter was useful as this methodology is classification result independent. It should be noted that the Bhattacharyya distance assumes a Gaussian distribution of the features of interest - an assumption that will not always hold for a wide range of earth observation applications.

2.3.3 Approaches to automated and semi-automated segment generation

The automatic generation of good quality image segments in the earth observation domain has recently attracted research attention due to the limitations of the time consuming and subjective manual tuning process. The metrics briefly discussed in Section 2.3.2 could be used in automatic segmentation generation systems. Lübker & Schaab (2009) caution that systems attempting to automate the generation of adequate segments in geographic object based image analysis should:

- be objective by using statistical methods such as the metrics discussed in Section 2.3.2;
- find optimized settings for each class or group of classes separately;
- cover a representative number of objects per class preferably in more than one test area;
- account for all degrees of freedom of the algorithm (parameter space);
- test parameter settings in small increments; and
- preferably be automated to a large extent.

All of the studies reported in Section 2.3.2 use metrics to compare algorithms for a specific application or to assess the capability of a specific algorithm for a narrow range of applications. Three limitations of these general approaches followed by the authors should be noted:

1. The parameter spaces of the algorithms were explored in a manual or semi-automatic manner.
2. The resolution of the exploration of the parameter space was generally very coarse. These studies typically evaluate a limited set of combinations of parameters. An

exception is the study by Lübker & Schaab (2009) that explores 2400 parameter combinations, in five study sites for four groups of classes.

3. The search in the parameter space was unsupervised, resulting in prohibitively long computation times, especially for large sets of algorithm parameters. (Lübker & Schaab (2009) is a fitting example). In an interactive system, prolonged computation time is a serious drawback as it prohibits the systematic exploration of parameter space.

A noteworthy caution is that the choice of the metric used be guided by its applicability to the features of interest and be backed by quantitative comparisons (Feitosa et al. 2010).

Two broad approaches to automated image segmentation in the domain of earth observation have been proposed recently (based on work originating in computer vision), namely a) techniques focussing on local segment optimisation and predictive behaviour, and b) a global optimisation approach making use of recent numerical optimisation techniques and adequate fitness functions for segment generation (more specifically, segmentation algorithm parameter exploration). The first approach is briefly described in the following subsection and the second one is expounded in the next subsection.

2.3.3.1 Local segment optimisation

Zhang & Maxwell (2006) put forward a semi-automated segmentation parameter tuning system operating on the region merging segmentation algorithm of Baatz & Shäpe (2000). The workflow starts by over segmenting an image with an arbitrary small scale parameter. All the segments comprising a feature of interest are selected by the analyst. A fuzzy inference system is used to predict the appropriate set of parameters that would result in the segmentation of the selected sub-segments as a single segment. The system has the advantage of not requiring accurate user delineated input from a digitising exercise.

Osman, Inglada & Christophe (2009) present a semi-automated segmentation generating system operating on local only features. This approach performs segmentation on input data transformed by a maximum margin classifier. The system asks the analyst to digitise a limited number of lines within the feature of interest, as well as a few lines on the outside of the feature. The system automatically transforms the input data (e.g. NDVI) followed by unsupervised clustering of the pixels in the input samples. An SVM is used to find an optimal

margin between the inside and outside of the feature (based on the pixels around the digitised lines). Finally, a region growing algorithm is applied with seed pixels in the feature of interest to generate the segment. A general limitation noted about these techniques is that they function on single features only and do not generate mechanisms that can be applied to unseen areas.

2.3.3.2 Numerical optimisation techniques for segmentation algorithm free parameter tuning

Population based search or optimisation techniques, especially evolutionary algorithms, are well suited to address problems of high dimensionality and complex search space behaviour as they typically have more favourable convergence times than linear search functions. An overview of such techniques (evolutionary search strategies) and the reasons why they are effective are given in Section 2.5.

Typically, evolutionary algorithms are employed in two distinct ways in the image segmentation process, namely as a tool for segmentation algorithm parameter selection and in pixel level image segmentation (Farmer & Shugars 2006; Woods 2007). Commonly segmentation algorithms have a set of inter-dependent free parameters that needs to be tuned. This search space is typically large and shows a complex behaviour regarding resultant segment quality. Bhanu, Lee & Ming (1995) first proposed the general use of evolutionary algorithms for segmentation algorithm free parameter tuning in the field of computer vision. Evolutionary algorithms can be employed to search this complex parameter space for an optimum in a supervised or unsupervised manner. See Zingaretti, Tascini & Regini (2003) for an example of an unsupervised implementation.

Feitosa et al. (2006) introduce a supervised search methodology operating on earth observation imagery (urban scene) and the Baatz & Shäpe (2000) region merging segmentation algorithm. They note that the relationship between algorithm parameters and segmentation results are complex and cannot be measured analytically. A standard genetic algorithm is invoked with an area based empirical discrepancy metric as fitness function (see Section 2.5). Due to the high computational load of a segmentation algorithm, a directed search strategy such as a genetic algorithm is suggested to allow reasonable segmentation algorithm parameters to be found in an acceptable time span.

Feitosa et al. (2006) noted that in a practical setting an analyst will delineate the ground truth areas to be segmented (using empirical discrepancy metrics), but in their initial experiments these authors chose segments generated with the algorithm as the ground truth. The methodology produced promising results by converging on parameters closely resembling the ground truth segments, thus recommending the utility of evolutionary search strategies for quick parameter space exploration applied to RS problems. They also noted that in some cases multiple suitable or "correct" parameters were obtained for some samples. The fitness function used (empirical discrepancy metric), is called the reference bounded segments booster, and it is expressed as:

$$F(S,P) = \frac{1}{n}\sum_{i=1}^{n}\frac{\#(S_i - O(P)_i) + \#(O(P)_i - S_i)}{\#(S_i)} \qquad (2.12)$$

where S denotes the set of segments under consideration and n the number of segments in the set S. S_i is the pixels of the i-th segment of the set S. $O(P)_i$ denotes the set of pixels belonging to the segment with the largest intersection with S_i among the segments produced by using P as parameter values of the segmentation algorithm. This metric is illustrated in Figure 2.9. The blue segment depicts the reference segment and the dashed red segment is a generated segment with the largest overlap with the reference segment. This metric is a multi-segment extension of equation 2.9.

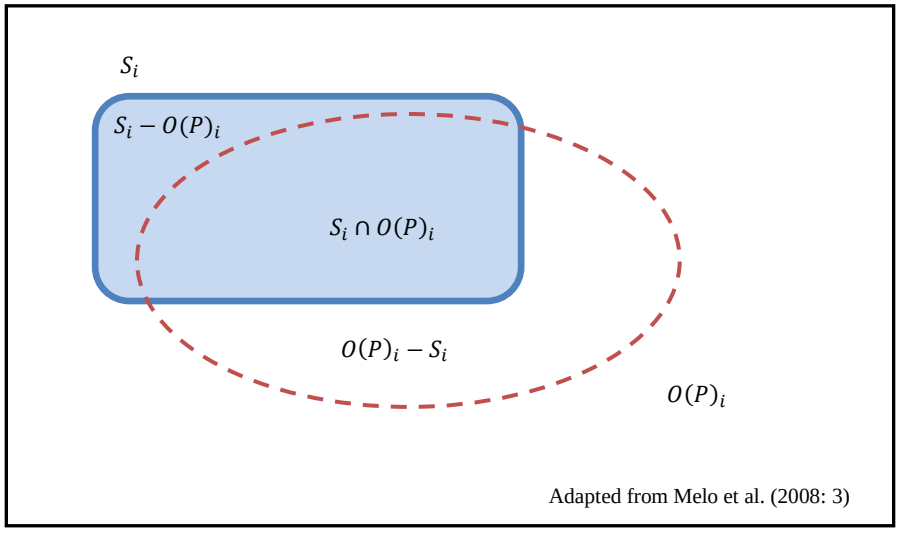


Figure 2.9: Reference bounded segments booster metric

Melo et al. (2008) present experiments with additional evolutionary search strategies to this general approach. They reported that by using a quantum inspired evolutionary algorithm and a differential evolution (DE) search strategy the computational load of the search was reduced

by 44% and 50% respectively for their case studies compared to a standard genetic algorithm. They also noted a slight increase in convergence accuracy by using these algorithms.

Fredrich & Feitosa (2008) extend this work by:

- introducing an additional metric that is not limited to matching one generated segment to one reference segment, entitled the larger segments booster (LSB); and
- introducing a post-segmentation heuristic to allow an analyst to automatically delineate inhomogeneous segments consisting of homogeneous sub-segments.

Figure 2.10 illustrates the components of the LSB metric. The LSB metric is:

$$F(S,P) = \frac{1}{n}\left[NS + \sum_{SO(P)_i \neq \emptyset} \frac{fp_i + fn_i + b_i}{\#(S_i)}\right], if\ NS < n \qquad (2.13)$$

where n denotes the number of reference segments delineated by the user. $\#(S_i)$ is the area (in pixels) of the i^{th} reference segment. $SO(P)_i$ is the set of segments produced by the segmentation algorithm and possessing at least half of its pixels intersecting S_i. P is the parameter set used in the segmentation algorithm. Further, fp_i is the number of pixels in $SO(P)_i$ that do not belong to the i^{th} reference segment, called the false positives (see Figure 2.10). And fn_i is the number of pixels in the i^{th} reference segment that do not belong to $SO(P)_i$, called the false negatives; b_i denotes the number of border pixels in $SO(P)_i$ that interstect the i-th reference segment area. NS denotes the number of empty $SO(P)_i$ (empty generated segments).

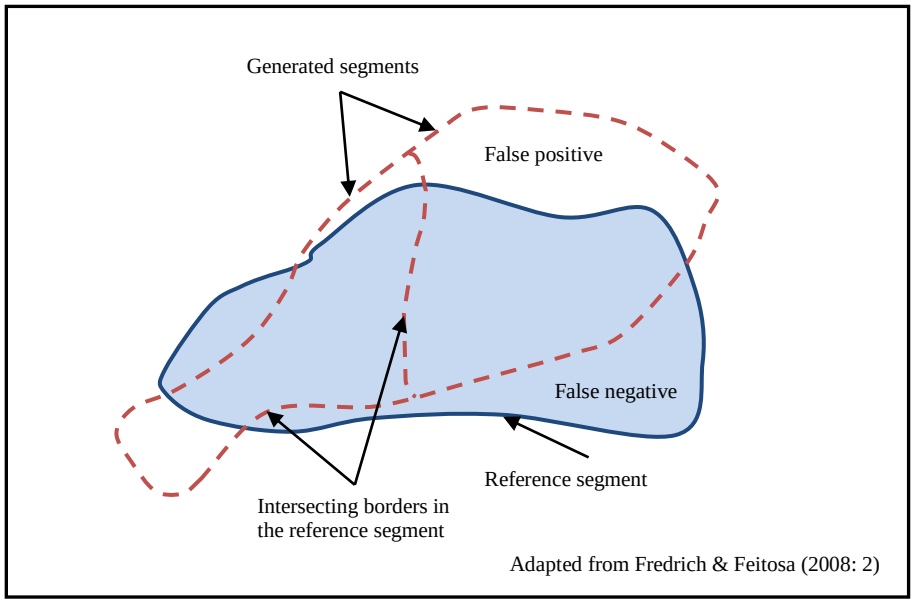


Figure 2.10: Larger segments booster (LSB) metric

This metric favours generated segments with a tight fit to the reference segment. The parameter *b* provides an offset to this notion by penalising a parameter set for over segmentation by counting the pixels in the intersecting borders of the reference segment (Cazes et al. 2008). A prime advantage of the LSB metric is that it does not assume that a single generated segment will be paired with a user created reference segment. Feitosa et al. (2009) presents a quantitative comparison of the abilities of the LSB and reference bounded segments booster metrics for segmenting urban land cover features (houses with discrete boundaries).

Although the general approach to parameter tuning advanced by Feitosa et al. (2006), Melo et al. (2008), Fredrich & Feitosa (2008) and Feitosa et al. (2009) shows promise, these authors stress a few limitations or considerations to this approach. First, such a system may suggest multiple parameters as very adequate, especially if the problem posed by the analyst is relatively simple. That is, numerous sets of vastly different segmentation parameters may be adequate for a specific (possibly very simple) application. Second, the authors warn that meaningful training samples are necessary to deliver good results. The training samples should all be delineated at a similar scale, to wit the samples should be similarly sized. Last, it is possible that no set of parameters is able to segment the features delineated by the analyst, either due to faulty feature delineation or purely the inability of the segmentation algorithm to generate the requested segments.

Note that the experiments conducted by the above authors have thus far focussed on very particular segmentation problems within an urban setting, namely segmenting buildings and buses. These objects display sharp and homogeneous boundaries. It is hypothesised that the accurate analyst delineation of features may be improved with an edge detection algorithm or a line matching heuristic to identify the true edges the analyst intends to delineate, for example by using a live wire algorithm (Falcao et al. 1998). Such an approach has not been extended to or tested on imagery of natural environments that have less distinct boundary characteristics (this notion is explored further in Chapter 5).

The choice of image band input (or colour channels when employing aerial photography) in the segmentation process also influences the segmentation results. This effect is explored next.

2.3.4 Influence of the choice of input bands on image segmentation results

The data input to a segmentation algorithm is usually a raster image with an arbitrary number of channels or input bands. Most segmentation is performed on grey scale images, or images consisting of only one band. Segmentation can also be performed on multi-band imagery (e.g. colour photographs or earth observation imagery) by treating the bands individually and summarising the results (see equation 2.6 for an example). Non-default colour spaces or input channels can be used as input to the segmentation algorithm and the default input channels can be subjected to linear or nonlinear transformations. The performance of an image segmentation algorithm is known to depend on the choice of input channels (Cheng et al. 2001; Macaire, Vandenbroucke & Postaire 2006; Pal & Pal 1993; Vandenbroucke, Macaire & Postaire 2003), in addition to the free parameters that usually needs to be set.

Cheng et al. (2001) and Vandenbroucke, Macaire & Postaire (2003) comment that the selection of an adequate colour space in image segmentation problems has attracted little attention in the (computer vision) literature. Although most reported studies in the field of computer vision use the standard RGB colour space as basis, Pal & Pal (1993) hold that a colour image is a special (simplified) case of a multispectral (multidimensional) image, and that algorithms and methodologies are interchangeable between these two modes of data representation.

Cheng et al. (2001) overview the common linear and nonlinear colour transformations (e.g. RGB, HSI, CIE, Nrgb) and their applications in typical image segmentation algorithms. They note that no colour transformation can be considered generally superior and that the choice of effective colour input depends on the image to be segmented and on the features of interest. They also affirm that the hue, saturation and intensity (HSI) transformation proves useful in segmenting features with strong highlights, shadow and texture interference - cases where the assumption of feature colour homogeneity does not hold. They warn that this nonlinear transformation may lead to anomalies in the data at low saturation levels.

Vandenbroucke, Macaire & Postaire (2003) address the choice of input channels by proposing a supervised colour space selection system. The proposed supervised system automatically creates hybrid colour spaces consisting of core components of basic and transformed colour spaces. These hybrid colour spaces are evaluated on their ability to effectively segment the

features of interest. This extended colour input space is explored by a sequential forward selection algorithm. The authors present promising results, although some limitations (it is a manual process and the colour space exploration is not thorough) to this system are noted. Busin et al. (2009) put forward an unsupervised variation of this approach that use a one dimensional histogram thresholding to determine the separability of the different hybrid colour spaces.

In the domain of applied RS, authors typically choose adequate bands for segmentation on trial and error, through expert knowledge of the nature of the features of interest or by plainly using the default bands of the sensor. For example, Mallinis et al. (2008) advocate the use of HSI and vegetation indices transformations as channel input for Mediterranean vegetation segmentation on QuickBird imagery. They note that different transformed colour spaces produced optimal results at different scales of observation when using a scale dependent region merging algorithm (the Baatz & Shäpe region merging algorithm). Trias-Sanz, Stamon & Louchet (2008) provide a case study of the use of transformed colour spaces for segmenting aerial photography of rural areas. They investigated the use of numerous colour transforms (and hybrid combinations thereof) in a stepwise forward selection process for landform segmentation (Guigues' scale sets hierarchical segmentation algorithm). They found that the use of transformed colour spaces rather than the standard image bands, improved the quality of the generated segments.

Texture can also be used as input to a segmentation process, although very few case studies exist where it is employed in earth observation image segmentation. Tzotsos, Losifidis & Argialas (2008) and Johansen et al. (2007) report promising segmentation results when employing texture bands in a region merging segmentation algorithm, while Triaz-Sanz, Stamon & Louchet (2008) report that using texture as additional input bands did not result in better segmentation.

This section has given a general review of the concept of input channel selection to segmentation, its effect on segmentation quality and several examples of proposed methodologies from the computer vision and RS domains. No study has yet attempted to investigate the usefulness of selecting optimal input channels automatically in conjunction with an automated segmentation parameter tuning system that employs empirical discrepancy methods. In a supervised parameter tuning system, the features or areas of interest are already

known or given a priori by the analyst, along with their spectral and boundary characteristics which suggests that the choice of data input could be explored in an automated manner. A simple methodology to investigate adequate segment algorithm input band selection is presented in Chapter 4.

2.4 IMAGE CLASSIFICATION TECHNIQUES

The performance of a supervised classifier depends on the interrelationship between the sample sizes, the number and quality of attributes and the classifier complexity (Jain, Duin & Mao 2000), necessitating careful consideration of the choice of a classifier for a specific application. It is well known that no single classifier or learning algorithm can outperform other algorithms over all datasets (Kotsiantis 2007). In this section an overview is given of the supervised statistical classification methods selected for the implementation and subsequent use in the comparative analyses. Synopses of other prominent methods are given to provide a qualitative comparison with the selected techniques.

2.4.1 Supervised classification approaches

In supervised image classification a set of training data is used to construct a classifier to predict the membership of unseen samples. Supervised classifiers can be categorised according to two different criteria, either by their use of parameters in estimating class probability (parametric and non-parametric classifiers) or in the methodology employed to model class locality (boundary based and density estimation methods) (Jain, Duin & Mao 2000).

Parametric classifiers assume an a priori distribution of the data (typically a Gaussian distribution). The parameters, for example mean vector and covariance matrix of the distribution, are computed from the training data. The common maximum likelihood (parametric, density estimation method) classification strategy uses these parameters to construct an estimated distribution for a class. The probability that a feature of interest, for example a pixel or object, belongs to a specific class can be calculated from this distribution. A criticism of parametric classifiers applied in earth observation problems is that the features of interest may not display a normal distribution (or the distribution of the features is difficult to predict), resulting in poor classification accuracies. Using a parametric classifier leads to

difficulties in integrating ancillary data with the spectral data (Lu & Weng 2007). Parametric classifiers are also very sensitive to the adequate selection of training data (Campbell 2002).

Non-parametric classifiers do not assume a known distribution of the data and are generally more universally applicable than parametric classifiers (Schowengerdt 2007). A frequently employed simple non-parametric density estimation classifier in the field of RS is the k-nearest neighbour classifier (KNN). For categorical data (classes) defined in Euclidian space, the definition of the KNN rule is extremely simple. A pixel or segment is assigned a label according to the majority label of the k-nearest neighbour training pixels or segments. The variable k is typically set by the analyst. The KNN classifier is often employed in operational RS applications (Franco-Lopez, Ek & Bauer 2001; Haapen et al. 2004) and it is used as a benchmark to evaluate more complex statistical or expert knowledge classifiers (Mallinis et al. 2008; Muñoz-Marí, Bruzzone & Camps-Valls 2007; Sanchez-Hernandez, Boyd & Foody 2007b). The KNN method may suffer, relative to other techniques, when insufficient training data are available or in cases of high dimensionality input data (Muñoz-Marí, Bruzzone & Camps-Valls 2007). In this study the KNN method is employed as a benchmark classifier.

Non-parametric classifiers can be further divided into two groups according to whether the classes can be separated linearly or not in the native input space (no dimensionality transformations) (Theodoridis & Koutroumbas 2006). The classes to be separated in earth observation imagery are not always linearly separable (at least in the default Euclidian input space). In the past two decades, a few promising non-parametric non-linear non-probability density supervised classification techniques have been applied in RS classification problems, with promising results. Three approaches are identified, namely decision tree methods, multilayered perceptrons (a single layer perceptron being a linear classifier) and support vector machines (being a non-linear classifier if employing a non-linear kernel).

Synopses of decision tree methods and multilayer perceptrons including their advantages and shortcomings are given. The family of non-statistical classification procedures (expert systems) is noted. Later an overview is given of support vector machines, the main statistical classification technique employed in this study, along with variants of this basic technique and comparative examples to other methods as applied in the domain of RS.

2.4.1.1 Decision trees

Decision tree induction has recently attracted much attention as an alternative classifier in RS due to some nominal advances over simpler methods. Three broad types of decision trees are identified, namely orthogonal, oblique and genetic decision trees (Huang et al. 2007). Univariate orthogonal decision trees are the most prevalent type employed in RS image classification, although some authors (e.g. Huang et al. 2007) note that the use of genetic decision trees is unexplored in RS image classification problems.

Decision trees are praised for their ability to handle data of different scales and types, their non-linear relationship between features and classes, their non-parametric nature, the comprehensibility of the results, relative autonomy of the technique, and robust classification performance (Kotsiantis 2007; Lewis 2000; Pal & Mather 2003; Skidmore 2002). A major shortcoming of decision trees is the high variance in tree structure – a small change in training data may lead to a totally different tree. Decision trees are also sensitive to the quality and quantity of the training samples (Pal & Mather 2003; Rogan et al. 2008; Theodoridis & Koutroumbas 2006). Mallinis et al. (2008) and Rogan et al. (2008) present case studies to determine the utility of employing decision trees in land cover classification using object based and pixel based methodologies respectively. They report promising results when compared to standard classification approaches.

2.4.1.2 Artificial neural networks

An artificial neural network (ANN) is a mathematical model consisting of non-linear computational elements, called neurons, operating in parallel and connected by links characterised by different weights (Pacifici et al. 2008). Two basic types of neural networks have been identified, namely feed forward and back propagation networks. In back propagation networks, connections are present that lead signals back into the network from the output. The most commonly applied ANN classifier in RS is the feed forward multilayer perceptron, an extension of the single layer perceptron that is able to handle non-linearly separable data (Lu & Weng 2007).

The performance of a multilayer perceptron network is highly dependent on the topology of the hidden layers (Pal & Mather 2003). Too few neurons can lead to overgeneralisation and too many can lead to an over fitting in predictions. ANNs are commended for their non-parametric nature, adaptability to different data structures, robust generalisation and

classification capabilities, and ability to handle non-linear relationships between input and output features (Lu & Weng 2007). General criticisms directed towards ANNs are the difficulty in designing the architecture of the classifier as well as a tendency of such a classifier to fall into local optimums (Pacifici et al. 2008). Pacifici et al. (2008) points out that SVMs may generally achieve similar classification accuracies as ANNs without the same amount of effort directed toward architecture design and parameter tuning.

It is important to note that decision trees and ANNs have structures based on heuristic (structural) arguments, thus lacking firm mathematical foundations for their general capabilities and performances (Oosthuizen 2008; Shawe-Taylor & Cristianini 2004).

2.4.1.3 Expert systems

Expert systems are a branch of applied artificial intelligence. Simply put, the basics of expert systems is that expertise, or the vast collection of knowledge regarding a classification problem, are transferred from an expert to a computer. This body of knowledge can subsequently be used by the computer in addressing various problems (Liao 2005). In dealing with RS classification problems, two prominent expert system based strategies are noted, namely rule based classification and fuzzy classification systems.

The literature abounds with case studies employing rule based fuzzy classification in object based RS applications, typically employing *eCognition* as the implementation tool (Blaschke 2010; Lucas et al. 2007). Although expert systems methodologies constitute a promising set of tools to tackle classification problems in cases where supervised statistical classifiers may struggle in finding patterns, an inherent limitation of such an approach is the completely manual and time consuming process of classifier construction which effectively rules out this approach for application in this research.

2.4.1.4 Combined expert system and statistical classifier approaches using ancillary data

In practical RS applications, the use of the available spectral data only for classification may not be sufficient to accurately identify the features of interest. Case studies consequently employ other data sources in conjunction with the spectral data in numerous ways to improve the overall accuracy of the results. Examples include the use of digital elevation model (DEM) data (and derivatives thereof) and context information as extra channel input in

statistical classification (Mallinis et al. 2008; Sesnie et al. 2008); DEM data/context information in a post statistical classification expert systems refinement process (Shrestha & Zinck 2001); context information (e.g. expert knowledge) in masking a scene to narrow the geographical extent of statistical classification (Trias-Sanz 2006); and DEM and GIS derived information as channel input to a segmentation algorithm (Kim, Xu & Madden 2006).

In the case study investigation of this thesis, a masking approach (using context information) will be followed to restrict the geographic extent on which the statistical classifiers operate in an attempt to improve the classification accuracy.

2.4.2 Support vector machines

As a modern discipline, pattern analysis has undergone three revolutions in algorithm design, namely the formulation of robust linear algorithms for vector data in the 1960s, the introduction of non-linear algorithms (artificial neural networks and decision trees) in the 1980s and the introduction of kernel based learning in the mid 1990s (Shawe-Taylor & Cristianini 2004). Kernel based learning is a statistical learning technique first proposed by Vapnik (1995). Kernel methods combine the theoretically well founded approaches of linear pattern analysis algorithms with the flexibility of non-linear algorithms such as artificial neural networks.

The support vector machine (SVM) classifier (Vapnik 1995) has emerged as a theoretically superior and popular binary, kernel based statistical classifier. Tax & Duin (1999) and Schölkopf et al. (2001) proposed one class or anomaly detector variants of the SVM. In this research the potential of a free parameter and attribute optimised one class kernel classifier, operating on object based data, for classifying sparse natural features is investigated. In this section the basic principles of kernel method classification and two one class variants thereof are set out. Kernel functions are briefly discussed and strategies for free parameter tuning are outlined. Case studies and comparative studies from the RS and non-RS literature that employs these techniques are commented on. Some open research questions regarding these techniques are reported and reasons are given for applying this classifier in this thesis.

2.4.2.1 Overview of the support vector machine classifier

A support vector machine is a binary statistical classifier composed of two parts. The first part is a component that uses quadratic or linear programming to find an optimal separating

hyperplane (line) between the two sets of samples in some defined feature space. The second part is a mapping function, also called a kernel function, which transforms the input space to an arbitrary higher dimensional feature space which allows a better linear discrimination to be constructed by the first component.

The first component is explained by example. Say a training set is represented by $[x_i,y_i]$, i = 1,2, ..., N, $y_i \in [-1,+1]$ and $x \in \mathbb{R}^2$. The symbol *x* denotes the samples and *y* their corresponding labels. Assume that the two classes are linearly separable. An SVM attempts to find an optimal line, also called a hyperplane, which separates the two populations. The hyperplane is defined by a standard linear function:

$$f(x) = w \cdot x + b = 0 \qquad (2.14)$$

such that $y_i(w \cdot x_i + b) \geq +1$ for all $y_i == +1$ and $y_i(w \cdot x_i + b) \leq -1$ for all $y_i == -1$. Although numerous linear functions may possibly separate the samples of the two classes, only one optimal hyperplane exists that separates the two classes with the greatest margin.

Figure 2.11 illustrates such an example with the optimal separating hyperplane shown by a blue line with samples of the +1 and -1 classes on either sides of this hyperplane. The brown lines define the bounds of the maximum margin hyperplane, called the support planes.

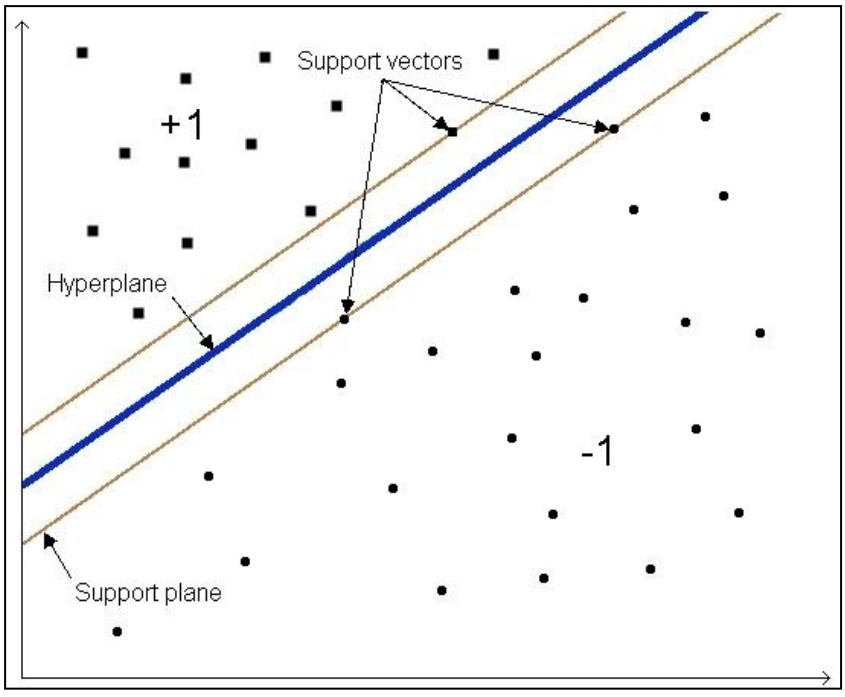


Figure 2.11: Optimal margin hyperplane separating the samples of two arbitrary classes

The aim of an SVM classifier is to find the optimal hyperplane, or hyperplane with the largest geometric margin between the samples. By using the optimal hyperplane as the objective function in classification, the most robust generalisation capability of the classifier is obtained, as opposed to an arbitrarily chosen sub-optimal hyperplane. The samples that lie on the support planes are called the support vectors. The algorithm only needs these vectors to construct the classifier (the classifier is aptly named a support vector machine). The support planes are arbitrarily defined as $w \cdot x_i + b = \pm 1$.

An SVM functions by maximising the distance between the support planes (with the aid of component two, described later) by slowly pushing the support planes outward until a few support vectors are encountered and no further maximisation of the hyperplane margin is possible. During this process it is required that samples remain on the correct side of the support planes following the requirement $w \cdot x_i - b \geq 1$ for samples of class +1 and $w \cdot x_i - b \leq -1$ for samples of class -1.

The margin between the support planes can be expressed as $\frac{2}{\|w\|}$ by a simple rescaling of the hyperplane parameters *w* and *b*. The problem of maximising this value can be rewritten as:

Minimise $\frac{1}{2}\|w\|^2$

Subject to $y_i(w \cdot x_i + b) \geq 1 \; for \; i = 0,1, \ldots N.$ (2.15)

This optimisation problem constitutes the basic definition of an SVM. The above problem can be rewritten as a Lagrange function. SVMs are typically expressed and solved using this mathematically simplified dual formulation due to its simpler constraints (Bennett & Campbell 2000), given by:

Maximise $\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j=1}^{N} \lambda_i \lambda_j y_i y_j (x_i \cdot x_j)$

Subject to $\sum_{i=1}^{N} \lambda_i y_i = 0 \; and \; \lambda_i \geq 0, i = 1,2, \ldots N$ (2.16)

where λ_i is the Lagrange multiplier.

Samples of two different classes are typically not linearly separable, with outlier samples possibly obstructing any form of linear separation. The basic form of the SVM is modified to

accommodate such an eventuality by introducing a relaxation heuristic to the problem in the form of a slack variable. Equation 2.15 is rewritten as:

Minimize $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} z_i$

Subject to $y_i(w \cdot x_i + b) \geq 1 - z_i \; for \; i = 0,1, \ldots N$ (2.17)

with *C* as a user defined variable, called the slack variable. This model simultaneously attempts to maximise the buffer of the hyperplane and minimise the amount of slack or samples removed from the problem. The Lagrangian dual is reformulated as:

Maximise $\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j=1}^{N} \lambda_i \lambda_j y_i y_j (x_i \cdot x_j)$

Subject to $\sum_{i=1}^{N} \lambda_i y_i = 0 \; and \; C \geq \lambda_i \geq 0, i = 1,2, \ldots N.$ (2.18)

This formulation of the SVM is extended to handle non-linearly separable problems by mapping the input space (in this example a two dimensional plane) to an arbitrary higher dimensional feature space (component two of the SVM formulation). A SVM attempts to construct a linear discriminant function in such an arbitrarily defined feature space. The non-linear mapping of the input space to a higher dimensional feature space is defined as:

$\theta(x): R^n \rightarrow R^{n'} \quad with \; n' \gg n.$ (2.19)

Equation 2.18 can be rewritten to incorporate such a non-linear mapping as:

Maximise $\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j=1}^{N} \lambda_i \lambda_j y_i y_j \theta(x_i) \cdot \theta(x_j)$

Subject to $\sum_{i=1}^{N} \lambda_i y_i = 0 \; and \; C \geq \lambda_i \geq 0, i = 1,2, \ldots N.$ (2.20)

The calculation of the new feature space is not necessary. According to Mercer's theorem for certain mappings θ and two arbitrary points x_i and x_j, the inner product in the feature space can be expressed as a function of the inner product in the input space (Theodoridis & Koutroumbas 2006). This is given by:

$\theta(x_i) \cdot \theta(x_j) = K(x_i, x_j)$ (2.21)

where K() denotes a valid kernel function. This is referred to as the kernel trick. Equation 2.20 can be rewritten to accommodate such a function, namely:

Maximise $\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j=1}^{N} \lambda_i \lambda_j y_i y_j K(x_i, x_j)$

Subject to $\sum_{i=1}^{N} \lambda_i y_i = 0 \; and \; C \geq \lambda_i \geq 0, i = 1,2, \ldots N.$ (2.22)

Note that using a standard linear kernel $K(x_i, x_j) = x_i^T x_j = x_i \cdot x_j$ results in no transformation of the input space and constitutes a linear evaluation of the objective function.

The power of the SVM classifier lies in the non-linear transformations enabled by the kernel function. A kernel function typically introduces additional free parameters to the optimisation problem, controlling the extent of the feature space transformation. These parameters need to be set in addition to the standard slack variable (*C*). An overview of routinely employed kernels is presented in Section 2.4.2.3. Strategies for selecting free parameters (*C* and any kernel function parameters) are considered in Section 2.4.2.4.

The SVM classification procedure is summarised by Bennett & Campbell (2000) as follows:

1. Select an adequate slack variable C. Select a kernel function and set any kernel function free parameters.
2. Solve equation 2.20 on the training data using a quadratic or linear programming algorithm (typically using this Lagranian formulation). Recover the threshold variable *b*.
3. Classify any new point x, using $f(x) = sign(\sum_i y_i \lambda_i K(x, x_i) - b)$.

The utility of an SVM can be extended to tackle multi-class problems by one-against-all or one-against-one strategies (Huang, Davis & Townsend 2002; Tzotsos 2006).

2.4.2.2 Anomaly detection variants of support vector machines

Anomaly detection or novelty detection is a classification scheme where the interest lies with a single class. It is not always possible or feasible to train a classifier on numerous classes when only one class is of interest. Novelty detection is a difficult classification task, with numerous models having been proposed in the literature. This difficulty arises from the fact that no information about the distribution of the 'other' class or background is available, resulting in poorer classification performance than an equivalent multi-class approach where adequate attention is given to the construction of the multi-class classifier. No model is generally superior to others and classification results depend on the underlying distribution of the data, as with multi-class classifiers (Markou & Singh 2003).

In a RS context, anomaly detectors are employed when there is a narrow focus of interest in a complex scene. Typically, multi-class classifiers need exhaustive samples of all possible

classes in a scene. The generation of ancillary classes that accurately model intra-class variability is a difficult and expensive task which is aggravated because no thematic interest is directed toward these classes (Muñoz-Marí, Bruzzone & Camps-Valls 2007; Sanchez-Hernandez, Boyd & Foody 2007b). In cases where the dataset is severely unbalanced, that is the class of interest is very sparse and has very few samples compared to the background class, the use of a multi-class strategy generally performs poorly compared to anomaly detector approaches (Zhuang & Dai 2006).

In anomaly detection, the classification process is seen as a domain description task. In domain description, the extent of the feature of interest is modelled and described to separate it from the background. New instances are evaluated against these measures and labelled either as belonging to this class (anomaly) or rejected. As with multi-class statistical classifiers, an assumption can be made about the distribution of the class of interest, for example a parametric classifier. When no distribution assumptions can be made, the extent of the anomaly in input space can be modelled either by a probability density estimation method (e.g. parzen windows or mixture of Gaussians) or by a boundary based method (e.g. one class SVM or SVDD). Boundary methods are generally preferred as they require less samples than probability density methods in accurately modelling a class (Sanchez-Hernandez, Boyd & Foody 2007b; Tax & Duin 1999).

Tax (2001) and Tax & Duin (1999; 2004) propose a boundary based anomaly detector inspired by SVMs, entitled the support vector domain descriptor (SVDD). In the authors' formulation, a hypersphere is constructed around the samples of the class of interest or anomaly. An SVDD attempts to minimise the volume of this enclosing hypersphere. This concept is illustrated in Figure 2.12. New samples (dots) are either accepted as anomalous if they fall within this sphere or rejected (squares) if they fall outside this sphere. The samples of the class of interest that lie on the hypersphere are the support vectors. As with SVMs, an SVDD classifier needs only these vectors in its construction.

An SVDD aims to minimise the radius (R) of the hypersphere according to the function:

$$F(R, a) = R^2 \qquad (2.23)$$

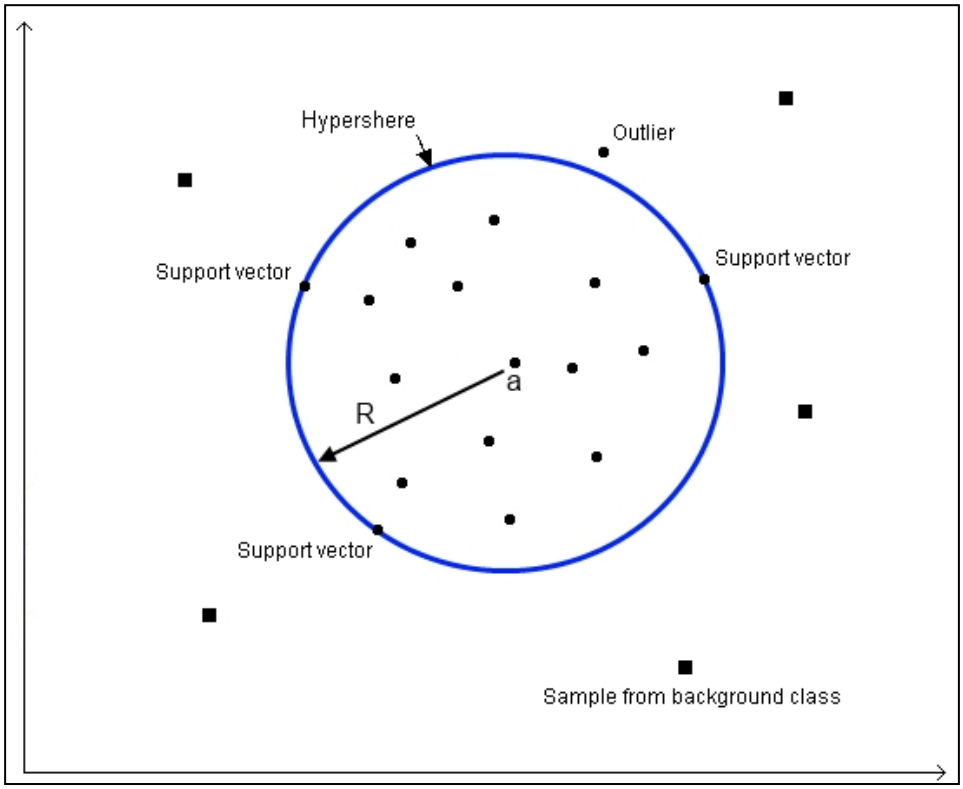


Figure 2.12: A hypersphere separating samples of the class of interest from all other samples in an SVDD

constrained by $\|x_i - a\|^2 \leq R^2$ for all i. As with SVMs a slack variable (C) is introduced to allow for the possibility of outliers, resulting in:

$$F(R, a, z) = R^2 + C \sum_i z_i \quad (2.24)$$

constrained by $\|x_i - a\|^2 \leq R^2 + z_i,\ z \geq 0$, for all i. The Lagrange formulation is written as:

$$L = \sum_i \lambda_i (x_i \cdot x_i) - \sum_{i,j} \lambda_i \lambda_j (x_i \cdot x_j) \quad (2.25)$$

constrained by $0 \leq \lambda_i \leq C$, $\sum_i \lambda_i = 1$ and $a = \sum_i \lambda_i x_i$. Similar to SVMs, a kernel function can be substituted for $(x_i \cdot x_j)$ in equation 2.25.

A new sample (z) is accepted if it falls within the sphere. If the distance of z to the centre a is smaller than the radius R, z is accepted. This decision function is written as:

$$(z \cdot z) - 2 \sum_i \lambda_i (z \cdot x_i) + \sum_{i,j} \lambda_i \lambda_j (x_i \cdot x_j) \leq R^2. \quad (2.26)$$

A new sample is accepted or labelled as belonging to the class of interest if equation 2.26 holds on the evaluation.

Tax & Duin (2004) present an additional heuristic to the SVDD algorithm that allows the inclusion of limited samples from outside the class of interest to refine the hypersphere. This process attempts to mitigate the general disadvantage of anomaly detectors by giving some shape to the background class. Tax & Duin (2004) note that employing such a mechanism will not always deliver superior results and requires careful optimisation of the *C* parameter. In one of the few RS case studies employing an SVDD, Sanchez-Hernandez et al. (2007b) report that using samples from outside the class of interest to increase the classification accuracy did not result in better results (compared to not delineating background samples).

Schölkopf et al. (2001) present a similar anomaly detector based on the principles of the SVM. They propose that a hyperplane separates the samples of the class of interest from the origin (Figure 2.13). This classifier attempts to maximise the area separating these samples from the origin. A parameter, called the rejection rate, is introduced that controls the ratio of points considered to be outliers and replaces the slack variable C found in SVDDs. This parameter is also a lower bound on the fraction of support vectors.

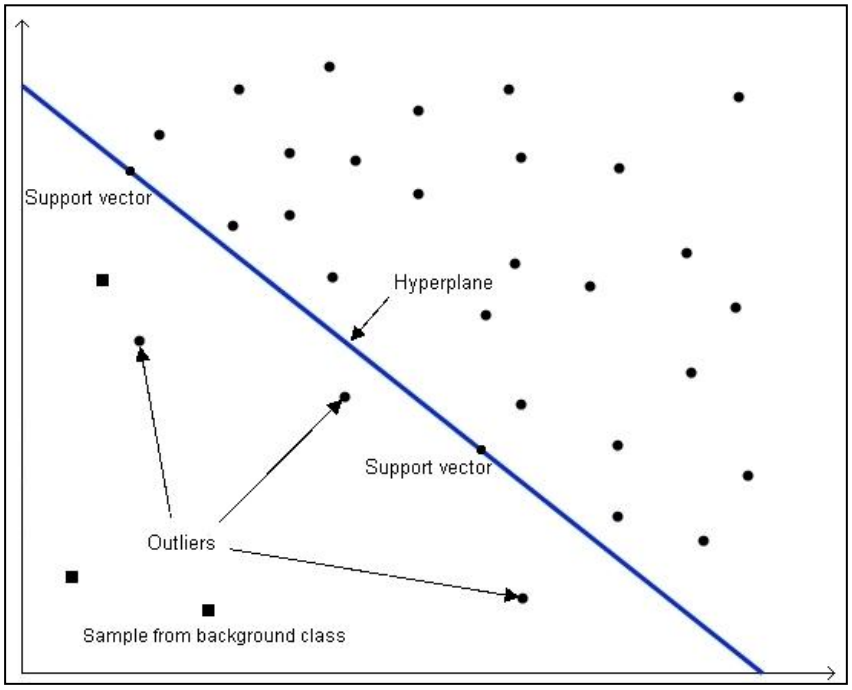


Figure 2.13: Basic principle of the one-class SVM showing that the area between the samples of the class of interest and the origin is maximised

Tax & Duin (2001) and Schölkopf et al. (2001) note that the one class SVM and SVDD approaches are identical under the assumption of a radial basis function (RBF) kernel. Zhuang & Dai (2006) present a comparative study of these two techniques, along with other parametric and probability density anomaly detectors, in object character recognition problems. Both the one class SVM and SVDD outperformed the other anomaly detectors when using the RBF kernel, while a very small margin in accuracy separated these two techniques. Case studies in RS typically use either the one class SVM (Guo et al. 2008) or the SVDD (Muñoz-Marí, Bruzzone & Camps-Valls 2007; Sanchez-Hernandez, Boyd & Foody 2007b) with an RBF kernel.

2.4.2.3 Kernel functions

A kernel function defines the dot product of two arbitrary points subject to some mapping function as equal to some other function (see equation 2.21). Data can be described in some arbitrary higher dimensional feature space by only evaluating the inner products in the input space because a kernel function dictates that the inner product will remain unchanged (kernel functions satisfy Mercer's theorem). The use of kernel functions can thus be considered a technique (the so called kernel trick) to reduce the computational load of an optimisation function, as it is not necessary to explicitly calculate the new feature space for an arbitrary evaluation; a computationally expensive task with large datasets. Shawe-Taylor & Cristianini (2004) give a detailed overview of the theory and formulations of kernel functions. Shawe-Taylor & Cristianini (2004) suggest that the selection of a kernel function should be guided by knowledge of the structure of the data. They also note that this is not always possible and that users are forced to select an appropriate kernel and subsequent free parameters by quantitative trial and error using the data at hand (see Section 2.4.2.4).

The polynomial and RBF kernels are popularly employed. The simplified (one parameter) polynomial kernel is written as:

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d \qquad (2.27)$$

where d defines the degree of the polynomial and is a user defined variable along with the slack variable C (when using an SVM or SVDD). As an example a two dimensional vector (x_i, x_j) in input space is mapped to a five dimensional vector in feature space

$(x_i, x_j, x_i x_j, x_i^2, x_j^2)$ when the degree of the polynomial is specified as two; for example $(x_i \cdot x_j + 1)^2$.

The RBF kernel is written as:

$$\mathrm{K}(\mathrm{x_i}, \mathrm{x_j}) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma}) \quad (2.28)$$

where gamma (σ) (also called sigma) denotes the width of the Gaussian kernel (how tight the boundary fits over the training data).

In an anomaly detection context, setting a tight gamma value may lead to numerous separate regions in the feature space that defines the class of interest. This allows an SVDD to model the class of interest as a multi-modal distribution. Huang et al. (2002) present a comparison of these two kernel types (polynomial and RBF) in per-pixel based RS SVM classification, stating that the polynomial kernel free parameter (*d*) has a finer grain of sensitivity or narrower extent of suitable values as opposed to the free parameter in the RBF kernel. Tzotsos (2006) and Munoz-Mari, Bruzzone & Camps-Valls (2007) employed the RBF kernel in RS applications (SVM and SVDD) and as reasons for its choice noted its generality (some other kernels are specific cases of the RBF kernel), its single parameter to tune (gamma) and its lack of numerical difficulties (results are constrained to [0...1]) compared to other kernels.

Tax & Duin (1999; 2004) offer quantitative arguments for the choice of kernel for the SVDD. They present a case study using the polynomial kernel in SVDD classification, noting that for larger polynomial degrees large regions in the feature space are generated that is void of support vectors or samples, resulting in inaccurate classifications. They note that the RBF kernel does not suffer from this problem and gives promising results. For a small gamma (RBF) the SVDD generalises to a parzen density estimation and for a large gamma a rigid hypersphere is obtained, generalising to a linear evaluation function only dependent on *C*. These authors suggest the use of the RBF kernel when using the SVDD. In contrast to this suggestion, Sanchez-Hernandez (2007b) found the polynomial kernel to be superior to the RBF kernel in a per-pixel based mapping of fenland from Landsat ETM+ imagery, citing the use of a cross-validated grid search (presumably using averaged user's accuracy as accuracy evaluation) on training data as quantitative measurement although no empirical evidence is presented. In this study the RBF kernel is selected in view of recommendations from the

literature. The resultant classifier has two free parameters (gamma and C/nu) that need to be set. In the next section free parameter tuning strategies are briefly considered.

2.4.2.4 Free parameter tuning strategies

Although SVMs are generally superior to simpler classification strategies (see Section 2.4.2.5), an implementation typically requires at least two interdependent free parameters to be tuned (excluding the linear kernel). This has a definite influence on the accuracy of the resultant classification. These parameters can be set by an analyst by manual trial and error, although the extension of such a system to be autonomous in terms of parameter tuning is trivial. The parameter tuning process consists of two components:

1. An evaluation step where the results of the classification on the training dataset using an arbitrary parameter set are evaluated.
2. A search procedure that manually or automatically traverses the parameter space in a linear or non-linear fashion, using the evaluation component to test the classifier on specific parameters. The search process typically stops when a sufficient level of accuracy is reached or when the parameter space is sufficiently explored.

