DXN008540

98804318

1199695

UNIVERSITY OF SURREY LIBRARY

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Mixed Pixel Classification in Remote Sensing

Panagiota Bosdogianni

Submitted in
Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
of
Surrey University

Department of Electronic and Electrical Engineering
University of Surrey
Guildford, Surrey GU2 5XH, U.K.

July, 1996

To Costas

Abstract

This thesis is concerned with the problem of mixed pixel classification in Remote sensing applications and attempts to find accurate and robust solutions to this problem.

The application we are interested in, is to monitor burned forest regions for a few years after the fire in order to identify the type of vegetation present in these areas and consequently assess the danger of desertification. The areas of interest are semi-arid where the vegetation tends to vary at smaller scales than the area covered by a single Landsat TM pixel, thus mixed pixels are quite common. In this thesis we considered whole sets of mixed pixels.

First, an overview of the methods currently used to solve the mixed pixel classification problem is presented, focused on the linear mixing model which is adopted in this thesis. Then a method that incorporates higher order moments of the distributions of the pure and the mixed classes is proposed. This method is shown to augment the number of equations used for the classification and theoretically it allows the specification of more cover classes than there are bands available, without compromising the accuracy of the results.

The problem of deterioration of the classification performance, due to inaccuracies in calculation of the statistics when outliers are present, is also examined. The use of the Hough Transform is proposed for the linear unmixing in order to provide robust estimates even in cases where outliers are present. The Hough transform method though, is an exhaustive method and therefore has higher computational complexity. Furthermore, its performance, in the absence of outliers, is not as good as the solution obtained by the Least Squares Error method. Hence, the Randomized Hough Transform is proposed in order to improve the computational speed and maintain the same level of performance, while the Hypothesis Testing Hough Transform is proposed to improve the accuracy of the classification results.

All the methods proposed in this thesis have been compared with the Least Squares Error method using simulated and real Landsat TM image data, in order to illustrate the validity and usefulness of the proposed algorithms.

Acknowledgements

I would like to thank my supervisors **Dr. Maria Petrou** and **Prof. Josef Kittler** not only for providing the opportunity for this research but also for their support, advice and encouragement.

I would like to express my gratitude to all the members in the **Vision, Speech and Signal Processing Group** who have generously given their advice and help during the course of my research. Especially I would like to thank **Dr. Nikolaos Georgis** and **Dr. Petros Papachristou** for their constructive comments. Also I would like to thank **Dr. Heiki Kälviäinen** for his cooperation concerning the Randomized Hough Transform.

Furthermore, I would like to thank **Constantinos Boukouvalas** not only for his scientific advice but mainly for his emotional support and encouragement during this research. Also I would like to thank my parents for always supporting and believing in me.

Finally the financial support of the CEC project 0025 under the Environment programme is gratefully acknowledged.

Contents

1	Introduction	1
1.1	Motivation of this work and application area	2
1.2	Scope of this work	6
1.3	Outline of the Thesis	8
2	Methods used for mixed pixel classification	11
2.1	Introduction	11
2.2	Vegetation index approach	12
2.3	Spectral mixture analysis approach	13
2.3.1	Unmixing using training data	15
2.3.1.1	Least Squares Error (LSE) method	17
2.3.1.2	Maximum Likelihood Estimation (MLE) method	19
2.3.1.3	Comparison of the LSE and MLE methods	22
2.3.2	Unmixing without using prior knowledge	23
2.3.3	Geometric models	25
2.4	Problems with the classical mixture model	27
2.5	Visualization	30
2.6	Conclusions	32
3	Mixed pixel classification using higher order moments	33
3.1	Introduction	33

3.2	The proposed method	34
3.3	Testing of the method with simulated data	38
3.3.1	Variation of the size of the distributions	40
3.3.2	Testing different proportion combinations for the mixed class .	44
3.3.3	Performance in the presence of noise	48
3.4	What if pure class sites are not available?	50
3.5	Application to real data	52
3.6	Discussion and conclusions	57
4	Mixed pixel classification using Robust Statistics	60
4.1	Introduction	60
4.2	Brief overview of the Hough Transform	61
4.3	The proposed method	62
4.4	Testing of the methods with simulated data	66
4.4.1	Time and accuracy concerns	67
4.4.2	Normal distributions	68
4.4.2.1	Testing different proportion combinations for the mixed class	69
4.4.2.2	Presence of outliers	70
4.4.3	Uniform distributions	75
4.5	What if pure classes are not available?	76
4.6	Application to real data	80
4.7	Discussion and conclusions	81
5	Mixed pixel classification using the Randomized Hough Transform	83
5.1	Introduction	83
5.2	The Randomized Hough Transform (RHT)	84
5.3	Testing of the methods with simulated data	85
5.3.1	Time and accuracy concerns	86