Strategies routinely employed in the evaluation step include the training error estimation, hold out testing, bootstrapping, k-fold cross-validation and the leave-one-out strategy. In k-fold cross-validation, the training data are subdivided into k subsets (Duda, Hart & Stork 2000). The classifier is trained on k-1 subsamples and evaluated using some notion of result accuracy on the subsample that was not used in the training. This procedure is repeated and the subsamples rotated, so that each subsample is used only once as the evaluation subset. The results over all k evaluations are averaged. The leave-one-out strategy is a specific case of k-fold cross-validation where k equals the number of samples. Such an evaluation strategy guards against the performance prediction of a classifier based on data seen a priori, typically leading to a lower generalisation capability of the classifier as it overfits the classifier on the training data. For highly unbalanced data or when using anomaly detection approaches (e.g. SVDD), more complex measures of classification accuracy are needed as typical evaluation measures will favour the dominant classes. This is discussed further in Section 2.6.2.3.

A k-Fold cross validated grid search, or more advanced variants thereof is the most frequently employed parameter space search strategy (Hsu, Chang & Lin 2008). In a grid search procedure the parameter space is arbitrarily divided into a number of grids. The classifier is

evaluated on the average of the parameters in every grid. The grid delivering the best results is repartitioned into smaller grids and the evaluation process is repeated. Grid search is a local search method and may be trapped in local optima (see Section 2.5). This method is employed in the few case studies presented in the RS domain as a manual parameter tuner, for example Muñoz-Marí, Bruzzone & Camps-Valls (2007), Sanchez-Hernandez, Boyd & Foody (2007b) and Tzotsos (2006) although Guo et al. (2008) use an automatic brute force search.

In addition to the free parameters that need to be set, the selection of attributes or features also need consideration. Hsu et al. (2008) suggest that in cases where the attribute set of the data is large, an attribute selection strategy may be required for better results. In a RS context, attributes or features may include mean pixel or object spectral response, texture measures and context information (spatial information/DEM derivations). In object based image analysis, the attribute set may be large and contain redundant or irrelevant information, especially in the case of hyperspectral imagery or multispectral imagery with numerous user defined data transforms (e.g. NDVI and HSI).

Unfortunately, the selection of an attribute subset is also interdependent with the tuning of the free parameters of the SVM (Fröhlich, Chapelle & Schölkopf 2003; Huang & Wang 2006; Lin et al. 2008). It has been argued that although SVMs are generally more robust to high dimensional, redundant and irrelevant data than other statistical classifiers, the Hughes phenomenon (loss in accuracy as the ratio of data dimensionality against training sample size increase) is still a constraint so that feature subset selection is called for (Fröhlich, Chapelle & Schölkopf 2003; Kotsiantis 2007; Pal & Foody 2010).

Recently, a few techniques have been proposed in the pattern recognition community that use population based search functions (genetic algorithm and swarm optimisation strategies) for simultaneous attribute subset selection and free parameter tuning applied to SVMs (also applicable to anomaly detectors with careful consideration of fitness evaluation) (Huang & Wang 2006; Lin et al. 2008). Such techniques offer promising advantages in autonomy and accuracy over simpler exclusive parameter tuning strategies, especially when the data dimensionality is high and training samples few. Population based search techniques have been shown to converge to global optimums, resulting in fast and accurate determination of attributes and free parameters in SVM classification, although some concerns exist about generalisation performance (see Section 2.6). It is hypothesised that such techniques may

hold utility in object based anomaly detection, assisting in creating a faster, autonomous and more accurate classifier. This utility is investigated as part of the primary aim of this research.

To facilitate further explanation of such approaches, evolutionary inspired numerical optimisation techniques are discussed in Section 2.5. Advanced approaches to feature selection and parameter tuning as applied to SVMs are discussed in Section 2.6. In the next two subsections the use and results of kernel methods in RS image classification are noted.

2.4.2.5 Case studies and comparative studies of SVMs in the RS domain

SVMs use boundaries and maximum margins to optimally separate classes. Other classifiers are based on some probabilistic decision function that decides which label a sample should receive, and do not consider the relative distances in feature space among all samples. This free parameter guided increased complexity of SVMs compared to other classifiers results in theoretically more accurate predictions and generalisation capabilities (Theodoridis & Koutroumbas 2006). Meyer, Leisch & Hornik (2003) compare SVMs with 16 other popular advanced classifiers, citing inconclusive evidence to attest to the general superiority of SVMs over other advanced approaches, necessitating data specific investigations. In the data specific context of earth observation imagery, operating on both per-pixel and object based data, SVMs have recently been applied in case studies as anomaly detectors, as multi-class classifiers and in semi-supervised learning. Active learning strategies were not considered in this study.

Huang, Davis & Townsend (2002) investigate the utility of SVMs in multi-class per-pixel classification on moderate resolution satellite imagery, finding SVMs to be superior to artificial neural networks and maximum likelihood classifiers. The effectiveness of SVMs increased more dramatically compared to neural networks as the sample sizes and data dimensionality increased. Foody & Mathur (2004) report similar results, with SVMs performing better in most cases compared with ANNs and DTs in per-pixel multi-class classifications on moderate resolution satellite imagery. They note the robustness of SVMs to high dimensional data.

Camps-Valls & Bruzzone (2005) investigate the effectiveness of SVMs and other kernel methods in high dimensional hyperspectral image classification. SVMs yielded the most

accurate classifications at the lowest computational costs. Camps-Valls et al. (2006) and Camps-Valls et al. (2007) subsequently present novel composite kernel methodologies, further enhancing the classification accuracies of SVMs on high dimensional hyperspectral and multi-source data. Muñoz-Marí et al. (2007) present comparative results of using composite kernels on high dimensional (multi-source) data for anomaly detection. They point out slight increases in accuracy using this proposed methodology compared to the standard practise of data stacking. This technique is noted, but it was not pursued further due to the nominal increase in implementation complexity and the small cited gain in accuracy.

The use of kernel based anomaly detection approaches in RS has recently attracted some attention. Banerjee, Burlina & Diehl (2006) compare an SVDD with other nominal anomaly detection approaches for hyperspectral per-pixel classification, noting that SVDD outperforms the other evaluated anomaly detectors. Muñoz-Marí et al. (2007) concluded similarly.

Some studies also compare an SVDD or binary SVM (used as an anomaly detector) approach with standard multi-class approaches. Taruvinga (2008) investigated the usefulness of a binary SVM (training samples obtained for the class of interest and the background class) in identifying erosion gullies in KwaZulu-Natal, South Africa. The SVM outperforms more standard multi-class approaches. Sanchez-Hernandez et al. (2007b) found the SVDD to be superior to other anomaly detectors and a multi-class maximum likelihood classification in identifying fenland (per-pixel moderate resolution imagery). In a similar study, Sanchez-Hernandez et al. (2007a) evaluated the use of an SVDD and binary SVM under similar conditions for mapping a specific vegetation class. The SVDD outperformed the SVM, exemplifying that a weak description of the background class when using a binary SVM as anomaly detector may lead to a loss in classifier accuracy (versus a pure anomaly detection approach).

Concerning supervised classification, object based image processing differs from the per-pixel based approach in the dimensionality of the data (object based data typically have a higher order of dimensionality), the number of samples available (due to segmentation, object based data have less samples) and the relationship among samples (per-pixel data has less obvious data relationships in high resolution imagery). Tzotsos (2006) and Tzotsos & Argialas (2008) propose the use of kernel based classifiers to address these limitations when dealing with object based data. They observed an improved classification performance of

SVMs compared to a nearest neighbour approach operating on object based data. Guo et al. (2008) investigated the use of an object based SVDD classification approach versus a per-pixel SVDD classification. They report the ability of the SVDD to handle the higher dimensionality of object based data. The per-pixel approach was unable to achieve similar accuracies, due to the complex nature of the high resolution imagery. These studies provide quantitative demonstrations of the qualitative notions why SVMs may be well suited, in contrast to other supervised methods, for high resolution object based image analysis.

The studies conducted to date that use kernel methods in the domain of RS have not placed emphasis on the influence of the selection of attributes and/or the accurate tuning of classifier free parameters. The above studies focus on the general utility of kernel method classifiers in a RS context and they employ simple free parameter tuning strategies. They also provide quantitative comparisons with other statistical classifiers. High resolution object based RS data can have uncharacteristic attribute sets, display complex data and cognitive information relationships (inhomogeneous objects), are extensive and are typically subject to sparse sampling conditions, all suggesting that interdependent attribute selection and parameter tuning processes may have a considerable influence on classifier performance and should therefore not be overlooked.

At the time of conducting this research, the literature revealed only two (Guo et al. 2008; Liu & Sun 2008) peer reviewed (conference papers) exploratory studies using supervised object based kernel based geographic anomaly detection techniques although they place little emphasis on attribute subset selection and free parameter tuning.

2.4.2.6 One class SVM as the chosen statistical classifier

SVMs have a range of proven advantages. They deliver robust classification and generalisation performance compared to most other popular classifiers; they have no problems with local minima (DTs and ANNs may suffer from falling in local minima); they have relatively few free parameters to set (only two when employing an RBF kernel); they are robust to sparse sampling compared to other classifiers; and they are less sensitive to irrelevant, redundant and interdependent data than DTs and ANNs (Bennett & Campbell 2000; Camps-Valls 2005; Kotsiantis 2007). Although SVMs cannot be considered overall superior to other advanced classifiers (Meyer, Leisch & Hornik 2003), the literature does

present case studies where SVMs outperform other classifiers (DTs, ANNs and KNN) in the data specific context of RS imagery.

Some authors (Raskutti & Kowalczyk 2004; Sanchez-Hernandez, Boyd & Foody 2007a; Zhuang & Dai 2006) advise that in cases of extremely unbalanced datasets, anomaly detectors give better classification performances. They submit that by discarding "distractive majorities" the domain of the minority class can be better modelled. This is especially true in cases where the background class is multi-modal, a condition that will almost always hold in RS imagery. A one class SVM/SVDD is a variant of an SVM and thus has the same advantages of an SVM, but only requires samples of the class of interest. As an anomaly detector, a one class SVM is also theoretically superior in cases of highly unbalanced datasets. Additionally, due to its non-linear data discrimination capabilities, a kernel based anomaly detector is well suited to model the multi-modal nature of the background class (Banerjee, Burlina & Diehl 2006) when careful consideration is given to parameter tuning.

Finally, the literature provides practical recommendations regarding the actual sampling in the classification process when using a kernel based classifier. Foody & Mathur (2004; 2006) investigate the influence of the quality and quantity of training samples in per-pixel SVM land cover classification. They propose that the training phase in an SVM based classification should be handled differently from the classically accepted way of finding pure samples that describe classes. Because a SVM models the separation of a class by support vectors, as opposed to modelling (e.g. probability density) a class in feature space, only the samples that generated those support vectors are needed. Samples that do not form support vectors are acquired unnecessarily. Foody & Mathur (2004; 2006) advocate a knowledge driven selective sampling approach that focusses on finding impure or border cases of the class of interest that will most likely form support vectors in the classifier.

Foody et al. (2006) extend the work by the above authors by measuring the effect of training sample size on classifier performance of SVM and SVDD per-pixel land cover classification against the popular 10-30p rule (Mather 2006) normally employed in training sample collection. The 10-30p rule states that $10 \times p$ to $30 \times p$ samples per class should be trained where p is the dimensionality of the input data. It was found that with the intelligent selection of impure or border training samples the classification accuracy did not nominally decrease as the number of samples fell below the 10-30p rule. Specifically, the good performance of the

SVDD with a small intelligently selected training sample was noted. As an example, for a specific class it was reported that a SVDD intelligently trained on only 30 samples of the class of interest – which constituted 6.7% of the samples required by the 30p rule achieved an accuracy that differed insignificantly from the SVDD trained on 30p samples. These results commend a kernel based anomaly detector as being well suited in cases with limited training samples and high dimensional (object based) input data in comparison to other approaches.

In the next section population based search strategies are considered, followed by an overview of recently proposed strategies for free parameter tuning and attribute subset selection.

2.5 EVOLUTIONARY ALGORITHMS AND SWARM INTELLIGENCE

Evolutionary computation is a subfield of computational intelligence. Algorithms developed and studied within evolutionary computation mainly address combinatorial optimisation problems (also referred to as search and minimisation problems), and to a lesser extent self organisation and self learning problems. The design of algorithms and heuristics in this subfield was inspired by two naturally occurring processes, Darwinian evolution (survival of the fittest) and the social behaviours (flocking/swarming) of bees, birds, ants and other social creatures (Eiben & Smith 2003). Evolution was observed to be effective in selecting properties of individuals or the population that aid in their survival in an environment. In the next section an overview is given of population based evolutionary and swarm inspired search and the reasons why they are effective. Consideration of specific variants used in this study is also given.

2.5.1 Overview of evolutionary inspired population based search algorithms

An evolutionary computation algorithm (ECA) consists of evolutionary algorithm and swarm intelligence algorithm subsets. These subsets have a few fundamental differences in their implementation but very similar performance characteristics, thus a general description applicable to both models is given. ECAs are iterative search algorithms operating on a population of items with an arbitrary number of attributes. The properties of these items gradually change through iterations due to interactions among themselves, with the aid of some notion of desirable item properties (Eiben & Smith 2003; Mishra 2006; Price, Storn & Lampinen 2005). This general process is illustrated in Figure 2.14.

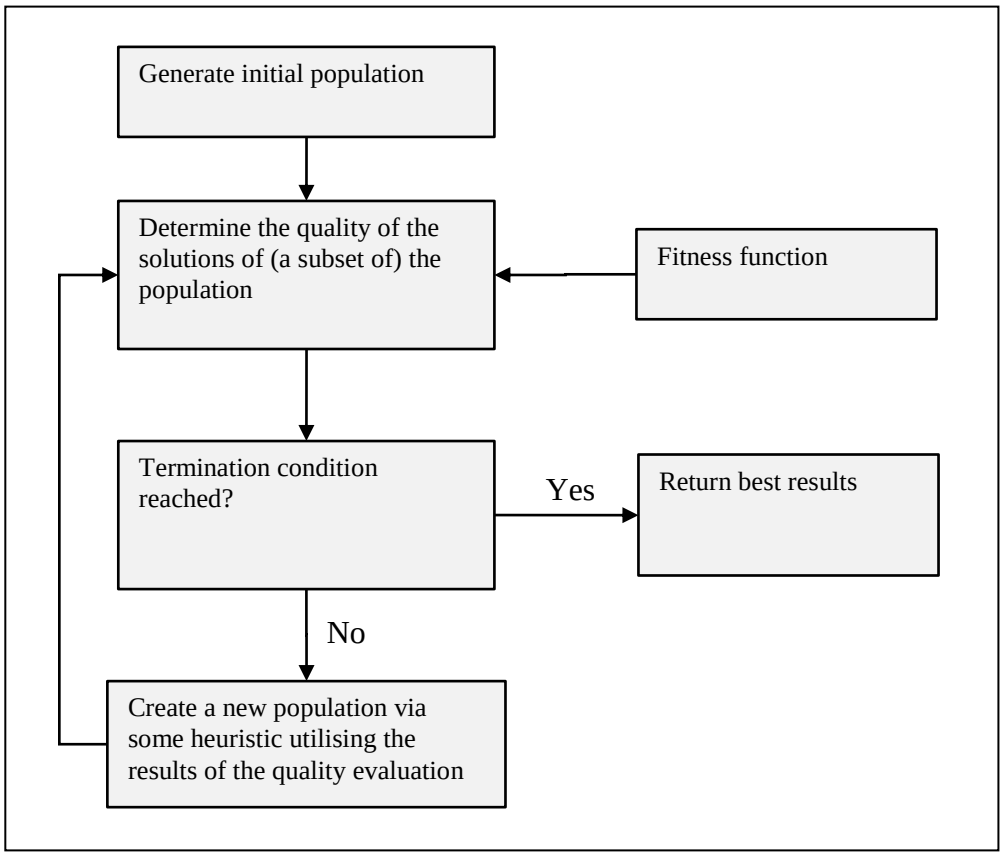


Figure 2.14: Generalised structure of a population based search algorithm

In an ECA an initial population is defined. A population is a collection of randomly generated sets of solution vectors to a specific problem. As an example, each individual, also called an agent, in the population may contain segmentation algorithm parameters or classification algorithm parameters as solution vectors. A fitness function is an evaluation process. A fitness function evaluates the quality or 'fitness' of an agent, or rather how the solution vectors of this agent perform in a given task. For example, a fitness function may contain a segmentation task (see Section 2.3.3.2) or a classification task (see Section 2.4). The solution vectors of the agent are used in such a segmentation or classification task. The fitness function returns the relative performance of the segmentation or classification task using the parameters/solution vectors provided by the agent. Staying with this example, quality measures could be metrics employed in segmentation evaluation (see Section 2.3.2) or general classification accuracy.

An ECA algorithm runs in parallel, meaning all individuals in a population are evaluated, and possibly changed, before the next iteration of the algorithm is executed. After completion of the fitness evaluation of the individuals in the population the population is transformed using

some heuristic observing the performances of the agents. ECAs are stochastic; some randomisation mechanism found in the heuristic changes the population in addition to the changes performed due to fitness evaluation. Typically, aspects of the well performing agents are transferred to agents in a subsequent iteration (also called generation). As the algorithm proceeds, agents with ineffective solution vectors will be mutated to be more suited to the environment or can be removed from the population (depending on the specific ECA implementation). An ECA algorithm has a termination heuristic.

Figure 2.15 illustrates a simple one dimensional optimisation problem tackled by an ECA. The solution vectors (red dots) of the agents are changed at each iteration of the algorithm. The search surface (solution to the search problem over all possible values generated by the fitness function) of this simple problem is indicated by a solid black line. The algorithm runs for an arbitrary number of iterations, with Figure 2.15(a) indicating the first iteration, Figure 2.15(b) an iteration halfway through the algorithm and Figure 2.15(c) the final iteration. The x axis denotes the search space while the y axis denotes the fitness of the agents. The solution vector of the fittest agent (highest value with respect to the y axis) is chosen as the output of the algorithm. Note the ability of agents to escape from suboptimal peaks on the search surface.

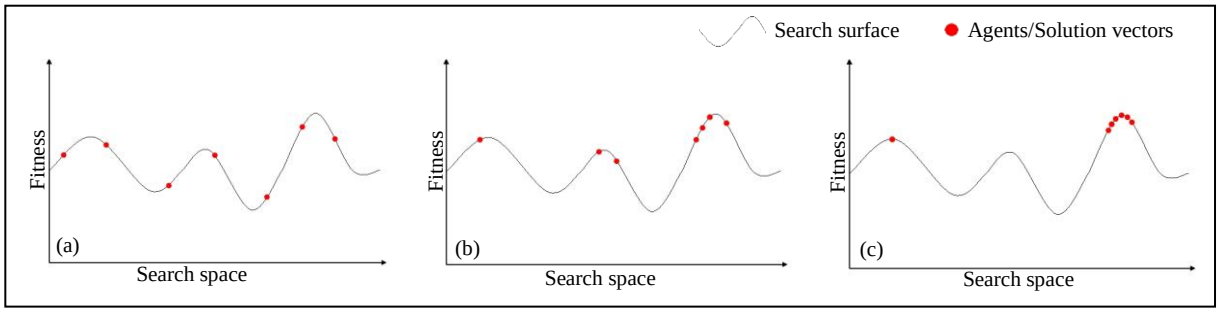


Figure 2.15: Agent behaviour of an ECA in a simple one dimensional problem

Eiben & Smith (2003) define five general components of an ECA that need consideration:

1. The method of representing individuals in the population.
2. The method of quality evaluation of an individual.
3. The structure of the population.
4. The method of changing individuals at each iteration of the algorithm.
5. The stopping criteria of the algorithm.

Component one refers to the method in which a population can be represented, also called its solution vector encoding. Solution vectors can be kept in their original form (referred to as genotype form), or transformed to another form (phenotype form). Evolutionary strategies, which are subsets of ECAs, typically encode parameters as real numbers (e.g. differential evolution) while genetic algorithms encode parameters in bit strings (discrete values) and are more suited to combinatorial optimisation problems (Price, Storn & Lampinen 2005). Strategies do exist that can combine solution vectors of different quantisations (e.g. discrete, Boolean, integer or real valued representations – such as particle swarm optimisation or ant colony optimisation), although not all algorithms are equally suited for such tasks due to internal structures (e.g. the real valued mutations of DE suggest it is not particularly suited for tasks with mixed solution vector quantisations).

The second component refers to the structure of the fitness function (also called objective function), or more specifically how the fitness function will use the solution vectors in generating a quality score. An ECA employs these quality measures in searching for an optimal solution. The construction of a viable and accurate fitness function in an evolutionary search process is an important consideration (Eiben & Smith 2003). A faulty fitness function may result in an ECA displaying slow and sporadic convergence. Some ECA algorithms are resistant to this type of noise.

Component three refers to the structure of the population (specifically to the number of individuals in the population) that is, if the size of the population changes through iterations and if individuals are spatially aware of their neighbours (e.g. particle swarm optimisation). Component four forms the core of an ECA. It describes how agents are changed at each iteration based on their fitness evaluation, the fitness of other agents in the current or past iterations, the spatial relationship among agents and some randomisation heuristic. This component defines how an ECA searches the feature space. Eiben & Smith (2003) note that a search heuristic (component four) has a trade-off between its ability to search relatively unexplored regions of the feature space and its ability to refine results in the vicinity of known good solutions.

The last component considers how an ECA terminates itself. An ECA terminates when some agent provides solution vectors resulting in a fitness evaluation exceeding the required user defined quality, when a certain number of iterations of the algorithm have passed or when

very little improvement in fitness is observed through a number of generations (see Figure 2.16 as a typical example of this phenomenon). Details of the ECAs used are given in later subsections.

2.5.2 Beneficial properties of evolutionary computation algorithms

This thesis describes a system that handles extensive, complex search spaces that need to be processed at near interactive speeds. ECAs have a few advantages over simpler search strategies in tackling complex optimisation tasks. To define these advantages an overview is given next of the basic categories of search methods.

Search algorithms can be either deterministic or stochastic. Deterministic search algorithms explore the feature space in a thorough, systematic way according to some exact structure or heuristic. Deterministic algorithms typically attempt to exploit the structure of the search space in finding an optimum. Stochastic search algorithms have a randomisation component added to the search heuristic. A stochastic search approach is better suited in cases where the structure of the search space cannot reliably be predicted or modelled. Simple deterministic or direct search strategies such as brute force search, branch and bound search and problem specific heuristics are guaranteed to find an optimal solution in any given search space, but may take an unacceptably long time to compute, depending on the data dimensionality and search surface extent (Eiben & Smith 2003).

Price, Storn & Lampinen (2005) describe the four basic categories of search algorithms listed in Table 2.1 along with relevant examples, although the authors note that not all search algorithms can be neatly categorised within this system. The algorithm categorisation is based on two principles: whether the algorithm uses a population or single agent in a search, and whether the search is guided by derivatives or not.

Table 2.1: The basic forms of search algorithms.

	Single agent algorithm	Multi-agent algorithm
Derivative based algorithm	Steepest descent Gradient search	Clustering techniques
Derivative free algorithm	Random walk Hill climbers	Evolutionary computation algorithms

Source: Price et al. (2005: 5)

Derivative based algorithms may not always be able to quantify an objective function that observes the search space gradients. A simple example of this phenomenon is a uni-modal (one good solution) search space with an observed gradient at a coarse scale, but with a terrace structure at a finer scale, creating problems in estimating gradients.

Single agent, derivative free algorithms, or local search algorithms such as hill climbers, are able to quickly find an optimum in the vicinity of the start location of the single agent. Such algorithms cannot guarantee that the provided solution is the global optimum and they are not recommended for multi-modal datasets.

ECAs are stochastic multi-agent derivative free search algorithms. By using multiple agents, ECAs sample the search surface at many different points. These points communicate and influence one another's search direction, possibly allowing the mutation of agents in a conflicting way to that which the local agent perceives as a more optimal direction for the search. This property of a communicating multi-agent framework allows agents to escape from local optimums, as illustrated in Figure 2.15. ECAs are more robust to multi-modal, noisy, large and discontinuous search spaces compared to other search techniques (Eiben & Smith 2003; Price, Storn & Lampinen 2005). Many ECAs do not guarantee to deliver a global optimum, as no mathematical proofs have been given of their assurance to do so (e.g. PSOs). Results are dependent on the number of agents, algorithm iterations and search algorithm free parameters. ECAs are not always capable of effectively traversing any given complex search surface. This notion is explored for a new problem type presented in Chapter 5.

In addition to these beneficial properties of effective search, ECAs are also able to find relatively good solutions very quickly owing to a distributed search heuristic (Storn & Price 1995). Eiben & Smith (2003) describe the universal behaviour of the fitness trace of ECAs. A fitness trace is the quality of the solution vectors of the best agent or the population as a whole, as a function of algorithm iterations (generations). The fitness trace of an ECA displays rapid progress in the initial iterations of the algorithm. In later iterations the fitness or quality gain rapidly decreases. This characteristic property of ECAs is illustrated in Figure 2.16.

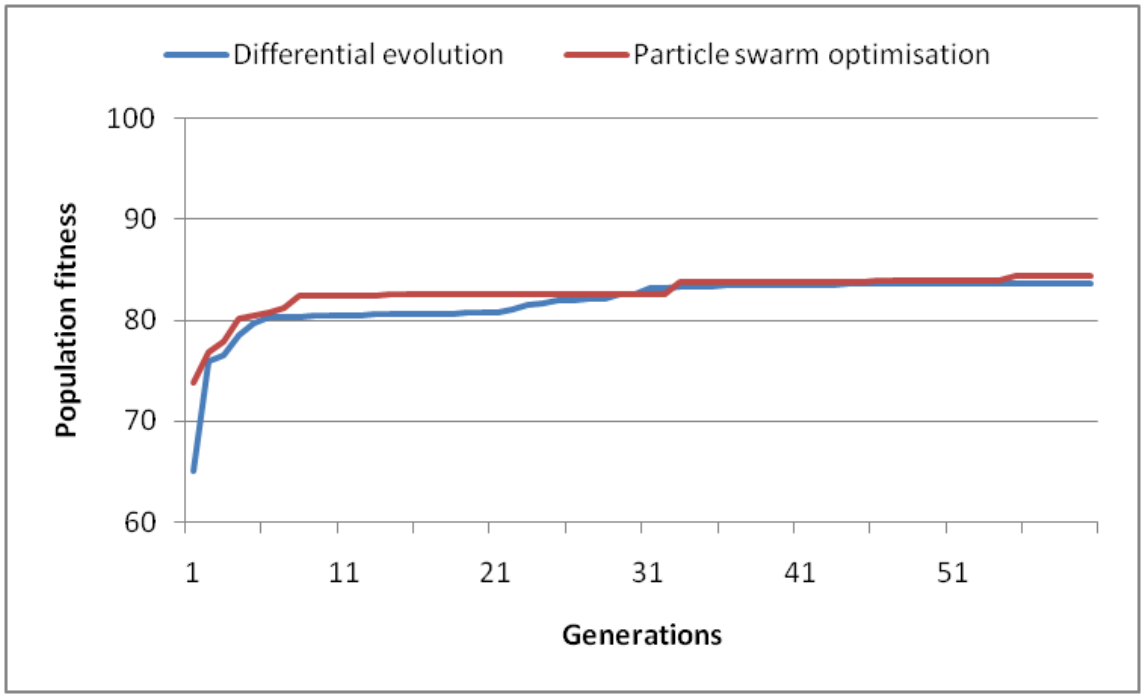


Figure 2.16: Typical fitness traces observed in ECAs

The technique proposed by Feitosa et al. (2006) was replicated in generating Figure 2.16, using the larger segment booster (equation 2.13) as fitness function to delineate an arbitrarily chosen agricultural field on a SPOT 5 scene. A fitness score of 100 means the algorithm and accompanying parameter set are able to segment the agricultural field exactly the same as the user delineated reference segment. Differential evolution and particle swarm optimisation search strategies (see below) were used to find segmentation algorithm parameters of a region merging algorithm. After ten iterations, both algorithms reached a point of near optimality. Another 50 iterations of the algorithms resulted in a minor increase in population fitness. This property of ECAs means that a relatively good solution, albeit suboptimal, can be obtained quickly. In this example, because of algorithm limitations and user expectations, the segmentation algorithm generated a very good but not perfect segmentation of the agricultural field, as indicated by the maximum population fitness of 85%. Note that LSB metric values are not directly correlated with percentage area overlap (see equation 2.13).

Search surfaces generated by RS data may contain numerous local optima and are characteristically extensive and noisy (depending on how the data are transformed and viewed), implying the use of ECAs for handling such data types. ECAs have been successfully employed in a variety of ways within the RS domain, for example in constructing a non-linear classifier using a genetic algorithm (Huang et al. 2007), in feature selection for a

non-linear classifier using a genetic algorithm (Van Coillie, Verbeke & De Wulf 2007) and in automatic segmentation parameter tuning (Feitosa et al. 2006; Fredrich & Feitosa 2008).

Differential evolution and particle swarm optimisation are ECA variants that have shown to have faster convergence rates and are more robust to a wide variety of problems than the classical genetic algorithm proposed by Holland (1975) (Price, Storn & Lampinen 2005; Vesterstroem & Thomsen 2004). These variants are chosen primarily for their robustness to a wide variety of search surfaces (Vesterstroem & Thomsen 2004) and they are briefly discussed next.

2.5.3 Differential evolution

Differential evolution (DE) is a numerical optimisation heuristic for continuous spaces originally proposed by Storn & Price (1995). It was designed to handle non-differentiable, non-linear, high dimensional and multi-modal fitness functions, to have good convergence properties and be parallelisable for quick execution with expensive fitness functions (Storn & Price 1995). The DE heuristic resembles the structure of a genetic algorithm, with a fundamental difference in the methodology for generating new solution vectors (component four as described in Section 2.5.1).

2.5.3.1 Overview of differential evolution algorithm

DE has a fixed population size. The algorithm can be terminated using any heuristic as described in Section 2.5.1. Figure 2.17 illustrates the solution vector mutation heuristic employed at each iteration of the algorithm.

During the evaluation of the population at an iteration of the algorithm, each agent is subjected to the population mutation heuristic illustrated in Figure 2.17. In this illustration agent 1 is arbitrarily chosen to be under observation. The population in every generation is expressed as:

$$x_i\ , i = 1,2, \dots , NP \qquad (2.2)$$

where *NP* denotes the number of agents or individuals in the population. *NP* is a user defined control parameter of the DE algorithm.

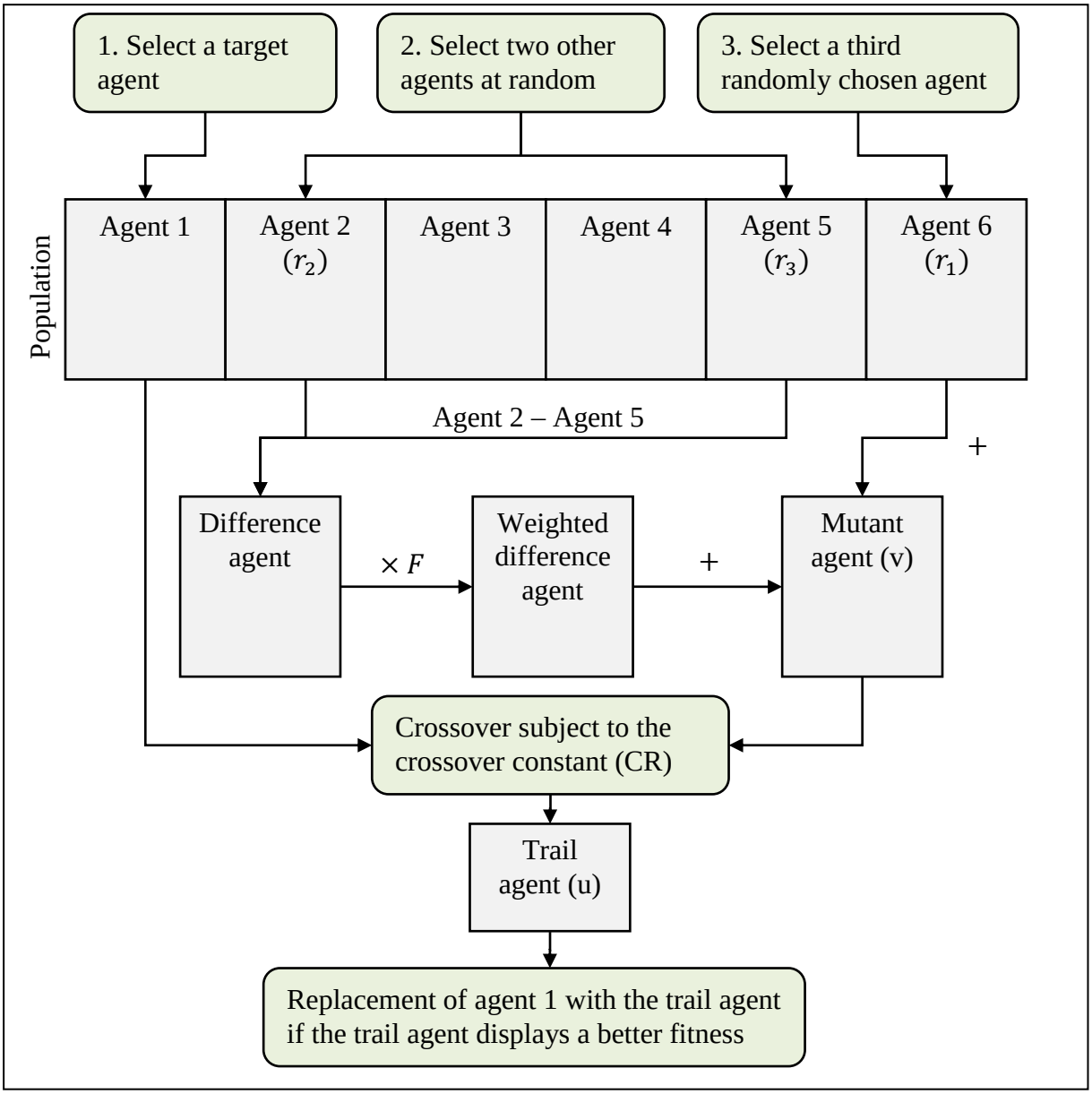


Figure 2.17: Population mutation heuristic of the DE algorithm

Three other agents are chosen at random to create the mutant agent (*v*) given by:

$$v_i = x_{r1} + F \times (x_{r2} - x_{r3}). \quad (2.30)$$

This mutation heuristic allows an arbitrarily chosen agent (x_{r1}) to be mutated with a differential subset of randomly chosen agents ($x_{r2} - x_{r3}$ in this example). *F* is a real valued constant in the range [0...2] that controls the amplification of the differential subset. *F* is a user defined control parameter of the DE algorithm. Intuitively, *F* has a great influence on the exploitation and exploration behaviours of the algorithm, with a small *F* resulting in a fast, possibly premature convergence (exploitation) (Price, Storn & Lampinen 2005) and a large *F* resulting in larger, more explorative mutations.

The mutant agent is not directly compared to the agent under observation (agent 1) as this would result in a loss of solution vector diversity. Agents displaying extremely good fitness behaviour may have a limited number of poorly performing solution vectors that would potentially never be mutated. The DE algorithm addresses this problem by creating a new (trail) agent that receives selective solution vectors from the mutant agent. The crossover operation generates a trail agent (u) by merging the mutant agent (v) and the agent under observation (x_1 in this example) on a solution vector basis via:

$$u_j = \begin{cases} v_j \; if \; (rand \leq CR) or \; j = rnbr \\ x_i \; if \; (rand > CR) and \; j \neq rnbr \end{cases} \tag{2.31}$$

where $J = 1,2,...,D$ is the set of solution vectors in the problem and $rand$ is a randomly generated number in the range [0...1]. $Rand$ receives a new value for each solution vector under observation. The element $rnbr$ is a randomly chosen solution vector index that ensures that at least one solution vector from the mutant vector will be carried over to the trail vector. This could happen if *rand* is never smaller or equal to *CR*. *CR* is the crossover constant in the range [0...1] and also a user defined variable. *CR* determines the probability that solution vectors from the mutant vector are carried over to the trail vector. Figure 2.18 portrays the crossover process having agents with six solution vectors (j = 6). Solution vectors two, four and six are carried over from the mutant agent, with vector four as a mandatory crossover.

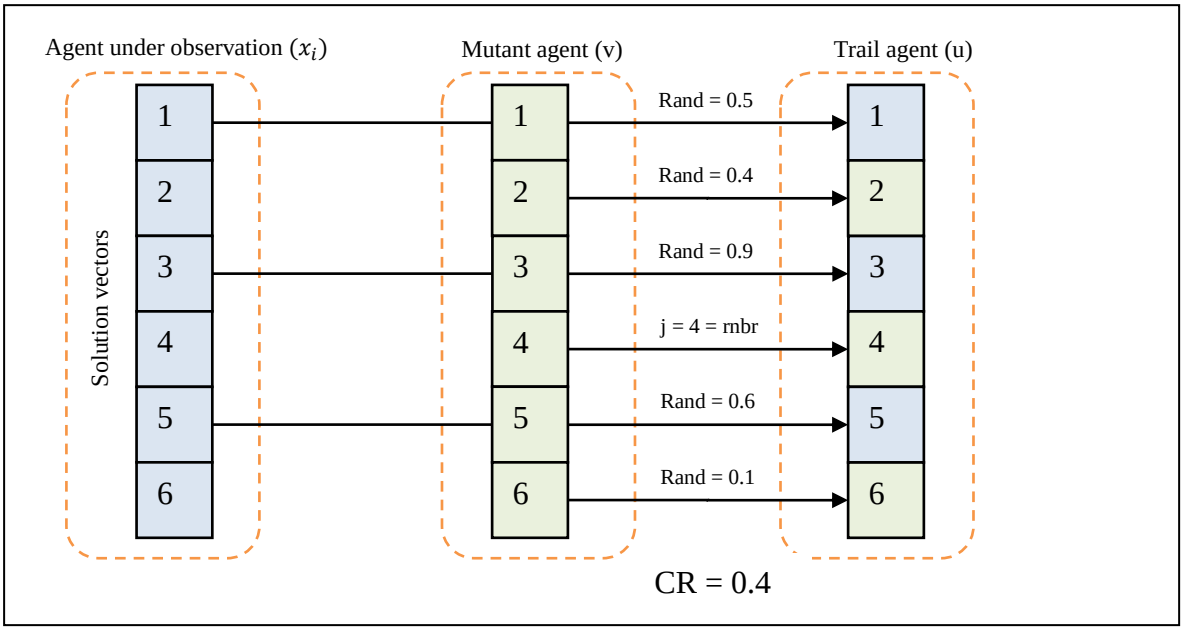


Figure 2.18: Crossover heuristic in the DE algorithm

Finally, the trail agent is compared to the agent under observation. If the trail agent has a better fitness compared to the agent under observation, the agent under observation is replaced by the trail agent. After the heuristic shown in Figure 2.17 is applied to all agents, a new iteration of the algorithm is initiated if a termination condition is not reached.

Numerous variations on this basic DE algorithm exists, for example using different strategies for selecting the number of difference agents, selecting which agents are observed in each iteration of the algorithm and varying details on the crossover scheme. In this thesis the general form described above, denoted 'DE/rand/1/' by Storn & Price (1995), is used (the mutant agent is chosen randomly, one difference agent is used and the crossover heuristic is based on independent binomial experiments).

Storn & Price (1995) offer case studies of the DE algorithm outperforming other prominent single agent and multi-agent search strategies. A fundamental difference between the genetic algorithm (GA) and DE is that a DE evaluates each agent with a mutated version of itself, as opposed to GAs that mutate individuals with the best individuals in the current generation. GAs also require inflexible user defined strategies or parameters to the mutation frequencies of individuals. This distributed mutation heuristic, or so called self organising scheme of DE is more dependent on the search surface structure than GA strategies and consequently allows a fast divergence of solution vectors in a flat search surface (exploration) without hampering the convergence to minima (Storn & Price 1995). An agent gets replaced by a mutated version of itself, suggesting an agent trapped at the peak of a local maximum will not continue its search within this maximum as it will never improve its fitness, except if it escapes.

Price et al. (2005) elaborate on the usefulness of DE in problem domains with multiple solution vector data types (real, integer and discrete values). DE cannot be used as an effective combinatorial optimiser due to its internal real valued functioning (Paterlini & Krink 2006; Price, Storn & Lampinen 2005).

2.5.3.2 Meta parameter tuning

The DE algorithm has three control parameters that require setting. Although Storn & Price (1995) observe that these parameters are relatively insensitive and easy to set, in their overview of studies that investigated the influence of DE meta-parameters on search results,

Brest et al. (2006) observe that meta-parameter tuning should not be overlooked. Applying the DE algorithm to a new problem domain (search surface) requires some consideration of this problem. Meta parameters have an influence on the convergence speed of an algorithm and its ability to find the true optimum (Brest et al. 2006; Eiben & Smith 2003).

The *NP* parameter in the DE algorithm defines the number of agents in the population and it can safely be set to five to ten times the dimensionality of the solution vector space (Storn & Price 1995). The selection of *F* (amplification factor of the difference agent) and *CR* (crossover constant) needs more detailed consideration.

According to Eiben & Smith (2003) control parameters of meta-heuristic algorithms (such as DE) can be tuned:

1. by means of deterministic conditions;
2. through an adaptive parameter tuning strategy that uses feedback from actual search runs (so called meta-optimisation strategies); and
3. via self adaptive strategies that incorporate the meta-parameters in the solution vectors.

Meta-optimisation is a search strategy whereby the agents in the search space represent optimisation problems with meta-parameters as solution vectors. See Pedersen (2010) for a summary of effective meta-optimisation strategies. Although effective, such a strategy can be prohibitively expensive regarding computation time required, in part due to computationally expensive fitness functions.

Effective self adaptive strategies in ECA algorithms are an area of active research with a few self adaptive DE being strategies recently proposed as well as criticised. Self adaptive DE methodologies encode the *CR*, *F* and *NP* parameters as solution vectors, giving bounds to their maximum ranges and evolving in a similar manner as the other solution vectors (Brest et al. 2007). Brest et al. (2006) offer a self adapting DE algorithm encoding the *CR* and *F* parameters in the agents as solution vectors. When an arbitrary agent is evaluated, the *CR* and *F* parameters are considered first, and possibly mutated. After this procedure, the remaining solution vectors are evaluated as in Figure 2.17 with the new *CR* and *F* values. A new *F* parameter for an agent is calculated as follows:

$$F_i = \begin{cases} F_l + rand_1 \times F_u, if\ rand_2 < T_1 \\ F_i\ if\ rand_2 \geq T_1. \end{cases} \quad (2.32)$$

And *CR* is calculated by:

$$CR_i = \begin{cases} rand_3, if\ rand_4 < T_2 \\ CR_i\ if\ rand_4 \geq T_2 \end{cases} \quad (2.33)$$

in which $rand_j, j \in [1,2,3,4]$ are random variables in the range [0...1]. F_l and F_u are the lower and upper bounds on the *F* parameter. T_1 and T_2 are in the range [0...1] and define the probabilities that *CR* and *F* will be changed. This heuristic, which is intended to automate the selection of the *CR* and *F* parameters, introduces an additional four user defined parameters. Brest et al. (2006) note that these parameters are insensitive and suggest $T_1 = T_2 = 0.1$, $F_l = 0.1$ and $F_l = 0.9$ as default values (see Brest et al. 2006 for a thorough explanation). They demonstrate the utility of this approach, noting the self adapting DE converge faster and more consistently on most test datasets, compared to a hand tuned DE.

This heuristic can be improved by incorporating additional boundary constraint heuristics (Brest et al. 2007) and through creating a flexible population size (Brest & Maucec 2008). Qin, Huang & Suganthan (2009) presents a methodology with adaptive *CR*, *F* and *NP* variables in addition to an adaptive trail vector generation strategy.

Pedersen (2010) disputes the research findings on adaptive DE, giving quantitative evidence comparing meta-optimised adaptive DE and meta-optimised DE and suggests that meta-optimized adaptive DE holds no accuracy advantages over a simpler meta-optimised DE. Although the adaptive DE strategies still lack convergence proofs (due to the difficulty introduced by variable meta-parameters), these strategies have shown good results on test datasets. The basic adaptive DE strategy presented by Brest et al. (2006) is used in this thesis due to the expensive fitness functions employed (segmentation) and a requirement for fast execution times (as opposed to a meta-optimisation process or arbitrary selection of control parameters).

2.5.4 Particle swarm optimisation

Particle swarm optimisation (PSO) is a population based stochastic search technique credited to Kennedy & Eberhart (1995). PSO simulates the social behaviour of flocking birds or schooling fish. Like other ECAs, PSO has a population of individuals that change position in every iteration of the search algorithm. Unlike GA strategies, agents in a PSO are not

subjected to crossover or mutation. Intuitively, agents in a PSO can be described as moving over the search surface. Agents are modified in every iteration of the algorithm by some notion of search direction and speed based on personal experience of the search space and that of the global best agent. In addition to storing the position of an agent in the search space, a PSO also stores the current velocity of an agent.

In a PSO, each agent remembers the best position it has occupied in the search space, called 'pbest'. The current position of the agent with the best fitness is known as 'gbest'. The position of an arbitrary agent x, in an arbitrary iteration of the algorithm, in the D-dimensional search space is expressed as $x = (x_1, x_2, \dots, x_D)$. The agent's current velocity is expressed as $v = (v_1, v_2, \dots, v_D)$. Pbest of the agent is written as $p = (p_1, p_2, \dots, p_D)$ and gbest as $g = (g_1, g_2, \dots, g_D)$. In each iteration of the PSO algorithm a particle updates its velocity through:

$$v^{new} = v^{previous} + c_1 r_2 (p - x) + c_2 r_2 (g - x) \quad (2.34)$$

where c_1 denotes the cognition learning factor and c_2 the social learning factor. The elements $c_{1,2}$ are commonly in the range [0...4] and typically set to $c_1 = c_2 = 2$ (Hu 2006). It can intuitively be suggested that $c_{1,2}$ controls the exploration and exploitation behaviours of the PSO algorithm. A PSO also has bounds on agent maximum velocities and an inertia weight (maximum allowed change in velocity from one generation to the next). A large c_1 compared to c_2 implies the agent will be influenced more by its own local behaviour than by the location of the global best agent. The variables $r_{1,2}$ are random numbers (weights) in the range [0...1]. An agent updates its new position according to its new velocity as follows:

$$x^{new} = x^{previous} + v^{new}. \quad (2.35)$$

This search heuristic of PSO is illustrated in Figure 2.19. The z axis (range [-1...1]) denotes the fitness values to this two dimensional search space. The position and velocity of an agent is updated by observing its own past best solution and the global best solution.

During initialisation, a PSO generates a number of agents at random locations in the search space. At the start of every iteration of the algorithm the fitness of all the agents is calculated and gbest is updated accordingly. Agents are altered with equations 2.34 and 2.35 in every iteration. As an ECA algorithm, a PSO can terminate in any method described in Section 2.5.1.

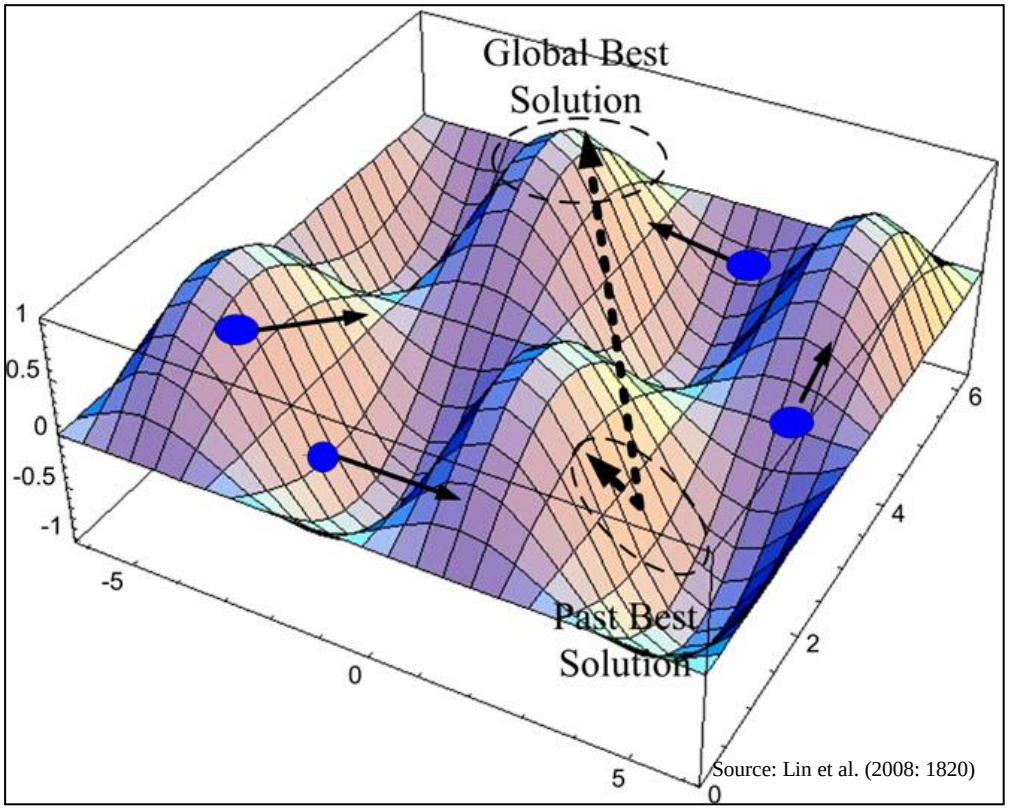


Figure 2.19: Illustration of the search heuristic found in particle swarm optimisation

Many variations of PSO have been suggested that aim to improve this basic variant in general conditions or problem specific domains. Del Valle (2008) and Poli, Kennedy & Blackwell (2007) submit details on recent developments in PSO research (e.g. on improved agent seeding strategies, adaptive PSO and advanced variants of PSO). The basic variant described above is used in this work.