5.3.2	Testing different proportion combinations for the mixed class .	90
5.3.3	Outliers	92
5.4	Application to laboratory data	94
5.5	Discussion and conclusions	98
6	Mixed pixel classification using the hypothesis testing Hough transform	99
6.1	Introduction	99
6.2	The proposed method	100
6.3	Testing of the method with simulated data	104
6.3.1	Time and Accuracy Concerns	105
6.3.2	Outliers	110
6.4	Application to real data	114
6.5	Conclusions	115
7	Discussion and conclusions	116
7.1	Major contributions of this thesis	116
	Bibliography	120

List of Figures

1.1	<i>Landsat TM image of Attica.</i>	3
1.2	<i>Test sites within the Pateras test area.</i>	4
1.3	<i>Test sites within the Barnabas test area.</i>	5
1.4	<i>Test sites within the Penteli test area.</i>	6
1.5	<i>Test sites within the Lavrio test area.</i>	7
2.1	<i>Ternary classification diagram.</i>	31
3.1	<i>A set of simulated distributions to represent the pure classes, projected onto the Band3 - Band4 plane.</i>	40
3.2	<i>Relative mean error of proportion estimation versus the size of the pure classes data sets. The mixed data sample size is 200 and the set of proportions tested is 0.1, 0.2, 0.3 and 0.4. One line represents the performance of a model that uses only the means and the other of a model that uses the means and the variances as well. The error bars represent the standard deviation of the errors for each method over 100 experiments performed for the same set of parameters but with different data sets. (a) Error for a=0.1, (b) Error for b=0.2, (c) Error for c=0.3, (d) Error for d=0.4. The top two curves follow the extreme errors that were observed in each set of 100 experiments, for each method.</i>	42

- 3.3 *Relative mean error of proportion estimation versus the size of the mixed data set. The pure class data sets were represented by 200 samples and the set of proportions tested is 0.1 0.2 0.3 0.4. The error bars represent the standard deviation of the errors for each method observed over 100 experiments. (a) Error for $a=0.1$, (b) Error for $b=0.2$, (c) Error for $c=0.3$, (d) Error for $d=0.4$. The top two curves follow the extreme errors that were observed in each set of 100 experiments, for each method.* 43
- 3.4 *Relative mean error of proportion estimation versus the size of the pure classes data sets. The mixed data sample size is 10,000 and the set of proportions tested is 0.1, 0.2, 0.3 and 0.4. One line represents the performance of a model that uses only the means and the other of a model that uses the means and the variances as well. The error bar represent the standard deviation of the errors for each method over 100 experiments performed for the same set of parameters but with different data sets. (a) Error for $a=0.1$, (b) Error for $b=0.2$, (c) Error for $c=0.3$, (d) Error for $d=0.4$. The top two curves follow the extreme errors that were observed in each set of 100 experiments, for each method.* 45
- 3.5 *Relative mean error of proportion estimation versus the size of the mixed data set. The pure class data sets were represented by 10,000 samples and the set of proportions tested is 0.1 0.2 0.3 0.4. The error bars represent the standard deviation of the errors for each method observed over 100 experiments. (a) Error for $a=0.1$, (b) Error for $b=0.2$, (c) Error for $c=0.3$, (d) Error for $d=0.4$. The top two curves follow the extreme errors that were observed in each set of 100 experiments, for each method.* 46

3.6	<i>Error in proportion estimation when different combinations of components a, b, c, d are examined.</i>	47
3.7	<i>Highest measured error for the elements of the mean vector for each site.</i>	54
3.8	<i>Derived means of the four pure classes. (a): 23 sites from Lavrio, Penteli and Varnavas. (b): 16 sites from Pateras.</i>	58
4.1	<i>CPU times estimated using an Alpha workstation versus the number of points representing each class.</i>	67
4.2	<i>Influence of the choice of datasets.</i>	68
4.3	<i>Some example sets used to represent the three "pure" classes.</i>	69
4.4	<i>Average maximum relative errors in estimating different proportions for TM, SHT and LSE.</i>	71
4.5	<i>Outliers placed at a given distance (3, 6, and 9 times the standard deviation) from the mean of the mixed set.</i>	72
4.6	<i>The mixed set contains points from two mixtures: half of the points belong to a mixture with composition 30-60-10 and the other half to a mixture with composition 30-10-60. The two clusters indicate the two mixtures.</i>	74
4.7	<i>The means of the pure classes derived from mixed sets of known composition when slightly erroneous mixing proportions for one of the sets were used in the derivation.</i>	79
5.1	<i>Some example sets used to represent the three "pure" classes.</i>	86
5.2	<i>CPU times estimated using a Sun SPARCstation 20 versus the number of points representing each class. The RHT allows for greater error (about 5% greater) than the SHT in this case.</i>	87

5.3	<i>Influence of the choice of datasets. The error bars indicate the standard deviation of the distribution of the observed errors with 100 different initializations of the random number generator, for fixed set of data.</i>	88
5.4	<i>Average relative error for proportion estimation, when we vary the number of iterations used. The error bars represent the standard deviation of the distribution of the observed errors with 100 different initializations of the random number generator, for fixed set of data. .</i>	89
5.5	<i>Influence of the choice of datasets when RHT is iterated 10 times. The error bars indicate the standard deviation of the distribution of the observed errors with 100 different initializations of the random number generator, for fixed set of data.</i>	90
5.6	<i>The area around the peak in the accumulator array. This is a part of the accumulator space with size 32×38</i>	91
5.7	<i>Average maximum relative errors in estimating different proportions for LSE, SHT and non-iterated RHT.</i>	92
5.8	<i>Average maximum relative errors in estimating different proportions for LSE, SHT and iterated RHT.</i>	93
5.9	<i>An image of the coloured chalks.</i>	95
5.10	<i>The two colour sets tested.</i>	96
5.11	<i>The three pure classes and the mixture with several random outliers. .</i>	97
6.1	<i>Some example sets used to represent the three "pure" classes.</i>	105
6.2	<i>Relative error in estimating a whose true value is 30% versus different sets of samples representing the same mixture.</i>	107
6.3	<i>Relative average error in proportion estimation when we vary the spread of the sets used to represent the pure classes.</i>	109

6.4	<i>Average maximum relative errors in estimating different proportions for HTHT, TM and LSE.</i>	110
6.5	<i>Effect of coherent outliers present in the mixed distribution. The strips represent the average errors plus or minus 1 standard deviation of these errors. The black dashed line is the mean error of the TM method minus one standard deviation.</i>	112
6.6	<i>Effect of random outliers present in the mixed distribution. The strips represent the average errors plus or minus 1 standard deviation of these errors. The black dashed line is the mean error of the TM method minus one standard deviation.</i>	113

List of Tables

3.1	<i>Estimated significance of the results shown in Figure 3.2.</i>	44
3.2	<i>Effect of noise when applied to the mixed distribution. The error at noise level 0% is the error due to subsampling.</i>	48
3.3	<i>Effect of biased error in the mixed distribution of 200 points. The numbers given are the errors with which the true proportions ($a = 10\%$, $b = 20\%$, $c = 30\%$ and $d = 40\%$) were estimated by the various methods.</i>	49
3.4	<i>Effect of biased error in a very well represented mixed distribution. The numbers given are the errors with which the true proportions ($a = 10\%$, $b = 20\%$, $c = 30\%$ and $d = 40\%$) were estimated by the various methods.</i>	50
3.5	<i>Comparison of the mixed model results with the ground data for Lavrio, Penteli and Varnavas training sites.</i>	55
3.6	<i>Comparison of the mixed model results to the ground data for Pateras.</i>	56
3.7	<i>Comparison of the mixed model results to the ground data for testing sites.</i>	57
4.1	<i>Effect of coherent outliers present in the mixed distribution. The numbers in the table are the average errors over 100 experiments conducted. Inside parentheses is the standard deviation of these errors.</i>	73