PSOs are acclaimed for their fast convergence to a global optimum (see Figure 2.16) (Vesterstroem & Thomsen 2004) and for their adaptability to different data quantisations (Poli, Kennedy & Blackwell 2007). PSO algorithms may have problems in escaping from local optima and subsequently may need numerous runs of an algorithm to obtain a robust result. PSOs are sensitive to the selection of free parameters as well as the distribution of the initial agent seeding (Vesterstroem & Thomsen 2004), although techniques addressing these problems have been proposed (see Del Valle et al. 2008). There is also a lack of convergence proofs for PSO (Poli, Kennedy & Blackwell 2007).

Despite these concerns, PSOs have recently received much attention as combined numerical and combinatorial optimisers (further discussed in Section 2.6), which is a scenario encountered and further investigated in this work.

2.5.5 Comparison of differential evolution and particle swarm optimisation

DEs and PSOs have different internal structures and consequently display different performance characteristics depending on the search surfaces (problem domain), algorithm settings and algorithm variant used. The application of different ECAs has only been investigated in segmentation parameter tuning in a RS context. Very few case studies exist where these two techniques are directly compared in general or for domain specific problems exist.

The performance of an ECA can be measured in three ways, namely:

1. In terms of its convergence speed, or how fast it can reach the global optimum.
2. In its ability to converge to the true optimum. This performance measure can be extended to measure the algorithm over multiple runs, resulting in a so called measure of robustness, or how often the algorithm converges to the true optimum (Vesterstroem & Thomsen 2004). Robustness is typically indicated by the standard deviation of the optimum obtained over multiple runs of the algorithm.
3. How the first two measures change with a change in population size, agent dimensionality or algorithm iterations.

Mishra (2006), Paterlini & Krink (2006) and Vesterstroem & Thomsen (2004) found that DE slightly outperforms simpler ECA strategies as well as the simple formulations of PSO over test and real world datasets, although data specific cases exist where PSO performed better than DE. The following general conclusions were formulated by these authors:

- PSOs have the fastest convergence rate;
- PSO is susceptible to local optima and is consequently not as robust over multiple runs;
- DE is the most apt in finding the true optimum;
- DE is the most robust algorithm (see above explanation);
- DE is the most stable as the agent dimensionality increases;
- DE is the least sensitive to accurate tuning of control parameters; and
- DE is susceptible to noise or a weakly defined fitness function.

These observations are considered in choosing appropriate ECA strategies for the different components that use an ECA in this work, and will be further investigated in later chapters.

2.6 TECHNIQUES FOR FEATURE SELECTION AND PARAMETER TUNING

Feature selection is the process of identifying a subset of features, attributes or properties that will perform most adequately as input to a classification task. Although a large set of features may be available for a given classification task, many features may hold no discriminative value, may be redundant due to strong linear correlations with other features or may be subjected to noise and detract from the classifier performance. In some cases the dimensionality of the feature set may be prohibitively large. Fewer features imply fewer samples are needed for a good classification (Guyon et al. 2006), may reduce computational load and potentially allow a classifier to generalise better to unseen data (Theodoridis & Koutroumbas 2006). In an object based RS context, features imply object attributes. Feature selection can be seen as a combinatorial optimisation problem.

As discussed in Section 2.4.2.4 with specific reference to SVMs, classifier control parameter tuning is the process of selecting adequate values for the classifier control parameters that will maximise the classifier performance. Control parameters can be mutually dependent. In such a case, classifier parameter tuning is the process of finding an optimal solution in a d-dimensional search space, where d is the number of control parameters and the values in the search space are typically the classifier accuracies with the corresponding control parameters. Control parameter tuning can be seen as a numerical minimisation problem and can be solved with any appropriate search technique.

Two broad strategies exists that can perform feature subset selection, depending on whether the subset selection is done by consulting the classifier/learning algorithm or not, namely the filter and wrapper approaches. In later chapters, segment attribute selection will imply feature selection.

2.6.1 Filter approach

The filter approach to feature subset selection is a pre-processing step to classification, as shown in Figure 2.20.

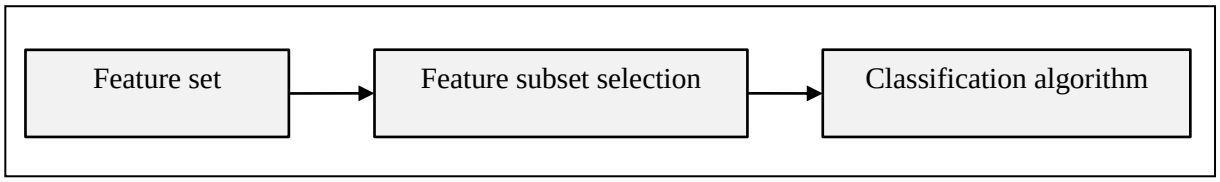


Figure 2.20: Filter approach to feature subset selection

A feature subset selection is done independently of the classifier, which is also its main disadvantage as such an approach disregards the effects of subsets on the performance of the classifier (Kohavi & John 1997). Because a filter approach does not employ the classifier, it is computationally less expensive than a wrapper approach. Examples of filter approaches to feature subset selection include statistical dimensionality reduction techniques such as principle component analysis, independent component analysis, the focus and relief algorithms, f-score, random forest and decision tree methods (Guyon et al. 2006; Kohavi & John 1997; Lin et al. 2008).

2.6.2 Wrapper approach

A wrapper approach to feature subset selection continuously use feedback from the classification algorithm in selecting an optimal subset, as illustrated in Figure 2.21. Classification accuracy feedback from the classification algorithm guides the search for a subset by a meta-heuristic algorithm. A wrapper encompasses a search space (feature space), agent(s) to traverse the search space, a heuristic for exploring the search space (search function) and an evaluation function (fitness function). Generally speaking, a wrapper approach is potentially slower than a filter approach, but may suggest an optimal or better feature subset than a filter approach.

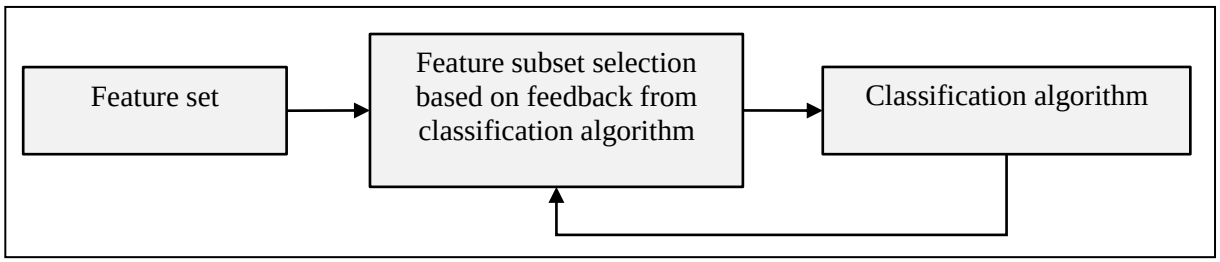


Figure 2.21: Wrapper approach to feature subset selection

2.6.2.1 ECAs for feature selection

An ECA can be used as a wrapper in feature subset selection. The selection of an appropriate feature subset for a given problem influences the accuracy of the resultant classifier, the time needed for training the classifier and the number of training samples needed for accurately training the classifier (Huang & Wang 2006). In such a strategy the solution vector encoding (component 1 of ECAs, see Section 2.5.1) is typically a binary string containing all features, denoting if they are selected or not (e.g. 1 implies a feature is selected and 0 implies it is not selected).

In a RS context, Van Coillie, Verbeke & De Wulf (2007) investigated the use of such an approach for object feature (attribute) selection in an object based RS application. A GA was used and 89 object features were encoded in a binary string as solution vectors. Feature selection was done for an ANN based classification of IKONOS imagery for forest identification. They found improved classification accuracies with reduced feature sets, especially as the number of training samples decreased.

Such a technique can be extended to theoretically superior ECAs. Binary variants of PSO have been proposed (Eberhart 1997; Lee et al. 2008). Lee et al. (2008) did a comparative study of using a pure binary PSO, a PSO with an encoded binary string and a continuous PSO on test datasets. The PSO with an encoded binary string displayed the best performance. Tu et al. (2007) demonstrate the successful use of a PSO for feature selection for SVM classification, using a binary string encoding. Such an approach is also illustrated with a DE (Khushaba, Al-Ani & Al-Jumaily 2008), albeit with a predefined fixed number of features to be selected and using a roulette wheel strategy to handle conflicts (DE cannot be directly used with binary encodings).

Al-Ani (2009) criticises existing feature selection strategies based on PSO, DE and GA, stating that these algorithms were not specifically designed to perform feature selection and consequently do not exploit any notion of feature interdependence. A wrapper strategy is proposed whereby pair wise feature dependence is continuously evaluated. The author notes faster convergence and more accurate results with this approach for feature selection opposed to binary encoded GA, PSO or DE strategies.

2.6.2.2 ECAs for feature selection and parameter tuning applied to SVMs

DE, PSO and GA algorithms have also been used for SVM free parameter tuning for two class and mutli-class problems with various measures of quality used as fitness functions (Kulkarni, Jayaraman & Kulkarni 2004; Lorena & Carvalho 2008). Boardman & Trappenberg (2006) criticise the generalisation performance of such approaches for noisy datasets as the optimal solution in free parameter space may be near a steep gradient to lesser values. They subsequently propose a technique to improve generalisation performance.

The feature subset selection and free parameters of an SVM classifier are interdependent components (Fröhlich, Chapelle & Schölkopf 2003; Huang & Wang 2006; Lin et al. 2008). Fröhlich, Chapelle & Schölkopf (2003) first investigated the worth of encoding the C parameter of an SVM classifier along with features in a GA based combined feature selection and parameter tuning process. They found such a process to be beneficial while employing the theoretical bounds on the generalisation error as fitness function (they suggest this process as it is less computationally intensive compared to a typical k-fold cross-validation). Huang & Wang (2006) extend this approach by adding the gamma parameter for a radial basis function (RBF) SVM to the binary encoding. A fitness function is proposed that incorporates overall classification accuracy, assigns weights to features and awards an encoding with fewer selected features. A ten-fold cross-validation strategy was employed to evaluate this fitness function.

Lin et al. (2008) transferred the approach of Huang & Wang (2006) to a PSO based system. The problem features (attributes) were encoded as continuous variables in the range [0...1], while the SVM free parameters remained continuous variables. They found very little performance increase compared to the technique proposed by Huang & Wang (2006). A significant increase in performance was observed by enabling feature selection as opposed to not using it in the PSO based system. Similar systems have been advanced that combine a binary PSO (feature selection) and continuous PSO (parameter tuning) for SVM classification (Huang & Dun 2008) and support vector regression (Guo et al. 2008). A one class variant (different classifier and different fitness function) of the approach presented by Lin et al. (2008) is used in this work (detailed in Chapter 6).

2.6.2.3 Fitness function considerations for ECA based parameter tuning and feature selection as applied to anomaly detection

The techniques mentioned in the previous section all use some notion of classifier accuracy as fitness function. Two broad approaches are followed, namely using the generalisation error on the theoretical bounds of the SVM (Vapnik & Chapelle 2000), or using some variant of classifier performance measurement (e.g. Lin et al. 2008). Use of the classifier accuracy approach is computationally more expensive. These studies investigated the usefulness of simultaneous feature selection and free parameter tuning on two class classification problems (SVMs). In this work, the focus is on an anomaly detection approach and such fitness functions are ineffective for estimating classifier performance on highly unbalanced datasets.

Tax & Müller (2004) propose a system for free parameter tuning applied to SVDD based on incrementally increasing classifier complexity (e.g. increasing gamma when utilising an RBF kernel) until the accuracy on the target class becomes unstable. Zhuang & Dai (2006) presents a free parameter tuning strategy (no feature selection) for one class SVM and SVDD that uses the geometric means metric. The geometric means metric proposed by Kubat & Matwin (1997) is expressed as:

$$g = \sqrt{acc_p * acc_n} \tag{2.36}$$

where $acc_p$is the accuracy of the class of interest, also called the sensitivity measure written as $\frac{true\ positives}{true\ positives + false\ \ negatives}$ and acc_n is the accuracy of the other class, written as $\frac{true\ negatives}{true\ negatives + false\ \ positives}$. Kubat & Matwin (1997) elaborate on the distorting effect of highly unbalanced datasets on typical accuracy metrics and suggest the use of the geometric means in such cases. Zhuang & Dai (2006) use this metric (and other metrics for unbalanced datasets) in an advanced grid based search system for free parameter tuning in a SVDD and one class SVM based classification, although they note its use can easily be extended to a more advanced and automated search procedure (such as an ECA). The geometric means accuracy is used as a metric in this study.

2.7 SUMMARY

Recent advances in the information sciences offer promising possibilities for the domain of applied RS. In this chapter an overview was given of RS principles and processing paradigms. Approaches to automating image segmentation have been investigated, with

empirical discrepancy methods showing promise for tailoring segmentation based on user defined features. In the domain of earth observation image processing, such approaches are in their infancy, compared to the maturity of the techniques proposed in the domains of computer vision and biomedical imaging.

Kernel based statistical classification is the newest paradigm to emerge from pattern analysis, and has recently been successfully applied to a diversity of RS image classification problems. Kernel based classifiers are robust to high dimensional data, are adept at predicting non-linear data relationships and are less sensitive to the Hughes phenomenon, suggesting its applicability in RS image processing. In addition, evolutionary computation algorithms can be employed to automate and enhance the capabilities of such classifiers. To date, few case studies exist that investigate the usage of recently proposed optimised kernel based classification techniques on high dimensional object based RS data.

Objective two of this thesis was addressed in this chapter. Findings from this chapter were applied in constructing a highly automated one class classification system. This system is introduced in Chapter 4 and profiled in Chapters 5 to 7. In the next Chapter an overview is given on the data, data transformations used, data pre-processing steps undertaken and the choice of study area.

CHAPTER 3: DATA DESCRIPTION

In this chapter the selection of an appropriate sensor and scene is discussed and a description is given of relevant geographical features of the study area. The pre-processing steps performed on the selected data are explained and the data transforms undertaken are recounted.

3.1 SPOT 5 SATELLITE IMAGERY

The SPOT 5 satellite (launched in 2002) carries, among other optical sensors, a high resolution imaging instrument. The important characteristics of the satellite and its high resolution geometric sensor (HRG) are summarised in Table 3.1.

Table 3.1: SPOT 5 satellite and HRG instrument details

Satellite/Sensor property	Description
Orbit	Sun synchronous
Equator crossing time	10:30 am
Spectral bands and spatial resolution (HRG)	Two stereoscopic panchromatic bands at 5 m resolution Three multispectral bands at 10 m resolution (green, red, near infrared) One short wave infrared band (20 m re-sampled to 10 m resolution)
Spectral range (HRG)	Panchromatic band: 0.48 µm - 0.71 µm Green band: 0.50 µm – 0.59 µm Red band: 0.61 µm – 0.68 µm Near infrared band (NIR): 0.78 µm – 0.89 µm Short wave infrared band (SWIR): 1.58 µm – 1.75 µm
Data quantisation	8-bit
Imaging swath	60 km x 60 km

The two stereoscopic panchromatic bands in SPOT 5 HRG scenes are fused to generate 2.5 m panchromatic data. The SWIR band data is re-sampled to ten metres to coincide with the other 10 m spatial resolution bands. Five band SPOT 5 HRG 60 km x 60 km scenes are distributed in completely raw form (level 0 pre-processing), having undergone basic radiometric correction (level 1A pre-processing) and basic radiometric geometric corrections

(level 1B pre-processing) or in geo-referenced (level 2 pre-processing) and fully pre-processed forms (level 3 pre-processing). SPOT 5 HRG imagery is part of the very high resolution (VHR) family of earth observation data products, along with imagery captured by other VHR sensors onboard satellites such as IKONOS, QuickBird and Geo-Eye.

Given its good sensor characteristics, SPOT 5 imagery has seen widespread use in many problem domains, including vegetation specific applications. The literature endorses the applicability of SPOT 5 HRG imagery for forest discrimination studies due to:

- its nominal spatial resolution closely resembling the optimal resolution for forest stand boundary delineation (see Section 2.2 for scale space considerations in image classification) (Radoux & Defourny 2007);
- it having a spatial resolution of adequate fidelity to allow texture measures to be used in forest structural discrimination (Van Coillie, Verbeke & De Wulf 2007), a technique found to be effective in Mediterranean vegetation discrimination (Mallinis et al. 2008); and
- the presence of a SWIR band, lacking in other prominent VHR imagery, useful in vegetation identification and biophysical parameter estimation (Shoshany 2000; Wunderle, Franklin & Guo 2007).

In addition to these properties, a SPOT 5 scene covers a large geographic extent (60 km x 60 km) circumventing to some degree problems associated with multiple scene pre-processing captured from sensors with a smaller swath.

A pilot study was conducted whereby the spatial fidelity of SPOT 5 HRG 2.5 m pan fused multispectral imagery was subjectively evaluated for application in this study. Randomly chosen Cape thicket patches were identified on the imagery from expert knowledge of the area. The immediate areas around the identified patches were subjected to two different image segmentation algorithms with a small set of varying segmentation algorithm parameters (see Figures 2.3 and 2.5). The identified Cape thicket patches ranged between 150 m^2 and 3000 m^2 in size with the smallest patch identified comprising roughly 9 x 9 pixels. Both segmentation algorithms were able to segment the identified patches to a satisfactory degree. Other data sources investigated but not used include Landsat 7 ETM+ imagery (too coarse spatial resolution) and aerial photography (extensive coverage of study area not available). Owing to these findings and the above recommendations, SPOT 5 HRG satellite

imagery was chosen as an appropriate data source for generic experimentation and for use in the case study.

3.2 STUDY AREA

The study area in this work is a majority subset of the extent of Cape thicket. Campbell (1985) suggested the differentiating features and floristic composition of this vegetation structural type, in addition to roughly sketching its habitat (see Figure 3.1). Cape thicket (and Mitchell thicket) are found on mountain slopes in the south western and central areas of the Cape Floristic Region (CFR) (Campbell 1985). An understanding of the biogeography of this area is necessary to comprehend the features readily identifiable on VHR RS imagery. Cape thicket and other single class features of this area are used as classes of interest in the experiments reported in subsequent chapters.

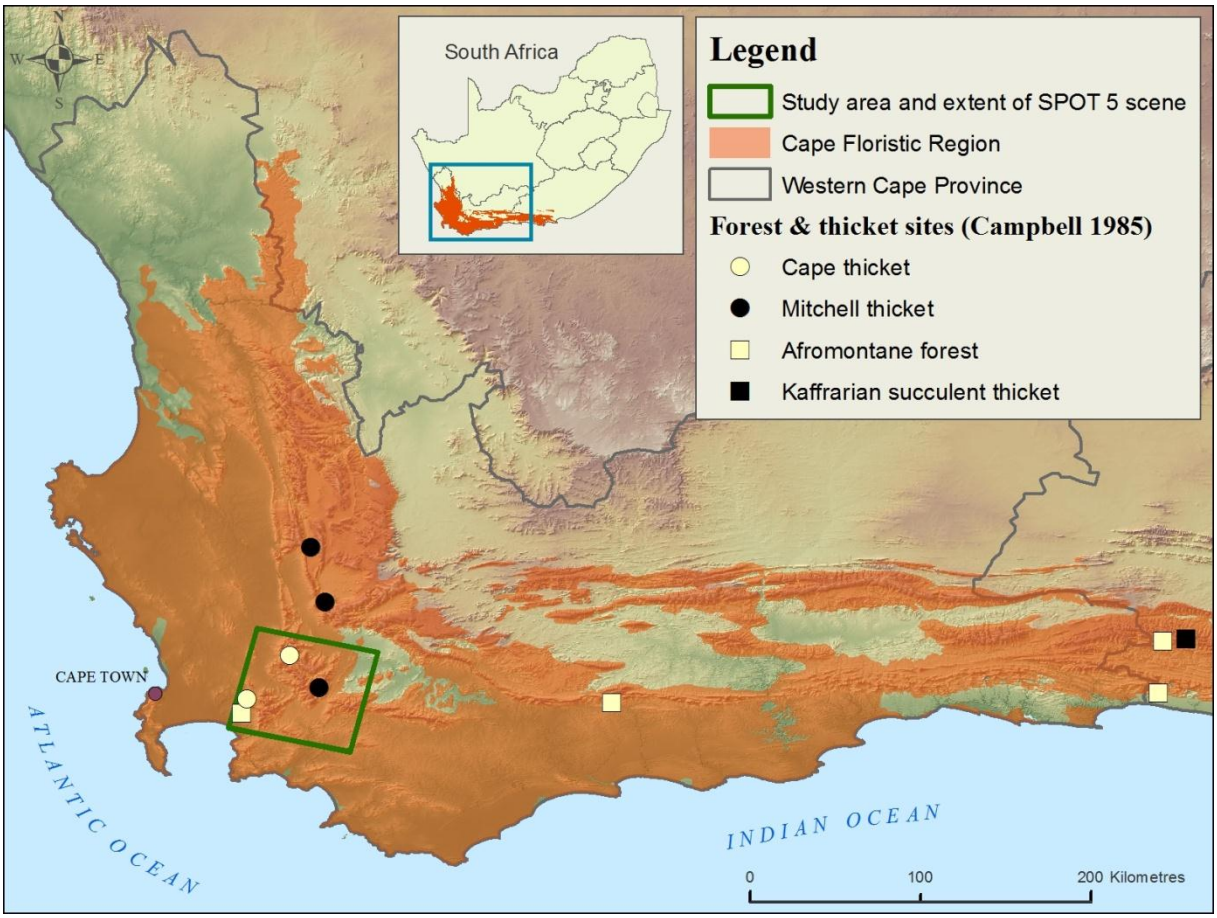


Figure 3.1: Study area and extent of the selected SPOT 5 scene

3.2.1 Major vegetation patterns

The CFR is a phyto-geographical unit that occupies one of the five Mediterranean climate regions in the world (see Figure 3.1). The CFR, predominantly confined to the Western Cape province of South Africa (70% of the region), is recognised as the floral kingdom with the highest species density and displays a high level of species endemicity (68%) (Cowling & Holmes 1992). The mountains in the south western and central areas of the CFR are part of the Cape Fold Belt, comprising primarily of metamorphic sandstones of the Table Mountain Group and Witteberg Group and a few Cape granite intrusions. The soils formed from these quartzitic sandstones, especially near the mountains, are acidic, nutrient poor, shallow and coarse grained (Deacon, Jury & Ellis 1992).

The study area has a Mediterranean climate, with annual rainfall averaging from 500 mm to 1600 mm a year (Campbell 1985). Geographic features such as steep coastal mountains (leeward-windward winds), north-south facing mountains (radiation imbalances), and cold (Benguella Current) and warm (Agulhas Current) ocean currents create climatic gradients that influence the microclimates and consequently the vegetation compositions in this region (Deacon, Jury & Ellis 1992).

Due to the poor soils and winter rainfall climate, the predominant vegetation type in the south-western mountainous inland areas of the CFR is fynbos. Fynbos is a closed schlerophyllous shrubland comprising mainly of proteas, ericas and restios (Cape reeds). An extremely high species richness and the presence of localised endemics prevents such a definition from holding in all cases, also resulting in structural classifications being preferred over floristic classifications in describing vegetation of the Fynbos Biome (Campbell 1985; Cowling & Holmes 1992).

Most nutrient poor mountain slopes in the study region are covered by variants of fynbos. Although fynbos displays high species diversity, due to its very homogeneous structural appearance, predominant schlerophyllous leaf structures and low cover on silica rich sandstone (displaying strong near infrared/mid infrared reflectance), little spectral differentiation is visible among fynbos veld types. An exception is high fynbos composed of dense stands of tall (2 m) Proteaceae, having a typical ‘greener’ appearance on remote sensing (RS) imagery. RS studies on Mediterranean type shrubland typically try to identify end

member components (species) by spectral mixture analysis, with chaparral and macchia receiving some attention in the literature (Roberts et al. 1998; Shoshany 2000).

Fynbos is a pyrophylic (fire loving) vegetation with its life strategies attuned to the seasonal fires common to the region. Due to the fire regimes, fynbos can display very complex structures regarding species composition and age of stands. Limited Afrotemperate forest patches (10 – 30 m in height), occur in fire protected ravines and kloofs (mesic sites). Cape thicket patches (1.5 – 5 m in height) occupy more xeric sites, typically on exposed slopes around sandstone screes, but they are also highly fire resistant (recall Figure 1.1 and see Appendix A). Afrotemperate forests and Cape thicket patches display identifiable boundaries to varying levels, to the encapsulating fynbos (see Figures 2.3 and 2.5 and the photographs in Appendix A) due to some nominal difference in leaf area index, vegetation biomass and structural height (from an RS image processing perspective).

The predominant land use in the study region is agriculture in the lowlands, with limited pine plantations on steeper slopes. Urban and agricultural expansion pose a threat to the endemic mountain vegetation species (Carr & Langhammer 2007).

Common land cover elements visible on an earth observation image in this mountainous area include exposed sandstone, variations of extensive, spectrally undifferentiated schlerophyllous shrubland (fynbos), stretches of invasive tree species, pine plantations, Afrotemperate forest patches, towns and rural land cover elements (agricultural fields, scattered buildings, ad hoc human introduced vegetation and transport infrastructure).

3.2.2 Selection of SPOT 5 scene

A single SPOT 5 scene covering a large portion of the habitat of Cape thicket as described by Campbell (1985) was identified. The scene was captured on 7 November 2006. Figure 3.1. shows the extent of the selected SPOT 5 scene along with Cape and Mitchell thicket habitat hotspots as defined by Campbell (1985). The scene covers the two mountain regional community complexes defined as the south-western area by Campbell, as well as one of the two central mountain regional community complexes.

The selected SPOT 5 multispectral scene (level 1A) identifier is: 5 119-418 06/11/07 08:49:26 2 J. Additionally a level 3 version of this scene was obtained and used in pre-

processing the level 1A scene. Although this level 3 scene was orthorectified, its band values were distorted for use in a mosaic and it displayed distortions in areas of sharp elevation gradients (due to the use of a coarse DEM in pre-processing) thus necessitating the pre-processing of a level 1A scene.

3.3 IMAGE PRE-PROCESSING

Earth observation image pre-processing entails the correction of satellite or aerial imagery for various distortion effects (geometric or electromagnetic). To achieve the most robust classifications it is necessary for images to display reflectance values that are free of these errors. The level of pre-processing required depends on the requirements of the project (Mather 2006). Variations in illumination angle (mountainous terrain) have an influence on reflectance values and subsequent image classification results (Dorren, Maier & Seijmonsbergen 2003; Richter, Kellenberger & Kaufmann 2009; Sheperd & Dymond 2003). This called for robust pre-processing of the data used in this study. The pre-processing carried out on the level 1A SPOT 5 scene used in this study is summarised in Figure 3.2.

3.3.1 Orthorectification

Orthorectification (geometric correction) is the process of referencing a scene having no coordinate system or point of reference to a known coordinate system. Orthorectification takes into account elevation and earth curvature. In orthorectification, a scene is georeferenced with the aid of the sensor geometric model, a geo-reference dataset and a georeferenced DEM. The level 1A SPOT 5 panchromatic scene (2.5 m spatial resolution) was orthorectified (to UTM 34S projection) using the 3D Canada Centre for Remote Sensing Parametric Model supplied with the scene metadata (Toutin's model for optical satellite sensors as implemented in the *PCI Geomatica v10.2* software). The Satellite Applications Centre (SAC) level 3 panfused SPOT 5 image (2.5 m spatial resolution) was used as the reference dataset. This was deemed the optimal orthorectification strategy as other reference data types, for example geo-referenced vector layers or geo-referenced aerial photography have less correlation with the image to be rectified. A 20 m DEM compiled from 20 m contour lines (Van Niekerk 2001) was used in orthorectification. Thirteen ground control points were used with a resulting 0.95 metres root mean square error. Nearest neighbour resampling was used. The level 1A multispectral scene (4 band, 10 m spatial resolution image) was orthorectified in the same manner, using the newly orthorectified SPOT 5

panchromatic scene as reference. Fourteen ground control points were used with a root mean square error of 1.53 metres.

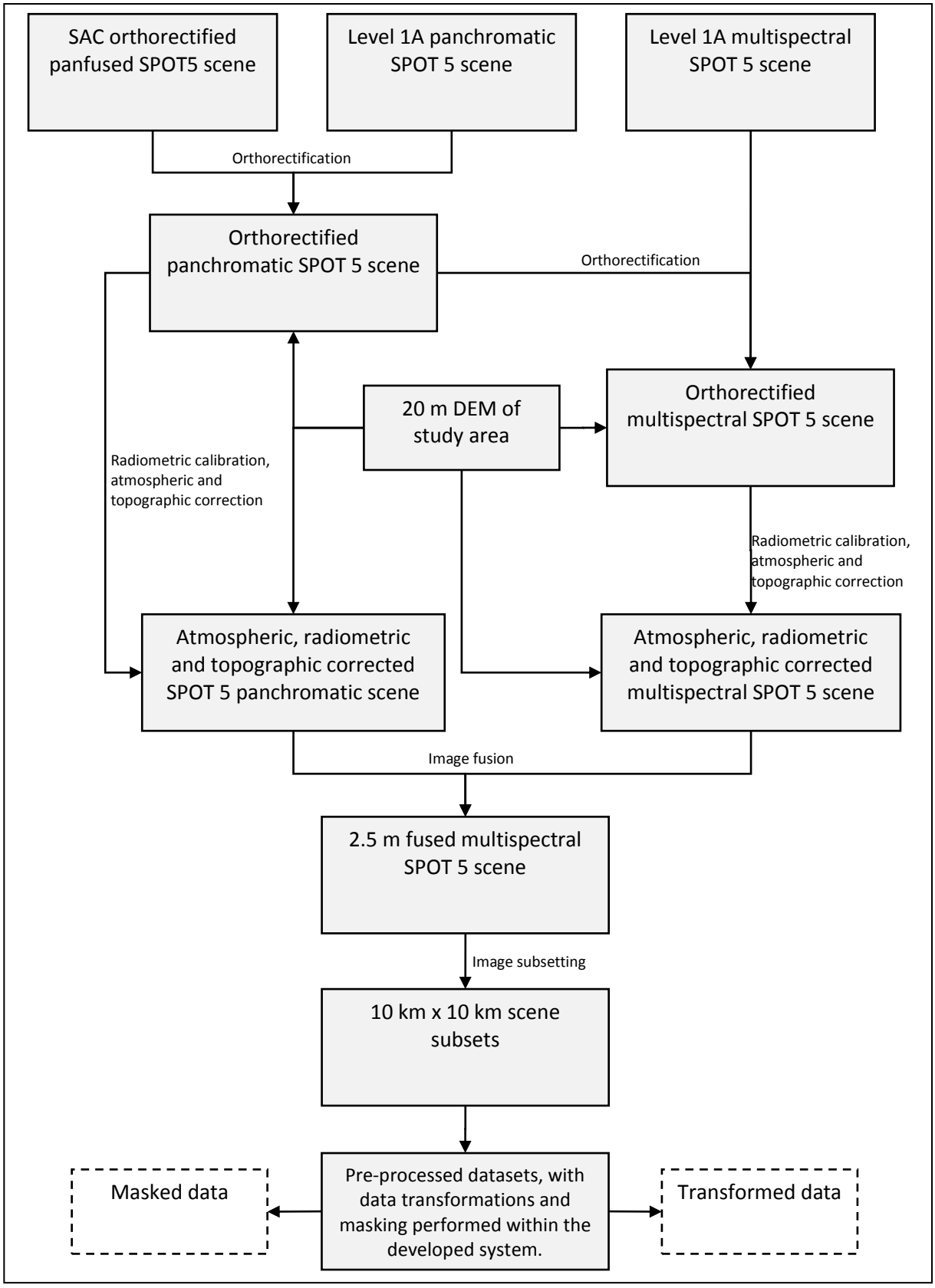


Figure 3.2: Data pre-processing steps undertaken on the SPOT 5 image

3.3.2 Radiometric correction

Radiometric correction (electromagnetic correction) is the process of correcting the pixel values in a scene for distortion effects introduced by the sensor, various atmospheric gases and by rugged topography of the landscape under observation. The ATCOR-3 (Richter 2010) pre-processing heuristics were employed to correct the radiometric distortion effects in the orthorectified SPOT 5 multispectral and panchromatic scenes. The scenes were radiometrically calibrated, corrected for atmospheric gas influences and for differences in illumination caused by topography.

3.3.2.1 Radiometric calibration

Sensors carried by satellites or aircraft capture electromagnetic radiation at specific wavelengths (image bands). The amplitude of incoming radiation is conventionally quantised to 8-bit or 11-bit values and needs to be scaled to fit in this domain (offset and slope). Radiometric calibration entails the calculation of surface radiance or reflectance values from these arbitrary digital numbers, using the gain and offset detail for each band of the sensor at the time the image was captured. The scenes were normalised to surface reflectance values, keeping the quantisation at 8-bit. See Richter (2010) for the formulation of digital numbers to radiance to surface reflectance.

3.3.2.2 Atmospheric correction

This study used a physical atmospheric correction technique based on the MODTRAN-4 radiative transfer model as implemented in ATCOR for atmospheric correction. Relevant scene metadata were used along with relevant input to the model for atmospheric and aerosol conditions (rural land cover, dry and arid conditions, image captured 11 November 2006 with little visible atmospheric influence). Radiative transfer models determine the effect that different, typically encountered, gas concentrations have on radiance/reflectance observed at the sensor according to prior knowledge of the radiance/reflectance of present surface features. Such models can readily be used to correct imagery for errors introduced by particles in the atmosphere (haze, pollution, etc) by using information on the elevation of the topography in the scene, sensor geometric information, atmospheric conditions at the time the image was captured and general information on the surface types encountered in the scene (Lück 2004).

3.3.2.3 Topographic correction

Topographic correction entails the normalisation of reflectance values of pixels, caused by differences in illumination angles. The local (per-pixel) illumination angle is calculated from the solar zenith and azimuth angles as well as the local pixel slope and aspect. The reflectance of this inclined pixel can be normalised to its corresponding horizontal reflectance. In this study, a modified version of the Minnaert normalisation process was used (Richter 2010; Richter, Kellenberger & Kaufmann 2009) to only correct the illumination of pixels having an illumination angle greater than 80 degrees (relative to the position of the sun) thus addressing shadows and very steep south facing slopes. The 80 degree threshold was calculated by querying the solar zenith angle in a set of empirical rules defined in Richter (2010).

3.3.3 Data fusion

Concerning RS image processing, data fusion entails the enhancing of the spatial resolution of the multispectral bands by using the information content of the higher spatial resolution panchromatic band. Panchromatic bands have a higher spatial resolution than the multispectral bands due to the trade-off between the instantaneous field of view of a sensor and its signal to noise ratio. The 10 metre multispectral bands of the SPOT 5 scene were inadequate for the purposes in this study, necessitating a sharpening of these bands. Many fusion techniques exist, including the Intensity, Hue and Saturation, principle components analysis, wavelet based fusion and the least squares pan sharp techniques. The latter technique, used for fusing the imagery in this study, do not suffer from colour distortions displayed by the intensity, hue and saturation and principle components analysis techniques and also preserves the histogram shape of the multispectral bands (Zhang 2002).

Figure 3.3 illustrates a subset comparison of the original level 1A multispectral 10 m spatial resolution scene with the fully pre-processed sharpened 2.5 m scene. The area shown is on the northern slopes of the mountains in the Jonkershoek Valley, near Stellenbosch. The pre-processed image displays a higher contrast due to atmospheric correction and sharpening. A slight anomaly (line) is visible a few metres north of the south facing cliffs, presumably due to the use of a coarser resolution and inaccurate DEM in topographic correction.

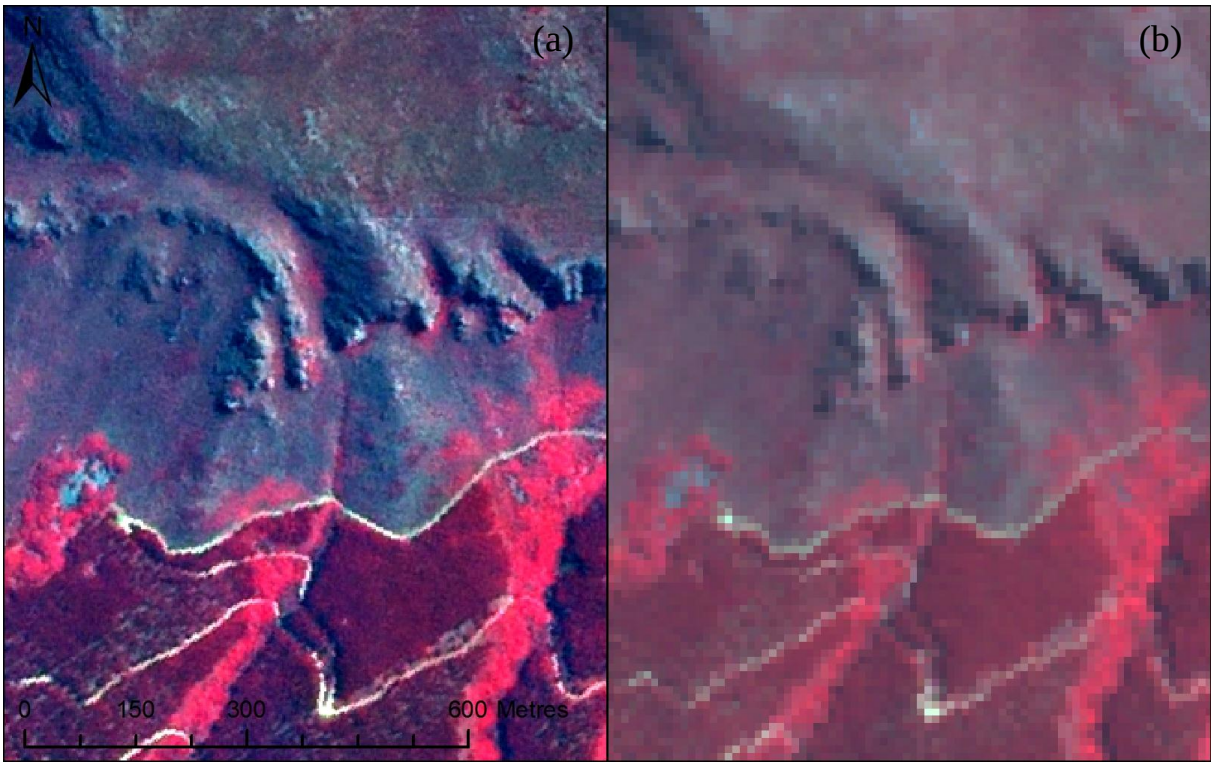


Figure 3.3: Subset of the fully pre-processed scene (a) compared to the original level 1A SPOT 5 scene (b)

3.3.4 Data subsetting

The techniques implemented in the system developed here are computationally expensive regarding memory and CPU time required (repeated segmentation of micro areas, numerous data transforms performed, graph matching, population based search), necessitating the use of smaller subsets of the pre-processed scene. The system can handle SPOT 5 2.5 m scene subsets of up to 3500 x 3500 pixels or 8.75 x 8.75 kilometres in size. The pre-processed SPOT 5 scene is subdivided in 7.5 km x 7.5 km grids for use in the case study presented in Chapter 7. Smaller arbitrarily sized subsets from the fully pre-processed scene and the SAC pre-processed scene are used in the experiments reported in Chapters 5 and 6. Areas within an image can be masked with a binary raster in the presented system.

3.4 DATA TRANSFORMATIONS

Data transformation entails the warping of the original image data, without discarding or losing any part of the data, into a form more suited for a particular information extraction application. Data can be transformed from their original spectral space into a feature space using a linear or non-linear transform function. The spatial interrelationships (called spatial statistics or texture measures) among data subsets can also be calculated and might hold

informative value. The derivation of spectral transforms and texture measures are routinely performed in RS workflows because such data transforms might hold higher discriminative power for the specific application (for classification or segmentation purposes). In this study a selection of spectral and textural transforms are calculated from the original data.

3.4.1 Spectral transformations

In addition to the four default multispectral bands of the panfused SPOT 5 scene, a few other transformed spectral bands are used in this work as guided by suggestions made in the literature, namely several multispectral ratio bands and one discrete colour space transformation.

Simple ratios of the original spectral bands are usually calculated and employed in RS workflows because some earth surface features are more readily identifiable in these stretched feature spaces (Mather 2006). Vegetation indices denote the ratio between observed red and near infrared light, due to vegetation that characteristically absorbs red light and reflects near infrared light. The ratio vegetation index (RVI) is written as:

$$RVI = \frac{Near\ infrared\ band}{Red\ band} \tag{3.1}$$

and the normalised vegetation index (NDVI) as:

$$NDVI = \frac{Near\ infrared\ band - Red\ band}{Near\ infrared\ band + Red\ band}. \tag{3.2}$$

NDVI is more sensitive to lower ratio values than the unmodulated RVI ratio. RVI and NDVI are indicators of vegetation biomass or vegetation vigour. Many problems exist with these ratios, for example oversaturation in densely vegetated areas and poor indicator performance of biomass in lightly vegetated areas, but they are still useful as a strong discriminative feature space when investigating vegetation and are consequently used in this work. Another six band-ratios (with a four channel scene) were also calculated and employed.

The hue, saturation and intensity (HSI) colour space, maps the red, green and blue colour space according to colour brightness (intensity), colour (hue), and colour purity (saturation) axes. In this work, HSI is calculated from the green, red and near infrared bands of the pan sharpened SPOT 5 scene. The HSI colour transform have been used successfully in

Mediterranean vegetation segmentation and classification (IKONOS, using the first three bands) (Mallinis et al. 2008), rural land cover segmentation (Trias-Sanz, Stamon & Louchet 2008), semi-arid object based vegetation classification (Yu et al. 2006) and object based rangeland classification (Laliberte & Rango 2008) confirming this colour space's discriminative power for such vegetation applications.

3.4.2 Texture measures

Spatial statistics or texture measures constitute information describing the interrelationships among pixels (or even image objects) in a scene. In a Mediterranean land cover scene, vegetation features to be differentiated may display similar spectral properties, necessitating the inclusion of texture measures in their discrimination (LLoyd et al. 2004; Mallinis et al. 2008). Johansen et al. (2007) found texture measures more useful in aiding the discrimination of vegetation structural types than vegetation indices from high resolution imagery. As with spectral properties, certain types of texture measures are more useful than others in the identification of the land cover features of interest and can be used in the same manner as spectral data in classification.

Three second order texture measures are calculated for object primitives (segments) in this work. A 3-bit gray level co-occurrence matrix (GLCM) is calculated for each object primitive, using the standard eastward (right side) neighbour rule with no spacing between neighbours. The texture measures (Haralick, Shanmugam & Dinstein 1973) calculated are:

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|^2} \quad (3.3)$$

$$\text{Contrast} = \sum_{i,j} |i-j|^2 p(i,j) \quad (3.4)$$

$$\text{Entropy} = \sum_{i,j} -\log\big(p(i,j)\big)\, p(i,j) \quad (3.5)$$

where $p(i,j)$ is the corresponding value in the GLCM. Homogeneity measures the closeness of the distribution to the GLCM diagonal. Entropy is a measure of textural smoothness. Contrast measures the intensity difference of pixel and pixel neighbour over the entire GLCM. Homogeneity, contrast and entropy are recommended in the literature as being relevant or important texture measures in the discrimination of vegetation structural types (Laliberte & Rango 2009; Laliberte et al. 2004; Wunderle, Franklin & Guo 2007; Yu et al. 2006).

CHAPTER 4: THE Geo-ND SYSTEM

This chapter introduces the proposed geographic object novelty detection system and an application embedding this system is presented. Chapters 5 and 6 elaborate on specific components of the developed system.

4.1 ANOMALY DETECTION WORKFLOW

In a RS context, anomaly or novelty detection entails the labelling of a specific type of land cover feature. A feature of interest has some universal spectral, textural and/or contextual properties that allow it to be discriminated from and identified among all other features present in a scene. In an object based image classification workflow, an anomaly detection approach has a few fundamental differences from a multi-class approach. These dissimilarities are briefly explained next.

4.1.1 Anomaly detection versus multi-class classification

In the context of object based data representation, the segmentation process should focus on accurately segmenting the features of interest with a disregard for other features present in the scene. Multi-class image segmentation entails the segmentation of a scene with numerous types of features having importance. In an anomaly detection approach, greater fidelity in segmentation can be achieved by using the shape and spectral properties of a few identified anomalous features as guides to obtain good segmentation algorithm parameters (see Section 2.3.3.2).

Classification of anomalous features, in a statistical supervised context, may be conducted by a dedicated anomaly detector (see Section 2.4.2.2) or through multi-class classification. Multi-class classification entails the collection of training samples of numerous classes not of interest to the specific application. Also, in anomaly detection, multi-class measures of classification accuracy cannot be used as an indication of classifier performance, as discussed in Section 2.6.2.3.

4.1.2 Proposed geographic object anomaly detection workflow

An object based anomaly detection workflow is proposed that attempts to minimise user interaction, attain high classification accuracy through automatic attribute subset selection and

free parameter tuning, and to provide a measure of quantitative credibility in the image segmentation process. This workflow is illustrated in Figure 4.1 which shows that a user is required to delineate a few anomalous features of interest to be used in guiding the generation of image segments. The system assumes that the features of interest are representable at a single nominal scale (one level representation). After the segmentation process, anomalous features are identified to be used in a one class statistical classifier.

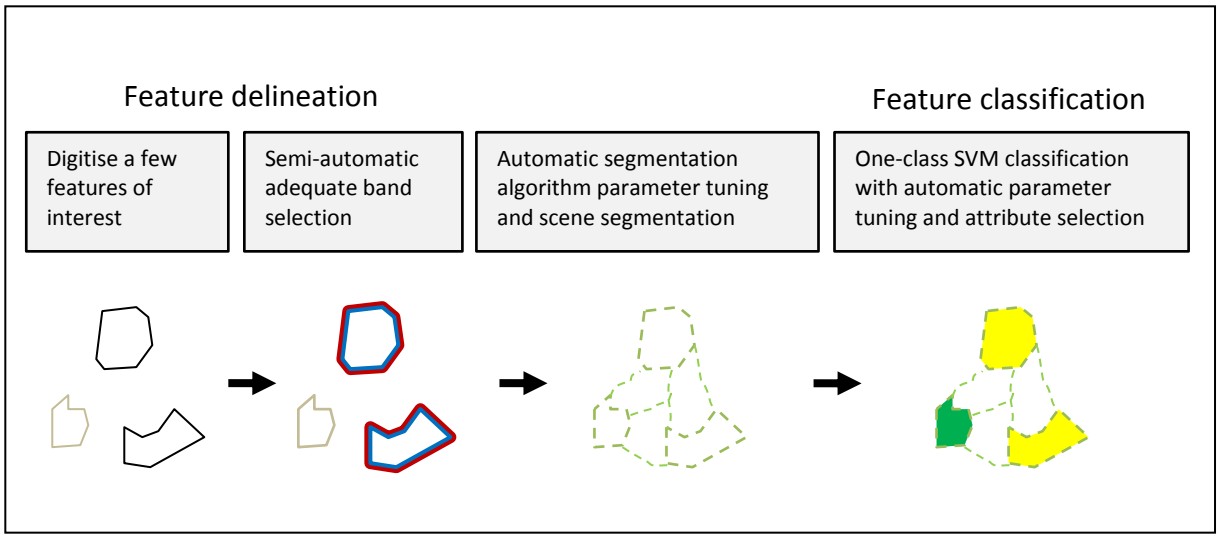


Figure 4.1: Proposed one class classification workflow

An empirical discrepancy metric is used, in conjunction with a band selection tool to automatically generate suitable image segments delineating the features of interest. These segments are subsequently classified with a one class SVM classifier, using a population based search technique for classifier free parameter tuning and attribute subset selection.