4.2	<i>Effect of random outliers present in the mixed distribution. The numbers in the table are the average relative errors over 100 experiments conducted, and inside parentheses is the standard deviation of these errors.</i>	73
4.3	<i>Effect of coherent outliers present in the mixed uniform distribution. The numbers in the table are the average errors over 100 experiments conducted. Inside parentheses is the standard deviation of these errors.</i>	75
4.4	<i>Effect of random outliers in the mixed uniform distribution. Mixed set composition $a = 30\%$, $b = 60\%$, $c = 10\%$.</i>	75
4.5	<i>Classification results obtained using TM, SHT and LSE when the end-members were derived with different methods.</i>	80
5.1	<i>Effect of coherent outliers present in in the mixed distribution. The numbers in the table are the average errors over 100 experiments conducted. Inside parentheses is the variance of these errors. RHT results were obtained with the iterating version of the algorithm. . . .</i>	94
5.2	<i>Effect of random outliers present in the mixed distribution. The numbers in the table are the average relative errors over 100 experiments conducted and inside parentheses is the variance of these errors. RHT results were obtained with the iterating version of the algorithm. . . .</i>	94
5.3	<i>Results on real images.</i>	96
5.4	<i>Results on real images.</i>	97
5.5	<i>Results on real image when outliers are present.</i>	97
6.1	<i>Effect of coherent outliers present in the mixed distribution. The numbers in the table are the average errors over 100 experiments conducted. Inside parentheses is the standard deviation of these errors. .</i>	111

6.2	<i>Effect of random outliers present in the mixed distribution. The numbers in the table are the average relative errors over 100 experiments conducted, and inside parentheses are the variances of these errors.</i>	112
6.3	<i>Classification results obtained using HTHT, TM and LSE when the endmembers were derived with different methods.</i>	115

Chapter 1

Introduction

Remote sensing is defined as the acquisition of information about the condition of a target (e.g. earth surface) by a sensor (e.g. satellite) that is not in direct physical contact with it [1]. Each recorded radiance measurement of the satellite provides a single picture element, or *pixel*, the size of which is determined by the *Instantaneous Field Of View* (IFOV) of the sensor. The spatially ordered assembly of pixels then provides the individual scan lines, and their assembly the satellite image.

Two basic processes are involved in the remote sensing of the Earth: *data acquisition* and *data analysis* [43]. The data acquisition process involves study of the energy sources, the propagation of energy through the atmosphere, the energy interactions with the Earth's surface, the airborne and spaceborne sensors, etc. the data analysis process involves the examination of the data with various viewing and interpretation devices or a computer to aid the analysis of any digital sensor data. This thesis is focused on computer assisted data analysis of satellite data.

The objective of this analysis is to automate the identification of features in a scene and thus replace the visual analysis of satellite data. This normally involves the analysis of multispectral data and application of statistical rules for determining the land cover identity of each pixel.

It is often the case in Remote Sensing that one wants to identify fractions of

diverse coverages in a region. Vegetation density varies at characteristic scales (a few meters) that are much smaller than the pixel scale of most satellite images. Moreover, some indispensable preprocessing procedures are likely to produce more mixed pixels. The proportion of pixels in an image that are mixed is dependent on the variability of the terrain and the spatial resolution of the sensor. Even with sensors of relatively high resolution, one expects to have to deal with mixed pixels. Generally, as the spatial resolution coarsens, the proportion of mixed pixels rises due to the variability of the terrain. On the other hand, when the spatial resolution becomes finer, additional detail may be resolved and the proportion of the mixed pixels may increase again [76]. Thus, the problem of *mixed pixel classification* is a major issue in Remote Sensing and Geography and many approaches have been developed to deal with it.

1.1 Motivation of this work and application area

The motivation of our work is to monitor burned forests for a few years after a fire so that the regeneration processes can be evaluated. In particular, we are interested in assessing the danger of desertification conditions ensuing in the site of a burned forest in the Mediterranean region. If the forest does not show signs of recovery a couple of years after the fire, it probably has to be artificially re-forested to prevent further erosion. Quite often, different types of vegetation grow in a burned region. It is usually the case that this new vegetation represents a deterioration of the quality of the flora of the region. Thus, for the purpose of our work, we are interested in assessing the degree of presence of each type of vegetation in a region using Landsat TM images.

The specific area of interest is located close to Athens, the capital of Greece, in the prefecture of Attica. An image of this area is presented in *Fig. 1.1*. From the climatic point of view, Attica is characterised as semi-arid. As a result the vegetation

1.1. MOTIVATION OF THIS WORK AND APPLICATION AREA 3

is adapted to aridity and there is a high risk of fire. The main types of vegetation in this area are: aleppo pine (*pinus halepensis*), maquis and phrygana. The Aleppo pine forests of Greece and in particular those of the prefecture of Attica have been burnt repeatedly during the last twenty years [64]. Remotely sensed data can be used to monitor forest regeneration processes, but semi-arid landscapes are not easy to model due to their variability at a subpixel level. So, mixed pixel classification techniques should be employed to deal with this sort of problem.

Figure 1.1: *Landsat TM image of Attica.*

In the prefecture of Attica, four test areas have been selected because there were forest fires in each of these areas within the last ten years. The test areas are located around 50–70km north-east (test areas of Penteli and Barnabas), west (test area of Pateras) and south-east of Athens (test area of Lavrio). Within the four test areas, 39 test sites, ranging between 1ha and 4ha, representative of different stages of regeneration after a forest fire, were selected by NARF (Institute of Mediterranean

1.1. MOTIVATION OF THIS WORK AND APPLICATION AREA 4

Forest Ecosystem - National Agricultural Research Foundation of Greece). These test areas are presented in *Figures 1.2-1.5*.

Figure 1.2: *Test sites within the Pateras test area.*

In this thesis, as it has already been mentioned, the images used were taken from the Landsat TM (Thematic Mapper). The Thematic Mapper scanner, carried on board by Landsat-4 and -5, has found a wide range of applications in arid area geomorphology. The Thematic Mapper senses in seven broad bands from the visible to the short - wave infrared and has an IFOV of $30m$ for the six reflective bands and $120m$ for the thermal band. The seven spectral bands of the TM are: blue (band 1: $0.45 - 0.52\mu m$), green (band 2: $0.52 - 0.60\mu$), red (band 3: $0.63 - 0.69\mu m$), near IR (band 4: $0.76 - 0.90\mu m$), two mid-IR (band 5: $1.55 - 1.75\mu m$ and band 7: $2.08 - 2.35\mu m$) and one thermal IR (band 6: $10.40 - 12.5\mu m$).

Each of these bands is more sensitive to certain materials of the ground. Band

Figure 1.3: *Test sites within the Barnabas test area.*

1 is placed to take advantage of the relationship between spectral radiances from vegetation which are determined in part by the chlorophyll and carotenoid concentrations. Band 2 is placed to record green region radiances. It is well chosen to maximise the spectral information content but is not as highly correlated with green vegetation as are bands 1, 3, and 4. Band 3 is well chosen from a green vegetation perspective. It is configured to be an excellent *in vivo* chlorophyll band. The wide bandwidth of band 4, coupled with the high levels of spectral reflectance characteristic of green vegetation for this region, should result in optimal remote sensing of vegetation density by band 4. Bands 5 and 7 are sensitive to water present in the leaves of the canopy, while band 6 exhibits thermal properties [77].

The Landsat TM images used in this thesis were collected in September 1993. These images were radiometrically and atmospherically corrected by the Institute for Digital Image Processing (DIBAG) at Joanneum Research, Graz, Austria.

Figure 1.4: *Test sites within the Penteli test area.*

1.2 Scope of this work

The purpose of this thesis is to propose algorithms that solve the problem of mixed pixel classification, efficiently and effectively under many possible circumstances.

All algorithms proposed in this thesis assume that the spectral reflectance of the examined terrain can be modelled by a *Linear Mixing Model*. According to the Linear Mixing model the reflectance of a mixed area is given by the linear superposition of the reflectances of the constituent components. In our case these components are the types of vegetation most likely to be present in these areas, namely: aleppo pine, maquis, phrygana and soil. Furthermore, in this thesis we consider not only the mean values of the sets of the mixed pixels we have to unmix, but also their statistical distribution.

Figure 1.5: *Test sites within the Lavrio test area.*

The algorithms proposed here are compared with the *Least Square Error* method, which is the most popular method for solving the Linear Mixing model due to its simplicity and effectiveness. The performance of these algorithms is first assessed using simulated data. The simulated data were created so as to be as close to real data as possible. The experiments on simulated data allow control over the expected results, and give an accurate point of reference to assess the performance of the proposed algorithms. As a middle stage of the evaluation some images were acquired in the laboratory, that contained materials of controlled composition and used to evaluate our algorithms as well. Finally, the proposed algorithms were applied to Landsat TM data.