4.2 GEO-ND

An image processing software application titled the Geographic Object Novelty Detector (Geo-ND) was developed that encapsulates the proposed classification workflow described above. Geo-ND was developed for a dual purpose:

1. To act as a proof-of-concept application for the proposed workflow, readily usable with new imagery and new anomaly detection problems. Geo-ND is used in the case study presented in Chapter 7.
2. To act as a platform for quantitative experiments in proposed techniques for automating the segmentation and classifier tuning processes. See Section 4.2.2 for details on the experimental designs.

4.2.1 Geo-ND overview

Geo-ND was mainly developed in C++ programming language with the exception of a few components programmed in MATLAB. A few general public license libraries were also used in the construction of Geo-ND, most notably Terralib (Terralib 2009), Qt (Qt 2009), libSVMplus (Russo 2009) and SwarmOps (Pedersen 2008). The application and sample data are available on the accompanying DVD, in addition to a detailed user guide (see Appendix C).

Geo-ND consists of six distinct panes. Each pane performs a specific task in the geographic object anomaly detection workflow. The panes are entitled "1. Digitise features of interest", "2. Select input bands for segmentation", "3. Parameter search", "4. Segment image", "5. Classify segments" and "6. Export results". A workflow proceeds through these panes from left to right. The panes are situated just below the program toolbar (see Figure 4.2). The application and workflow are briefly discussed below.

4.2.1.1 Digitise features of interest pane

Figure 4.2 shows the content of the first pane found in Geo-ND. As an example, a user is interested in identifying Cape thicket patches on a SPOT 5 scene. In this pane a user digitises a few anomalous features of interest to be used in an automatic segmentation algorithm free parameter tuning process. The more features a user delineates, the more directed the results will be: at the expense of computation time. Two methods of feature delineation are available, namely digitising features with polygons, and digitising edges within a user defined region of interest. The polygon features are used together with the larger segments booster (LSB) metric (discussed in Section 2.3.3.2) in a segmentation parameter tuning process. This technique, proposed by Fredrich & Feitosa (2008) and using the LSB metric, is replicated in Geo-ND and used as a benchmark to evaluate an edge based system. The white semi-opaque areas around the digitised features indicate the regions of interest to be used in segmentation parameter tuning. The edge delineation method uses an edge metric for segmentation algorithm free parameter tuning and is discussed further in Chapter 5.

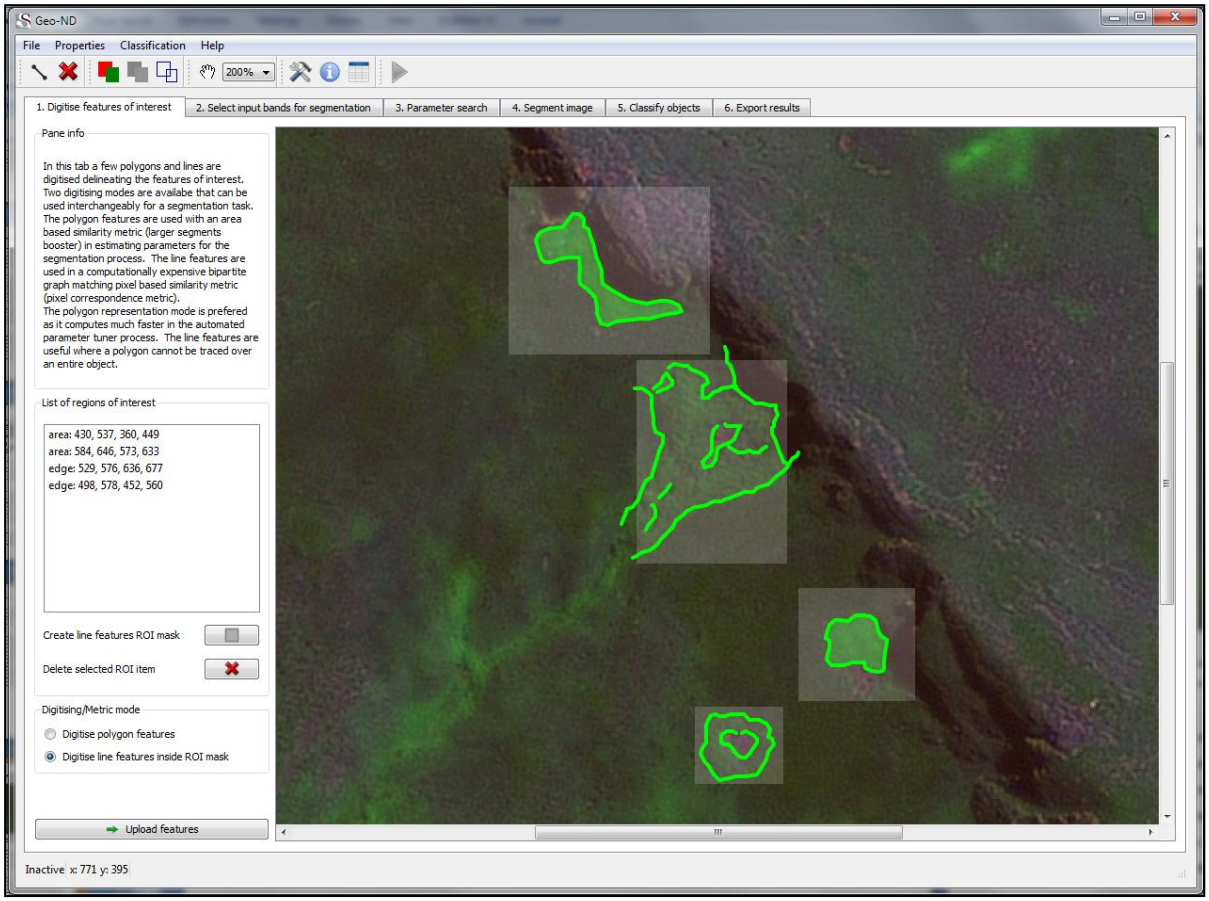


Figure 4.2: First pane in Geo-ND

4.2.1.2 Input band selection pane

Figure 4.3 shows the second pane found in Geo-ND. In this pane the spectral separability of the spatial neighbourhood of the digitised features of interest is investigated. Fifteen spectral transforms are generated over each digitised feature (see Section 3.4.1). The spectral separability, measured by the Jefries-Matusita distance (a bounded derivative of the Bhattacharyya distance in the range [0...2]), is calculated over a certain user defined buffer distance from the digitised lines and polygons.

This simple tool assists users to explore and select adequate image bands to be used in the segmentation process by suggesting the use of the three bands with the greatest spectral separability. The region merging segmentation algorithm implemented in Geo-ND uses a measure of spectral separability in deciding if two segments should be merged (the mean distance between two normal distributions, see equation 2.1). The rationale for this tool is that the spectral bands with a high spectral seperability in the boundary regions of the features of interest are more suited as input to the segmentation process (as measured by the Jefries-Matusita distance which combines a notion of the mean distance between two distributions

and the standard deviations of said distributions). As an example, the near-infrared band, NDVI transformation and intensity transformation (HSI) were found to deliver the greatest spectral separability over an eight pixel buffer distance around the delineated features as shown in Figure 4.3.

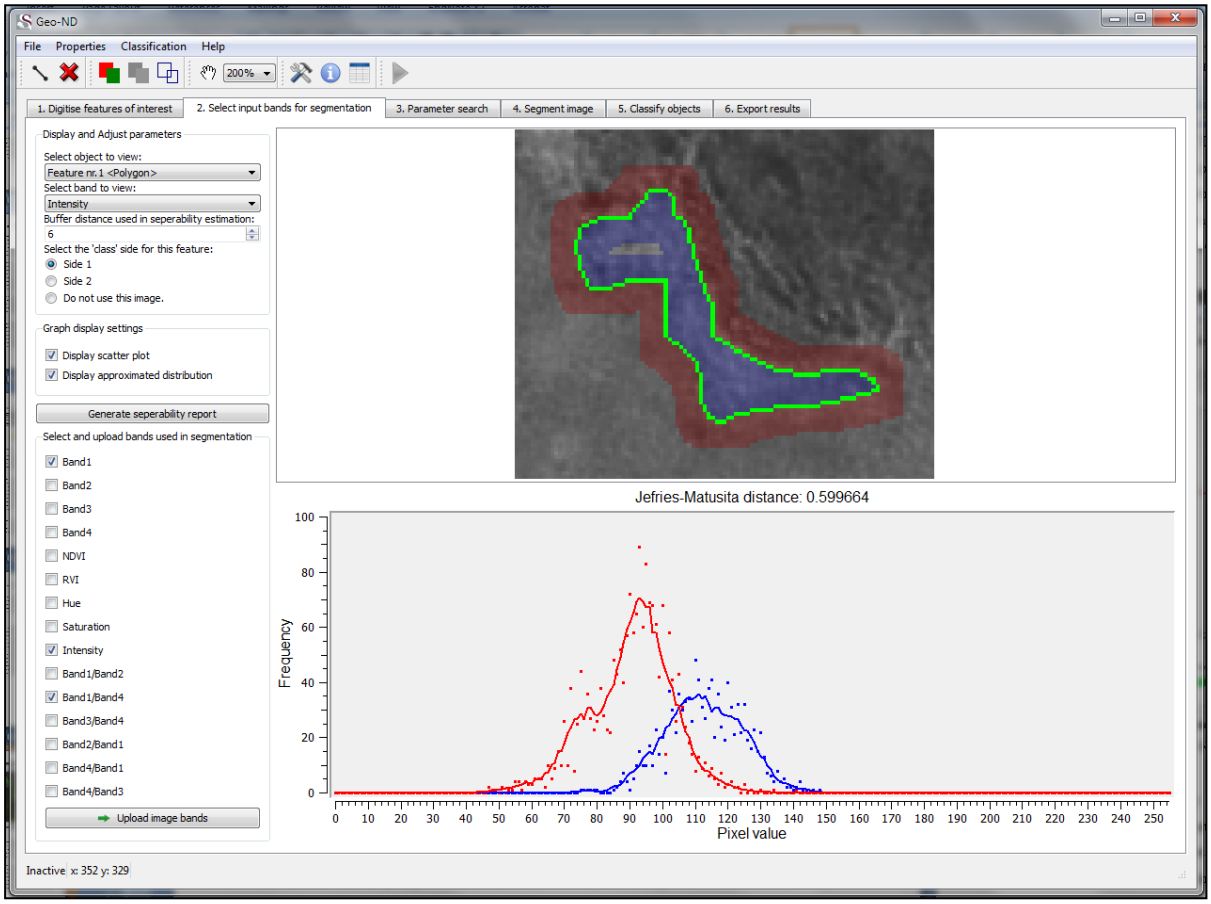


Figure 4.3: Second pane in Geo-ND

This tool does not attempt to be a quantitative method of automated adequate band selection for the segmentation process. Automatically determining adequate buffer distances is not a trivial task, nor would a buffer of continuous width be appropriate in all cases. This tool merely aids a user in qualitatively exploring segmentation algorithm input bands. Adequate segmentation algorithm input band selection has an influence on the quality of the generated segments (see Section 2.3.4). The accompanying user guide provides more detail about the use of this tool (see Appendix C).

4.2.1.3 Parameter search pane

The third pane in Geo-ND is illustrated in Figure 4.4. In this pane adequate segmentation algorithm parameters are determined by the technique described in Section 2.3.3.2 using the LSB metric for polygon features and the pixel correspondence metric for edge features (see Section 4.3). Either a differential evolution or particle swarm heuristic can be used for the search. A user can adjust the length of the population based search (by number of iterations). More iterations allow for a more thorough exploration of the parameter space at the expense of computing time. The parameter search heuristic uses all digitised features when searching the parameter space.

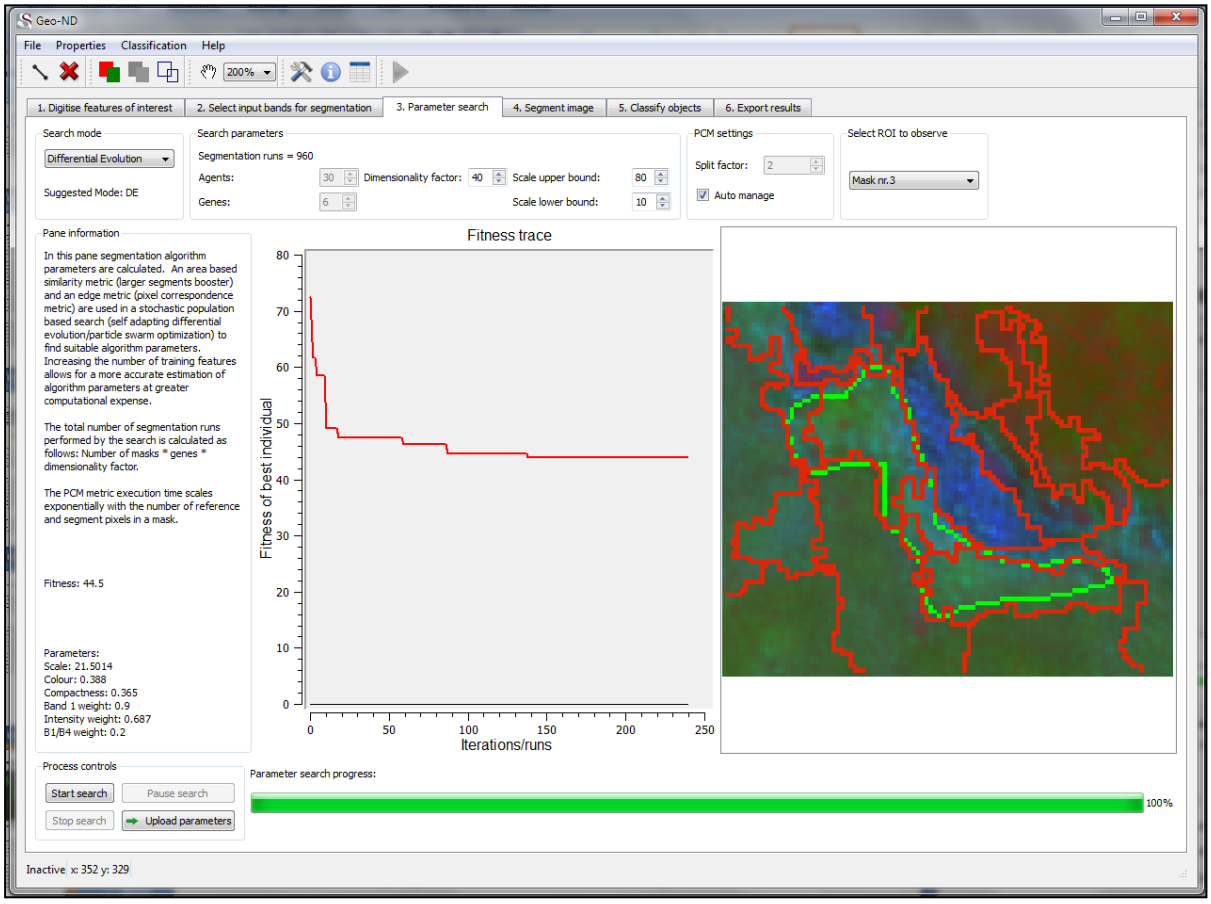


Figure 4.4: Third pane in Geo-ND

The graph in the pane illustrates the fitness of the best individual as the algorithm proceeds in its processing. The image on the right side of this pane illustrates the segments generated with the previous parameter set in the evaluation (a user can choose which region of interest to view). It is continuously updated as the algorithm proceeds. Algorithm generated segments and user polyline input are transformed to border/edge pixels by a Roberts filter (Roberts 1965) to be used by the similarity metrics.

4.2.1.4 Segment image pane

The three fittest parameter sets are presented to a user as results in the fourth pane (Figure 4.5). These three sets of parameters can be averaged, a specific set selected or edited before subjecting the entire image to the segmentation algorithm with the given parameter set. The area and edge features used in segmentation algorithm parameter search can be reviewed in this tab, along with the generated segments.

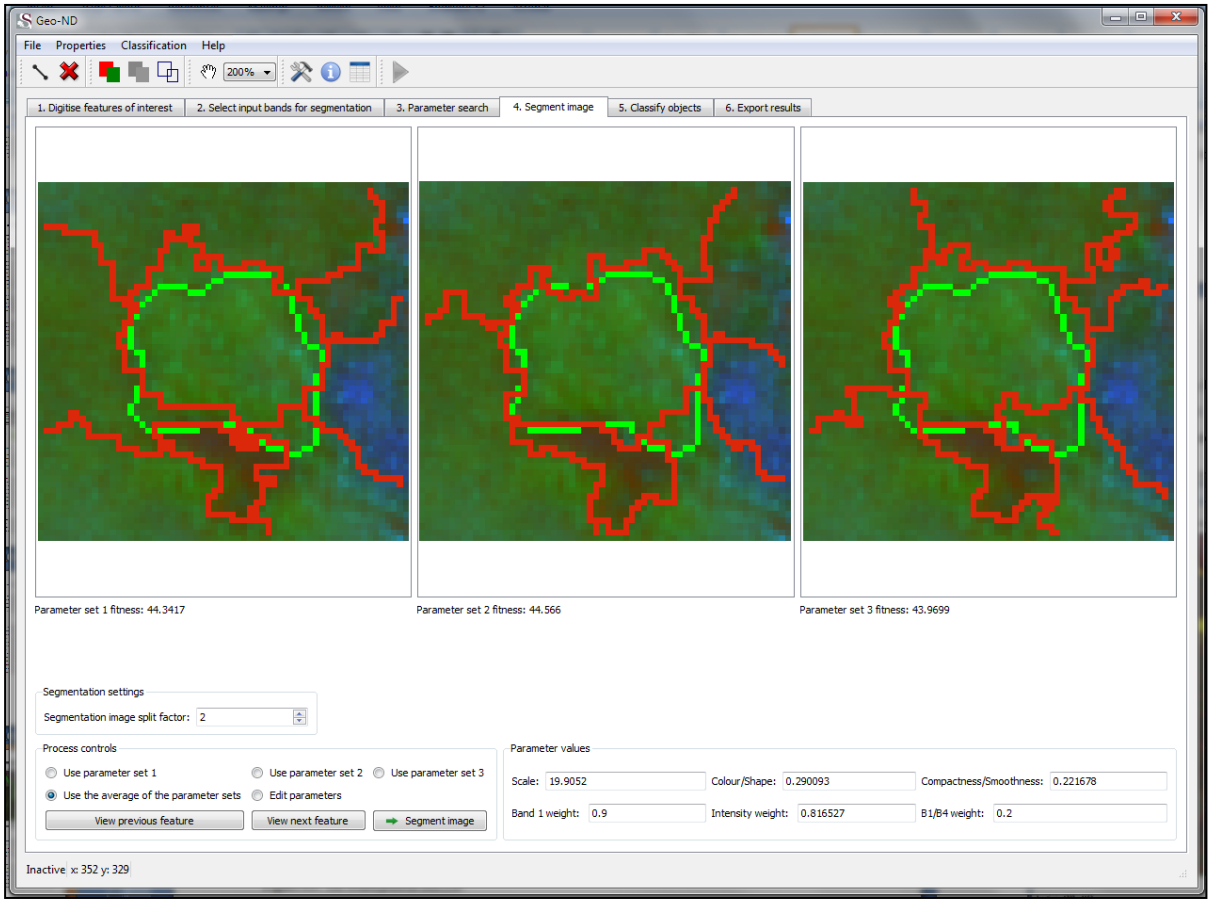


Figure 4.5: Fourth pane in Geo-ND

4.2.1.5 Classification pane

Figure 4.6 reproduces the fifth pane in Geo-ND. In this pane a user can either perform a manual supervised object based one class SVM classification of the entire image or, alternatively, consider the entire image as part of training data used as input to an automated classifier free parameter tuning and attribute subset selection heuristic.

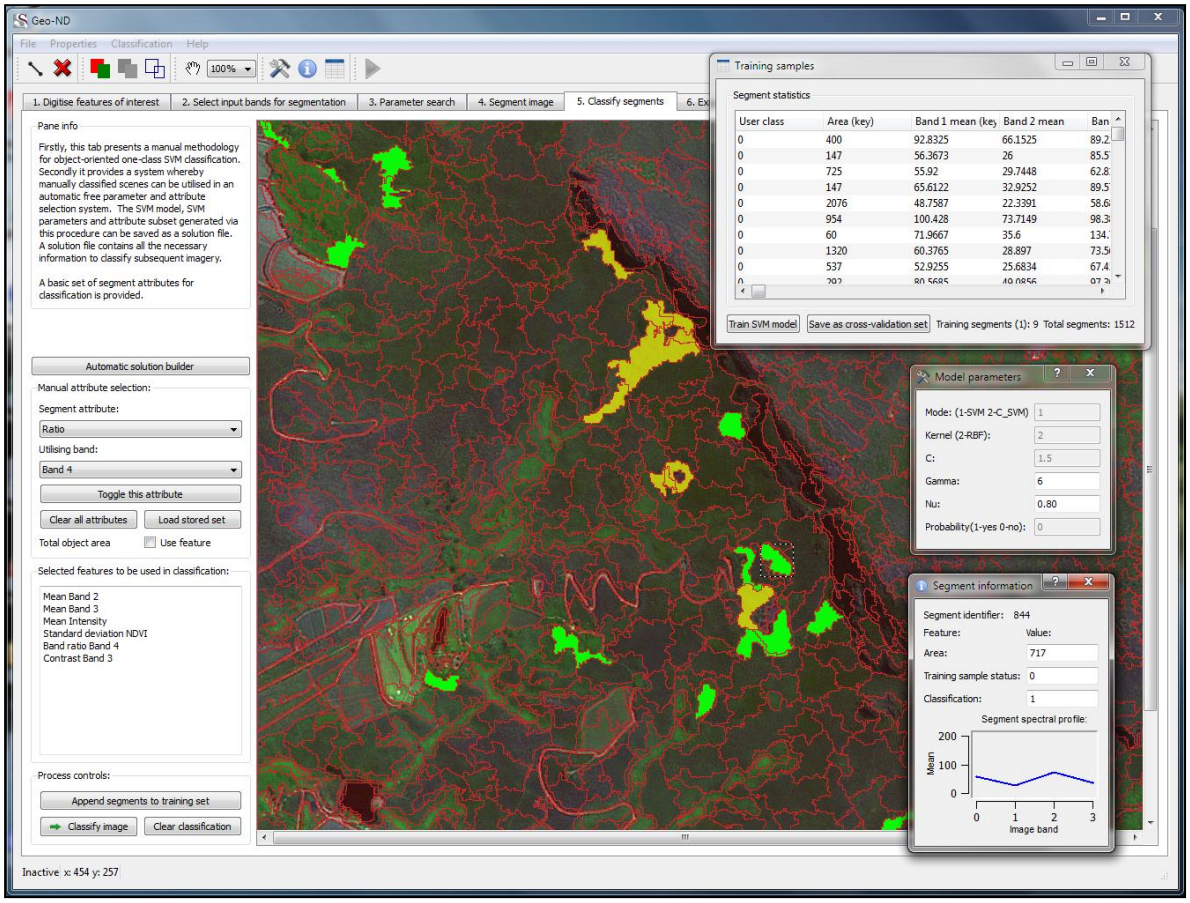


Figure 4.6: Fifth pane in Geo-ND

This fifth pane consists of an interactive mapping canvas, displaying the earth observation image and generated segments as red lines. A segment attribute selection tool is anchored on the left side of this pane. A few basic segment attributes are provided, including the mean object value, standard deviation, ratio to bands, maximum difference, GLCM homogeneity, GLCM entropy and GLCM contrast. These attributes are calculated for the 15 spectral data bands described in Section 3.4.1, except for the ratio to bands attribute that is only calculated for the original four loaded image bands. In total, 95 segment attributes are available, including the segment surface area. A user can adjust which attributes are used when conducting a manual classification, or specify an initial set of attributes for the automatic free parameter tuning and attribute subset selection system.

Accompanying this pane are three floating windows, titled “Training samples”, “Model parameters” and “Object information”. The training samples window displays the attribute statistics for all segments, indicates training sample status and provides functionality for a user to train the one class SVM model (manual classification mode) or save the training data to be used in the automatic free parameter tuning and attribute subset selection system. The

model parameter window allows a user to tune the one class SVM model parameters (gamma and nu for an RBF one class SVM). The object information window displays basic information on the currently selected segment.

The manual classification mode in the fifth pane of Geo-ND involves selecting anomalous segments of interest, selecting segment attributes to be used in the classification and tuning the classifier free parameters. Figure 4.6 illustrates such an approach with yellow segments selected as training areas. The attribute subset and one class SVM parameters used were set through trial and error testing. The green segments are classified as belonging to the class of interest. Negatively classified segments remain translucent.

As discussed in Section 2.6.2.2, the selected data attribute subset and SVM classifier free parameters are interdependent features and can be optimised with a search heuristic. Such a methodology is implemented in Geo-ND. Geo-ND allows a user to use the segments in the entire image as ground truth data and requires that all anomalous (class of interest) segments are identified and selected. The data from such an image are saved as a cross-validation training set. New cross-validation training data sets can be generated by loading new imagery and re-starting the classification workflow. The fifth pane in Geo-ND provides a tool, entitled the automatic solution builder, that uses these datasets to derive an adequate attribute subset and one class SVM parameter set. This information, which includes the attribute subset, one class SVM parameters and training data, can be saved and loaded as a solution file. A solution file contains all the necessary information to identify anomalous features on new scenes. The accompanying user guide provides more detail about using these two classification approaches.

Due to the nature of the segmentation algorithm used in Geo-ND (positions of region seeding points in imagery of varying dimensions), the exact segmentation results over the regions of interest as viewed in pane three will not match the same regions in the entire scene in the fifth pane. Currently, computation constraints prohibit the use of an entire scene as input to a segmentation algorithm that forms part of a fitness function in an evolutionary search heuristic.

The sixth pane in Geo-ND allows a user to export the results of the classification and/or the generated segments as raster imagery. Such output can be disseminated further in a GIS.

4.2.2 Geo-ND components experimental design and evaluation

The aim of this thesis is to evaluate new techniques for creating a more accurate and autonomous geographic object novelty detection workflow. An evaluation of the literature led to the construction of Geo-ND which consists of two core components adopted from recently proposed techniques in the computer vision and machine learning domains. The two core components of Geo-ND are the empirical discrepancy method for segmentation algorithm parameter tuning credited to Feitosa et al. (2006) and Fredrich & Feitosa (2008) and the search heuristic for attribute subset selection and classifier free parameter tuning, from Lin et al. (2008).

The first core component is computationally expensive and requires accurate user delineated features. An edge metric, originally proposed by Prieto & Allen (2003) is introduced as a fitness function to this process. This technique has only been investigated with area based metrics. The feasibility of this edge metric is tested as a fitness function and is qualitatively and quantitatively compared with the LSB metric regarding resulting segments and resulting classification accuracies (see Chapter 5).

The value of the second core component has not been investigated in the context of earth observation image processing, nor anomaly detection. The approach by Lin et al. (2008) is adopted for object based one class earth observation image classification using an appropriate one class fitness function. In Chapter 6 the utility of this approach for application on this domain specific data is investigated.

CHAPTER 5: Geo-ND SEGMENTATION COMPONENT

In this chapter an edge metric is introduced, evaluated and compared to an area metric for use in the empirical discrepancy methodology for segmentation algorithm free parameter tuning as implemented in Geo-ND.

5.1 SEGMENTATION ALGORITHM SELECTION AND METRIC CHARACTERISTIC CONSIDERATIONS

Although no segmentation algorithm is generally superior, some authors (Meinel & Neubert 2004; Neubert, Herold & Meinel 2006; 2008) propose a combined qualitative and quantitative evaluation approach for segmentation algorithms applied to earth observation data. In the comparative studies of Meinel & Neubert (2004), Neubert, Herold & Meinel (2006; 2008) and Hay et al. (2003), the region merging algorithm proposed by Baatz & Shäpe (2000) and the baseline method proposed by Castilla (2003) display promising and consistent results. A qualitative subjective visual investigation of these two algorithms (using different parameter sets) was conducted to assess their ability to segment the features of interest in this work (see Figures 2.3 and 2.5). The region merging algorithm proposed by Baatz & Shäpe was selected for the implementation in Geo-ND due to the inability of the baseline method to detect weak edges, which is a common occurrence in the model problem. Detail on the Baatz & Shäpe region merging algorithm was presented in Chapter 2.

The Baatz & Shäpe region merging algorithm has six parameters that need to be tuned (scale, colour/shape, compactness/smoothness and three image band weights) by the empirical discrepancy method implemented in Geo-ND. These parameters all have an influence on the characteristics of the generated segments. Due to the complex nature of the segment merging heuristic in this region merging algorithm any attempt at describing the relationship between this segmentation algorithm's parameters and the generated segments for a given application is not a trivial task (Feitosa et al. 2006). The scale parameter might be easily selected through trial and error, but finding an adequate combination of scale, colour/shape and compactness/smoothness parameters for a given land cover feature in an objective manner is more difficult.

In Sections 2.3.2.1 and 2.3.2.2 overviews were given of common discrepancy methods for segmentation quality evaluation. Different metrics place different emphases on what is

regarded as good quality. The ability of a search heuristic to find a suitable parameter set for a given task depends on the ability of such discrepancy measures to correctly – as perceived by the end users of the segments – gauge the quality of the generated segments. A specific metric (e.g. edge, area, data clustering or a hybrid) may not be optimal for all types of features (Weidner 2008). The amount and quality of training information also has an influence on the performance of the search heuristic (Feitosa et al. 2006). Results also depend on the intrinsic ability of the search heuristic to effectively traverse the search surface created by the fitness function (see Section 2.5).

Area metrics, such as the LSB (Fredrich & Feitosa 2008), measure segment area correspondence. Feitosa et al. (2006) demonstrated the ability of genetic algorithms to explore the parameter space, using an area metric (reference bounded segments booster, see Chapter 2), on features with discrete boundaries. It is hypothesised that an edge metric, such as the pixel correspondence metric (PCM) (Prieto & Allen 2003), might be useful in situations where land cover features display less distinct boundaries resulting in subjective area delineations with an area metric. For a given area, an edge metric measures all generated segment borders, irrespective of the features of interest. When using an edge metric, a user is required to delineate the perceived optimal segments in a specific area, with some knowledge of the behaviour of the specific algorithm, as opposed to an area metric that requires only the accurate delineation of features of interest, an exercise not always objectively possible.

In the next section the PCM metric is introduced, followed by an evaluation of this metric as a suitable discrepancy measure in an evolutionary search heuristic to find adequate segmentation algorithm parameters for the Baatz & Shäpe region merging algorithm operating on earth observation imagery.

5.2 OVERVIEW OF THE PIXEL CORRESPONDENCE METRIC

The pixel correspondence metric, proposed by Prieto & Allen (2003), is a correspondence metric designed to evaluate the similarity of two edge images. Compared to other common edge metrics, the PCM incorporates a measure of pixel offset for edges that do not match precisely. The metric also allows for weighted or greyscale edge matching. In this work a binary version of this metric is implemented.

The idea behind the PCM is to match pixels from the ground truth image with pixels from the generated image. Pixels that are not matched from either the ground truth or generated images are counted as errors by the metric. If multiple pixels from an image match with one pixel from the other image, the pixel pairing that minimises the metric is selected. The PCM also takes into account the spatial offset of matched pixels.

The cost of a possible pixel match between a hypothetical ground truth and generated pixel is defined as:

$$C\big((i,j),(k,l)\big) = E(\max(|k-i|,|l-j|)) \quad (5.1)$$

where i and j are the row and column numbers of the ground truth image with (i,j) representing the corresponding pixel. The same holds for k and l. The value returned by $\max(|k-i|,|l-j|)$ is the chessboard distance between the two given pixels. E is a scaling function in the range [0...1]. In this work pixels within a five pixel chessboard distance of each other are considered for matching. E scales this value accordingly. A cost value of 0 denotes an exact match of a pixel pair while a cost value of 1 denotes a match with a chessboard distance separation of five between pixels.

The PCM metric is defined as:

$$PCM(g,r) = 100\left(1 - \frac{C(M_{opt}(g,r))}{|g \cup r|}\right) \quad (5.2)$$

where g denotes the ground truth image and r the reference (generated) image. The element $|g \cup r|$ denotes the number of pixels in both (numerical 'or') imagery that are not zero/blank. And $C(M_{opt}(g,r))$ denotes the cost of the optimal matching of all pixels in the images g and r. A PCM value of 100 implies no matching between the imagery and a value of 0 implies a perfect match.

Finding the optimal matching (M_{opt}) that will minimise the PCM constitutes a bipartite graph matching problem (Prieto & Allen 2003). The nodes of a bipartite graph denote pixels and vertices denote the cost of a possible match. In bipartite graph matching, a node may only be attached to one vertex. The aim of solving such a problem is to match all nodes using the least expensive vertices. In this work the Hungarian algorithm (Munkres 1957) (also called the Munkres algorithm) is used to search for the optimal matching of pixels. Bipartite graph matching is computationally expensive so that less precise but faster matching strategies

(approximation algorithms) may be needed for operational use of such a metric in an evolutionary search procedure. Prieto & Allen (2003) present an approximated graph matching approach which uses the Gabow & Tarjan bipartite graph matching algorithm with little loss in metric accuracy. This methodology is not implemented in this work as the focus is on technique evaluation and not on improving operational capability.

5.3 METRIC SUITABILITY TEST

The PCM metric has not been tested or used as part of a fitness function to converge to adequate segmentation algorithm parameters. This metric is tested as a fitness function in an evolutionary search heuristic for its ability to effectively guide the search procedure to a ground truth in applicable problems.

5.3.1 Experimental design

The PCM is tested for its ability to converge to the true optimum in segmentation evaluation, using two prominent search heuristics. Such a test is conducted by providing the segmentation algorithm parameter search heuristic with a ground truth generated by the segmentation algorithm itself to ensure the ground truth can be accurately replicated by the algorithm (as opposed to a user delineated ground truth). Feitosa et al. (2006) applied this experimental methodology in evaluating an area metric in a genetic algorithm based search, although the details of the experiment were not reported.

Two search heuristics, namely self adapting DE (JDE) and PSO are employed in this experiment. Using these two different search heuristics allows them to be compared for this type of problem in addition to creating a more objective evaluation. Both heuristics were given 60 agents and 15 generations to explore the parameter space resulting in 900 segmentation and evaluation runs. Storn & Price (1995) suggest a $D \times 10$ population size for the DE heuristic, where D is the dimensionality of the problem (in this case a six parameter segmentation algorithm). The meta-parameters of JDE (Brest et al. 2006) and PSO (Vesterstroem & Thomsen 2004) were set to default values suggested in the literature for unseen problems.

Two regions are investigated, namely an area with agricultural/viticultural fields displaying distinct edges on most features, and a mountain slope with Cape thicket patches displaying

less distinct edges on arbitrary land cover elements of interest. Figure 5.1 illustrates the regions selected in this experiment.

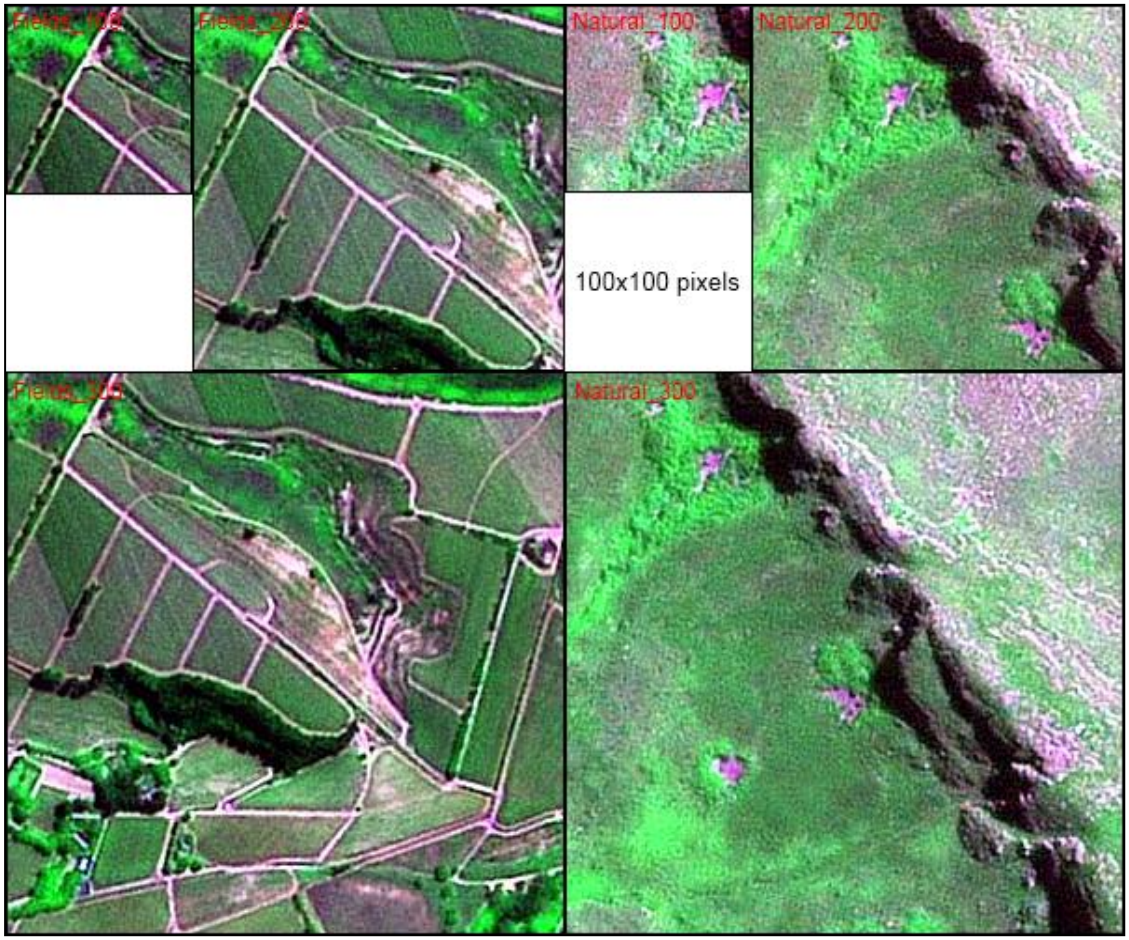


Figure 5.1: "Fields" and "natural" areas with differently sized subsets used in the PCM suitability test

Three different image subsets (with 300 x 300, 200 x 200 and 100 x 100 pixel dimensions) of these areas are used to investigate the effect of ground truth sample size on the performance of the search heuristic. The subsets are named fields_100, fields_200, fields_300, natural_100, natural_200 and natural_300 corresponding to the physical area under observation and the size of the image. The suffix DE or PSO is added to denote the type of search heuristic used. A lossy image splitting function is used to reduce the computational load of the bipartite graph matching on the larger imagery. Conducting graph matching with a theoretical upper limit of 90 000 (300 x 300 pixel image) vertices at each iteration of the search algorithm are computationally expensive and would probably not be conducted in practise. Agents were reduced to thirty for the 300 x 300 imagery to reduce computational loads. The fitness trace of the fittest agent is recorded as the search proceeds in each

experiment. Each experiment is conducted five times to investigate the robustness of the search heuristics.

To conduct this experiment, Geo-ND was modified to use algorithm generated segments instead of user delineated features as ground truth in the search procedure implemented in pane three of the application. The segmentation algorithm ground truth parameters chosen arbitrarily and used are: scale: 30, colour/shape: 0.5, smoothness/compactness: 0.5 and original SPOT 5 image bands 1 to 3 with a weight of 0.7 for all three bands. The segments generated by this parameter set are taken as the optima for a hypothetical problem, irrespective of land cover elements present in the images. The search bounds for the scale parameter were set to [20...50] and for the other five parameters to [0.2...0.9]. In total, 60 parameter search experiments were executed (10 experiments for each scene, using two different search algorithms).

5.3.2 Results

The fitness values obtained in the 60 experiments are summarised in Table 5.1.

Table 5.1: Mean fitness values and segmentation algorithm parameters obtained running the segmentation algorithm search heuristic over the six experimental images.

Experiment	Fitness	Scale	Colour	Compactness	Band 1 weight	Band 2 weight	Band 3 weight
Fields_100_DE	14.85 ±1.96	25.64 ±2.51	0.57 ±0.15	0.51 ±0.14	0.62 ±0.24	0.48 ±0.21	0.56 ±0.16
Fields_100_PSO	16.18 ±1.39	21.17 ±1.73	0.38 ±0.23	0.35 ±0.12	0.57 ±0.22	0.63 ±0.26	0.70 ±0.25
Fields_200_DE	**15.12 ±5.90**	**29.97 ±2.37**	**0.56 ±0.07**	**0.50 ±0.08**	**0.77 ±0.15**	**0.71 ±0.20**	**0.76 ±0.17**
Fields_200_PSO	17.67 ±5.19	28.18 ±3.61	0.58 ±0.25	0.49 ±0.17	0.73 ±0.21	0.67 ±0.29	0.74 ±0.20
Fields_300_DE	17.11 ±6.63	32.24 ±2.42	0.68 ±0.16	0.65 ±0.07	0.70 ±0.12	0.59 ±0.20	0.65 ±0.14
Fields_300_PSO	19.55 ±1.99	27.74 ±4.21	0.53 ±0.25	0.59 ±0.14	0.55 ±0.30	0.63 ±0.29	0.59 ±0.28
Natural_100_DE	29.62 ±2.34	29.38 ±1.78	0.56 ±0.16	0.50 ±0.06	0.73 ±0.13	0.67 ±0.21	0.70 ±0.11
Natural_100_PSO	30.54 ±4.53	24.47 ±5.36	0.38 ±0.23	0.48 ±0.10	0.44 ±0.15	0.73 ±0.15	0.54 ±0.21
Natural_200_DE	30.70 ±1.91	29.06 ±2.20	0.62 ±0.19	0.58 ±0.07	0.62 ±0.13	0.61 ±0.15	0.55 ±0.03
Natural_200_PSO	36.46 ±3.43	26.16 ±3.42	0.58 ±0.22	0.55 ±0.13	0.65 ±0.27	0.63 ±0.18	0.65 ±0.23
Natural_300_DE	44.59 ±3.03	28.24 ±6.28	0.49 ±0.19	0.56 ±0.08	0.52 ±0.13	0.61 ±0.12	0.47 ±0.09
Natural_300_PSO	41.92 ±1.42	26.62 ±3.76	0.68 ±0.20	0.59 ±0.17	0.54 ±0.31	0.73 ±0.23	0.60 ±0.27
Ground truth	0.0	30	0.5	0.5	0.7	0.7	0.7

Note: The suffix value, preceded by a '±' denotes the standard deviation of the mean values obtained.

The segmentation algorithm parameters are also given in Table 5.1. The values are the average of five experiments, with a suffix value denoted by '±' indicating the standard deviation over the five experiments. The ground truth values of the segmentation algorithm parameters are also listed as references. The row with bold values indicates the average parameter sets obtained with the closest match to the ground truth. A fitness value of 0 corresponds to the ground truth.

Figure 5.2 illustrates the average fitness traces of the fittest agent at each iteration of all the experiments conducted on the "fields" images. The dotted line in the graph denotes the fitness trace corresponding to the lowest fitness value obtained over all 60 experiments. Figure 5.3 shows the corresponding fitness traces for the six "natural" image segmentation experiments.

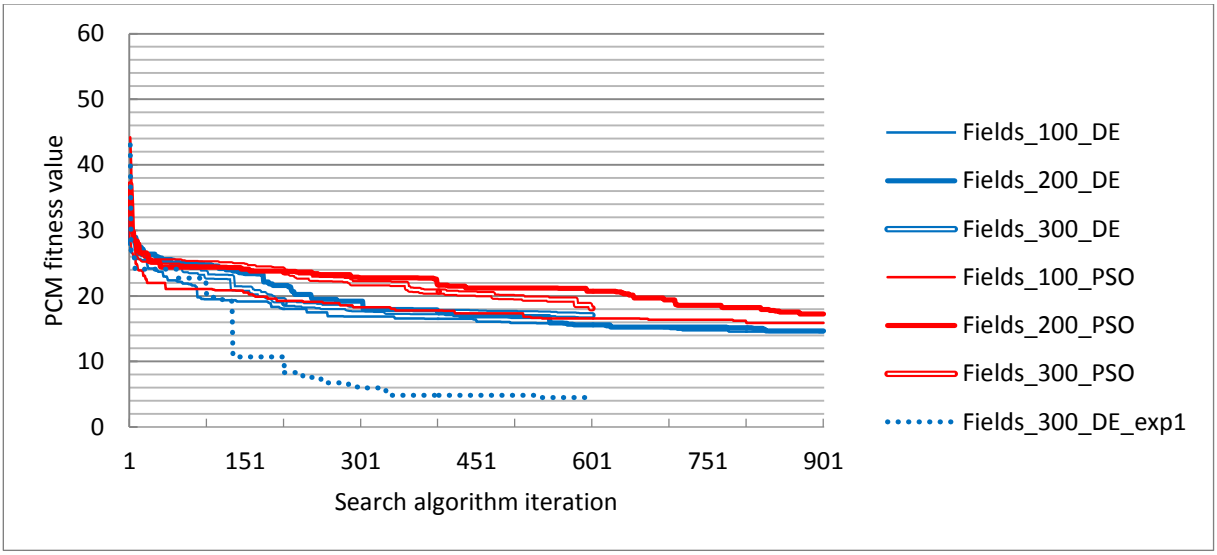


Figure 5.2: Average fitness traces of the fittest agents of PSO and DE using the three "fields" experimental images

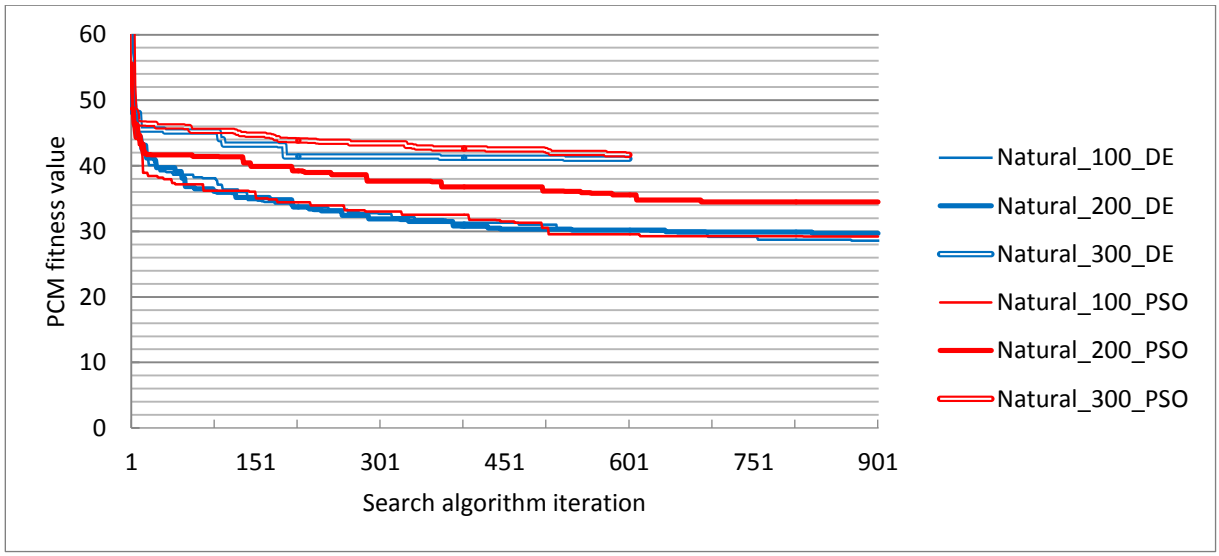


Figure 5.3: Average fitness traces of the fittest agents of PSO and DE using the three "natural" experimental images

As an example of the segmentation results of the experiments, Figure 5.4 illustrates the "fields_300" image layered with the ground truth segments and segments generated with a parameter set that obtained a fitness of 16.2 with a scale parameter of 33.6 with DE in the search heuristic. The ground truth is represented by the green segments. The segmentation results are represented by the red segments. The image on the left is layered with the ground truth on top (green segments) while the image on the right is layered with the generated segments on top (red segments). Similarly, Figure 5.5 illustrates the "natural_300" image layered with the ground truth (green lines) and segments generated with a parameter set that obtained a fitness of 43.1 with a scale parameter of 29.8 with the PSO heuristic. The blue areas in Figure 5.5 denote Cape thicket found within both the ground truth and generated segments. The yellow areas denote Cape thicket found in either the ground truth or generated segments.