First a novel algorithm that utilises *higher order moments* to represent the pure

and the mixed classes, is proposed. This algorithm incorporates the variability of the terrain and expands the number of equations used for the regression. It was proved to perform well, in cases that no outliers were present. The problem of outliers, especially when the classification extends to subpixel level, is not widely explored in remote sensing, even though their existence is almost inevitable. In this thesis, a novel method is proposed to handle the outliers. This method is based on the *Hough transform*. Various enhancements and extensions of the basic algorithm that improve the accuracy and speed are examined. A detailed outline of the components of this thesis is given next.

1.3 Outline of the Thesis

Starting from chapter 2, an overview of the methods currently used to solve the mixed pixel classification problem is given. Especially the *linear mixing model*, which has been adopted in this thesis, is presented. The problem with the methods used so far to solve the linear mixing model, is that they restrict the number of classes that are able to identify to the number of the uncorrelated bands and that they are either very complex or the simplified versions do not model the variability of the terrain. Therefore, in chapter 3 a novel method for mixed pixel classification is proposed to overcome these problems.

According to this method the classification of groups of mixed pixels is achieved by taking into consideration the higher order moments of the distributions of the pure and the mixed classes. The equations expressing the relationship between the *higher order moments* are used to augment the set of equations that express the relationship between the means only. It is shown that by appropriately weighting these equations we can avoid making the set of equations available less reliable. As a consequence, the number of equations can be increased and thus more classes than available bands can be identified. The method is exhaustively tested using

simulated data and is also applied to real Landsat TM images for which ground data are available.

The algorithms discussed so far require the calculation of various statistics (e.g. mean and variance). The problem with this type of statistics is that they are not accurately calculated in cases where outliers are present. In real applications though one expects to have to deal with outliers, so a method that uses robust statistics is proposed in chapter 4.

In this chapter a novel method for mixed pixel classification is presented, where the *Hough transform* and the *trimmed means* methods are used to classify small sets of pixels. The performance of these methods is compared with the *Least Squares Error* method and it is shown that in the presence of outliers the trimmed means method is far more reliable than the traditional Least Squares Error method. The method is exhaustively tested using simulated data and it is also applied to Landsat TM data.

Unfortunately, the Hough Transform is computationally a demanding process so in chapter 5 the *Randomized Hough Transform* algorithm is proposed for use with large datasets (for which the deterministic Hough is prohibitively slow) and in the presence of outliers (i.e. in cases that the classical Least Squares Error method cannot cope). It is shown that the Randomized Hough is of constant CPU time, irrespective of the size of the data sets and can be made even more accurate than the deterministic Hough for datasets that consist of more than about 50 sample points. The results are demonstrated both with simulated and laboratory real data.

Robust methods, as the Hough Transform, are known to perform well in the presence of outliers, but their performance when no outliers are present is inferior to the performance of the Least Squares Error method which gives the optimal solution. That is why in chapter 6 we propose the use of the *Hypothesis Testing Hough Transform* to classify sets of mixed pixels in order to increase the accuracy of the obtained results. The method is demonstrated using simulated data and proved

to perform equally well in the presence and in the absence of outliers. It is also applied to real Landsat TM data.

Finally, we conclude in chapter 7 where the main contributions of this thesis are clearly outlined.

Chapter 2

Methods used for mixed pixel classification

2.1 Introduction

The optical remote sensing of an object (e.g. vegetation) is based on the assumption that the radiation received by the remote sensor, which is the result of the interaction of solar radiation with the given object, can be deciphered to obtain the characteristics of the object.

The problem of assessing the vegetation characteristics from its spectral signature is not trivial. It is further complicated by the fact that the signal received by a remote sensor also carries with it the signatures of external factors such as soil and atmosphere.

There is a variety of methods for deriving the information content of the spectral reflectance data acquired by remote sensing sensors. The choice of method depends on the type of information sought and the complexity of the underlying processes that are depicted by the data. Two approaches are used for this purposes: calculation of indices, and spectral mixture analysis [1].

2.2 Vegetation index approach

One approach to the representation of spectral knowledge is to organize or synthesize information from multispectral bands into information that can be associated with the physical characteristics of ground classes. This process allows to compress the information from the original spectral bands into a smaller more manageable number of features.