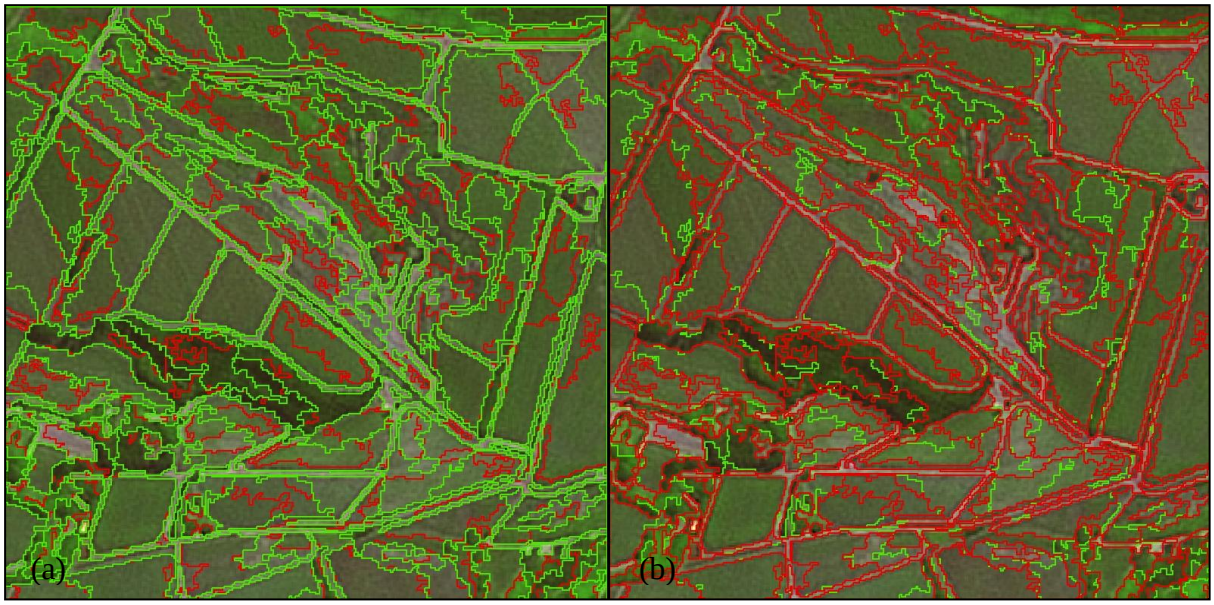


Figure 5.4: The "fields_300" image segmented with a parameter set that obtained a fitness of 16 with the PCM corresponding to the ground truth parameters. A) Illustrates the ground truth segments layered on top of the generated segments while b) shows the generated segments layered on top of the ground truth.

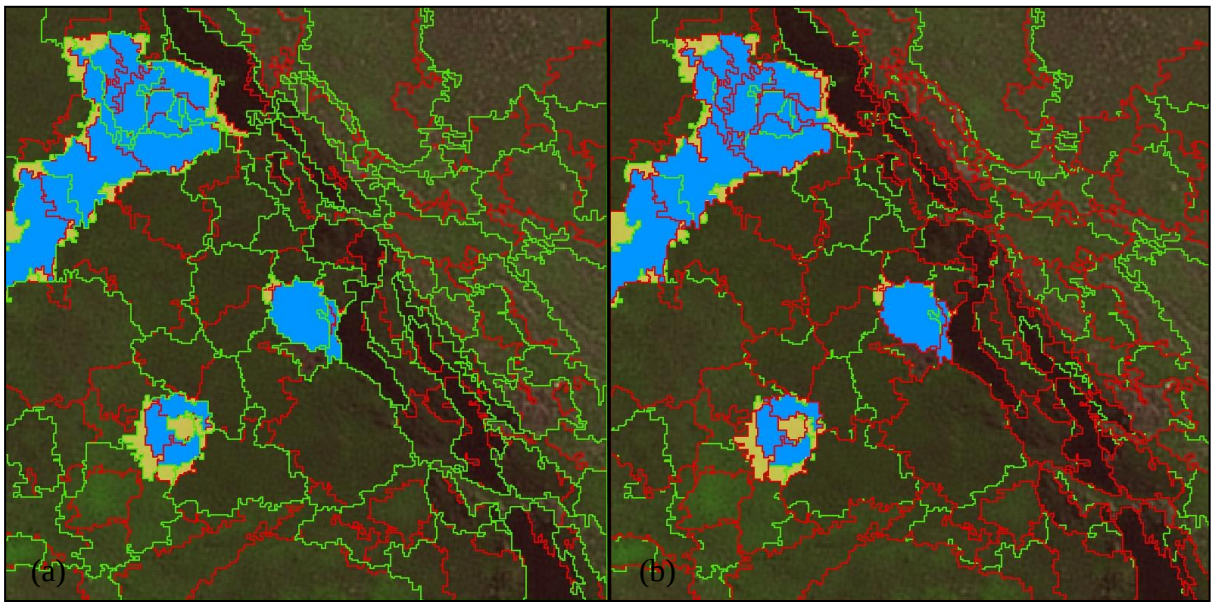


Figure 5.5: The "natural_300" image segmented with a parameter set that obtained a fitness of 43.1 with the PCM corresponding to the ground truth parameters. A) Illustrates the ground truth segments layered on top of the generated segments while b) shows the generated segments layered on top of the ground truth.

5.3.3 Discussion

The two images subjected to the experiments display very different land cover element characteristics. The "fields" image has more distinct edges while the only detectable edge in the "natural" image were that of the cliff face shadows and to a lesser extent the boundaries of

the Cape thicket patches. As observed in Table 5.1 the "fields" imagery experiments obtained substantially lower (better) fitness values over all three image sizes compared to the "natural" images. As shown in Figure 5.4, the search heuristics were able to converge to a parameter set closely resembling the ground truth in the "fields" images. The PCM with the given search heuristics was not able to replicate segment lines cutting through fields. Trial and error experimentation with the "fields" images revealed that extremely small changes in parameter values (fourth order decimals) resulted in vastly different segments in land cover elements with weak or no edges.

The phenomenon of the search heuristic being unable to match edges in areas with no or weak edges is prominent in the "natural" images, resulting in lower PCM scores. The Baatz & Shape region merging algorithm can generate a vastly different segment set with relatively similar parameter sets in images with weakly defined edges (see Figure 5.5). Despite this disparity between ground truth and generated segments, the few land cover features with identifiable edges (thicket and shadows) allowed the search algorithm (using the PCM) to converge to a scale parameter closely resembling the ground truth for these two types of features. Subjectively, the average achieved fitness value of 15-20 for the "fields" images and a fitness of 40 for the "natural" images correspond to good parameter sets which accurately delineate all features with identifiable boundaries. In figure 5.5 the blue areas denote Cape thicket found within both the ground truth and generated segments. The yellow areas denote Cape thicket found in either the ground truth or generated segments.

The PCM was able to converge closely to the ground truth scale parameter (30) in most of the test images except for fields_100 and to some extent natural_100. This can be attributed to a small training area, with a larger range of different parameter sets able to replicate the ground truth. For example, the fields_100_pso image obtained a fitness of 16 and scale parameter of 21 over five consecutive experiments with very little standard deviation.

Note that although the critical scale parameter was consistently approximated, the remaining five parameters displayed less consistency in approximated values. The fields_200_DE image attained a parameter set closely resembling the ground truth (Table 5.1). All the other "fields" images attained similarly low fitness values and a scale parameter close to 30, but displayed greater variation in values of the other five parameters. This could be attributed to the weak edge phenomenon, where an extremely small numerical margin is needed to

accurately replicate the ground truth. In most cases, both the DE and PSO heuristics were unable to converge to the absolute ground truth in the allocated 900 algorithm iterations, generating parameters (excluding the scale parameter) with relatively high standard deviations.

From Table 5.1 and Figures 5.2 and 5.3 it is deducible that the DE search heuristic slightly outperforms the PSO search heuristic concerning convergence, accuracy and convergence speed. Over all six experimental images, the DE search heuristic converged to a scale parameter closer to the ground truth than the scale parameters generated by the PSO heuristic. The DE heuristic attained a lower mean fitness value compared to the PSO heuristic in all but one experiment (fields_300). On three occasions the DE heuristic achieved a considerably lower fitness value (as indicated by a high standard deviation in Table 5.1). Through the course of the experiments, the DE heuristic attained a fitness of seven and eight in the fields_200 images and a fitness of four in the fields_300 image. The fitness trace with a resulting fitness value of four is illustrated in Figure 5.2 (dotted line). The lowest fitness score achieved by the PSO heuristic was ten. Figures 5.2 and 5.3 suggest that for this domain specific problem, despite both algorithms performing adequately, the DE search heuristic is preferred to the PSO search heuristic, thus confirming the general remarks found in the literature as discussed in Section 2.5.5. Note that by performing meta-parameter tuning (to both the DE and PSO search heuristics) improved convergence speed and accuracy can be achieved.

As the size of the images increases, so do the average fitness values, a direct relationship attributable to more weak edges cutting fields or fynbos that could not be replicated. Prieto & Allen (2003) suggest this type of phenomenon relates to the robustness of a metric and note some other edge metrics used in their studies not displaying this behaviour. Bear in mind that the remarks by these authors are not related to convergence experiments but that they are qualitative observations suggesting that the PCM is a more accurate edge metric than other edge metrics.

Interestingly, Melo et al. (2008) observed when using the LSB metric and the Baatz & Shape region merging segmentation algorithm that near optimal fitness values are normally achieved after 400 iterations (10 agents) of the search heuristic. A similar number of iterations is necessary to achieve a near optimal fitness when using the PCM (30-60 agents), as illustrated

in Figures 5.2 and 5.3. This suggests that 400 iterations is a safe lower minimum default value in the practical use of such a system.

It is concluded that the PCM with the DE and PSO search heuristics were able to effectively traverse the search surfaces and converge to adequate parameters if enough training samples (larger than 100 x 100 pixel images) are available. The PCM with DE and PSO were not able to replicate weak edges, something which might not be a problem in the practical use of such an approach to segmentation algorithm parameter tuning.

5.4 COMPARISON OF THE LSB AND PCM METRICS FOR SEGMENTING HETEROGENEOUS, VARIABLY SIZED LAND COVER ELEMENTS

The PCM and LSB metrics are compared qualitatively and quantitatively for the particular problem of generating segments over non-homogeneous features with weak edges displaying some variation in feature size. The LSB has been shown to be an effective metric for similarly sized features with relatively distinct boundaries (Feitosa et al. 2009; Fredrich & Feitosa 2008). The LSB can be used for such features in Geo-ND. As discussed in Section 2.2.2.2 the relative scale of segments in a single scale representation framework compared to object based image analysis has a substantial influence on the resulting classification accuracy. This section explores this notion.

5.4.1 Experimental design

In this experiment the range of segmentation algorithm parameters generated by the PCM and LSB metrics for an arbitrary anomaly detection problem on a SPOT 5 scene subset is investigated. A metric used as a fitness function can be evaluated on the basis of generated segments (Feitosa et al. 2006) as well as indirectly via classifier accuracy using the metric generated segments. The LSB and PCM segments are evaluated for segmenting the features according to the desired output. The segments generated by the LSB and PCM metrics are subjected to a ground truth supervised classification.

The experimental procedure is illustrated by Figure 5.6. Geo-ND is used in this experiment. Ten Cape thicket patches of varying geometric characteristics (region of interest sizes varying from 50 x 50 to 150 x 150 pixels) are identified on a SPOT 5 scene subset depicting a small region directly east of Stellenbosch (3171 x 2265 pixel dimensions). These ten features are

digitised using both areas and lines (for use with the area and edge metrics). Figure 5.7 illustrates five of the anomalous features delineated with edges for the PCM and areas for the LSB. The other five features are more dispersed geographically. Ten features are considered sufficient for both metrics to converge to adequate parameter sets.

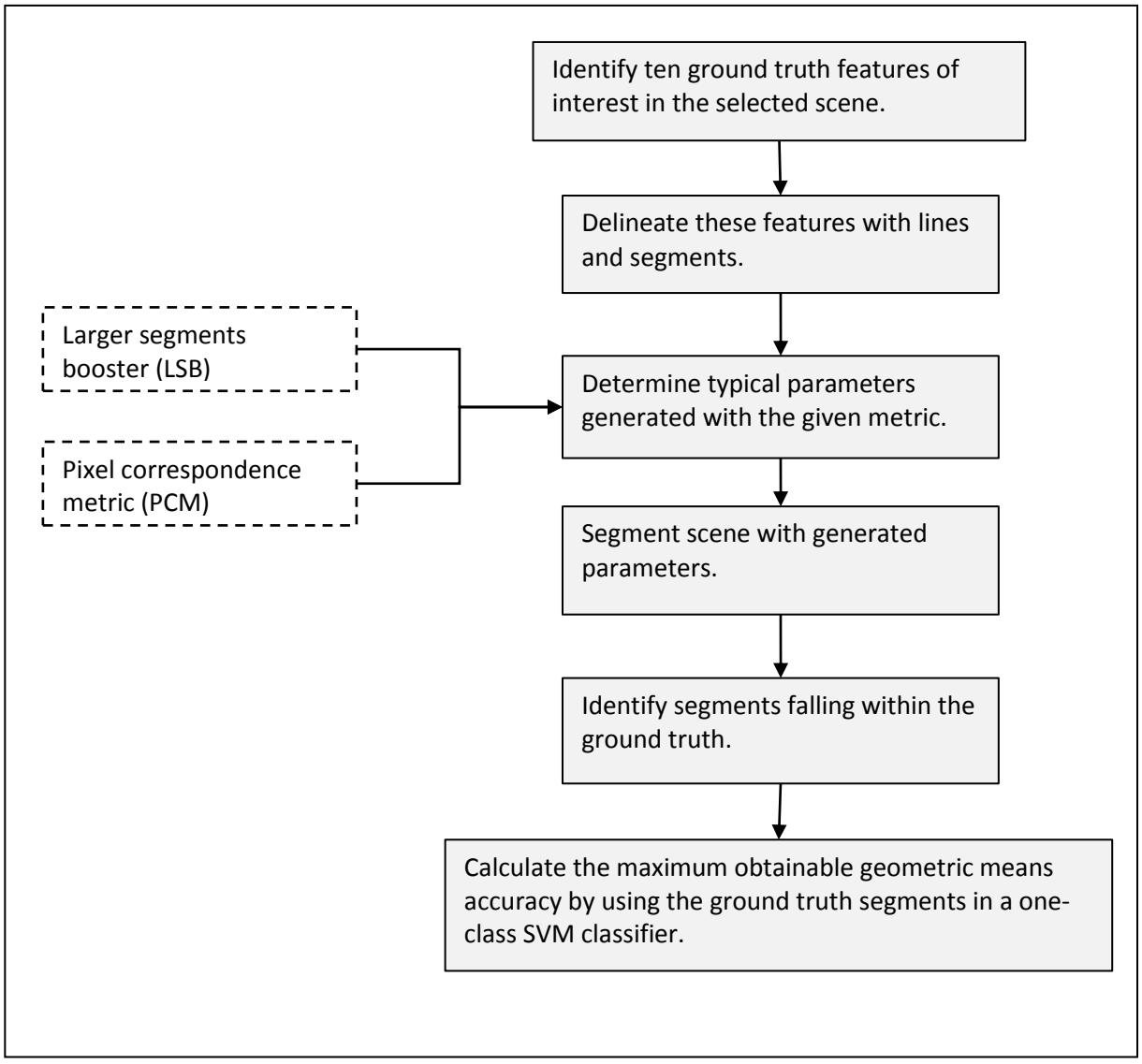


Figure 5.6: Methodology for comparing the PCM and LSB metrics used for automatic segmentation algorithm parameter tuning - in terms of resultant classifier accuracy

The self adapting differential evolution search heuristic is used to search for parameter sets using both the LSB and PCM metrics given the ten digitised features. Thirty agents are involved and given twenty generations to evolve, resulting in 6000 segmentation and evaluation runs. For each agent at each generation the ten user delineated areas are segmented and evaluated. The search range for the scale parameter is set to [5...80]. The

search range for all other parameters is set to [0.2...0.9]. The band selection tool in Geo-ND suggested that the NDVI, intensity and B3/B4 bands as best suited among the available bands for this arbitrary problem. These three bands displayed the highest spectral separability over an eight pixel buffer distance, as automatically suggested by Geo-ND. The search process using both metrics is repeated 25 times to obtain an indication of the specificity of the generated parameter sets. In addition, for every run the results of the three fittest agents are stored. The entire scene subset is segmented with the parameter sets generated by using these two metrics. The parameter sets generated by the two different metrics are compared regarding specificity and perceived quality compared to the metric input.

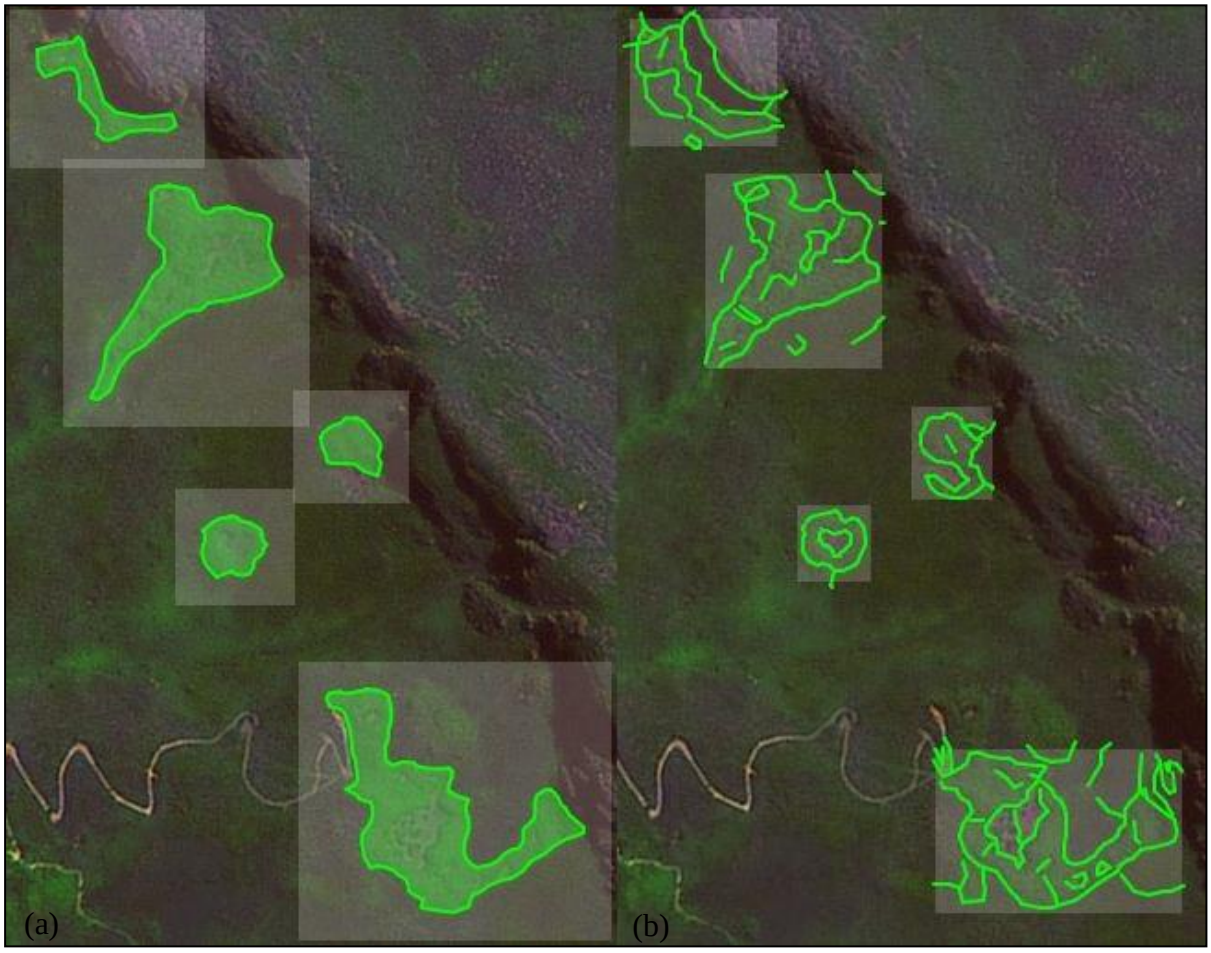


Figure 5.7: Five of the digitised features used in the comparison of an area based (a) and edge based (b) metric for automatic tuning of the segmentation algorithm parameters

All of the Cape thicket and riparian thicket in the scene subset, including the ten Cape thicket patches used in segmentation algorithm parameter tuning, were identified and designated as ground truth. All segments from both the PCM and LSB segmented scenes falling within this

ground truth (50% overlap needed) were tagged as anomalous features in Geo-ND's fifth pane (classification pane). Segments falling within urban areas were masked and not used.

The automatic classifier parameter tuning and attribute subset selection tool (discussed in Chapter 6), operating with both fixed and variable attribute sets, is used to obtain an indication of maximum achievable classifier accuracy (geometric means, discussed in Section 2.6.2.3), using the PCM and LSB generated parameter sets/segments. A particle swarm search heuristic using ten-fold cross-validation, thirty agents and 27 generations is employed. The search range for gamma (RBF kernel) was set to [0.05...50] and for nu it was set to [0.01...0.99]. Seven segment attributes judged to hold discriminative value for this arbitrary problem were selected for use in this experiment. The classification experiment was conducted 25 times. The PCM and LSB metrics are compared indirectly according to the classification performance of their resulting segment sets.

5.4.2 Segmentation and classification results

Table 5.2 details the parameters obtained by the search algorithm using the two metrics over the ten delineated features. The values are the means of the 75 runs (25 runs with the data from the three fittest agents collected) and the standard deviation is indicated by a '±'.

Table 5.2: Generated segmentation algorithm parameters using the LSB and PCM metrics.

	Scale	Colour	Compactness	NDVI	Intensity	B3/B4
LSB	13.95 ±2.24	0.77 ±0.15	0.22 ±0.04	0.62 ±0.21	0.54 ±0.24	0.30 ±0.16
PCM	39.27 ±13.99	0.63 ±0.23	0.37 ±0.17	0.51 ±0.21	0.65 ±0.21	0.47 ±0.21

Note: Each experimental run was performed 25 times. In each experiment the data of the three fittest agents were stored. The values are the means of the 75 runs with the standard deviation denoted by '±'.

Figure 5.8 shows segmentation results of using the mean parameter values generated by the two different metrics. Figure 5.9 illustrates the majority of the scene subset in the fifth pane in Geo-ND as used in this experiment. The yellow segments are those identified as Cape thicket and riparian thicket and are used as ground truth. The segments on display are those generated by the PCM. An identical ground truth selection exercise was conducted for the LSB generated segments. Urban and agricultural land were identified and masked for this experiment (gray areas).

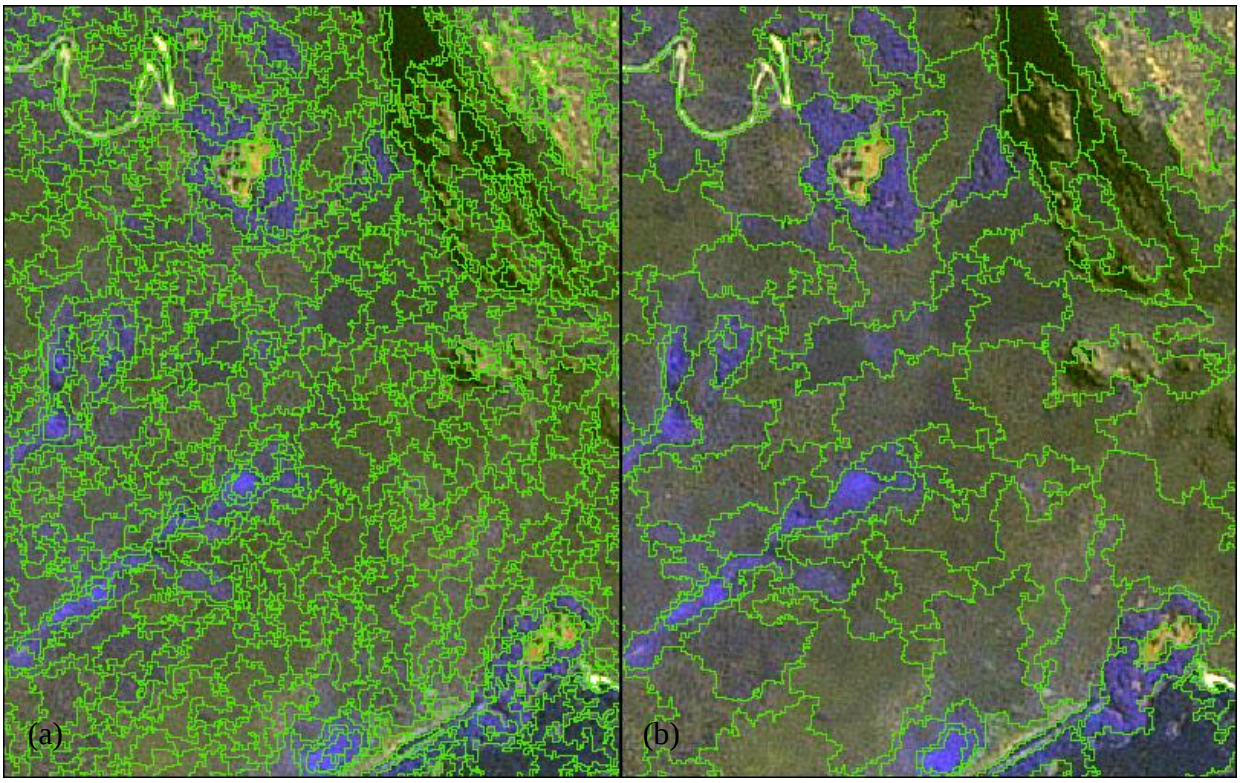


Figure 5.8: A SPOT 5 scene subset segmented with the parameters generated by using the LSB metric (a) and the PCM (b)

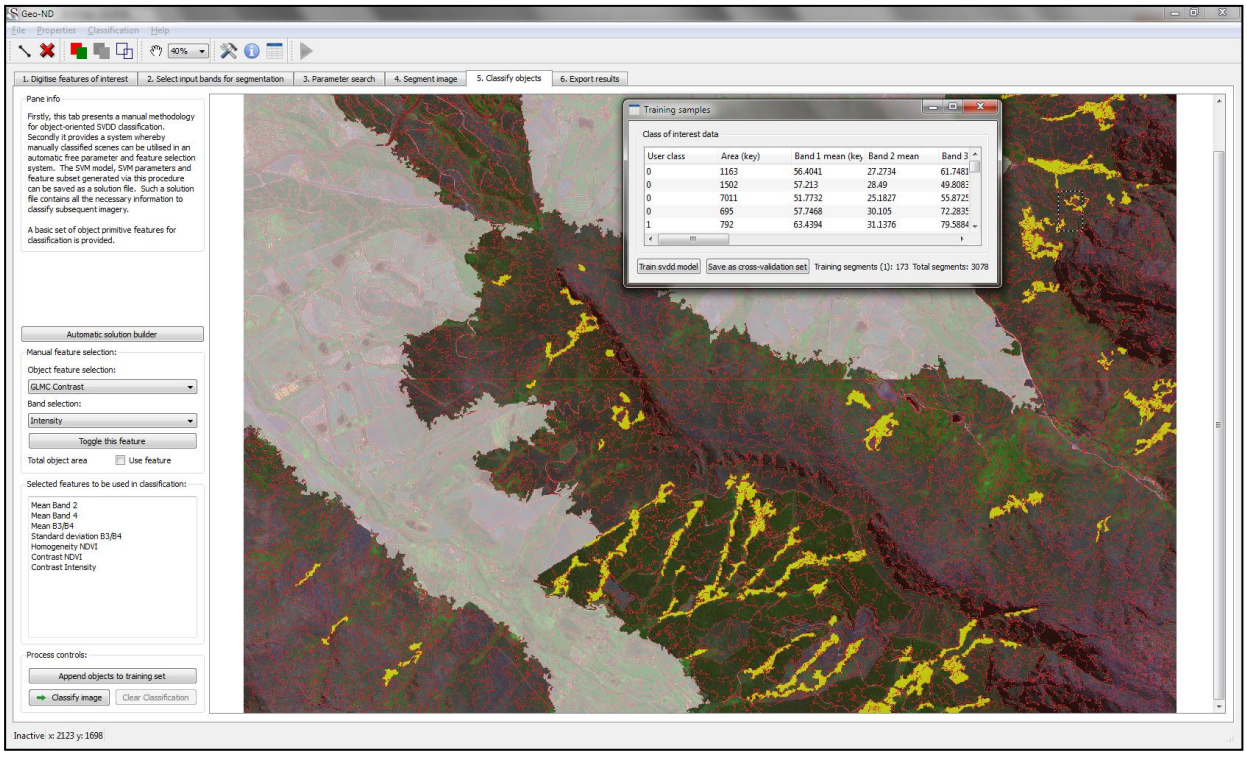


Figure 5.9: The fifth pane in Geo-ND highlighting the Cape and riparian thicket patches

Table 5.3 summarises the number of identified Cape and riparian thicket ground truth segments in the LSB and PCM generated segmented scenes.

Table 5.3: Number of ground truth training segments identified as Cape and riparian thicket in the PCM and LSB generated segmented scenes.

	Training segments	Total segments	Training segment percentage
LSB	978	19518	5.01%
PCM	173	3078	5.62%

Table 5.4 lists the achieved accuracies of classifying Cape and riparian thicket using the LSB and PCM generated segments.

Table 5.4: Classification results of Cape thicket using both PCM and LSB generated segments with variable and fixed attribute sets

	Gamma	Nu	Accuracy	M B2	M B4	M B3/B4	S B3/B4	H NDVI	C NDVI	C Int	# Attr
LSB fixed	5.08 ±3.78	0.17 ±0.02	78.99 ±0.0027	25	25	25	25	25	25	25	7
LSB variable	15.88 ±17.01	0.17 ±0.03	79.58 ±0.0050	23	21	25	1	19	16	12	4.68 ±0.80
PCM fixed	3.29 ±1.59	0.27 ±0.11	80.34 ±0.0042	25	25	25	25	25	25	25	7
PCM variable	5.11 ±10.71	0.31 ±0.08	**83.06 ±0.0053**	22	13	0	3	24	19	11	3.68 ±1.07

Note: M = mean; S = standard deviation; H = GLCM homogeneity; C = GLCM contrast.

The mean segment value of band 2, band 4 and band3/band4, the standard deviation of band3/band4, the GLCM homogeneity of NDVI and the GLCM contrast of NDVI and intensity attributes were presented to the classifier. The values indicated are the mean of 25 runs with the standard deviation signified by '±'. The full results for this experiment are recorded in Appendix B. Take note that the high standard deviation obtained for some gamma values are due to the presence of one or two outliers. The attribute columns denote the number of times the specific attribute was selected by the classifier over the 25 runs. The last column on the right in Table 5.4 lists the average number of attributes selected by the classifier when operating with variable attribute sets. The differences between "LSB fixed" and "PCM fixed" and between "LSB variable" and "PCM variable" are statistically significant (paired t-test with <0.0005 confidence). The boldface value indicates the metric/segments and attribute selection strategy delivering the highest classification accuracy.

5.4.3 Discussion

The range and specificity of the parameters generated by the two metrics vary considerably (see Table 5.2). The LSB consistently generated a scale parameter around 14, while the PCM generated a scale parameter around 40 with a substantial standard deviation. Both metrics

favoured a high colour weight opposed to shape and a high smoothness weight as opposed to compactness (see Section 2.3.1.1 for details on the function of these parameters). No apparent trend in preference for the band weights was observed, possibly due to all three bands having considerable significance for the problem (as suggested by the band selection tool implemented in Geo-ND).

The LSB penalises over segmentation by counting intersecting segment lines within the features of interest, in addition to applying a notion of overlapping area (see Section 2.3.3.2 for more detail on the functioning of the LSB). With this specific land cover feature, a relatively small scale parameter was found to be optimal with a very high consistency rate. In this experiment a standard deviation of less than three scale parameter units was obtained by observing the three fittest agents in each run, suggesting a very specific search surface (see Figure 5.8). It can be assumed that a larger scale parameter resulted in areas in some of the ten delineated features with a very low LSB value, possibly due to Cape thicket sub-areas being agglomerated with fynbos.

As illustrated in Figure 5.8, the PCM generated segmentation algorithm parameters with a substantially larger scale parameter, but with a higher standard deviation, although calculated over the three fittest agents per experiment. This suggests that, in the ten regions of interest the search surface generated by the PCM was not as specific as that of the LSB. A variety of parameter sets were able to approximate the user delineated ground truth. Note that although the numeric values were not as specific as those generated by the LSB metric, different combinations of parameter values may result in very similar segments. This phenomenon is peculiar to the Baatz & Shäpe region merging segmentation algorithm (see for example Van Coillie et al. (2010)). As reported in Section 5.3, the specificity of the PCM increases as the size of the training set increases. The lack of strong edges in the regions of interest also has an influence on the performance of the PCM (Section 5.3). Although not as specific as the LSB in generated parameters, the PCM delivered segments closely resembling user delineated input (Figure 5.8).

Recalling Table 5.3, the lower percentage of LSB generated segments as ground truth (compared to the total segments present) can be attributed to non-Cape and riparian thicket land cover elements displaying heterogeneous spectral and textural properties at a finer scale of observation, resulting in a slightly higher density of segments over these features when

using a relatively small scale parameter compared to that of a scene segmented with a larger scale parameter. It is assumed that this difference in percentage of training samples (0.61%) does not influence the classification as segments of both PCM and LSB sets were selected based according to membership of the ground truth.

The PCM generated segments resulted in statistically significant higher classification accuracies than the LSB generated segments (see Table 5.4 and Appendix B) with a very low standard deviation. A difference of three per cent in geometric means accuracy separates the classifiers using the PCM and LSB segments with variable attribute subsets. Use of variable attribute subsets also resulted in higher classifier accuracies compared to using all seven given attributes. This observation is investigated more formally in Chapter 6.

The PCM segments classification also used substantially fewer attributes than the LSB segments classication when given the opportunity to discard attributes. This suggests that the fewer discriminative values (due to smaller segments) the more attributes are needed by the LSB segments classification to obtain the maximum classifier accuracy (and potentially suffer due to the Hughes phenomenon). Interestingly, the mean of band three divided by band four was found to be important when using the smaller segments, but was discarded by the classifier using the larger PCM generated segments.

Examination of the gamma (tightness of the domain description) and nu (upper bound on outliers and lower bound on support vectors) parameters given in Table 5.4 lead to the conclusion that larger segments result in a more concise classifier description with more support vectors and a looser gamma (the classifier description when using small segments are more complex). On observing experimental run seven in Table B.3 (Appendix B), a high nu and low gamma are suggested by the search heuristic for a PCM segment generated classification with fixed attributes. Run seven resulted in the lowest geometric means accuracy of the 25 runs, suggesting that for this specific run the initial particle distribution could not escape the local optima in this nu/gamma region.

In Sections 2.2.2.2 and 2.2.3 it was established that the scale of observation/segmentation has an influence on the resultant classifier accuracy. In this case study, segments that more closely resemble the scale of observation of the features, as generated by the PCM, resulted in a classification with higher accuracy, although increasing the scale of

segmentation/observation runs the risk of under segmentation (see Section 2.3.2). This experiment's results are in accord with those presented by Kim & Madden (2006) and Kim, Xu & Madden (2006) regarding the influence of the scale parameter found in the Baatz & Shape region merging algorithm on classifier performance (as segment size increases and approaches the size of the features, so does the classification accuracy). The small scale parameter suggested by the LSB metric resulted in safer segments (no risk of under segmentation), at the expense of lower classifier accuracy.

From these results it is clear that the LSB is not suited for automatic segment generation for segmenting heterogeneous land cover elements. It seems that metrics that focus on geometry and edge delineation with no notion of under or over segmentation, such as the PCM, are more suited for such features, notwithstanding the risk of under segmentation.

The PCM and LSB metrics are both implemented in Geo-ND and can be used interchangeably. The LSB metric is recommended for discrete features and when computation time is limited. The PCM is recommended for delineating features with weaker edges and where uncertainty exists about the exact extent of the features of interest.

5.5 SUMMARY

In this chapter an edge metric, entitled the pixel correspondence metric, was evaluated as a fitness function of segmentation algorithm free parameter tuning. By employing the PCM the common PSO and DE search heuristics were able to converge to ground truth parameters in typical earth observation imagery when adequate training data were available. The search functions display some difficulty in detecting segment lines on weak edges.

PCM was compared with an area based metric, namely the LSB, for segment algorithm parameter generation on heterogeneous features with non-discrete boundaries. The PCM delivered segmentation algorithm parameters that corresponded closest to the user input. Additionally, for an arbitrary problem, a subsequent classification with the PCM generated segments delivered superior classifier results compared to the LSB generated segments.

CHAPTER 6: Geo-ND CLASSIFICATION COMPONENT

This chapter gives details on the one class classification component embedded in Geo-ND. Experimental results are presented to evaluate various aspects and working conditions of the proposed one class classification process.

6.1 OVERVIEW OF THE CLASSIFICATION COMPONENT

In the proposed anomaly detection workflow, feature classification follows the automatic segment generation process (see Figure 4.1). A user is required to select some segments (in pane five in Geo-ND) for use in a manual classifier workflow or to identify all segments belonging to the class of interest in the currently segmented scene to be used in an automated parameter tuning and attribute subset selection (PTASS) system.

A user provides the PTASS system with one or more sets of training data consisting of segments identified as belonging to the class of interest. By using this data the system attempts to construct a one class SVM classifier with an optimal two dimensional free parameter set and segment attribute subset for the specific problem. The complex interdependent free parameter and attribute space is simultaneously explored with a population based search heuristic, namely a particle swarm optimiser (PSO).

The proposed system is based on a SVM optimised with a PSO strategy advanced by Lin et al. (2008), but functions with an anomaly detector instead of a two class SVM. Consequently, an appropriate fitness function is used in the PSO search heuristic that can cope with highly unbalanced data, that is the geometric means accuracy metric. This approach (a one or two class SVM optimised with a PSO) has not been investigated in the data specific domain of earth observation image processing (object based or pixel based classification).

Figure 6.1 portrays the architecture of the implemented PTASS system. The system accepts one or more training sets (*.tsf files in Geo-ND), agglomerated as a master set, as input. The PSO is provided with the user selected meta-parameters and the value of k, which denote the number of cross-validation partitioning sets to be generated.

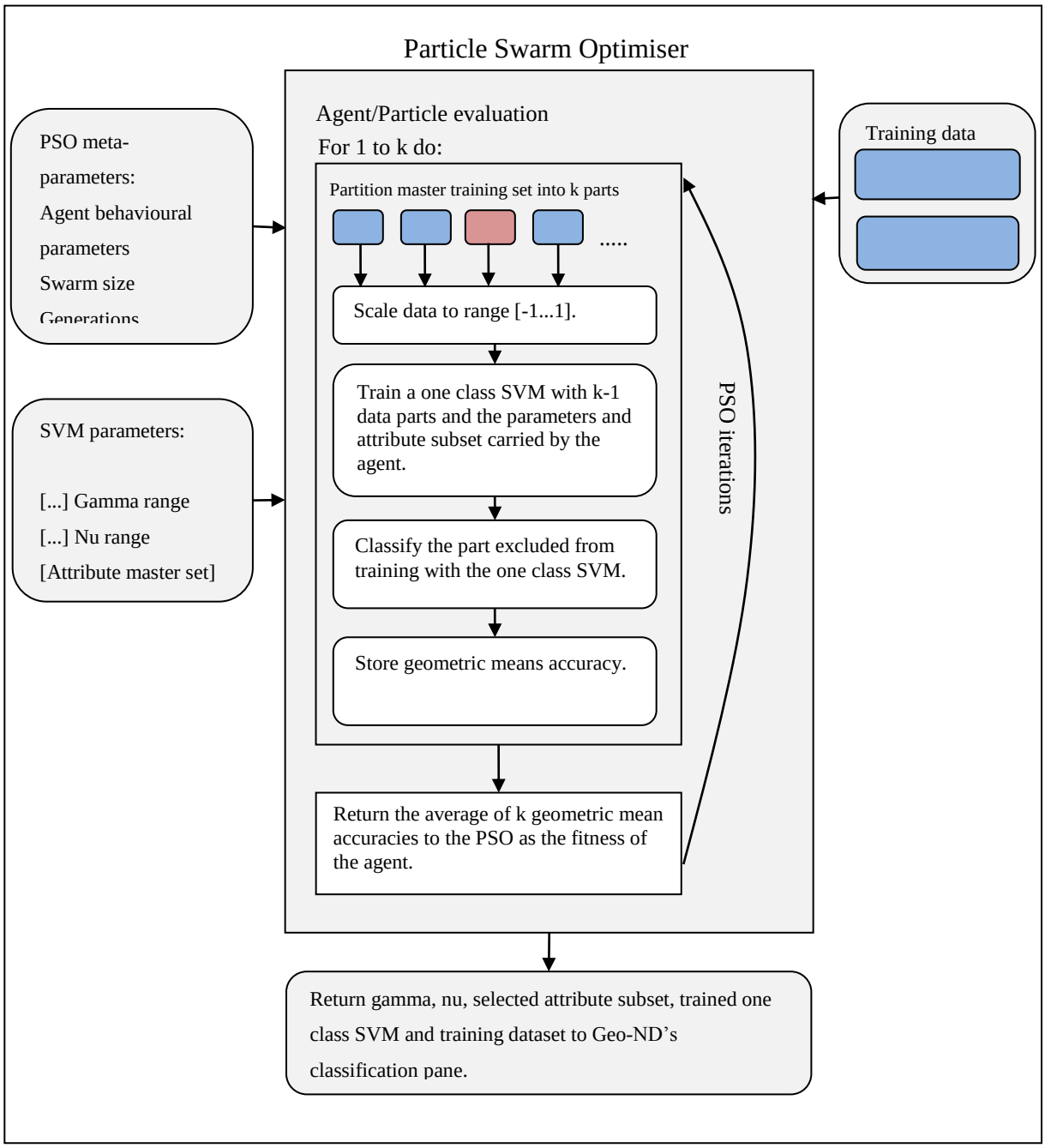


Figure 6.1: The architecture of the automatic free parameter tuning and attribute subset selection system implemented in Geo-ND

The k-fold cross-validation method for accuracy assessment is commonly employed in estimating the performance of a classifier on unseen data (Huang & Wang 2006; Sanchez-Hernandez, Boyd & Foody 2007b; Tu et al. 2007; Vapnik 1995; Zhuang & Dai 2006) and is recommended for problems with small sample sizes as opposed to the hold-one-out method (Huang & Wang 2006; Tu et al. 2007). A 10-fold cross-validation strategy was found sufficient for typical object based classification problems. Increasing k results in unnecessary computational load. A suitable gamma (0.0001 to 32) and nu (0.01 to 0.99) parameter space

range is also provided to the system in addition to a selection of segment attributes (called the master attribute set). Each agent/particle in the PSO contains a gamma and nu value along with attributes values. An attribute is represented in the agent as a real value with bounds [0...1]. If an attribute has a value greater than or equal to 0.5, it is considered selected/carried by the agent. Figure 6.2 illustrates the encoding of the agents along with their numerical bounds.

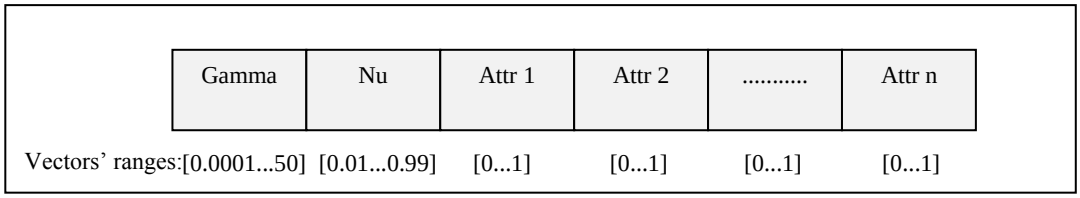


Figure 6.2: The agent encoding of the combined continuous and binary PSO as used in Geo-ND for simultaneous free parameter tuning and attribute subset selection

During the execution of the PSO, each agent is evaluated in the following manner (see Figure 6.1):

1. Partition the master training data set into k subsets each containing even proportions of samples from the class of interest as well as from the background class.
2. Scale the numerical range of the data to [-1...1] to avoid attributes in greater numeric ranges from overpowering attributes with less numeric range (Hsu, Chang & Lin 2008).
3. Train the one class SVM model with k – 1 subsets, using the gamma and nu values as well as the attribute subset carried by the agent as input (illustrated by the small blue boxes in Figure 6.1).
4. Classify the data subset not used for training with the one class SVM model (illustrated by the small red box in Figure 6.1).
5. Store the resulting geometric means accuracy.
6. Repeat this process until all k subsets have had an opportunity to act as the testing subset.
7. Return the average of the k geometric means accuracies as the fitness of the agent to the PSO.

Initial experimentation with inertia weights and social and personal cognition factor parameters (explained in Section 2.5.4) disagrees with values quoted in the literature (no

quantitative results presented in the literature) for similar implementations (Huang & Dun 2008; Lin et al. 2008). An inertia weight of 0.8 is specified, a personal cognition learning factor of 1.6 and social cognition factor of 1. Drastically increasing the number of agents holds little consequence in an SVM with PSO implementation (Huang & Dun 2008). The PSO algorithm terminates when a certain number of generations/algorithm iterations has passed. The classification pane in Geo-ND is updated with all relevant information generated by this process.

Figure 6.3 illustrates the interface for this tool as implemented in Geo-ND (accessible from pane five). As an example, five training sets were loaded (training set files) in which all dams were identified. As output, the system gave a one class SVM model with accompanying parameters and attribute subset resulting in a (near) optimal geometric means accuracy. This tool also allows a user to force the PSO to use all the attributes presented (by the "Force use of all input attributes" checkbox). See the accompanying userguide for detail on using this tool (Appendix C).

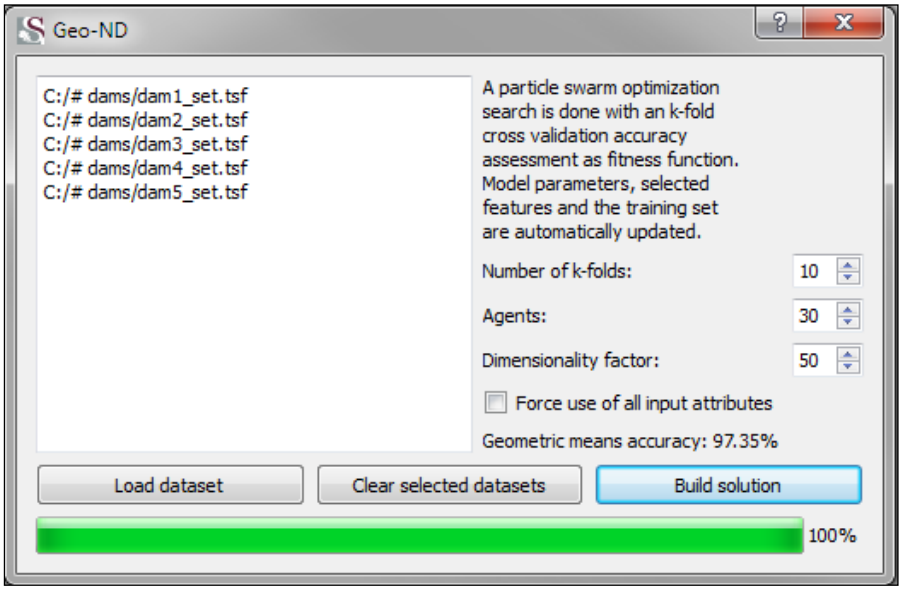


Figure 6.3: Interface for the automatic attribute subset selection and free parameter tuning tool accessible from Geo-ND's fifth pane

6.2 METHODOLOGICAL CONSIDERATIONS ON THE SIMULTANEOUS FREE PARAMETER TUNING AND ATTRIBUTE SUBSET SELECION SYSTEM

Huang & Wang (2006) and Lin et al. (2008) report on generally favourable results of using a SVM optimised with a genetic algorithm (GA) and a SVM optimised with particle swarm

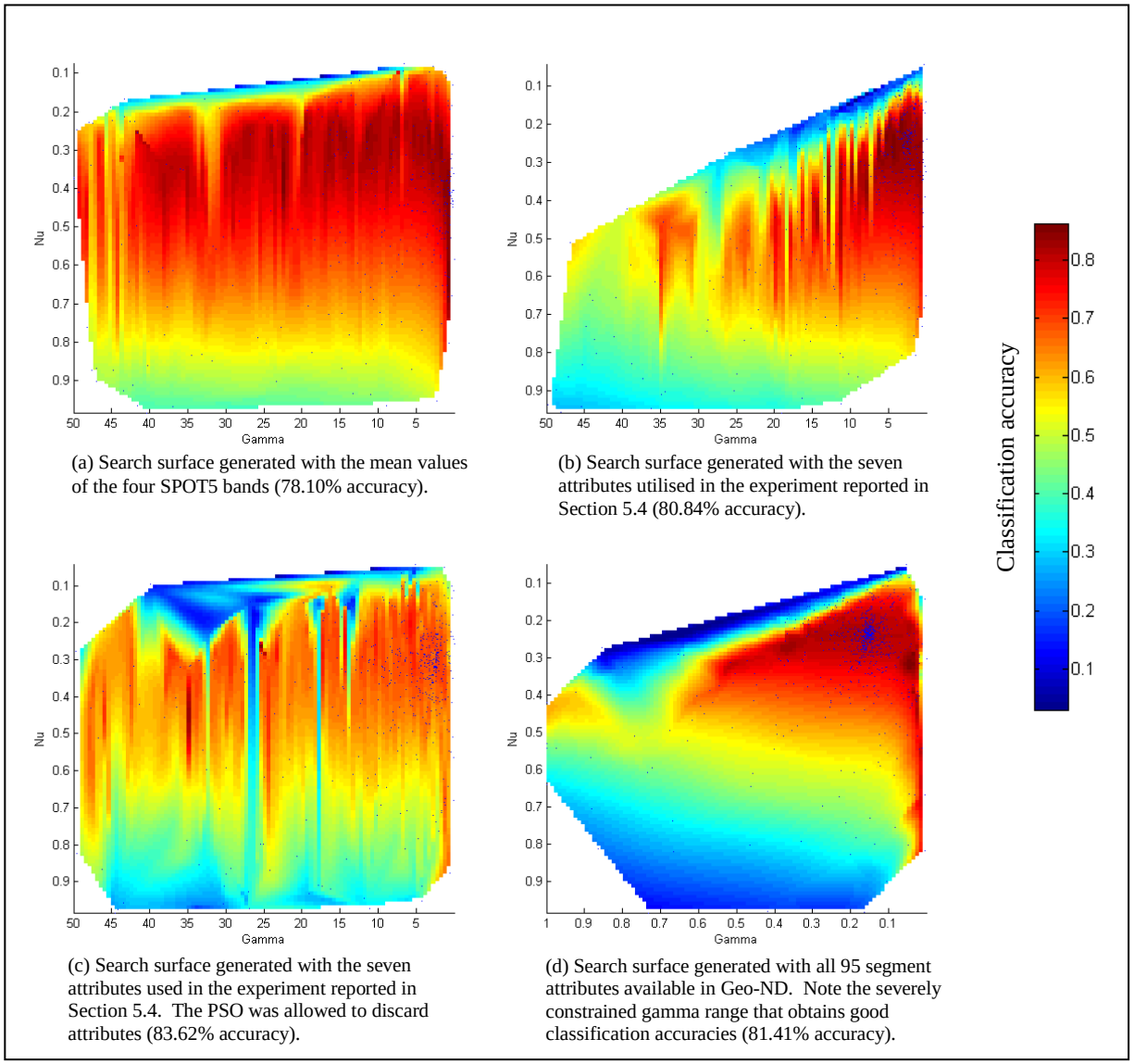
optimisation (PSO). Applications on some datasets do not produce statistically significant results and a free parameter tuning and attribute subset selection (PTASS) approach can even lead to degradation in classifier accuracy compared to an exclusive free parameter tuning only (PTO) approach. In these reported studies datasets with relatively low attribute dimensionality (compared to data with 96 attributes as generated by Geo-ND) were used. These studies do not report on the influence of change in sample size and attribute dimensionality on the performance of the search heuristic.

The PTASS system, as applied to earth observation object based anomaly detection, should be profiled under certain numerical constraints to find safe bounds in which it can function. It is also necessary to investigate whether such an approach can consistently deliver a useful increase in classifier accuracy.

Figure 6.4 elaborates on this concern by illustrating the two dimensional search surfaces generated by the Cape and riparian thicket PCM dataset used in Chapter 5 (see Figure 5.9 and Table 5.3 for more details on the collected data). The small blue dots in the four figures indicate positions that agents have occupied in the two dimensional gamma and nu parameter space at some point during the execution of the search heuristic. This dataset contains 173 samples from the class of interest, allowing high dimensional attribute spaces to be used. Note the decreasing area of good classifier accuracies as the good gamma range decreases, resulting in looser descriptions with more available attributes.

Figure 6.4(a) illustrates the search surface generated by this dataset when using the original four SPOT 5 bands (segment mean values) as the attribute space. Figure 6.4(b) portrays the search surface obtained from using seven intelligently selected attributes as input (the attributes used in Section 5.4). Figure 6.4(c) illustrates the search surface generated by these seven attributes when the PSO was allowed to discard attributes altogether (note the extreme discontinuities in accuracy due to constant attribute swapping). Figure 6.4(d) portrays the search surface generated by using all 96 available attributes in Geo-ND. Note that the high accuracy surface area in Figure 6.4(d) is much smaller (restricted gamma range) compared to the other three search surfaces. The use of seven intelligently selected attributes with the option given to the PSO to discard attributes resulted in the highest geometric means accuracy. Note further that the right edges of the figures (gamma < 0.005) are areas of abrupt change in classifier accuracy. These four figures make it clear that selecting the optimal

number of attributes and samples is not a clear cut task and that results also depend on the intrinsic ability of the used search heuristic.



(a) Search surface generated with the mean values of the four SPOT5 bands (78.10% accuracy).

(b) Search surface generated with the seven attributes utilised in the experiment reported in Section 5.4 (80.84% accuracy).

(c) Search surface generated with the seven attributes used in the experiment reported in Section 5.4. The PSO was allowed to discard attributes (83.62% accuracy).

(d) Search surface generated with all 95 segment attributes available in Geo-ND. Note the severely constrained gamma range that obtains good classification accuracies (81.41% accuracy).

Figure 6.4: Search surfaces generated with the PTASS system using different attribute subsets and operating modes

In an ANN based classification approach using a GA attribute selection search heuristic, it has been shown that the importance of attribute subset selection increase as the size of the training sample decreases (Van Coillie, Verbeke & De Wulf 2007). The effect of attribute subset selection should be profiled under different attribute set size conditions for a one class SVM classifier. It might not be advisable to arbitrarily use all available attributes in an SVM based system, although an SVM is more robust to high dimensional data compared to an ANN classifier.

The effect of the Hughes phenomenon lowers classifier accuracy as data dimensionality increases and sample size decreases, potentially resulting in a lower robustness of the PSO search. The search bounds of the PSO had to be severely constrained to effectively search the region of good classifier accuracies in the search surface illustrated in Figure 6.4(d). A PSO search strategy might falter after certain numerical bounds (sample size or attribute set size) are transgressed.

To summarise, the PTASS system should be profiled in the context of common object based earth observation image classification tasks. Specifically the usefulness of such an approach should be investigated in terms of consistently delivering superior results compared to a PTO strategy. The PTASS system should also be profiled under different sample- and attribute set size conditions.

6.3 INVESTIGATION OF SIMULTANEOUS FREE PARAMETER TUNING AND ATTRIBUTE SUBSET SELECTION OVER DIFFERENTLY SIZED ATTRIBUTE SETS AND TRAINING SAMPLE SIZES

In this section the performance of the free parameter tuning and attribute subset selection system is profiled on common earth observation anomaly detection problems according to the concerns mentioned in the previous section.

6.3.1 Data description

Six different land cover types are investigated, namely dams, Cape thicket, pine plantations, fynbos, built land and agricultural fields. These classes were chosen as the focus of arbitrary anomaly detection problems and they vary considerably regarding the spectral and textural homogeneity of their member elements. The classes are arbitrarily defined informational classes (they hold some value to the users of the data). For each land cover class, between six and ten areas measuring between 1000 and 3000 pixels in width and height were selected which contain good representations of these classes. The pine plantations and Cape thicket class image subsets were extracted from the fully pre-processed SPOT 5 scene, while the image subsets used for identifying the other four classes were extracted from the L2 (SAC) pre-processed scene (see Chapter 3). These image subsets were segmented with appropriate image input bands and segmentation algorithm parameters as determined by the segmentation

component in Geo-ND. Figure 6.5(a)-(f) illustrates samples of the agricultural fields, dams, fynbos, pine plantation, built land and Cape thicket class respectively.

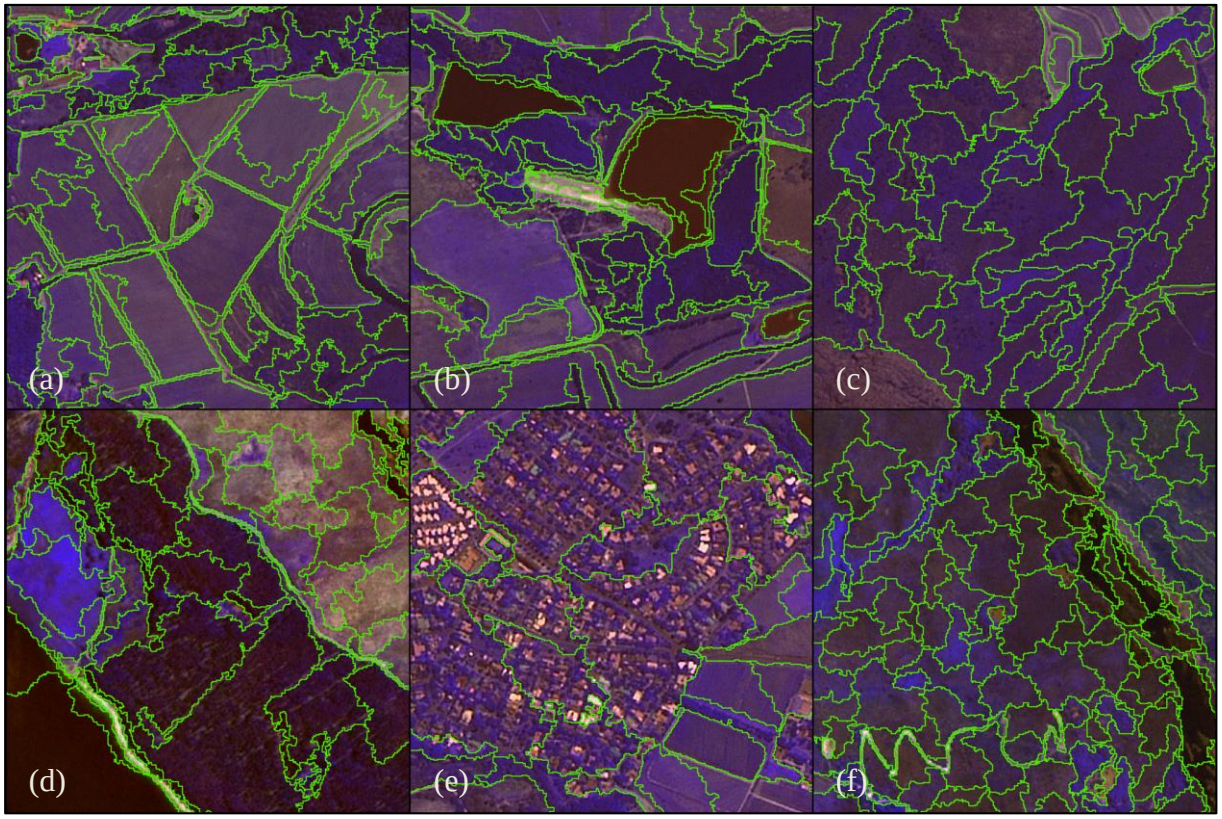


Figure 6.5: Samples and segments used for (a) agricultural fields, (b) dams, (c) fynbos, (d) pine plantation, (e) built land and (f) Cape thicket class respectively

The agricultural fields class (Figure 6.5(a)) encapsulates all cultivated and ploughed fields, vineyards and livestock camps. Degraded fields, displaying natural vegetation growth (fynbos) are not considered part of this class. The pine plantations class (Figure 6.5(d)) encapsulates all segments having any pine species present, irrespective of stand age and/or density. The fynbos class (Figure 6.5(c)) encapsulates all fynbos veld types, including high fynbos (tall Proteaceae) but excludes segments judged to have a majority representation of exposed rock.

The built land class (Figure 6.5(d)) comprises land cover elements displaying high levels of spectral heterogeneity. Images used for investigating the built land class were segmented using a scale parameter of 50 and a colour/shape parameter of 0.2 to allow larger segments to form over these land cover elements. A segment containing any discernable habitable structure belongs to this class (urban land, farmsteads and any other structures).

The scenes used contain numerous mountain face shadows, causing some challenge to a classifier in differentiating dams (Figure 6.5(b)). No context attributes are used.

Samples for classes were selected on unbiased criteria for their membership of the class of interest (informative class sampling), irrespective of differing spectral and textural characteristics of the segments. All land cover elements of interest within these scene subsets were identified. Subsequent experiments were conducted on this dataset. The sample sizes and images used to identify these six classes are summarised in Table 6.1. In all subsequent experiments the PSO is allocated 30 agents and at least 15 generations to explore the search surfaces. Between 15 and 40 generations were found adequate for all the experiments.

Table 6.1: Sample sizes and total investigated surface areas of the classes of interest

Class	Positive samples	Negative samples	Imagery surface area
Agricultural fields	2707	21 677	12.07 km²
Dams	374	20 913	9.89 km²
Fynbos	1119	6625	9.44 km²
Pine plantations	1669	23 246	13.54 km²
Cape thicket	404	12 278	7.96 km²
Built land	2766	10 457	11.48 km²

6.3.2 The effect of feature subset selection on classifier accuracy

In this experiment the usefulness of attribute subset selection in conjunction with free parameter tuning is evaluated on typical earth observation anomaly detection problems under random training sample and attribute set sizes. The variation of the results, in terms of standard deviation, specificity, sensitivity and statistical significance, generated by theses two techniques is profiled. In two subsequent experiments more results are presented under different sample size and attribute set size conditions respectively.

6.3.2.1 Experimental design

For each of the land cover class training datasets listed in Table 6.1, a random number of samples are selected. For each class, the four mean values of the original SPOT 5 bands are included as base attributes while a random number of additional attributes (between one and

forty) were randomly selected. Using these samples and attribute subsets the PTASS system as described in Section 6.1 is employed to build a one class SVM model. Ten fold cross-validation is employed with 30 generations to explore the parameter and attribute space. The cross-validation strategy ensures that the results obtained will generalise to unseen data. For each class the PSO is executed in two modes: the first mode allows the PSO to discard any attributes (PTASS) while the second forces it to use all available attributes (PTO).

To obtain a measure of statistical significance according to the Wilcoxon rank sum test, each experiment is conducted 25 times. In total, each land cover class is classified 50 times (two experiments with 25 runs in each experiment).

6.3.2.2 Results

Table 6.2 contains the results of this experiment. The "Positive/Negative samples" column lists the number of samples used from the class of interest and from the background (random sampling). The "Attributes/Selected attributes" column lists the number of original available attributes (the four mean band values plus a random number of randomly chosen attributes) and the average number of attributes selected by the PSO when it is allowed to discard some attributes. The "Sensitivity, Specificity and Geometric means" columns list the respective forms of classifier accuracies obtained. The values followed by '±' indicate the standard deviation obtained over the allocated 25 runs.

Table 6.2: Classifier accuracies over the six land cover classes using both fixed and variable attribute sets.

Feature	Positive/ Negative samples	Attributes/ Selected attributes	Sensitivity	Specificity	Geometric means	Wilcoxon rank sum test
Agri fixed	1759/11539	16+4/	77.06 ±0.01	78.94 ±0.01	77.36 ±0.00	0.0000001
Agri variable		10.04 ±1.77	81.06 ±1.44	81.26 ±1.38	80.56 ±0.93	
Dams fixed	111/4180	37+4/	71.71 ±0.47	91.76 ±0.58	79.49 ±0.03	0.00004
Dams variable		16.96 ±2.75	81.46 ±5.52	90.80 ±2.32	85.03 ±3.83	
Fynbos fixed	697/3711	16+4/	77.61 ±13.81	44.58 ±17.93	54.04 ±7.56	0.0000001
Fynbos var		8.76 ±1.69	77.78 ±2.43	71.20 ±3.16	73.55 ±2.63	
Pine fixed	838/10194	8+4/	85.44 ±0.20	85.09 ±0.20	85.05 ±0.01	0.0004
Pine variable		7.04 ±1.34	85.88 ±1.10	85.68 ±1.27	85.59 ±0.42	
Thicket fixed	404/12278	11+4/	77.82 ±0.45	77.43 ±0.43	77.40 ±0.02	0.0000001
Thicket var		8.48 ±1.48	81.10 ±1.68	77.63 ±1.46	79.08 ±0.50	
Built fixed	563/4216	40+4/	77.69 ±2.06	54.20 ±11.28	63.61 ±5.43	0.00000005
Built variable		16.32 ±2.61	75.95 ±2.47	77.84 ±2.14	76.34 ±1.50	

Note: Agri = agricultural fields; Pine = pine plantations; Thicket = Cape and riparian thicket; Built = built land.

6.3.2.3 Discussion

In all of the six experiments conducted on the different land cover classes, the PTASS system obtained superior results to those of the PTO strategy. In all cases the differences are statistically significant with a very high confidence level (maximum of 0.0004% for the pine plantations class) which contrasts with the variable results obtained in similar implementations for datasets from other problem domains (Huang & Wang 2006; Lin et al. 2008). Note that the attributes used show some amount of overlap in information value (spectral data transforms) and it is assumed that the test datasets from the experiments reported by these authors only used attributes with no inter attribute correlation. The experiments showed very little deviation over the allocated 25 runs when variable attribute subsets are allowed, indicating that the PSO consistently obtained the same maxima.

In only two cases, measured by sensitivity or specificity, did the variable attributes selection strategy perform slightly worse than the fixed attribute strategy. The pine plantations and Cape thicket classes show little difference in classifier accuracy between the two evaluated strategies. Both these classes/experiments received ample training samples and relatively few attributes, illustrating that attribute subset selection provides little benefit when used under such conditions (this phenomenon is observed again in the experiment presented in Section 6.3.4).

The dams experiment had relatively few samples and a large attribute set. In the training sample collection phase, all types of dams were sampled which ranged from shallow turbid dams to dams partly covered by vegetation. Subsequently, the samples from the dams class did not display spectrally pure samples, resulting in a sensitivity of 72% using fixed attributes and 81% if about half of the attributes were discarded. When using a variable attribute subset, the dams class achieved a 5.5% increase in classifier accuracy.

The fynbos and built land classes attained significantly superior results when the classifier was allowed to discard attributes. This demonstrates that careful selection of attributes is of utmost importance for such difficult problems. For the fynbos class 12 of the 25 experimental runs became trapped in local optima which indicates that the PSO failed to effectively explore the search space. In these instances a gamma of 0.0001 and nu of 0.99 (46% accuracy) were returned. The other 13 runs achieved consistent gamma, nu and accuracy values. These accuracies are much lower than those gained with the variable attribute strategy (61%

accuracy). Similarly, the built land class failed on 19 of the 25 runs, with the remaining 6 runs delivering consistent results. The PSO did not experience similar problems when using variable attribute subsets on these two classes. It is evident that the results are also dependent on the ability of the PSO to effectively search the parameter and attribute space, especially when the attribute dimensionality increases (discussed in Section 6.2). Using a variable attribute strategy, some attributes are initially unselected when the search starts, allowing the PSO to place more focus on searching the gamma and nu dimensions.

Note that attributes will typically be selected intelligently according to expert knowledge of its discriminative power. The PSO free parameter tuning and attribute subset selection strategy could then be used to refine such a selection (as the current implementation stands). This experiment did not attempt to achieve good classifier accuracies by selecting the most discriminative attributes but to evaluate the effect of attribute subset selection on typical problems by using random sampling of training segments and attributes. The results in this experiment demonstrate that the specificity of the PTASS system is high for this domain specific data (low standard deviation of results and a very low Wilcoxon rank sum test compared to the PTO strategy). Under high attribute set size conditions, the PSO has trouble in exploring the search surface. This notion is further investigated in a later experiment.

6.3.3 Influence of training sample size on classifier accuracy with and without attribute subset selection

In this experiment the effect of changing the training sample size is investigated in relation to the performances of the one class SVM optimised by a PSO with (PTASS) and without (PTO) variable attribute subsets.

6.3.3.1 Experimental design

The one class SVM optimised with PSO system is used to classify all six classes from the test dataset (Table 6.1). For each class, 20 sample groups are created, the first group containing 5% of the available training samples and the last group 100% of the available training samples. Groups are incremented with a 5% increase in training sample size. Each sample group is classified five times with and without attribute subset selection. In the previous experiment it was demonstrated that the variability of classifier results is small, suggesting that five runs are sufficient to obtain an average value. A small attribute subset is selected. The 15 attributes used for this experiment are the mean of band 1, mean of band 2, mean of

band 3, mean of band 4, mean of NDVI, mean of the hue, mean of the saturation, mean of the intensity, standard deviation of band 1, standard deviation of band 3, standard deviation of the saturation, homogeneity of RVI, homogeneity of the intensity, contrast of NDVI and contrast of the intensity.

6.3.3.2 Results

Figures 6.6 to 6.11 display the classification accuracies when using differently sized sample sets. The striped lines indicate the classifier accuracies by using attribute subset selection (PTASS) and the solid lines the accuracies achieved without conducting attribute subset selection (PTO).

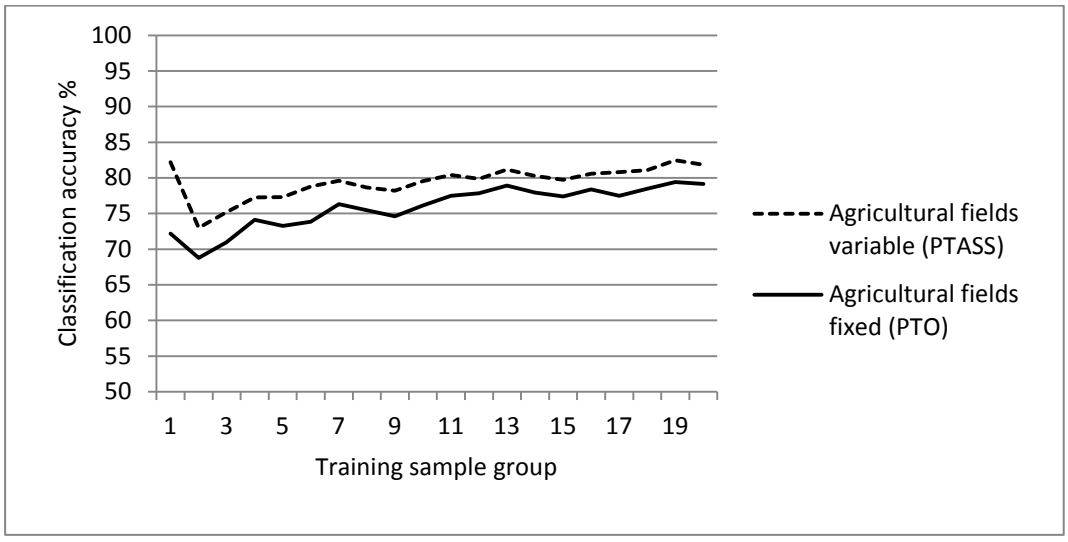


Figure 6.6: Classifier accuracies of the agricultural fields class with different sample sizes

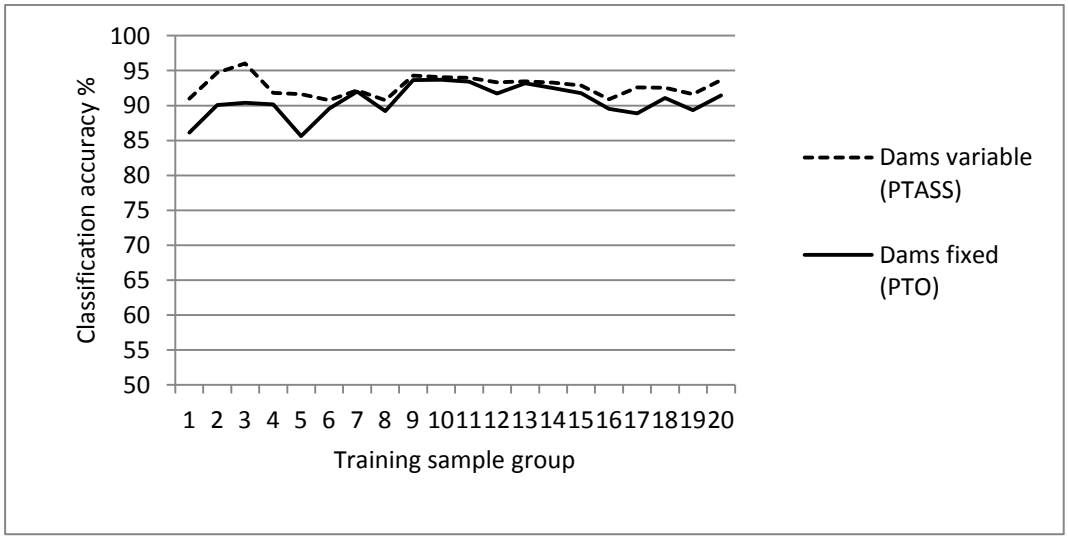


Figure 6.7: Classifier accuracies of the dams class with different sample sizes

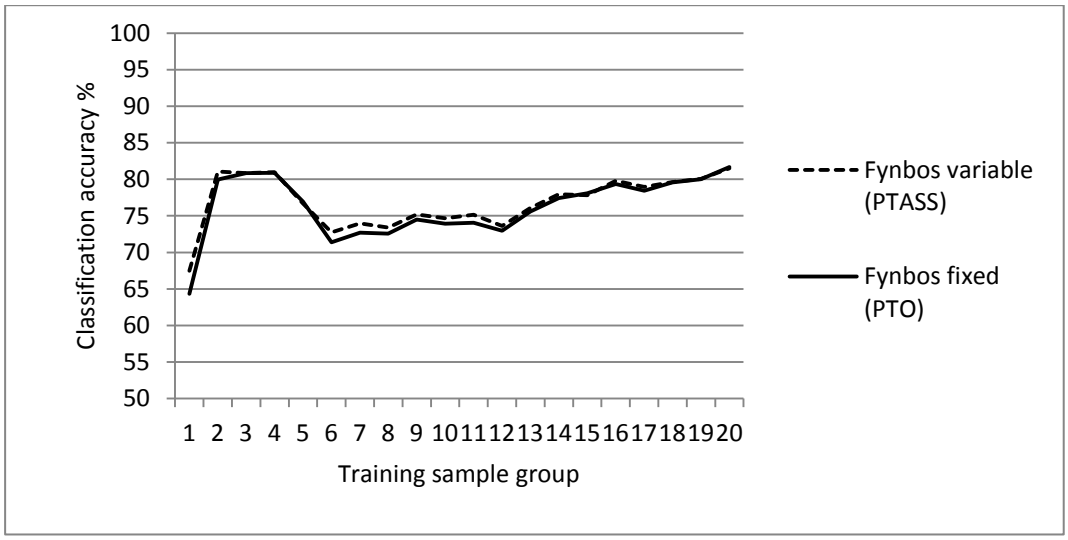


Figure 6.8: Classifier accuracies of the fynbos class with different sample sizes

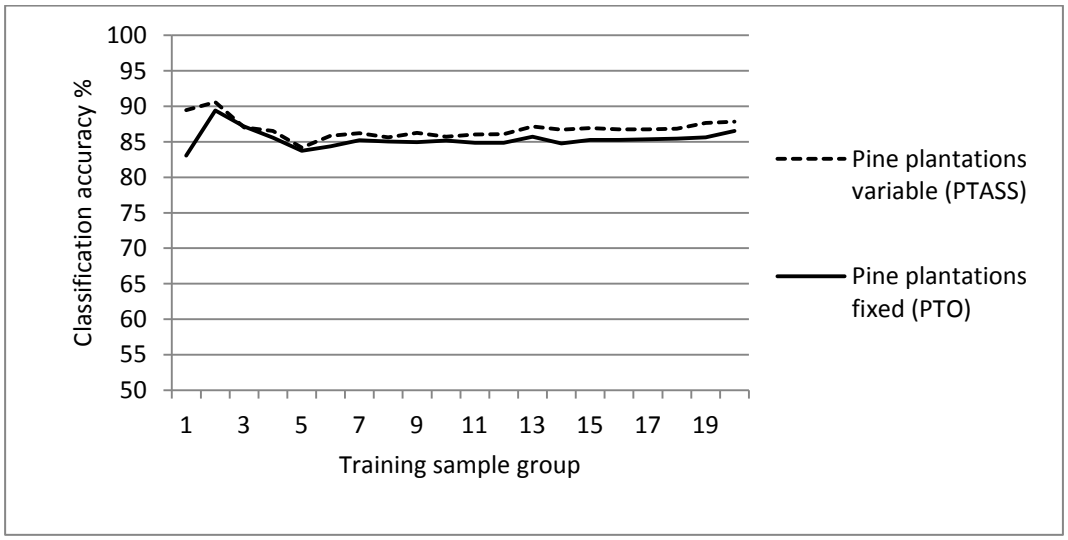


Figure 6.9: Classifier accuracies of the pine plantations class with different sample sizes

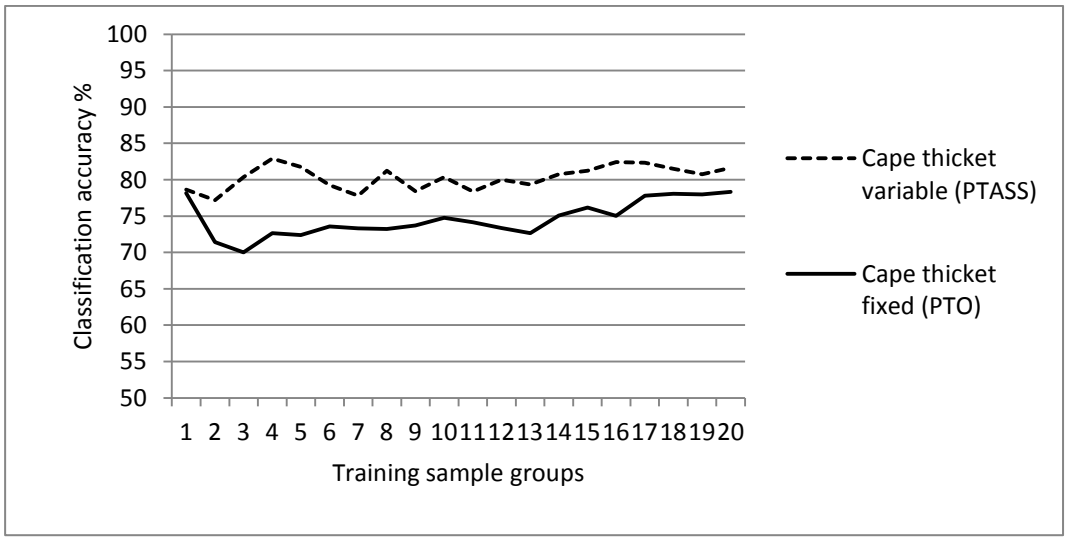


Figure 6.10: Classifier accuracies of the Cape thicket class with different sample sizes

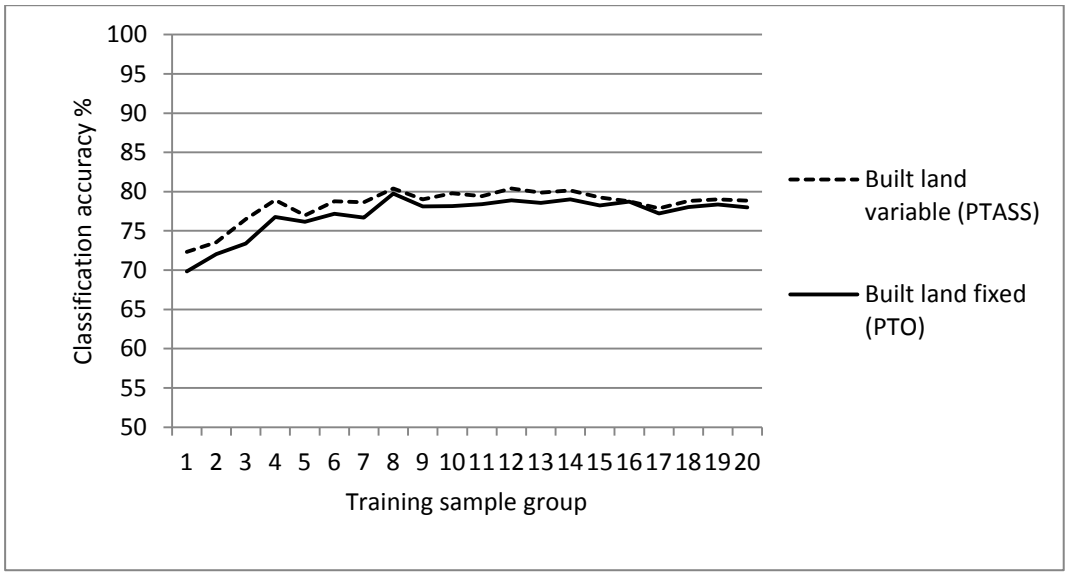


Figure 6.11: Classifier accuracies of the built land class with different sample sizes

Table 6.3 logs the average, minimum and maximum accuracy gains from using the differently sized sample sets with and without attribute subset selection in the classification system.

Table 6.3: The average, minimum and maximum classifier accuracies obtained by using different sample subsets in the six sample set size experiments.

Class	Minimum accuracy gain	Maximum accuracy gain	Average accuracy gain
Agricultural land	2.01%	10.01%	3.49%
Dams	0.21%	6.00%	2.10%
Fynbos	-0.26%	3.12%	0.62%
Pine plantations	-0.16%	6.40%	1.44%
Cape thicket	0.50%	10.38%	5.72%
Built land	0.04%	3.09%	1.29%

6.3.3.3 Discussion

On studying Figures 6.6 through 6.11, no apparent trend is visible regarding the relationship between training sample size and classifier accuracy in contrast to results presented for a similar system operating with an ANN classifier (Van Coillie, Verbeke & De Wulf 2007). In five of the six experiments the greatest gain in classifier accuracy by using attribute subset selection is achieved when very few training samples are available (5% for sample group one). The effect of a variable attribute subset is less noticeable as training sample size increases. The Cape thicket class did not display this behaviour. The classification system showed variation in accuracies when using smaller training sample sets due to variation in the

quality or usefulness of said small sample sets. The larger sample groups display less variation in the accuracies achieved.

Over all the sample groups and experiments (120 experiments), except for five cases, the PTASS system performed better than the PTO system. Within the fynbos class, four sample groups performed worse when using attribute subset selection, although only marginally (minimum of -0.26% accuracy loss). One experiment in the pine plantations class delivered a degradation of -0.16% when using attribute subset selection. Three experiments, using sample groups from the Cape thicket and agricultural land classes yielded an accuracy gain greater than 10% (see Table 6.3).

These results demonstrate that the usefulness of attribute subset selection is not related to training sample size. There is no apparent relationship regarding the number of attributes selected by the presented system and classifier performance. In all 120 experiments the number of attributes selected was between 33% and 66% of the original pool of fifteen attributes. It should be noted that in this experiment fifteen attributes were arbitrarily chosen and not tailored to any specific class. Use of a different attribute subset may result in substantially different results as presented in Figures 6.6 through 6.11. These results demonstrate that the benefit of utilising variable attribute subsets is not severely influenced by the quality or size of the samples used in training the classifier.

6.3.4 Robustness of classifier results in relation to attribute space dimensionality

In this experiment the usefulness of the PTASS system is profiled over differently sized attribute subsets.

6.3.4.1 Experimental design

For each class from Table 6.1 a random training sample group is created containing between one third and two thirds of the available samples. These training sample groups are classified with six different sizes of available attribute sets. The first attribute set contains the four mean values of the original SPOT 5 input bands. The other five attribute sets have 15, 30, 45, 60 and 75 attributes respectively. Each attribute set contains all the attributes from the set just smaller in size than itself, with an additional 15 attributes selected randomly. Table 6.4 lists the attributes used in these attribute sets. The columns indicate the respective attribute sets

and the rows the corresponding attribute properties. The numbers in the cells are keys that correspond to the spectral bands used to derive the attributes.

Table 6.4: The attributes used in the differently sized attribute subsets.

	4 Attribute set	15 Attribute set	30 Attribute set	45 Attribute set	60 Attribute set	75 Attribute set
Mean band value	1, 2, 3, 4	1, 2, 3, 4, 12, 14	1, 2, 3, 4, 12, 14, 15	1, 2, 3, 4, 10, 12, 14, 15	1, 2, 3, 4, 5, 6, 10, 12, 14, 15	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15
Standard deviation		2, 3, 8	2, 3, 4, 8, 13	2, 3, 4, 6, 8, 13, 14	2, 3, 4, 6, 8, 11, 13, 14	2, 3, 4, 5, 6, 7, 8, 11, 13, 14, 15
Band ratio		3	3	1, 3	1, 3	1, 2, 3, 4
Maximum difference		4, 7	1, 4, 7, 9, 12	1, 4, 5, 7, 9, 10, 12, 13, 14	1, 3, 5, 7, 9, 10, 12, 13, 14	1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
GLCM Contrast			4, 8, 9, 14	4, 5, 8, 9, 14, 15	4, 5, 7, 8, 9, 10, 13, 14, 15	3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 15
GLCM Homogeneity		9	2, 7, 9, 12, 15	1, 2, 7, 8, 9, 12, 13, 15	1, 2, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15
GLCM Entropy		3, 12	1, 3, 12	1, 3, 4, 12, 13	1, 2, 3, 4, 7, 8, 9, 12, 13, 14	1, 2, 3, 4, 7, 8, 9, 12, 13, 14

Note: 1 = Band 1, 2 = Band 2, 3 = Band 3, 4 = Band 4, 5 = NDVI, 6 = RVI, 7 = Hue, 8 = Saturation, 9 = Intensity, 10 = Band 1/Band 2, 11 = Band 1/Band 4, 12 = Band 3/Band 4, 13 = Band 2/Band 1, 14 = Band 4/Band 1, 15 = Band 4/Band 3.

The six classes are classified by using these six attribute sets, with (PTASS) and without (PTO) employing the attribute subset selection strategy. Each experiment is repeated 25 times to obtain a measure of classifier robustness. Thus for each class 300 particle swarms are invoked (each swarm constructs 15000 one class SVMs, resulting in 4.5 million one class SVMs constructed and evaluated per land cover class).

6.3.4.2 Results

Figures 6.12 - 6.17 illustrate the classifier accuracies achieved over the differently sized attribute subsets. The error margin lines indicate the standard deviation of the results obtained over the allocated 25 runs. No error margin lines are displayed for the PTO strategy as the standard deviations observed were insignificantly small.

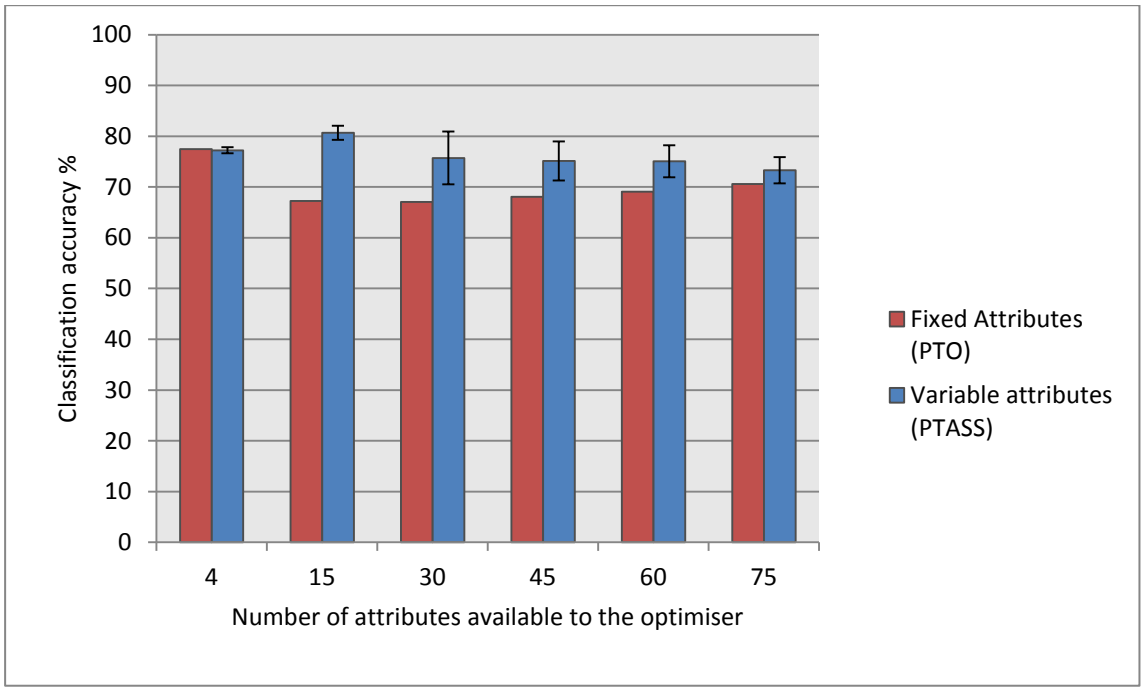


Figure 6.12: Classification accuracies for the fynbos class with the PTASS and PTO strategies respectively

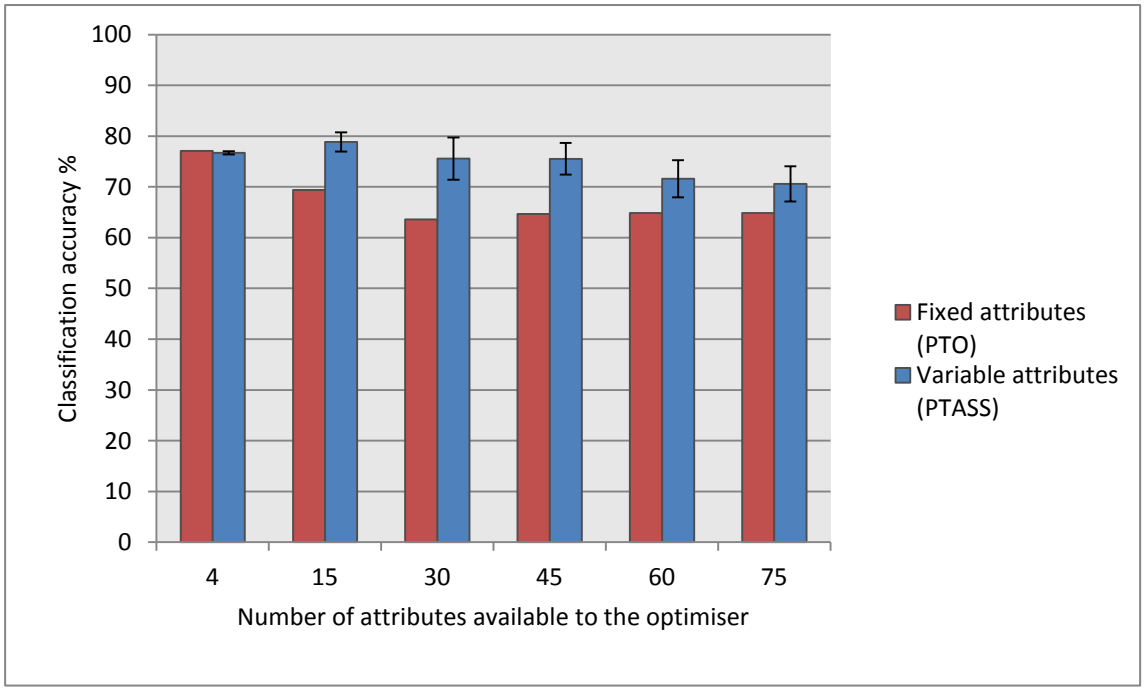


Figure 6.13: Classification accuracies for the agricultural land class with the PTASS and PTO strategies respectively

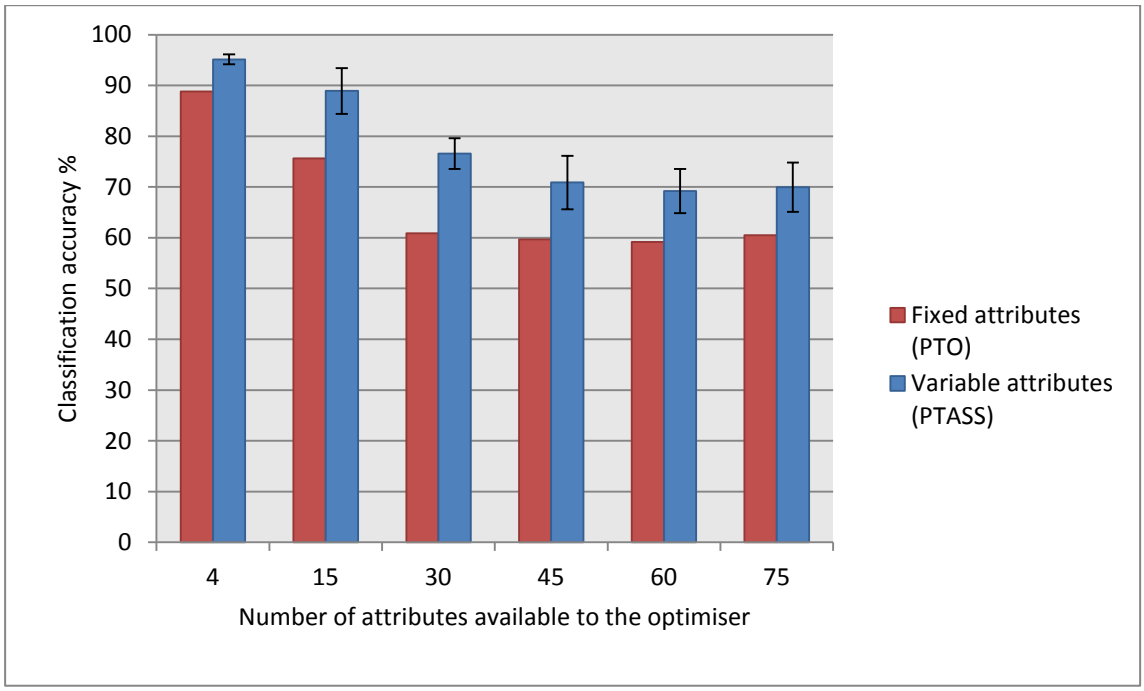


Figure 6.14: Classification accuracies for the dams class with the PTASS and PTO strategies respectively

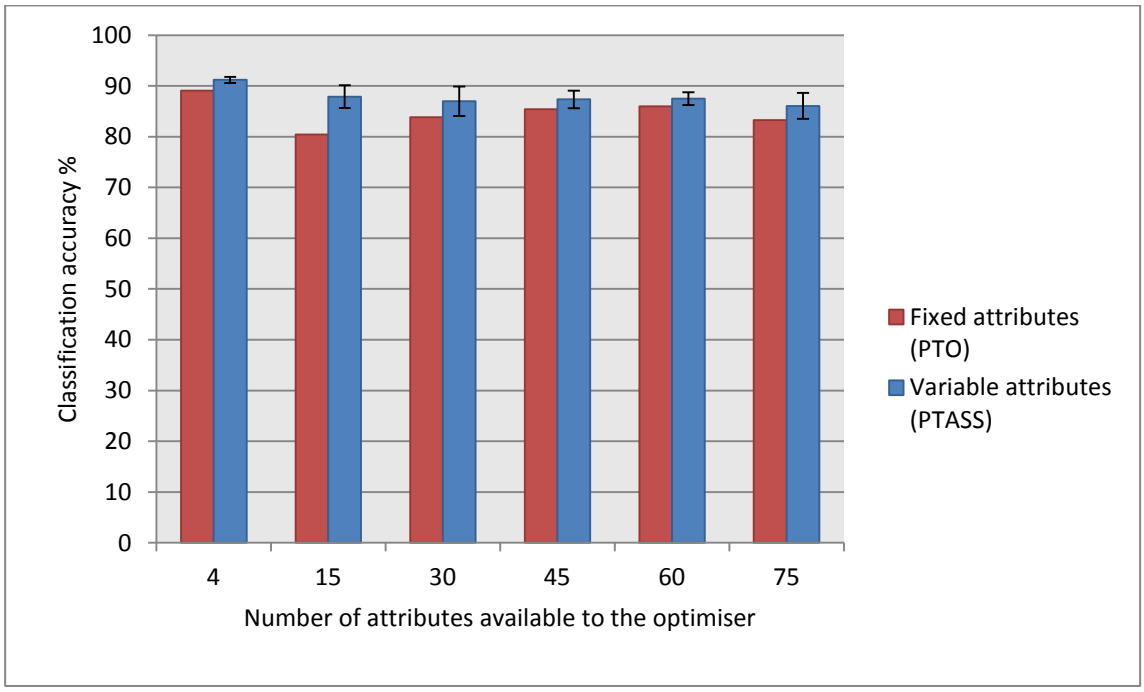


Figure 6.15: Classification accuracies for the pine plantations class with the PTASS and PTO strategies respectively

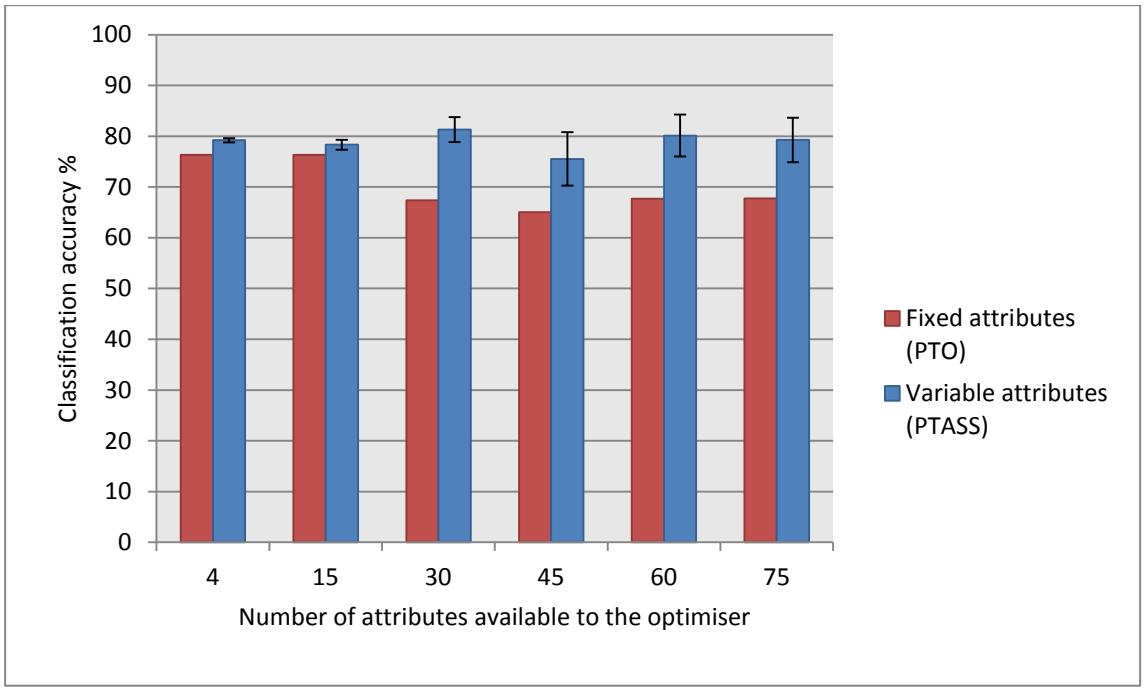


Figure 6.16: Classification accuracies for the Cape thicket class with the PTASS and PTO strategies respectively

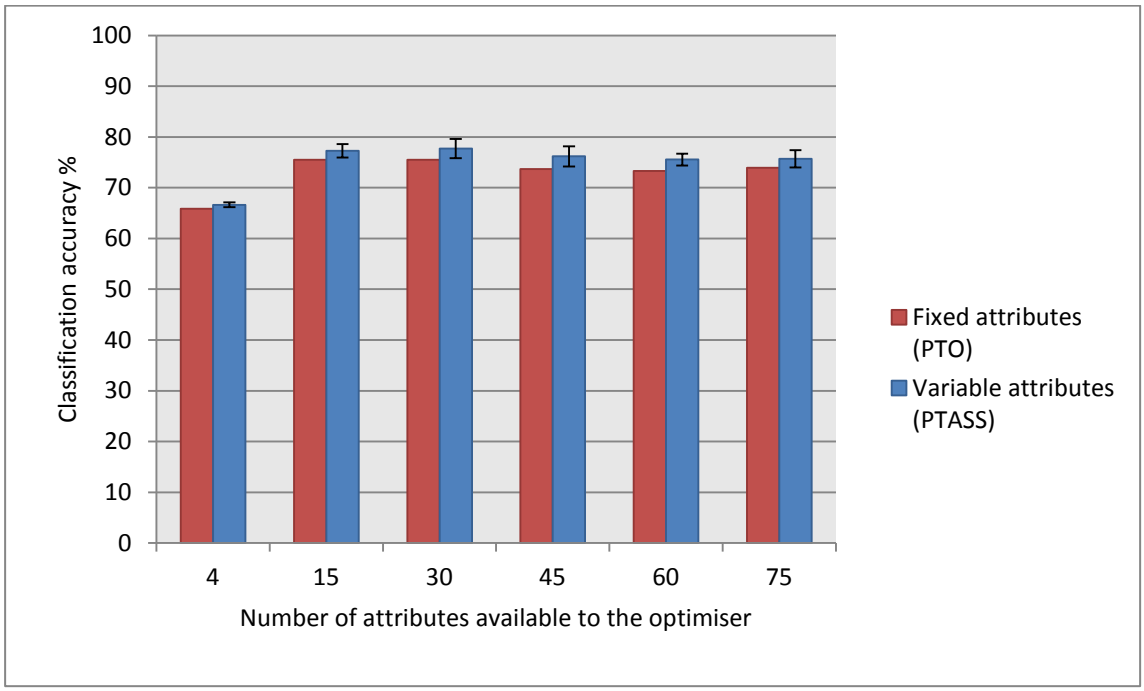


Figure 6.17: Classification accuracies for the built land class with the PTASS and PTO strategies respectively

6.3.4.3 Discussion

The PTASS strategy performed better than the PTO strategy for all land cover classes and over all the attribute groups. The only exceptions were when the four mean original SPOT 5 band values were used as input for identifying fynbos (Figure 6.12) and agricultural land (Figure 6.13). The PTASS strategy achieved an accuracy gain of more than 10% for four of the six land cover classes investigated.

The investigated land cover classes displayed unique accuracy profiles when the different attribute sets are used. In general, however, classifier performance increases as more descriptive attributes become available. The increase in the availability of descriptive attributes usually coincides with larger attribute sets. Conversely, the larger attribute sets also resulted in the classifier being hampered by irrelevant or redundant attributes. One such example is dams (Figure 6.14). Generally, dams are considered a relatively easy classification problem, with only shadows providing some difficulty for the classifier (see Figure 6.5). The one class SVM found one or two mean band values to be sufficient to classify dams. The lower dimensionality of the attribute set allowed the SVM to construct a more discriminative hyperplane using a few very discriminative attributes. As more attributes were added to the classifier, having no relevance (or discriminative power) to the problem, the ability of the SVM to construct a very discriminative hyperplane diminished. Using the PTASS strategy the PSO was able, to some extent, to discard these irrelevant attributes.

In contrast to the dams class, the built land class was a more difficult classification problem to solve. As evident from Figure 6.17 the attributes from the first two groups (as a collective) had less discriminative power compared to the attribute set from the third group. It is possible that no small group of attributes would be effective in identifying built land and as such a SVM constructed using numerous attributes (and thus a more general hyperplane not heavily influenced by any specific attributes) performed best. Consequently, the Hughes phenomenon will have a more profound effect when classifying such classes where no single/small group of attributes hold very discriminative value to the problem.

PTASS strategy found all four attributes useful in identifying fynbos segments (Figure 6.12) but only three in identifying agricultural land (Figure 6.13). In most runs the PTASS strategy discarded the mean value of the fourth SPOT 5 band (mid infrared) in identifying the agricultural lands class. A possible explanation for this is that some segments displaying very

low fynbos cover have a nominal mid infrared reflectance (see Figure 2.1). The mid infrared band is needed to differentiate fynbos from segments with exposed sandstone or soil, hence its inclusion in most of the PSO agents. Due to the "greener" spectral profiles of the agricultural class, the fourth SPOT 5 band was found to hold less discriminative value and was often (perhaps erroneously) discarded by PSO agents due to the stochastic nature of the search heuristic.

When using the three largest attribute groups (45, 60 and 75 attributes groups) in the agricultural land class experiments the PSO failed to find the true optimum in a number of runs. This was only observed when using the PTO strategy. Specifically, when using the 45 attributes group the PSO was trapped in a local optimum three times, seven times for the 60 attributes group and eight times for the 75 attributes group (out of 25 runs). These cases were excluded in generating Figure 6.13. The search surfaces in these instances had a similar structure to the search surface presented in Figure 6.4(d), with the PSO being trapped in a very low gamma (0.0001) and very high nu (99%) local optimum. The true optimum had a slightly higher gamma (±0.001) and a nu value of 0.58.

Reviewing Figures 6.12 - 6.17 it is evident that the ability of the PSO to effectively perform combinatorial optimisation faltered as the dimensionality of the combinatorial optimisation problem increased, as highlighted by Al-Ani (2009) (Figure 6.14 is a prominent example). Higher dimensionality also provided difficulty for the two dimensional numerical component of the encoded PSO as it fell into local optima in a few cases when using the free parameter tuning only strategy. This aspect was not addressed by Lin et al. (2008) and Huang & Wang (2006) as their datasets had relatively low dimensions compared to the 77 dimensional optimisation problem addressed in this section. Although not perfect, the PTASS strategy based on the PSO assisted (in many cases) in increasing the classifier accuracies significantly. It should be stressed, however, that a SVM classifier is not immune to the effects of the Hughes phenomena (Pal & Foody 2010) and that a decrease in classifier accuracy is expected as attribute dimensionality increase.

It can be concluded from this experiment that the numerical bounds that can be imposed on such a system are dependent on the ability of the search heuristic. For this PSO implementation, with this domain specific data and hand tuned meta-parameters, the best results are obtained when fewer than 30 attributes are provided and when the system is

allowed to discard attributes (PTASS). Note that attributes were selected randomly and that better classifier results could be obtained if a small attribute set (5 to 30 attributes) is intelligently selected. With the current implementation the PTASS system would be effective in further refining such an intelligent selection of attributes.

As part of the tutorial in the accompanying user guide (see Appendix C) a course is detailed that replicates the comparison (in a small, qualitative experiment) of the PTASS and PTO systems detailed in this section specifically for the identification of dams. Results similar to those presented in Figure 6.14 are observed.

6.4 SUMMARY

This chapter introduced a system that simultaneously performs free parameter tuning and attribute subset selection for an object based anomaly detector. The system was profiled under different sample size, class type and attribute subset size conditions. Under most conditions, the system provided a significant increase in classifier accuracy by discarding redundant or irrelevant attributes. Under some operating conditions (typically when very few attributes are available) a small negative increase in classifier accuracy was observed, this being attributable to the stochastic nature of the search heuristic.

It is recommended that such a system can readily be used for object based earth observation image processing to increase classifier accuracy without much risk of a degradation of the results. The system is especially useful in situations with small sample sizes and uncertainty about the usefulness of segment attributes for a specific application, thus endorsing its value to save time and costs (regarding training sample collection and manual attribute subset selection). Reducing the number of used attributes may also save costs (in terms of computing power) when classifying new areas because segment attribute generation is computationally the most expensive part of the segmentation/classification workflow (even more so than the actual segmentation process itself, as evident in the system developed here).

The favourable results of this system are dependent on the ability of the implemented search heuristic. The search heuristic used in this system is the original PSO (Kennedy & Eberhart 1995) with hand tuned meta-parameters. Other search heuristics that potentially perform better as combinatorial optimisers (e.g. ant colony optimisation) may increase the effectiveness of such a system by searching the attribute space more effectively.

The accuracies attained with this system in the experiments conducted in this chapter mainly ranged between 70% and 90% geometric means accuracy. Emphasis was placed on the effect of the variable attribute subset component of the system and not on obtaining maximum classifier accuracy by intelligently selecting or formulating (spectral, textural and context) attributes, nor on creating easily identifiable classes or using data with high discriminative power. Classifier accuracy can potentially be increased by defining more appropriate attributes and/or using more discriminative data.

CHAPTER 7: VEGETATION STRUCTURAL CASE STUDY

In this chapter the proposed anomaly detection system is used to map broadleaf vegetation of the thicket and forests structural types on a subset of a SPOT 5 scene. To enable a comparison of the system's thematic accuracy and workflow efficiency, a KNN multi-class classification is also performed to identify the same features.

7.1 CLASSIFICATION METHODOLOGY

A 22.5 km x 22.5 km subset of the fully pre-processed 60 km x 60 km SPOT 5 scene is classified to identify broadleaf vegetation of the thicket or forest structural types. The area on the image subset mainly encompasses the Drakenstein and Hottentots Holland Mountains. Stellenbosch is located in the north-west corner of this area and Somerset West in the south-west. Afrotemperate forest, Cape thicket, Mitchell thicket (for distinction see Campbell 1985) and riparian thicket display very similar spectral and textural profiles when using moderate to high resolution multispectral data. Consequently, constrained by the data (data suitability for vegetation mapping was addressed in Chapter 2), these classes are aggregated and are considered as the class of interest. This information class is defined as any broadleaved vegetation with a structural height greater than three meters. This classification step can be seen as a first one to a greater discrimination workflow to identify Cape thicket. The results of this classification step can be refined with additional modelling and expert knowledge to separate Cape thicket from Afrotemperate forests, Mitchell thicket and riparian thicket.

A portion of the scene subset is used as training data. These training data are used to classify the rest of the subset with the methodology presented in Geo-ND and with a multi-class nearest neighbour classifier as benchmark.

7.1.1 Sampling strategy

Figure 7.1 shows the area investigated in this case study. Nine equally sized grids (7.5 km x 7.5 km) were created from this subset for more efficient processing in Geo-ND. The subset covered with a mesh (see Figure 7.1) is used as a training area. The gray areas denote a mask created to cover most of the urban and agricultural land present in this scene subset. These areas are not used in this study because vegetation in these areas may easily be incorrectly

identified as belonging to the class of interest. This is done as a contextual classification step to remove unnecessary data.

The segmentation component in Geo-ND was used to find suitable input bands for segmentation and segmentation algorithm parameters (see Chapter 5). The original near infrared, NDVI and intensity bands were used. Geo-ND generated a scale parameter of 40, colour/shape parameter of 0.6, compactness/smoothness parameter of 0.4 and band weights of 50% (band 3), 65% (NDVI) and 47% (intensity). The entire scene was segmented with these parameters, tailored to the features of interest. These segments are used in both classification methodologies. The area investigated contains 47 502 segments (excluding the masked areas).

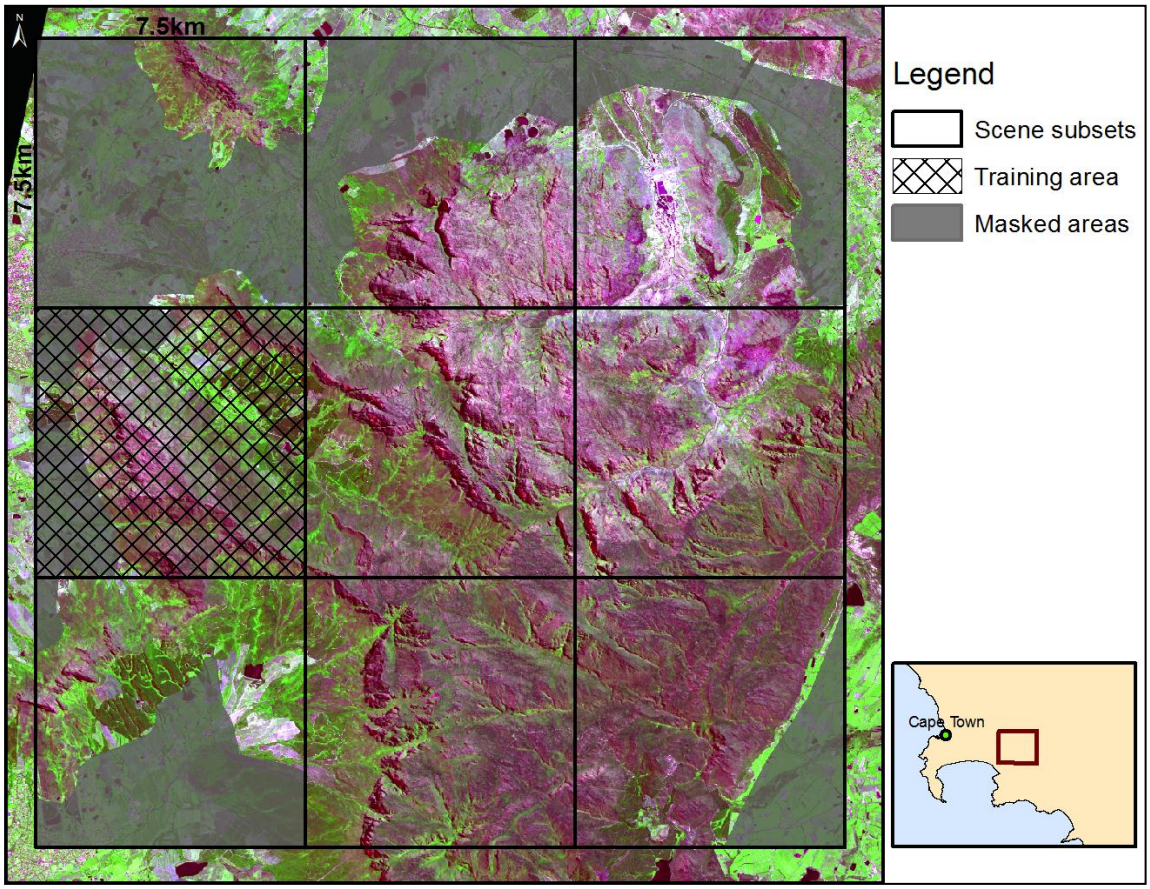


Figure 7.1: The extent of the SPOT 5 scene subset used in this case study.

All Cape thicket, Mitchell thicket, riparian thicket and Afrotemperate forests were identified on the used SPOT 5 subset in this area. If a segment was judged to contain more than 50% of any of these classes, it was selected as part of the ground truth sample set. Note that coniferous vegetation, vegetation on urban peripheries, seeps/wetlands, high fynbos and

Waboomveld are not part of this class. Third party very high resolution satellite imagery of this area (GeoEye-1/Google Earth) was used to verify every sample collected from the SPOT 5 image. In addition, the prominent sites of the class of interest in the mountainous areas in six of the nine subsets (the six left most subsets) were subjected to ground truth evaluation to confirm the sampling (see Appendix A). In total, 1167 segments belonging to the class of interest were identified in this 22.5 km x 22.5 km area.

7.1.2 Geo-ND classification

A one class SVM model with appropriate parameters was generated in Geo-ND (see Chapter 6) using the data in the training area subset. The training area, as depicted in Figure 7.1 contains 219 segments of the class of interest, out of a total of 8007 segments. The data from this area were used in the automatic free parameter tuning and attribute subset selection system. The generated model and accompanying data were stored as a solution file. The remaining eight subsets were classified by means of the one class SVM model. The ground truth data from the remaining eight subsets along with the classification results were used to generate confusion matrices.

Table 7.1 lists the attributes used for this case study. The attributes were selected subjectively from personal experience and their general usefulness in object based vegetation discrimination studies (Mallinis et al. 2008; Van Coillie, Verbeke & De Wulf 2007). The attribute set was limited to 13 attributes, based on the recommendations made on attribute set size in Chapter 6. The training/classification was performed with both fixed (PTO) and variable (PTASS) attribute subsets while given a large number of generations and agents to explore the search surface. No repeated runs of the PTASS and PTO strategies were performed.

Table 7.1: Segment attributes used in the case study

Segment property	Band used
Mean	Band 2
Mean	Band 3
Mean	Band 4
Mean	NDVI
Mean	Intensity
Standard deviation	NDVI
Standard deviation	Intensity
Band ratio	Band 3
GLCM homogeneity	Band 3
GLCM homogeneity	NDVI
GLCM contrast	Band 2
GLCM contrast	Band 4
GLCM entropy	Intensity

7.1.3 KNN Classification

Seven land cover classes, tailored specifically by the nature of the area under investigation, were formulated. These classes are listed in Table 7.2. For each class, 80 samples were identified within the training area depicted in Figure 7.1 so that 560 samples were available to build a KNN classifier.

Table 7.2: Land cover classes generated for use in the KNN classifier.

Land cover class
Thicket and Afrotemperate forests
Dams/Shadows
Pine plantations/Forests
Fynbos (all types)
Exposed rock/Land
Agricultural land
Hillslope/ravine seeps (granite or sandstone fynbos)

The k parameter in a KNN classifier denotes the number of neighbours observed in deciding the membership of a sample. The KNN classifier was trained with the 560 training samples by a 10-fold cross-validation strategy for k values in the range 1 to 100. This process was conducted with both normalised and non-normalised segment attributes to determine if it has a significant influence on classifier accuracy (to allow the KNN to achieve maximum classifier performance).

Although multiple classes are used in the KNN classification, results are only reported on the users' and producers' accuracies of the class of interest (thicket/Afrotemperate forest). The other classes are grouped into a single class, entitled "other". The same attributes used in the Geo-ND classification process were used in the KNN classifier. Due to the degree of difficulty of the problem and the spectral and spatial limitations of the data, and to allow the KNN to perform as well as possible, the entire scene was classified with k values ranging from 1 to 40. The ground truth data in these eight subsets, together with the results of the classification with the parameter k delivering the best results, were used to generate a confusion matrix. The classifier accuracy of the KNN with the optimal k generated with the training data (not all the data) is also calculated. The results of the KNN were biased to perform as well as possible by using knowledge of the entire scene a priori in creating the classifier.

7.2 CLASSIFICATION RESULTS

Figure 7.3 exhibits the users' and producers' accuracies of the thicket/Afrotemperate forest class achieved with the KNN classifier for a range of k values over the entire scene. Table 7.3 is the confusion matrix produced by the biased KNN classifier when using a six nearest neighbours rule which delivered the best classification results.

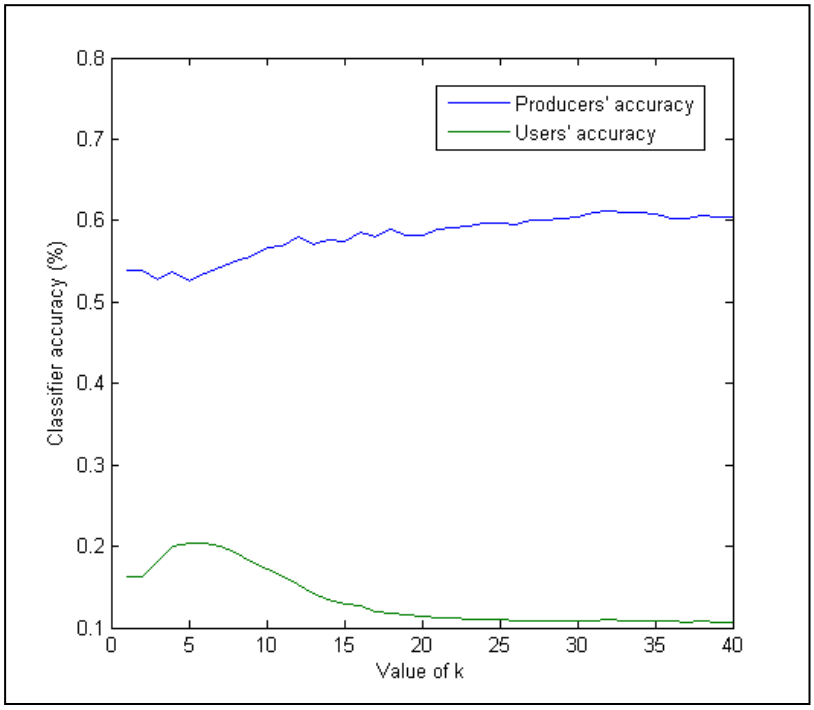


Figure 7.3: Users' and producers' accuracies achieved with the KNN classifier over a range of k values

Table 7.3: Confusion matrix of the biased KNN classifier with k delivering the best results (k = 6)

	Thicket	Other	Producers' accuracy
Thicket	502	435	53.58%
Other	1959	44 606	
Users' accuracy	20.40%		

During the training phase of the Geo-ND classifier, an accuracy of 80.03% was obtained by forcing the PSO to use all 13 attributes. An accuracy of 81.61% was obtained when using variable attributes. Six attributes were selected by the search process, namely the means of bands 2, 4, NDVI and intensity, the homogeneity of NDVI and the entropy of intensity. The system delivered a gamma of 0.13 and nu of 0.41 operating in fixed attribute mode and a gamma of 0.51 and nu of 0.30 operating with variable attributes. Note the tighter (higher) gamma obtained when using fewer attributes.

Tables 7.4 and 7.5 are the confusion matrices generated by the Geo-ND classification process operating with fixed and variable attributes respectively. Table 7.6 lists the geometric means accuracies obtained with the biased KNN, KNN, Geo-ND with fixed attributes and Geo-ND with variable attributes respectively.

Table 7.4: Confusion matrix of the Geo-ND classification strategy using a fixed attribute set

	Thicket	Other	Producers' accuracy
Thicket	575	362	61.37%
Other	2657	43 908	
Users' accuracy	17.79%		

Table 7.5: Confusion matrix of the Geo-ND classification strategy using a variable attribute set

	Thicket	Other	Producers' accuracy
Thicket	679	258	72.47%
Other	4055	42 510	
Users' accuracy	14.34%		

Table 7.6: Geometric means accuracies obtained with the four evaluated classification strategies

Classifier	Geometric means accuracy
Biased KNN	0.3305
KNN	0.2538
Geo-ND fixed	0.3304
Geo-ND variable	0.3223

7.3 DISCUSSION

From an examination of Tables 7.3 through 7.5 and Figure 7.3 it is evident that in this case study very low classification accuracies were reached with all the classification strategies. A substantial disparity exists between users' and producers' accuracies. The Geo-ND classification strategies delivered superior producers' accuracies while not suffering from a very low users' accuracy as with the KNN classifier as evident in Figure 7.3 (for the KNN trained on only the training data).

From Table 7.6 it is clear that the Geo-ND classification strategies performed better than the (non-biased) KNN strategies. Only one KNN classification strategy (k = 6) performed marginally better than the Geo-ND strategies. This is attributable to the biased nature in which that specific k was calculated. The Geo-ND strategies were not given the same liberty.

Interestingly, the variable attribute subset selection strategy performed worse (geometric means) than the fixed attribute strategy. Although the variable attribute strategy recorded a much higher producers' accuracy, numerous non-thicket/forest segments were classified as

thicket/forest by this model. This phenomenon is an indication that the samples used in the randomly selected training area were not representative of samples in the other subsets.

The low overall classifier accuracies are attributable to the inability of the data to discriminate the features (multispectral moderate to high resolution data). Numerous fynbos hillslope seeps/wetlands having very similar spectral characteristics to the Thicket/Afromontane forest class are present in the area. The differentiating factor between these two classes is sporadic darker pixels present in areas of thicket/Afrotemperate forest (canopy shadows) that were not properly modelled as an attribute. The subset randomly selected as the training area contained numerous riparian thicket strips among pine plantations which display a spectral profile very similar to seeps. Many riparian thicket sites were encapsulated by segments extending some distance into the neighbouring fynbos, creating some confusion for the classifiers.

More discriminative data (hyperspectral imagery or aerial photography) would probably greatly improve the accuracies. Better results could also have been obtained if more elaborate context attributes, such as DEM derived information, were used in the classification.

In addition to the Geo-ND classification strategy performing better (except for one value of k) than the simpler KNN strategy, the Geo-ND strategy only required samples from the class of interest. Finding samples for all the classes of the KNN classification strategy constituted the greatest workload in this mapping exercise.

7.4 SUMMARY

Geo-ND was used to identify a single class of interest in a 22.5 km x 22.5 km mountainous area. Adequate segmentation algorithm input bands and parameters were determined with Geo-ND. This area was subdivided into nine parts for easier processing. One of the parts was randomly chosen and acted as training data to train a one class SVM classifier. The remaining eight parts of this area were classified with this model. As a benchmark, a multi-class KNN classifier was built on the training data and used to classify the remainder of the scene.

Although general classification accuracies with both strategies were poor due to the general discriminative capabilities of the data, the Geo-ND based system performed slightly better

than the simpler KNN classification strategy. In addition, the Geo-ND strategy only required samples from the class of interest, suggesting its preference over the multi-class KNN classifier when only one type of feature is of interest and when the available training samples are few.

CHAPTER 8: CONCLUSION AND RECOMMENDATIONS

In this chapter the aim of the research is revisited, conclusions are drawn about the specific components of the developed system and ideas for future research are presented.

8.1 REVISITING THE RESEARCH AIM

In the introductory chapter of this thesis, the concept of automating information extraction from earth observation imagery was introduced as a topic of active research. Contemporary aims in geoinformatics research include creating more autonomous workflows, delivering more accurate results and delivering said results in a shorter timespan. This is necessitated by the real world needs of society and the burgeoning costs of traditional manual image processing workflows.

Geoinformatics is not a primary field of scientific investigation, but an applied field servicing a specific problem domain. Appropriately, geoinformatics research commonly involves theory and techniques from a wide variety of other disciplines such as the computing sciences (e.g. computational intelligence, theoretical computer science and machine learning), statistics, the image processing fields and the core related physical disciplines (e.g. physical geography, geology, ecology, botany and climatology) (Blaschke & Lang 2006; Hay & Castilla 2008). The techniques developed in these fields are not always, or in good time, investigated in the context of problems addressed in the field of geoinformatics.

In this research a particular type of problem was formulated and addressed, namely the identification of sparsely distributed land cover elements on earth observation imagery. This problem was approached in the context of a contemporary aim in geoinformatics research, namely creating more accurate and autonomous workflows (Blaschke 2010). This research aimed to investigate and formulate techniques that would satisfactorily service such a type of problem. Fields related to geoinformatics were investigated to find promising techniques to create an accurate and highly automated workflow.

A geographic object anomaly detection system was formulated, developed and evaluated. This system combined several components tailored to automate the classification process and increase classifier accuracy. Most notably, it consists of a component to automate good image segment generation in an anomaly detection context and a component to increase

classifier accuracy through simultaneously tuning classifier free parameters and the used attribute subset.

An underlying aim of this system is to automate the generation of numerous parameters found in the image processing workflow by using computational intelligence and empirical discrepancy techniques. Automating such parameter tuning processes (simplifying the information extraction workflow) that do not result in degradation in the accuracy of the final product, may allow a greater uptake of such technologies by a wider community of practise, in addition to the main benefits of saving time and costs.

This thesis has an applied focus. New theoretical algorithms or heuristics were not formulated. The thesis reports on a novel system that was formulated, combining existing algorithms and heuristics, applying them to a new data specific problem domain and evaluating their use for solving real world problems. Some quantitative results are presented on how this system as a whole, and its individual components, can aid in creating a more autonomous workflow and deliver more accurate results in an earth observation image processing context. The aim of the research was thus successfully addressed.

8.2 CONCLUSIONS

In this section the two main components of the developed anomaly detection system are reviewed and the main quantitative findings are summarised.

8.2.1 Image segmentation in the context of geographic object anomaly detection

The developed system contains an automatic segmentation parameter tuning module. A user provides the system with a delineation of some of the features of interest, which the system uses to automatically generate suitable segmentation algorithm parameters and subsequently image segments.

An edge metric, titled the pixel correspondence metric, was introduced as a fitness function to an evolutionary search process to find optimal segmentation algorithm parameters for a region merging segmentation algorithm. The metric was evaluated as a fitness function by segmenting two different types of scenes. One scene contained numerous sharp edges (agricultural landscape), while the second scene depicted a natural environment. Segments

generated by the segmentation algorithm were used as the desirable ground truth to ensure that the ground truth can be reached by the algorithm. Two search heuristics, namely differential evolution and particle swarm optimisation were employed to search the parameter space. In all cases the algorithms performed sufficiently with the pixel correspondence metric as fitness function but failed to find exact matches to the ground truth segments. This is due to the small numerical change in parameter values that results in vastly different segmentations in areas predominated by weak edges. This phenomenon is specific to the region merging algorithm used. The search heuristics could not find these exact parameters, although in practise it would possibly not be necessary. Even so, strong edges were always segmented correctly. The self adapting differential evolution strategy performed slightly better than the particle swarm optimisation strategy.

The pixel correspondence metric (PCM) has the advantage that closed polygons are not necessary to delineate the features of interest, but suffers from a lack of object recognition. This may lead to possible under segmentation. The PCM was compared with an area based metric, namely the larger segments booster (LSB), for segmenting Cape thicket. The LSB created much finer segments, while the PCM delivered segments closer to the natural scale of the features of interest. The PCM suffered from a very shallow search surface, resulting in a greater range (in a numerical sense) of suitable segmentation algorithm parameters.

Subsequent classification showed that the segments generated by the LSB resulted in inferior classifier accuracy, suggesting that the edge based metric may be an alternative for automating segment algorithm parameters on features that the segmentation algorithm finds difficult to segment, such as natural heterogeneous surface types. It is noted that such a segmentation algorithm free parameter tuning system is still dependent on the intrinsic ability of the segmentation algorithm to be able to segment the features of interest.

The input bands used for the segmentation process also influence the ability of the segmentation algorithm to accurately segment the features of interest. To address this problem, a novel band selection tool was implemented. This tool explores the spectral separability using the Jeffries-Matusita distance of the neighbouring pixels on both sides of the features of interest. Bands that display a higher spectral separability are deemed to be more suited as input for segmentation. The generalisability and effectiveness of this tool were not investigated quantitatively and this is recommended as an avenue for future research.

8.2.2 Segment classification in the context of geographic object anomaly detection

The classification component embedded in the developed system consists of a one class SVM classifier optimised with a search heuristic. The one class SVM uses a radial basis function kernel. The classifier meta-parameter space is two dimensional. A new strategy for simultaneous free parameter tuning and feature subset selection was adapted from the literature. This combined continuous and combinatorial optimisation strategy, based on the particle swarm optimiser was used to optimise the classifier with good gamma and nu values as well as a good attribute subset selection. The geometric means accuracy measure was used to evaluate the accuracies.

The system was quantitatively profiled under different numerical constraints for typical object based earth observation feature identification problems. The utility of such a system performing attribute subset selection and free parameter tuning simultaneously compared to free parameter tuning only was evaluated. In particular, the statistical significance of allowing the system to discard attributes was investigated. It was found that the search algorithm consistently delivers statistically significant results when allowed to discard attributes. The system was also profiled under different sample size conditions and different sizes of used attribute sets. It was found that the selection of adequate attributes has a significant influence on the resulting classifier accuracy for earth observation domain specific data.

The system shows promise for use in the context of object based earth observation feature identification and is especially useful in situations where uncertainty exists about the usefulness of attributes for a specific application. Under some conditions, very small attribute sets were used, and due to the stochastic nature of the search algorithm, a very small decrease in classifier accuracy was observed.

The developed system allows for 96 attributes to be used, consisting of basic object information such as object mean, object standard deviation and band ratio values. In addition, a few second order texture measures were made available. These object properties were calculated on the basic image input bands in addition to a few basic spectral transforms. Currently, the implemented attribute set does not contain more elaborate context information

such as elevation derived information, segment neighbour properties or more elaborate texture measures.

Although the implemented free parameter tuning and attribute subset selection system provided some improvement in classifier accuracy, results are still dependent on the discriminative capabilities of the data. A case study was performed to map a specific vegetation structural type as a first step to differentiating a vegetation subtype. A KNN classifier was compared to the classification methodology implemented in the developed system. The one class SVM based classification strategy performed better, although both classifiers delivered unsatisfactory results due to the discriminative capabilities of the data. Results can be improved by means of manual or automatic post-classification classification processes (using the Geo-ND classification as the basis for further refinement).

8.3 RECOMMENDATIONS FOR FUTURE RESEARCH

The system presented in this thesis, along with its specific components, can be improved in many ways. Suggestions are presented on how the system can be refined to obtain higher classifier accuracies or compute faster through straightforward extensions. Possible research avenues inspired by this work for geographic object based anomaly detection are also proposed.

8.3.1 Proposed extensions or modifications to the system

The 96 segment attributes formulated for the initial implementation of the system are limited in the object properties they address. As a straightforward improvement of the classification approach, it is suggested that more elaborate attributes are formulated and tested. Specific features of interest display different discriminative characteristics at different scales of observation/segmentation. At a finer scale, neighbourhood context attributes might be more useful, and at a larger scale within segment texture measures might hold more discriminative power.

Computational intelligence is rapidly progressing field, with new algorithms and heuristics constantly being proposed that improve older variants. In this work a standard self adapting differential evolution search algorithm and standard particle swarm optimisation algorithm were used. More advanced search strategies can be implemented that can potentially increase

the obtained segmentation and classification accuracies and/or reduce the computing time. Strategies such as quantum inspired search (Melo et al. 2008) or any advanced variant of particle swarm optimisation may yield better results. In addition, in this work the meta-parameters for these algorithms were manually set. Automated meta-optimisation, though computationally very expensive, may further improve results (Pedersen 2010).

The generalisation capabilities of statistical classifiers are a well known concern (Boardman & Trappenberg 2006). The optimal parameter and/or attribute combinations of a classifier may not generalise optimally due to the proximity of such points to, geometrically speaking, areas of very poor results (recall Figure 6.4). A generalisation heuristic, such as proposed by Boardman & Trappenberg (2006) may alleviate this problem and potentially increase the generalisation capability of the classifier.

Currently, the developed system can accommodate imagery of up to 3000 x 3000 pixels in size. This limitation is forced due to the extreme computational requirements of the implemented search heuristics, kernel based classifier, the image segmentation algorithm and segment statistic generation processes. Some of the experiments in this work computed for several hours on 30 standard desktop computers. Desktop parallel computing is becoming more commonplace due to the proliferation of low cost, mass produced general purpose graphics processing units.

A literature search revealed that all the computationally intensive components used or developed in this work, excluding the specific region merging segmentation algorithm, have freely available (open source) parallel implementations developed for general purpose graphics processing units. At the time of writing this thesis, speed increases quoted were usually in the 20 to 30 times range compared to standard desktop central processing units (pre-Fermi architectures). Running the computationally intensive components of the developed system in a parallel framework would not only result in much faster executions, but also indirectly allow for the exploration of techniques with larger imagery, bigger and more complex attribute sets and much larger (thus more accurate) training data for the automatic segmentation algorithm parameter tuning component.

Finally, it should be noted that the system (Geo-ND) is a prototype, meaning that excessive detail on the components of the system is presented in the implemented application. In

addition, the system is not tested on all types of data or for all eventualities. Creating a more streamlined and polished application would be possible.

8.3.2 Avenues for future research

In this work a basic area and edge metric were investigated in the segmentation algorithm parameter tuning process. More elaborate segment metrics exist that can be evaluated as fitness functions. Metrics that combine different notions of segment quality such as the metric proposed by Polak, Zhang & Pi (2009) might result in faster convergence times and/or more accurate matching to the ground truth.

Evaluating fitness functions and thus metrics based on the resultant classifier accuracies rather than the perceived quality of the segments they generate is an interesting research perspective that has not been fully explored. Gao et al. (2007) present an initial investigation addressing this notion. They defined a segment evaluation function (based on variance and autocorrelation measures) that correlates with obtained classifier accuracies.

The type of segment attributes used in such an investigation should also be considered carefully as classification performance may change as the nature of the generated segments change and so the usefulness of certain attributes. An investigation considering attribute utility through the scale space could possibly build on topics such as the effect of the scale of observation on classifier accuracy (Wulder et al. 2008) and the integration strategies of segmentation and classification as discussed earlier.

The band selection tool presented as part of the classification methodology was a small qualitative experiment in automatic input band selection. An apparent problem faced by such a system that attempts to use spectral information in and around the features of interest for band selection, is determining the extent of the area that should be investigated. A possible solution could be to use principles of cost distance flooding to restrict the extent of areas investigated from the edges of the features of interest. A more thorough investigation of the band selection tool working in conjunction with the segmentation parameter tuning component is called for. The selection of adequate bands for segmentation may be encoded (e.g. as a transformation matrix) into the search heuristic exploring the segmentation algorithm parameters.

In this work only spectral and some textural attributes were used and investigated. It should be of great interest to see some new perspectives on the use of more elaborate contextual data, cellular automata modelling strategies or contextual grammars (such as SPARK, discussed in Chapter 2) operating on segments rather than pixels. Techniques such as these hold value for earth observation image anomaly detection, especially in a Mediterranean landscape where spectral properties alone are rarely sufficient to identify vegetation land cover elements.

REFERENCES

Addink EA, De Jong SM & Pebesma EJ 2007. The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. *Photogrammetric Engineering and Remote Sensing* 73: 905-912.

Al-Ani A 2009. A dependency-based search strategy for feature selection. *Expert Systems with Applications* 36: 12392-12398.

Andrew ME & Ustin SL 2008. The role of environmental context in mapping invasive plants with hyperspectral image data. *Remote Sensing of Environment* 112: 4301-4317.

Atkinson PM 2004. Resolution manipulation and sub-pixel mapping. In De Jong SM & Van der Meer FD (eds) *Remote sensing and digital image processing*, 51-70. Dordrecht: Kluwer Academic Publishers.

Atkinson PM & Aplin P 2004. Spatial variation in land cover and choice of spatial resolution for remote sensing. *International Journal of Remote Sensing* 25: 3687-3702.

Baatz M, Hofmann P & Willhauck G 2008. Progressing from object-based to object-oriented image analysis. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 29-42. Berlin: Springer.

Baatz M & Schäpe A 1999. Object-oriented and multi-scale image analysis in semantic networks. Paper presented at the 2nd international symposium: Operationalization of remote sensing held 16 - 20 August 1999. Enschede, The Netherlands.

Baatz M & Schäpe A 2000. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung* 12: 12-23.

Banerjee A, Burlina P & Diehl C 2006. A support vector method for anomaly detection in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 44: 2282-2291.

Barlow J, Martin Y & Franklin SE 2003. Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia. *Canadian Journal of Remote Sensing* 29: 510-517.

Bennett KP & Campbell C 2000. Support vector machines: Hype or hallelujah? *SIGKDD Explorations* 2: 1-13.

Benz UC, Hofmann P, WillHauck G, Lingenfelder I & Heynen M 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing* 58: 239-258.

Bhanu B, Lee S & Ming J 1995. Adaptive image segmentation using a genetic algorithm. *IEEE Transactions Systems, Man, and Cybernetics* 25: 1543-1567.

Bhattacharyya A 1943. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematics Society* 35: 99-110.

Bins LS, Fonseca LMG, Erthal GJ & Mitsuo F 1996. Satellite imagery segmentation: A region growing approach. Paper presented at the VIII Brazillian symposium on remote sensing held 14-19 April 1996. Salvador, Brasil.

Blaschke T 2010. Object-based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2-16.

Blaschke T & Lang S 2006. Object based image analysis for automated information extraction - A synthesis. Paper presented at the MAPPS/ASPRS Speciality Conference held 6-10 November 2006. San Antonio, United States of America.

Blaschke T, Lang S & Hay GJ 2008. Preface. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, V-VII. Berlin: Springer.

Boardman M & Trappenberg T 2006. A heuristic for free parameter optimization with support vector machines. Paper presented at the IEEE international joint conference on neural networks held 16-21 July 2006. Vancouver, Canada.

Bock M, Xofis P, Mitchley J, Rossner G & Wissen M 2005. Object-oriented methods for habitat mapping at multiple scales: Case studies from Northern Germany and Wye Downs, UK. *Journal of Nature Conservation* 13: 75-89.

Brest J, Boskovic B, Greiner S, Zumer V & Maucec MS 2007. Performance comparison of self-adaptive and adaptive differential evolution algorithms. *Soft Computing - A Fusion of Foundations, Methodologies and Applications* 11: 617-629.

Brest J, Greiner S, Boskovic B, Mernik M & Zumer V 2006. Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *IEEE Transactions on Evolutionary Computation* 10: 646-657.

Brest J & Maucec MS 2008. Population size reduction for the differential evolution algorithm. *Applied Intelligence* 29: 228-247.

Burnett CN & Blaschke T 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling* 168: 233-249.

Busin L, Shi J, Vandenbroucke N & Macaire L 2009. Color space selection for color image segmentation by spectral clustering. Paper presented at the conference on spectral and fuzzy clustering techniques: Application to signal and image segmentation held 18-19 November 2009. Kuala Lumpur, Malaysia.

Campbell BM 1985. A classification of the mountain vegetation of the Fynbos Biome. *Memoirs of the Botanical Survey of South Africa* 50: 1-121.

Campbell JB 2002. *Introduction to remote sensing*. 3rd ed. London: Taylor & Francis.

Camps-Valls G 2005. Kernel methods in remote sensing: Introduction, applications and research opportunities [online]. Universitat de València. Available from: http://www.uv.es/gcamps/presentations/presentation_tubingen05.pdf. [Accessed 23 August 2009].

Camps-Valls G 2008. Kernel classifiers in remote sensing tutorial. Academic course material. Valencia: Department of Electronics Engineering, Universitat de Valencia.

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

Camps-Valls G, Gomez-Chova L, Muñoz-Marí J, Rojo-Alvarez JL & Martinez-Ramon M 2007. Kernel-based framework for multi-temporal and multi-source remote sensing data classification and change detection. *IEEE Transactions on Geoscience and Remote Sensing* 46: 1822-1835.

Camps-Valls G, Gomez-Chova L, Muñoz-Marí J, Vila-Francés J & Calpe-Maravilla J 2006. Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* 3: 93-97.

Cardoso JS & Corte-Real L 2005. Toward a generic evaluation of image segmentation. *IEEE Transactions on Image Processing* 14: 1773-1782.

Carleer AP, Debeir O & Wolff E 2005. Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing* 71: 1285-1294.

Carr J & Langhammer P 2007. Biodiversity hotspots: Cape Floristic Region - Human impacts [online]. Conservation International. Available from: http://www.biodiversityhotspots.org/xp/Hotspots/cape_floristic/Pages/impacts.aspx. [Accessed 4 March 2010].

Castilla G 2003. Object-oriented analysis of remote sensing images for land cover mapping: Conceptual foundations and a segmentation method to derive a baseline partition for classification. Doctoral dissertation. Madrid: Polythenic University of Madrid, Department of Economics and Management.

Castilla G & Hay GJ 2008. Image objects and geographic objects. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 91-110. Berlin: Springer.

Castilla G, Hay GJ & Ruiz JR 2008. Size-constrained region merging (SCRM): An automated delineation tool for assisted photointerpretation. *Photogrammetric Engineering and Remote Sensing* 74: 409-419.

Cazes TB, Costa GAOP, Feitosa RQ, Fredrich CMB, Gobbi HVS & Palmero BC 2008. Segmentation parameters tuner user guide [online]. PUC-Rio: Computer Vison Lab. Available from: http://www.lvc.ele.puc-rio.br. [Accessed 20 May 2009].

Chabrier S, Emile B, Rosenberger C & Laurent H 2006. Unsupervised performance evaluation of image segmentation. *EURASIP Journal on Applied Signal Processing* Volume 2006: 1-12.

Chen D, Stow DA & Gong P 2004. Examining the effect of spatial resolution and texture window size on classification accuracy: An urban environment case. *International Journal of Remote Sensing* 25: 2177-2192.

Chen Y, Su W, Li J & Sun Z 2009. Hierarchical object-oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research* 43: 1101-1110.

Cheng HD, Jiang XH, Sun Y & Wang J 2001. Color image segmentation: Advances and prospects. *Pattern Recognition* 34: 2259-2281.

Corcoran P & Winstanley A 2007. Segmentation evaluation for object-based remotely sensed image analysis. Paper presented at the geographical information science research UK conference held 11-13 April 2007. Maynooth, Ireland.

Corcoran P & Winstanley A 2008. Using texture to tackle the problem of scale in land-cover classification. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 113-132. Berlin: Springer.

Cowling RM & Holmes PM 1992. Flora and vegetation. In Cowling RM (ed) *The ecology of Fynbos: Nutrients, fire and diversity*, 23-61. Cape Town: Oxford University Press.

De Jong SM, Van der Meer FD & Clevers JGPW 2005. Basics of remote sensing. In De Jong SM & Van der Meer FD (eds) *Remote sensing image analysis: Including the spatial domain*, 1-15. Dordrecht: Kluwer Academic Publishers.

De Kok R 2006. Sequential processing of image objects and its consequences for automatic analysis. Paper presented at the 1st international conference on object-based image analysis held 4-5 July 2006. Salzburg, Austria.

De Kok R, Schneider T & Ammer U 1999. Object-based classification and applications in the alpine forest environment. Paper presented at the joint ISPRS/EARSel workshop held 3-4 June 1999. Valladolid, Spain.

De Kok R & Wezyk P 2008. Principles of full autonomy in image interpretation. The basic architectural design for a sequential process with image objects. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 697-710. Berlin: Springer.

Deacon HJ, Jury MR & Ellis F 1992. Selective regime and time. In Cowling RM (ed) *The ecology of Fynbos: Nutrients, fire and diversity*, 6-22. Cape Town: Oxford University Press.

Del Valle Y, Venayagamoorthy GK, Mohagheghi S, Hernandez J & Harley RG 2008. Particle swarm optimization: Basic concepts, variants and applications in power systems. *IEEE Transactions on Evolutionary Computation* 12: 171-195.

Dorren LKA, Maier B & Seijmonsbergen AC 2003. Improved landsat-based forest mapping in steep mountainous terrain using object-based classification. *Forest Ecology and Management* 183: 31-46.

Duda RO, Hart PE & Stork DG 2000. *Pattern Classification*. 2nd ed. New York: Wiley.

Eberhart RC 1997. A discrete binary version of the particle swarm algorithm. Paper presented at the IEEE international conference on intelligent systems, man and cybernetics held 12-15 October 1997. Piscatawary, United States of America.

Eiben AE & Smith JE 2003. *Introduction to evolutionary computing*. Berlin: Springer.

Espindola GM, Camara G, Reis IA, Bins LS & Monteiro AM 2006. Parameter selection for region-growing image segmentation algorithms using autocorrelation. *International Journal of Remote Sensing* 27: 3035-3040.

Falcao AX, Udupa JK, Samarasekera S, Sharma S, Hirsch BE & Lotufo RA 1998. User-steered image segmentation paradigms: Live wire and live lane. *Graphical Models and Image Processing* 60: 233-260.

Fan J, Zeng G, Body M & Hacid M 2005. Seeded region growing: An extensive and comparative study. *Pattern Recognition Letters* 28: 1139-1156.

Farmer ME & Shugars D 2006. Application of genetic algorithms for wrapper-based image segmentation and classification. Paper presented at the IEEE congress on evolutionary computation held 16-21 July 2006. Vancouver, Canada.

Fassnacht KS, Cohen WB & Spies TA 2006. Key issues in making and using satellite-based maps in ecology: A primer. *Forest Ecology and Management* 222: 167-181.

Feitosa RQ, Costa G, Fredrich CMB, Camargo FF & de Almeida CM 2009. Uma avaliacao de metodos geneticos para ajuste de parametros de segmentacao. Paper presented at the Brazillian simposium of remote sensing held 25-30 April 2009. Natal, Brazil.

Feitosa RQ, Costa GAOP, Cazes TB & Feijo B 2006. A genetic approach for the automatic adaptation of segmentation parameters. Paper presented at the 1st international conference on object-based image analysis held 4-5 July 2006. Salzburg, Austria.

Feitosa RQ, Ferreira RS, Almeida CM, Camargo FF & Costa G 2010. Similarity metrics for genetic adaptation of segmentation parameters. Paper presented at the Geographic Object-Based Image Analysis (GEOBIA) 2010 conference held 29 June - 2 July 2010. Ghent, Belgium.

Foody GM, Mather PM, Sanchez-Hernandez C & Boyd DS 2006. Training set size requirements for the classification of a specific class. *Remote Sensing of Environment* 104: 1-14.

Foody GM & Mathur A 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42: 1335-1343.

Foody GM & Mathur A 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment* 103: 179-189.

Forster M & Kleinschmit B 2008. Object-based classification of QuickBird data using ancillary information for the detection of forest types and NATURA 2000 habitats. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 275-290. Berlin: Springer.

Franco-Lopez H, Ek AR & Bauer ME 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment* 77: 251-274.

Franklin SE, Wulder MA & Lavigne MB 1996. Automated derivation of geographic window sizes for use in remote sensing digital image texture analysis. *Computers and Geosciences* 22: 665-673.

Fredrich CMB & Feitosa RQ 2008. Automatic adaptation of segmentation parameters applied to inhomogeneous objects detection. Paper presented at the conference on geographic object-based image analysis held 6-7 August 2008. Calgary, Canada.

Freixenet J, Munoz X, Raba D, Marti J & Cufi X 2002. Yet another survey on image segmentation: Region and boundary information integration. In Goos G, Hartmanis J & Van Leeuwen J (eds) *Lecture notes in computer science*, 408-422. Berlin: Springer.

Fröhlich H, Chapelle O & Schölkopf B 2003. Feature selection for support vector machines by means of genetic algorithms. Paper presented at the 15th IEEE international conference on tools with artificial intelligence held 3-5 November 2003. Sacramento, United States of America.

Gao Y, Kerle N, Mas JF, Navarrete A & Niemeyer I 2007. Optimized image segmentation and its effect on classification accuracy. Paper presented at the 5th international symposium on spatial data quality held 13-15 June 2007. Enschede, The Netherlands.

Goetz SJ, Wright RK, Smith AJ, Zinecker E & Schaub E 2003. IKONOS imagery for resource management: Tree cover, impervious surfaces, and riparian buffer analyses in the mid-Atlantic region. *Remote Sensing of Environment* 88: 195-208.

Goldman L, Adamek T, Vajda P, Karaman M, Morzinger R, Galmar E, Sikora E, O'Connor NE, Ha-Minh T, Ebrahimi T, Schallauer P & Huet B 2008. Towards fully automatic image segmentation evaluation. In Blanc-Talon J, Philips W, Popescu D & Scheuners P (eds) *Advanced concepts for intelligent vision systems*, 566-577. Berlin: Springer.

Gómez-Chova L, Camps-Valls G, Muñoz-Marí J & Calpe J 2008. Semisupervised image classification with laplacian support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 5: 336-340.

Govender M, Chetty K, Naiken V & Bulcock H 2008. A comparison of satellite hyperspectral and multispectral remote sensing imagery for improved classification and mapping of vegetation. *Water SA* 34: 147-154.

Guo Q, Du G, Liu Y & Liu D 2008. Integrating object-based classification with one-class support vector machines in mapping a specific land class from high spatial resolution images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37: 1159-1164.

Guyon I, Gunn S, Nikravesh M & Zadeh LA 2006. *Feature extraction: Foundations and applications*. Warsaw: Springer.

Haapen R, Ek AR, Bauer ME & Finley AO 2004. Delineation of forest/non-forest land use classes using nearest neighbor methods. *Remote Sensing of Environment* 89: 265-271.

Hall O, Hay GJ & Marceau DJ 2004. Multiscale object-specific analysis: Scale problems and multiscale solutions. Paper presented at the 12th international conference on geoinformatics - geospatial information research: Bridging the pacific and atlantic held 7-9 June 2004. University of Gävle, Sweden.

Haralick RM, Shanmugam K & Dinstein I 1973. Textural features for image classification. *IEEE Transactions on Systems Man and Cybernetics* 3: 610-621.

Hay G & Marceau DJ 2004. Multiscale object-specific analysis (MOSA): An integrative approach for multiscale landscape analysis. In De Jong SM & Van der Meer FD (eds) *Remote sensing image analysis: Including the spatial domain*, 71-91. Dordrecht: Kluwer Academic Publishers.

Hay GJ, Blaschke T, Marceau DJ & Bouchard A 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing* 57: 327-345.

Hay GJ & Castilla G 2008. Geographic object-based image analysis (GEOBIA): A new name for a new discipline. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 75-90. Berlin: Springer.

Hay GJ, Castilla G, Wulder MA & Ruiz JR 2005. An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation and Geoinformation* 7: 339-359.

Hay GJ, Marceau DJ, Dubé P & Bouchard A 2001. A multiscale framework for landscape analysis: Object-specific analysis and upscaling. *Landscape Ecology* 16: 471-490.

Hay GJ, Niemann KO & Goodenough DG 1997. Spatial thresholds, image-objects, and upscaling: A multiscale evaluation. *Remote Sensing of Environment* 62: 1-19.

Holland JH 1975. *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Horning N 2004. Remote sensing guides [online]. American Museum & Natural History. Available from: http://biodiversityinformatics.amnh.org/content.php?content_id=124. [Accessed 11 November 2009].

Hsu CW, Chang CC & Lin CJ 2008. A practical guide to support vector machines. Office report. Taipei: Department of Computer Science, National Taiwan University.

Hu X 2006. PSO Tutorial [online]. Available from: http://www.swarmintelligence.org/tutorials.php. [Accessed 8 September 2009].

Huang C, Davis LS & Townsend JRG 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* 23: 725-749.

Huang C & Dun J 2008. A distributed PSO-SVM hybrid system with feature selection and parameter optimization. *Applied Soft Computing* 8: 1381-1391.

Huang C & Wang C 2006. A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications* 31: 231-240.

Huang M, Gong J, Shi Z, Liu C & Zhang L 2007. Genetic algorithm-based decision tree classifier for remote sensing mapping with SPOT-5 data in the HongShiMao watershed of the loess plateau, China. *Neural Computing and Applications* 16: 513-517.

Jain AK, Duin RPW & Mao J 2000. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22: 4-37.

Jiang X, Marti C, Irniger C & Bunke H 2006. Distance measures for image segmentation evaluation. *EURASIP Journal on Applied Signal Processing* 2006: 1-10.

Johansen K, Nicholas CC, Gergel SE & Stange Y 2007. Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sensing of Environment* 110: 29-44.

Johansen K & Phinn S 2006. Mapping structural parameters and species composition of riparian vegetation using IKONOS and Landsat ETM+ data in Australian tropical savannahs. *Photogrammetric Engineering and Remote Sensing* 72: 71-80.

Ju J, Gopal S & Kolaczyk ED 2005. On the choice of spatial and categorical scale in remote sensing land cover classification. *Remote Sensing of Environment* 96: 62-77.

Kennedy J & Eberhart RC 1995. Particle swarm optimization. Paper presented at the IEEE international conference on neural networks held 26-28 June 1995. Cambridge, United Kingdom.

Kermad CD & Chehdi K 2002. Automatic image segmentation system through iterative edge–region co-operation. *Image and Vision Computing* 20: 541-555.

Ketting RL & Landgrebe DA 1976. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics* GE-14: 19-26.

Key T, Warner TA, McGraw JB & Fajvan MA 2001. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of

individual tree species in a temperate hardwood forest. *Remote Sensing of Environment* 75: 100-112.

Khushaba RN, Al-Ani A & Al-Jumaily A 2008. Differential evolution based feature subset selection. Paper presented at the international conference on pattern recognition (ICPR'08) held 8-11 December 2008. Tampa, United States of America.

Kim M & Madden M 2006. Determination of optimal scale parameters for alliance-level forest classification of multispectral IKONOS images. Paper presented at the 1st international conference on object-based image analysis held 4-5 July 2006. Salzburg, Austria.

Kim M, Xu B & Madden M 2006. Object-based vegetation type mapping from an orthorectified multispectral IKONOS image using ancillary information. Paper presented at the 1st international conference on object-based image analysis held 4-5 July 2006. Salzburg, Austria.

Kohavi R & John GH 1997. Wrappers for feature subset selection. *Artificial Intelligence* 97: 273-324.

Kotsiantis SB 2007. Supervised machine learning: A review of classification techniques. *Informatica* 31: 249-268.

Kubat M & Matwin S 1997. Addressing the curse of imbalanced training sets: One-sided selection. Paper presented at the 14th international conference on machine learning held 8-12 July 1997. Nashville, United States of America.

Kulkarni A, Jayaraman VK & Kulkarni BD 2004. Support vector classification with parameter tuning assisted by agent-based technique. *Computers and Chemical Engineering* 29: 311-318.

Laliberte AS & Rango A 2008. Incorporation of texture, intensity, hue, and saturation for rangeland monitoring with unmanned aircraft imagery. Paper presented at the 2nd international conference on geographic object-based image analysis held 5-7 August 2008. Calgary, Canada.

Laliberte AS & Rango A 2009. Texture and scale in object-based analysis of subdecimeter resolution unmanned aerial vehicle (UAV) imagery. *IEEE Transactions on Geoscience and Remote Sensing* 47: 761-770.

Laliberte AS, Rango A, Havstad KM, Paris JF, Beck RF, McNeely R & Gonzalez AL 2004. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of Environment* 93: 198-210.

Lamonaca A, Corona P & Barbati A 2008. Exploring forest structural complexity by multi-scale segmentation of VHR imagery. *Remote Sensing of Environment* 112: 2839-2849.

Lang S 2008. Object-based image analysis for remote sensing applications: Modeling reality – dealing with complexity. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 3-28. Berlin: Springer.

Lang S, Albrecht F & Blaschke T 2006. OBIA: Introduction to object-based image analysis v1.0 tutorial. Academic course material. Salzburg: University of Salzburg.

Lang S & Langanke T 2006. Object-based mapping and object-relationship modeling for land use classes and habitats. *PFG - Photogrammetrie, Fernerkundung* 6: 5-18.

Leckie DG, Gougeon FA, Tinis S, Nelson T, Burnett CN & Paradine D 2005. Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sensing of Environment* 94: 311-326.

Lee S, Soak S, Oh S, Pedrycs W & Jeon M 2008. Modified binary particle swarm optimization. *Progress in Natural Science* 18: 1161-1166.

Leprince S, Barbot S, Ayoub F & Avouac J 2007. Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. *IEEE Transactions on Geoscience and Remote Sensing* 45: 1529-1558.

Lewis RJ 2000. An introduction to classification and regression tree (CART) analysis. Office report. Torrance: Department of Emergency Medicine, Harbor-UCLA Medical Center.

Li P & Xiao X 2007. Multispectral image segmentation by a multichannel watershed-based approach. *International Journal of Remote Sensing* 28: 4429-4452.

Liang S 2004. *Quantitative remote sensing of land surfaces*. Hoboken: John Wiley & Sons.

Liao S 2005. Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Systems with Applications* 28: 93-103.

Lin S, Ying K, Chen S & Lee Z 2008. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications* 35: 1817-1824.

Lindenberg T 1994. *Scale-space theory in computer vision*. Dordrecht: Kluwer Academic Publishers.

Liu Z & Sun Z 2008. Active one-class classification of remote sensing image. Paper presented at the international conference on earth observation data processing and analysis held 28-30 December 2008. Wuhan, China.

LLoyd CD, Berberoglu S, Curran PJ & Atkinson PM 2004. A comparison of texture measures for the per-field classification of Mediterranean land cover. *International Journal of Remote Sensing* 25: 3943-3965.

Lorena AC & Carvalho ACPLF 2008. Evolutionary tuning of SVM parameter values in multiclass problems. *Neurocomputing* 71: 3326-3334.

Louw JH & Scholes M 2002. Forest site classification and evaluation: A South African perspective. *Forest Ecology and Management* 171: 153-168.

Lu D & Weng Q 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 28: 823-870.

Lübker T & Schaab G 2009. Optimization of parameter settings for multilevel image segmentation in GEOBIA. Paper presented at the high resolution earth imaging for geospatial information workshop held 2-5 June 2009. Hanover, Germany.

Lucas R, Bunting P, Paterson M & Chisholm L 2008. Classification of Australian forest communties using aerial photography, CASI and HyMap data. *Remote Sensing of Environment* 112: 2088-2103.

Lucas R, Rowlands A, Brown A, Keyworth S & Bunting P 2007. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* 62: 165-185.

Lucieer A 2004. Uncertainties in segmentation and their visualisation. Doctoral dissertation. Utrecht: Universiteit Utrecht, International Institute for Geo-Information Science and Earth Observation.

Lück W 2004. Remote sensing: A powerful but labour intensive tool for precision forestry. How can information extraction be automated? Office report. Stellenbosch: Centre for Geographical Analysis, Stellenbosch University.

Macaire L, Vandenbroucke N & Postaire J 2006. Color image segmentation by analysis of subset connectedness and color homogeneity properties. *Computer Vision and Image Understanding* 102: 105-116.

Mallinis G, Koutsias N, Tsakiri-Strati M & Karteris M 2008. Object-based classification using QuickBird imagery for delineating forest vegetation polygons in a Mediterranean test site. *ISPRS Journal of Photogrammetry and Remote Sensing* 63: 237-250.

Marcal ARS & Rodrigues AS 2008. A framework for the evaluation of multi-spectral image segmentation. Paper presented at the conference on geographic object-based image analysis held 6-7 August 2008. Calgary, Canada.

Markou M & Singh S 2003. Novelty detection: A review - Part 1: Statistical approaches. *Signal Processing* 83: 2481-2497.

Mather PM 2006. *Computer processing of remotely-sensed images: An introduction*. 3rd ed. Chichester: John Wiley & Sons.

Matinfar HR, Sarmadian F, Alavi Panah SK & Heck RJ 2007. Comparisons of object-oriented and pixel-based classification of land use/land cover types based on Landsat 7, ETM+ spectral bands (Case study: Arid region of Iran). *American-Eurasian Journal of Agricultural & Environmental Sciences* 2: 448-456.

McRoberts RE & Tomppo EO 2007. Remote sensing support for national forest inventories. *Remote Sensing of Environment* 110: 412-419.

Meinel G & Neubert M 2004. A comparison of segmentation programs for high resolution remote sensing data. *International Archives of the ISPRS* 35 (Part B, Commission 4): 1097-1105.

Melo LM, Costa GAOP, Feitosa RQ & Abs da Cruz AV 2008. Quantum-inspired evolutionary algorithm and differential evolution used in the adaptation of segmentation parameters. Paper presented at the conference on geographic object-based image analysis held 6-7 August 2008. Calgary, Canada.

Meyer D, Leisch F & Hornik K 2003. The support vector machine under test. *Neurocomputing* 55: 169-186.

Mishra SK 2006. Global optimization by differential evolution and particle swarm methods: Evaluation on some benchmark functions [online]. *Social Science Research Network and Parallel Computing* http://ssrn.com/abstract=933827.

Möller M, Lymburner L & Volk M 2007. The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation* 9: 311-321.

Mucina L & Geldenhuys CJ 2002. How to classify South African indigenous forests: Approach, methods, problems, perspectives. Paper presented at the 3rd natural forest & savanna woodlands symposium held 6-9 May 2002. Kruger National Park, South Africa.

Mucina L & Rutherford MC (eds) 2006. *The vegetation of South Africa, Lesotho and Swaziland*. Pretoria: South African National Biodiversity Institute.

Munkres J 1957. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics* 5: 32-38.

Muñoz-Marí J, Bruzzone L & Camps-Valls G 2007. A support vector domain description approach to supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 45: 2683-2692.

Muñoz-Marí J, Camps-Valls G, Gomez-Chova L & Calpe-Maravilla J 2007. Combination of one-class remote sensing image classifiers. Paper presented at the international geoscience and remote sensing symposium held 23-28 July 2007. Barcelona, Spain.

Munoz X, Freixenet J, Cufi X & Marti J 2003. Strategies for image segmentation combining region and boundary information. *Pattern Recognition Letters* 24: 375-392.

Navulur K 2007. *Multispectral image analysis using the object-oriented paradigm*. Boca Raton: CRC Press.

Neubert M, Herold H & Meinel G 2006. Evaluation of remote sensing image segmentation quality: Further results and concepts. *International Archives Photogrammatic Engineering and Remote Sensing* 36: 6-11.

Neubert M, Herold H & Meinel G 2008. Assessing image segmentation quality - concepts, methods and application. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image*

analysis: Spatial concepts for knowledge-driven remote sensing applications, 769-784. Berlin: Springer.

Newton AC 2007. *Forest ecology and conservation: A handbook of techniques*. Oxford: Oxford University Press.

Nussbaum S, Niemeyer I & Canty MJ 2006. SEaTH–A new tool for automated feature extraction in the context of object-based image analysis. Paper presented at the 1st international conference on object-based image analysis held 4-5 July 2006. Salzburg, Austria.

Oosthuizen S 2008. Variable selection for kernel methods with application to binary classification. Doctoral dissertation. Stellenbosch: Stellenbosch University, Department of Statistics.

Oruc M, Marangoz AM & Buyuksalih G 2004. Comparison of pixel-based and object-oriented classification approaches using Landsat-7 ETM spectral bands. Paper presented at the 20th ISPRS congress held 12-23 July 2004. Istanbul, Turkey.

Osman J, Inglada J & Christophe E 2009. Interactive object segmentation in high resolution satellite images. Paper presented at the IEEE international geoscience & remote sensing symposium held 12-17 July 2009. Cape Town, South Africa.

Ozdarici A & Turker M 2006. Field-based classification of agricultural crops using multi-scale images. Paper presented at the 1st international conference on geographic object-based image analysis held 4-5 July 2006. Salzburg, Austria.

Pacifici F, Frate FD, Solimini C & Emery WJ 2008. Neural networks for land cover applications. In Grana M & Duro RJ (eds) *Computational intelligence for remote sensing*, 267-294. Berlin: Springer.

Pal M & Foody GM 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing* 45: 2297-2307.

Pal M & Mather PM 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment* 86: 554-565.

Pal NR & Pal SK 1993. A review on image segmentation techniques. *Pattern Recognition* 26: 1277-1294.

Paterlini S & Krink T 2006. Differential evolution and particle swarm optimisation in partitional clustering. *Computational Statistics & Data Analysis* 50: 1220-1247.

Pavlidis T & Horowitz SL 1974. Segmentation of plane curves. *IEEE Transactions on Computers* 23: 860-870.

Pedersen MEH 2008. SwarmOps [online]. Available from: http://www.hvass-labs.org/projects/swarmops. [Accessed 2 May 2010].

Pedersen MEH 2010. Tuning & simplifying heuristical optimization. Doctoral dissertation. Southampton: University of Southampton, School of Engineering Sciences.

Perveen F, Nagasawa R & Husnain SA 2008. Evaluation of ASTER spectral bands for agricultural land cover mapping using pixel-based and object-based classification approaches. Paper presented at the 2nd international conference on geographic object-based image analysis held 5-7 August 2008. Calgary, Canada.

Polak M, Zhang H & Pi M 2009. An evaluation metric for image segmentation of multiple objects. *Image and Vision Computing* 27: 1223-1227.

Poli R, Kennedy J & Blackwell T 2007. Particle swarm optimization: An overview. *Swarm Intelligence* 1: 33-57.

Price KV, Storn RM & Lampinen JA 2005. *Differential evolution: A practical approach to global optimization*. Berlin: Springer.

Prieto MS & Allen AR 2003. A similarity metric for edge images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25: 1265-1273.

Qin AK, Huang VL & Suganthan PN 2009. Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Transactions on Evolutionary Computation* 13: 398-417.

Qt 2009.[online]. Nokia. Available from: http://qt.nokia.com/products. [Accessed 2 May 2010].

Radoux J & Defourny P 2006. Influence of image segmentation parameters on positional and spectral quality of the derived objects. Paper presented at the 1st internation conference on geographic object based image analysis held 4-5 July 2006. Salzburg, Austria.

Radoux J & Defourny P 2007. A quantitative assessment of boundaries in automated forest stand delineation using very high resolution imagery. *Remote Sensing of Environment* 110: 468-475.

Raskutti B & Kowalczyk A 2004. Extreme re-balancing for SVMs: A case study. *ACM SIGKDD Explorations Newsletter* 6: 60-69.

Richter R 2010. ATCOR-2/3 User Guide V7.1 [online]. ReSe Applications Schläpfer. Available from: http://www.rese.ch/download.html. [Accessed 14 April 2010].

Richter R, Kellenberger T & Kaufmann H 2009. Comparison of topographic correction methods. *Remote Sensing of Environment* 1: 184-196.

Roberts DA, Gardner M, Church R, Ustin S, Scheer G & Green RO 1998. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment* 65: 267-279.

Roberts LG 1965. Machine perception in three-dimensional solids. In Tippett JT (ed) *Optical and Electro-Optical Information Processing*, 159-197. Cambridge: MIT Press.

Rogan J, Franklin J, Stow DA, Miller J, Woodcock CE & Roberts D 2008. Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sensing of Environment* 112: 2272-2283.

Russo V 2009. libSVMplus [online]. GitHub. Available from: http://neminis.org/software/libsvm-plus. [Accessed 2 May 2010].

Sanchez-Hernandez C, Boyd DS & Foody GM 2007a. Mapping specific habitats from remotely sensed imagery: Support vector machine and support vector data description based classification of coastal saltmarsh habitats. *Ecological Informatics* 2: 83-88.

Sanchez-Hernandez C, Boyd DS & Foody GM 2007b. One-class classification for mapping a specific land-cover class: SVDD classification of fenland. *IEEE Transactions on Geoscience and Remote Sensing* 45: 1061-1073.

Schiewe J, Tufte L & Ehlers M 2001. Potential and problems of multi-scale segmentation methods in remote sensing. *GeoBIT/GIS* 6: 34-39.

Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ & Williamson RC 2001. Estimating the support of a high-dimensional distribution. *Neural Computation* 13: 1443-1471.

Schott JR 2007. *Remote sensing: The image chain approach*. 2nd ed. Oxford: Oxford University Press.

Schowengerdt RA 2007. *Remote sensing: Models and methods for image processing*. 3rd ed. Tucson: Academic Press.

Sesnie SE, Gessler PE, Finegan B & Thessler S 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment* 112: 2145-2159.

Shaban MA & Dikshit O 2001. Improvement of classification in urban areas by the use of textural features: The case study of Lucknow city, Uttar Pradesh. *International Journal of Remote Sensing* 22: 565-593.

Shawe-Taylor J & Cristianini N 2004. *Kernel methods for pattern analysis*. Cambridge: Cambridge University Press.

Sheperd JD & Dymond JR 2003. Correcting satellite imagery for the variance of reflectance and illumination with topography. *International Journal of Remote Sensing* 24: 3503-3514.

Shoshany M 2000. Satellite remote sensing of natural Mediterranean vegetation: A review within an ecological context. *Progress in Physical Geography* 24: 153-178.

Shrestha DP & Zinck JA 2001. Land use classification in mountainous areas: Integration of image processing, digital elevation data and field knowledge (application to Nepal). *International Journal of Applied Earth Observation and Geoinformation* 3: 78-85.

Skidmore A 2002. *Environmental modelling with GIS and remote sensing*. New York: Taylor & Francis.

Sluiter R 2005. Mediterranean land cover change. Doctoral dissertation. Utrecht: Universiteit Utrecht, Department of Physical Geography.

Sonka M, Hlavac V & Boyle R 2007. *Image processing, analysis and machine vision*. 3rd ed. Toronto: Thomson.

Storn R & Price K 1995. *Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces*. International Computer Science Institute. Berkeley.

Stuckens J, Coppin PR & Bauer ME 2000. Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sensing of Environment* 71: 282-296.

Taruvinga K 2008. Gully mapping using remote sensing: Case study in KwaZulu-Natal, South Africa. Masters thesis. Waterloo: University of Waterloo, Department of Geography.

Tax DMJ 2001. One-class classification. Doctoral dissertation. Delft: Delft University of Technology, Department of Mathematics and Computer Science.

Tax DMJ & Duin RPW 1999. Support vector domain description. *Pattern Recognition Letters* 20: 1191-1199.

Tax DMJ & Duin RPW 2004. Support vector data description. *Machine Learning* 54: 45-66.

Tax DMJ & Muller K 2004. A consistency-based model selection for one-class classification. Paper presented at the 17th international conference on pattern recognition held 23-26 August 2004. Cambridge, United Kingdom.

Terralib 2009.[online]. Brazilian National Institute for Space Research. Available from: http://www.dpi.inpe.br/terralib. [Accessed 2 May 2010].

Theodoridis S & Koutroumbas K 2006. *Pattern recognition*. 3rd ed. London: Academic Press.

Tomppo EO, Olsson H, Stahl G, Nilsson M, Hagner O & Katila M 2008. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment* 112: 1982-1999.

Toutin T 2004. Geometric processing of remote sensing images: Models, algorithms and methods. *International Journal of Remote Sensing* 25: 1893-1924.

Trias-Sanz R 2006. Semi-automatic rural land cover classification from high-resolution remote sensing images. Doctoral dissertation. Paris: Universite' Paris, Research Institute of Advanced Mathematics.

Trias-Sanz R, Stamon G & Louchet J 2008. Using colour, texture, and hierarchial segmentation for high-resolution remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 63: 156-168.

Tseng M, Chen S, Hwang G & Shen M 2008. A genetic algorithm rule-based approach for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 63: 202-212.

Tu C, Chuang L, Chang J & Yang C 2007. Feature selection using PSO-SVM. *International Journal of Computer Science* 33: 1-18.

Tzotsos A 2006. A support vector machine approach for object-based image analysis. Paper presented at the international conference on object-based image analysis held 4-5 July 2006. Salzburg, Austria.

Tzotsos A & Argialas D 2006. MSEG: A generic region-based multi-scale image segmentation algorithm for remote sensing imagery. Paper presented at the ASPRS 2006 annual conference held 1-6 May 2006. Reno, United States of America.

Tzotsos A & Argialas D 2008. Support vector machine classification for object-based image analysis. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 663-678. Berlin: Springer.

Tzotsos A, Losifidis C & Argialas D 2008. A hybrid texture-based and region-based multiscale image segmentation algorithm. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 221-236. Berlin: Springer.

Van Coillie FMB, Van Camp NAF, De Wulf RR, Bral L & Gautama S 2010. Segmentation quality evaluation for large scale mapping purposes in Flanders, Belgium. Paper presented at the conference on geographic object-based image analysis held 29 June - 2 July 2010. Utrecht, Netherlands.

Van Coillie FMB, Verbeke LPC & De Wulf RR 2007. Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sensing of Environment* 110: 476-487.

Van Niekerk A 2001. Western Cape digital elevation model: Product description. Office report. Stellenbosch: Centre for Geographical Analysis, Stellenbosch University.

Vandenbroucke N, Macaire L & Postaire J 2003. Color image segmentation by pixel classification in an adapted hybrid color space: Application to soccer image analysis. *Computer Vision and Image Understanding* 90: 190-216.

Vapnik VN 1995. *The nature of statistical learning theory*. New York: Springer.

Vapnik VN & Chapelle O 2000. Bounds on error expectation for support vector machines. *Neural Computation* 12: 2013-2036.

Vesterstroem J & Thomsen R 2004. A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. Paper presented at the IEEE conference on evolutionary computation held 20-23 June 2004. Portland, United States of America.

Vogiatzakis IN, Mannion AM & Griffiths GH 2006. Mediterranean ecosystems: Problems and tools for conservation. *Progress in Physical Geography* 30: 175-200.

Wang L, Sousa WP & Gong P 2004. Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International Journal of Remote Sensing* 25: 5655-5668.

Wang Z, Bovic AC & Lu L 2002. Why is image quality assessment so difficult? Paper presented at the IEEE international conference on acoustics, speech and signal processing held 13-17 May 2002. Orlando, USA.

Wannenburg A & Mabena S 2002. National indigenous forest inventory. Paper presented at the 3rd natural forest & savanna woodlands symposium held 6-9 May 2002. Kruger National Park, South Africa.

Weidner U 2008. Contribution to the assessment of segmentation quality for remote sensing applications. Paper presented at the International Society for Photogrammetry and Remote Sensing annual conference: Silk road for information from imagery held 3-11 July 2008. Beijing, China.

Weinke E, Lang S & Preiner M 2008. Strategies for semi-automated habitat delineation and spatial change assessment in an Alpine environment. In Blaschke T, Lang S & Hay GJ (eds) *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 711-732. Berlin: Springer.

Woodcock CE, Macomber SA & Kumar L 2002. Vegetation mapping and monitoring. In Skidmore A (ed) *Environmental modelling with GIS and remote sensing*, 97-120. New York: Taylor & Francis.

Woodcock CE, Strahler AH & Jupp DLB 1988. The use of variograms in remote sensing: Scene models and simulated images. *Remote Sensing of Environment* 25: 323-348.

Woods K 2007. Genetic algorithms: Colour image segmentation. Masters thesis. Cape Town: University of Cape Town, Department of Computer Science.

Wu J 2004. Effects of changing scale on landscape pattern analysis: Scaling relations. *Landscape Ecology* 19: 125-138.

Wu J & David JL 2002. A spatially explicit hierarchical approach to modeling complex ecological systems: Theory and applications. *Ecological Modelling* 153: 7-26.

Wu J & Loucks OL 1995. From the balance-of-nature to hierarchical patch dynamics: A theoretical framework shift in ecology. *The Quarterly Review of Biology* 70: 439-466.

Wuest B & Zhang Y 2009. Region based segmentation of QuickBird multispectral imagery through band ratios and fuzzy comparison. *ISPRS Journal of Photogrammetry and Remote Sensing* 64: 55-64.

Wulder MA, White JC, Hay GJ & Castilla G 2008. Towards automated segmentation of forest inventory polygons on high spatial resolution satellite imagery. *The Forestry Chronicle* 84: 221-230.

Wunderle AL, Franklin SE & Guo XG 2007. Regenerating boreal forest structure estimation using SPOT-5 pansharpened imagery. *International Journal of Remote Sensing* 28: 4351-4364.

Yu Q, Gong P, Clinton N, Biging G, Kelly M & Schirokauer D 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing* 72: 799-811.

Zhang HG, Fritts JE & Goldman SA 2008. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding* 110: 260-280.

Zhang Y 2002. Problems in the fusion of commercial high-resolution satellite images as well as Landsat 7 images and initial solutions. Paper presented at the symposium on geospatial theory, processing and applications, ISPRS held 8-12 July 2002. Ottawa, Canada.

Zhang Y & Maxwell T 2006. A fuzzy logic approach to supervised segmentation for object-oriented classification. Paper presented at the ASPRS 2006 annual conference held 1-5 May 2006. Reno, Nevada.

Zhang YJ 1996. A survey on evaluation methods for image segmentation. *Pattern Recognition* 29: 1335-1346.

Zhang YJ 1997. Evaluation and comparison of different segmentation algorithms. *Pattern Recognition Letters* 18: 963-974.

Zhang YJ 2001. A review of recent evaluation methods for image segmentation. Paper presented at the the international symposium on signal processing and its applications held 13-16 August 2001. Kaula Lumpur, Malaysia.

Zhuang L & Dai H 2006. Parameter optimization of kernel-based one-class classifier on imbalance learning. *Journal of Computers* 1: 32-40.

Zingaretti P, Tascini G & Regini L 2003. Optimizing the colour image segmentation. Paper presented at the 8th congress of the Italian association for artificial intelligence held 23-26 September 2003. Pisa, Italy.

APPENDICES

Appendix A: Case study ground truth photographs

These photographs aim to orientate the reader to the nature of the class of interest investigated in the case study presented in Chapter 7. The photographs were taken at some of the sites visited as part of a ground truthing exercise (and further afield) of the training samples used.

Figure A1: Cape thicket patches on the slopes of the Klein Drakenstein mountains, near Paarl

Figure A2: Riparian Afrotemperate forest in a kloof in the Hottentots Holland Mountains

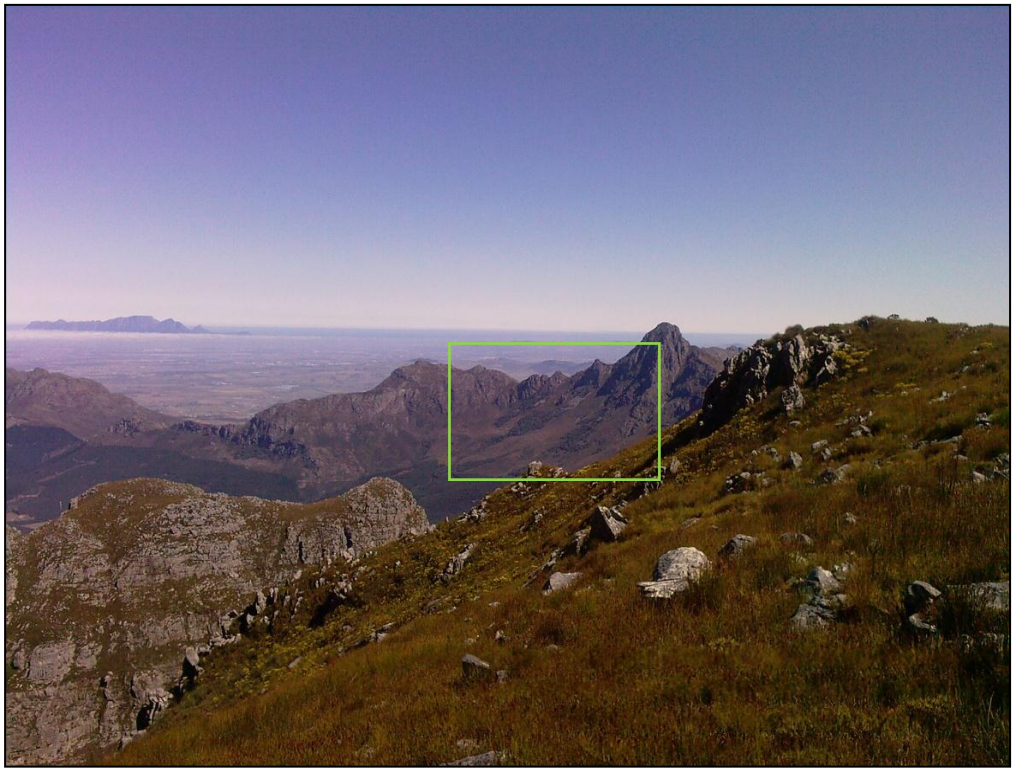


Figure A3: Cape thicket patches under the southern crest of the Helderberg, near Somerset West

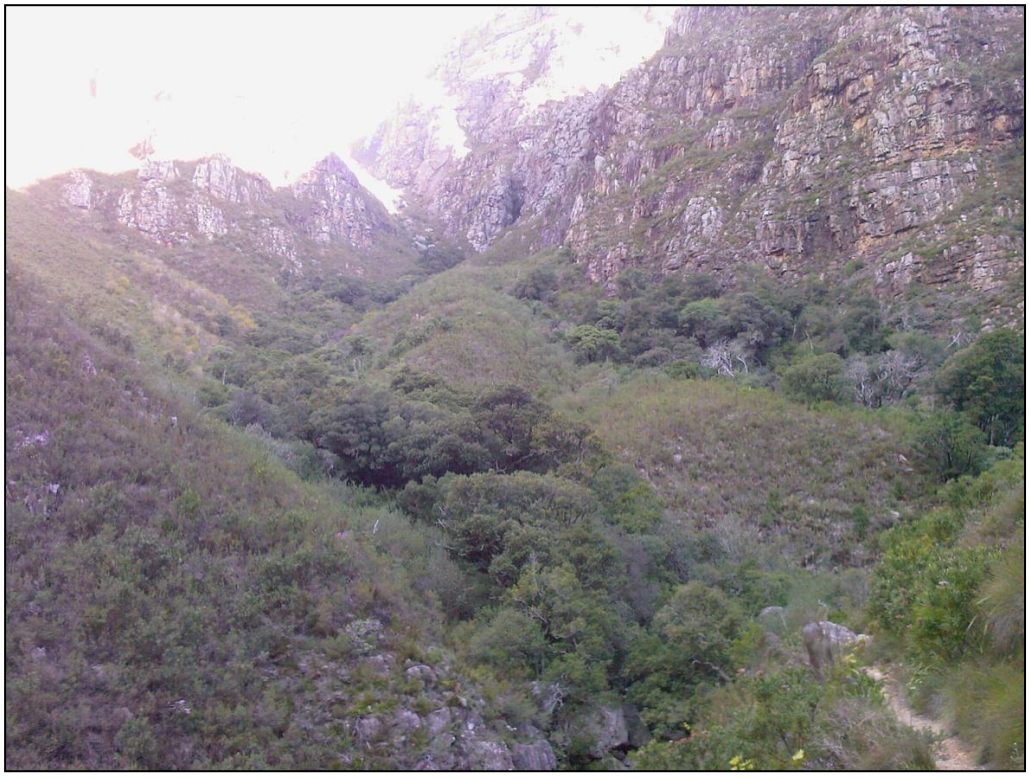


Figure A4: Riparian Afrotemperate forest/thicket in a kloof in the Jonkershoek Valley

Appendix B: Classification results with LSB and PCM generated segments

Tables B1 to B4 list the resulting gamma, nu, accuracy and attribute subset selection of the experiment conducted in Section 5.4. A concise summary of these results is presented in Section 5.4.

Table B1: Classification accuracies over 25 runs with the LSB metric generated segments. Seven fixed attributes are used.

	Gamma	Nu	Accuracy	M2	M4	M3/4	S3/4	HNDVI	CNDVI	CI	# attr
1	2.4283	0.1731	0.7879	1	1	1	1	1	1	1	7
2	8.5362	0.1863	0.7879	1	1	1	1	1	1	1	7
3	2.4173	0.1482	0.7904	1	1	1	1	1	1	1	7
4	3.3846	0.1537	0.7907	1	1	1	1	1	1	1	7
5	19.8296	0.2101	0.7798	1	1	1	1	1	1	1	7
6	8.9109	0.1734	0.7875	1	1	1	1	1	1	1	7
7	4.8625	0.1761	0.7919	1	1	1	1	1	1	1	7
8	2.7168	0.1524	0.7909	1	1	1	1	1	1	1	7
9	2.2689	0.1530	0.7903	1	1	1	1	1	1	1	7
10	5.9875	0.1773	0.7907	1	1	1	1	1	1	1	7
11	3.9844	0.1518	0.7903	1	1	1	1	1	1	1	7
12	3.3810	0.1666	0.7913	1	1	1	1	1	1	1	7
13	3.4211	0.1789	0.7897	1	1	1	1	1	1	1	7
14	3.0188	0.1594	0.7910	1	1	1	1	1	1	1	7
15	4.8723	0.1712	0.7918	1	1	1	1	1	1	1	7
16	4.5373	0.1778	0.7920	1	1	1	1	1	1	1	7
17	3.0163	0.1566	0.7912	1	1	1	1	1	1	1	7
18	3.8816	0.1522	0.7904	1	1	1	1	1	1	1	7
19	11.2381	0.2036	0.7851	1	1	1	1	1	1	1	7
20	3.9938	0.1611	0.7918	1	1	1	1	1	1	1	7
21	1.9471	0.1537	0.7903	1	1	1	1	1	1	1	7
22	4.7267	0.1621	0.7919	1	1	1	1	1	1	1	7
23	4.9059	0.1658	0.7918	1	1	1	1	1	1	1	7
24	4.4471	0.1841	0.7908	1	1	1	1	1	1	1	7
25	4.2897	0.1512	0.7902	1	1	1	1	1	1	1	7
Sum				**25**	**25**	**25**	**25**	**25**	**25**	**25**	
Average	**5.0802**	**0.1680**	**0.7899**								**7**
Standard deviation	**3.7790**	**0.0164**	**0.0027**								**0**

Note: M = mean; S = standard deviation; H = GLCM homogeneity; C = GLCM contrast; I = intensity; # attr = number of used attributes

Table B2: Classification accuracies over 25 runs with the LSB metric generated segments. Seven variable attributes are used.

	Gamma	Nu	Accuracy	M2	M4	M3/4	S3/4	HNDVI	CNDVI	CI	# attr
1	5.2110	0.1396	0.8008	1	1	1	0	1	1	0	5
2	6.0767	0.1667	0.7955	0	1	1	0	1	1	1	5
3	4.6084	0.1343	0.8002	1	1	1	0	1	1	0	5
4	20.4473	0.2082	0.7869	1	0	1	0	1	1	0	4
5	5.2453	0.1580	0.7987	1	1	1	0	1	0	1	5
6	2.6104	0.1465	0.7946	0	1	1	0	1	0	1	4
7	4.8730	0.1424	0.8005	1	1	1	0	1	0	0	4
8	41.6274	0.2062	0.7932	1	1	1	0	0	1	1	5
9	4.8036	0.1393	0.8003	1	1	1	0	1	0	0	4
10	5.1588	0.1389	0.8005	1	1	1	0	1	1	0	5
11	10.3884	0.1982	0.7864	1	0	1	0	1	1	1	5
12	4.5393	0.1411	0.8001	1	1	1	0	1	0	0	4
13	41.9983	0.2058	0.7931	1	1	1	0	0	1	1	5
14	50.0000	0.2489	0.7880	1	0	1	0	0	1	0	3
15	4.9084	0.1408	0.8002	1	1	1	0	1	0	0	4
16	46.6183	0.1975	0.7927	1	1	1	0	0	1	1	5
17	1.1733	0.1587	0.7895	1	1	1	1	1	0	0	5
18	41.6964	0.2054	0.7931	1	1	1	0	0	1	1	5
19	5.9421	0.1597	0.7989	1	1	1	0	1	1	1	6
20	4.6679	0.1387	0.8002	1	1	1	0	1	0	0	4
21	5.4407	0.1602	0.7995	1	1	1	0	1	1	1	6
22	27.2049	0.2207	0.7876	1	0	1	0	1	0	0	3
23	5.3082	0.1598	0.7994	1	1	1	0	1	1	1	6
24	5.4181	0.1397	0.8008	1	1	1	0	1	1	0	5
25	41.1122	0.2055	0.7932	1	1	1	0	0	1	1	5
Sum				**23**	**21**	**25**	**1**	**19**	**16**	**12**	
Average	**15.8831**	**0.1704**	**0.7958**								**4.68**
Standard deviation	**17.0066**	**0.0332**	**0.0050**								**0.80**

Note: M = mean; S = standard deviation; H = GLCM homogeneity; C = GLCM contrast; I = intensity; # attr = number of used attributes

Table B3: Classification accuracies over 25 runs with the PCM metric generated segments. Seven fixed attributes are used.

	Gamma	Nu	Accuracy	M2	M4	M3/4	S3/4	HNDVI	CNDVI	CI	# attr
1	4.1533	0.2392	0.8065	1	1	1	1	1	1	1	7
2	4.9372	0.2632	0.8038	1	1	1	1	1	1	1	7
3	5.2641	0.2586	0.8035	1	1	1	1	1	1	1	7
4	1.3698	0.2334	0.8008	1	1	1	1	1	1	1	7
5	4.8648	0.2513	0.8039	1	1	1	1	1	1	1	7
6	3.7470	0.2333	0.8074	1	1	1	1	1	1	1	7
7	0.0571	0.8093	0.7906	1	1	1	1	1	1	1	7
8	1.7243	0.2765	0.7966	1	1	1	1	1	1	1	7
9	2.4758	0.2300	0.8074	1	1	1	1	1	1	1	7
10	0.4358	0.2809	0.7973	1	1	1	1	1	1	1	7
11	1.4331	0.2389	0.7999	1	1	1	1	1	1	1	7
12	2.4783	0.2340	0.8078	1	1	1	1	1	1	1	7
13	1.7565	0.2356	0.8038	1	1	1	1	1	1	1	7
14	4.3985	0.2286	0.8072	1	1	1	1	1	1	1	7
15	3.8228	0.2285	0.8077	1	1	1	1	1	1	1	7
16	3.1755	0.2316	0.8085	1	1	1	1	1	1	1	7
17	4.9288	0.2641	0.8038	1	1	1	1	1	1	1	7
18	3.7321	0.2129	0.8042	1	1	1	1	1	1	1	7
19	4.4390	0.2667	0.8024	1	1	1	1	1	1	1	7
20	5.9712	0.2573	0.8027	1	1	1	1	1	1	1	7
21	5.2969	0.2552	0.8039	1	1	1	1	1	1	1	7
22	2.9655	0.2698	0.8010	1	1	1	1	1	1	1	7
23	3.0428	0.2550	0.8053	1	1	1	1	1	1	1	7
24	2.0958	0.2156	0.8014	1	1	1	1	1	1	1	7
25	3.7313	0.2351	0.8073	1	1	1	1	1	1	1	7
Sum				**25**	**25**	**25**	**25**	**25**	**25**	**25**	
Average	**3.2919**	**0.2682**	**0.8034**								**7**
Standard deviation	**1.5895**	**0.1142**	**0.0042**								**0**

Note: M = mean; S = standard deviation; H = GLCM homogeneity; C = GLCM contrast; I = intensity; # attr = number of used attributes

Table B4: Classification accuracies over 25 runs with the PCM metric generated segments. Seven variable attributes are used.

	Gamma	Nu	Accuracy	M2	M4	M3/4	S3/4	HNDVI	CNDVI	CI	# attr
1	2.3407	0.2700	0.8283	1	1	0	1	1	1	0	5
2	3.8697	0.2539	0.8337	1	1	0	0	1	1	0	4
3	1.6924	0.3115	0.8297	1	0	0	0	1	1	1	4
4	2.0739	0.2457	0.8312	1	1	0	0	1	1	1	5
5	2.0108	0.3278	0.8331	1	0	0	0	1	0	0	2
6	50.0000	0.2677	0.8214	1	0	0	0	1	1	0	3
7	2.1692	0.2426	0.8310	1	1	0	0	1	1	1	5
8	2.5367	0.2918	0.8366	1	0	0	0	1	1	0	3
9	0.7726	0.4237	0.8291	0	1	0	0	1	1	1	4
10	0.5257	0.4877	0.8347	1	1	0	0	1	1	1	5
11	2.8649	0.2905	0.8368	1	0	0	0	1	1	0	3
12	0.4535	0.5645	0.8342	1	1	0	0	1	1	0	4
13	2.8747	0.2894	0.8367	1	0	0	0	1	1	0	3
14	2.3926	0.2698	0.8247	1	1	0	1	1	0	0	4
15	2.5506	0.2577	0.8286	0	1	0	0	1	0	1	3
16	0.5740	0.4512	0.8330	1	1	0	0	1	0	1	4
17	3.1097	0.2887	0.8367	1	0	0	0	1	1	0	3
18	1.5315	0.2751	0.8274	1	1	0	1	1	1	1	6
19	2.2445	0.3076	0.8324	1	0	0	0	1	0	0	2
20	3.2909	0.2609	0.8279	1	0	0	0	1	0	1	3
21	2.4491	0.2480	0.8310	1	1	0	0	1	1	1	5
22	3.0203	0.2818	0.8360	1	0	0	0	1	1	0	3
23	27.9684	0.2586	0.8141	1	0	0	0	0	1	0	2
24	1.7416	0.3082	0.8297	1	0	0	0	1	1	1	4
25	2.8061	0.2523	0.8265	0	1	0	0	1	1	0	3
Sum				**22**	**13**	**0**	**3**	**24**	**19**	**11**	
Average	**5.1146**	**0.3091**	**0.8306**								**3.68**
Standard deviation	**10.7143**	**0.0829**	**0.0053**								**1.07**

Note: M = mean; S = standard deviation; H = GLCM homogeneity; C = GLCM contrast; I = intensity; # attr = number of used attributes

Appendix C: Geo-ND DVD

DVD containing the Geo-ND application, the Geo-ND user guide and sample data (see back cover).