Kauth and Thomas [35] determined that the data space distribution of soil reflectance variation in Landsat data is confined to a line in the 2-D space defined by two spectral components or a plane in the 3-D space defined by three spectral components. Usually the vegetation index relates the Landsat brightnesses in the near-infrared/red spectrum. This approach is based on the observation that in a near-infrared/red scatter diagram all pixels falling on a line (*cover line*) parallel to the soil line, have the same percentage of vegetation. Such lines of constant vegetation divide the plot area into regions that can be used as a lookup table to determine the percentage of each component contributing to the pixel reflectance e.g. soil, vegetation, etc [62, 31, 32]. The decision boundaries for low, medium and high-cover vegetation were determined arbitrarily.

Thus, a measure of the distance of a candidate point from the soil line can be used as an index of the vegetation amount of that point. Most commonly used are the *normalized difference vegetation index* (NDVI) and the *perpendicular vegetation index* (PVI). The NDVI is defined as:

$$NDVI = \frac{\text{Near Infrared band} - \text{Red band}}{\text{Near Infrared band} + \text{Red band}}$$

while PVI is defined as

$$PVI = \sqrt{(\text{soil in NIR} - \text{point in NIR})^2 + (\text{soil in R} - \text{point in R})^2}$$

While the vegetation index approach works well in some vegetation types, problems arise where there are variations in soil brightness from one area to another. Vegetation cover is usually less than 100%, so a given amount of vegetation on a dark soil background will appear in the plot differently (closer to the origin) from the same vegetation on a bright soil. Position on the cover line will also vary with the brightness of the vegetation cover. The cover line is therefore best used to detect change over time in an area with a particular soil background rather than to determine absolute cover values across a range of soil types [57]. Even though the importance of the vegetation index is not questioned, its classification performance is found, in some applications, to be inferior to the performance of spectral mixture analysis [58].

2.3 Spectral mixture analysis approach

A desirable estimator would be based on a model of how a given mixture of components gives rise to the pixel's signal and how we can invert the procedure to unmix the signal. These types of models are called *mixing models* and are best suited to Landsat Thematic Mapper or imaging spectrometer data with a large number of spectral bands [57].

The mixing models were first used to identify components in chemical mixtures [38]. Mixture modelling has been used as a remote sensing tool to identify minerals on the lunar and Martian surfaces [3], to map regional vegetation and geologic substrates [74], to estimate fluvial suspended sediment concentrations (a non-linear mixing problem) [49], to estimate the snow-covered area [66], to estimate the albedo for climatological studies [61], etc.

There are two varieties of spectral mixtures: *macroscopic* and *intimate*. Macroscopic mixtures result when a given photon will be reflected (or absorbed) by only one type of material. The reflected radiances are combined in an additive fashion

to produce a linear set of equations and the corresponding models are called *linear mixture models*. Intimate mixtures occur when individual photons interact with more than one type of material, for example the light might penetrate through the component of interest (e.g. salt) and interact with the component beneath. Intimate mixtures give non linear equations and the corresponding models are called *non linear mixture models* [21].

Smith et al. [74] showed that in Owens Valley the linear mixing assumption is valid for the waxy, semi-opaque foliage of arid environments despite wide variation in soil brightness, and that the best set of reference endmembers does not change with seasonal illumination and canopy density [66]. So, since the application we are concerned with in this thesis, examines semi-arid regions, we believe that the linear mixing is a good modelling approximation for our case.

Therefore, we shall restrict ourselves here to the case of linear mixture models. The mixture model can be expressed as:

$$y_k = \sum_{j=1}^J (p_{kj} x_j) + e_k$$

Where

y_k represents the spectral intensity of a mixed pixel in the k^{th} band.

p_{kj} represents the spectral intensity of the j^{th} constituent component in the k^{th} band.

x_j represents the proportion of the j^{th} constituent component present in the mixed pixel.

e_k represent the residual between the mixture model and the data for band k concerning the mixed pixel due to random errors in the data and inadequacies in the linear mixture model.

The basic step to mixture analysis is the unmixing, or else the inversion of the linear mixing model equations in order to estimate the proportions. There are various methods used for the unmixing, that in general depend on what kind of information is available to us.

2.3.1 Unmixing using training data

One very popular method is to use training data to find the constants of the mixed model and then use the model to estimate proportions of the basic components in other pixels [70]. The constants of this model are the p_{jk} that can be extracted from the training data. The spectral intensities of the mixed pixels y_k are extracted from the image and the aim is to estimate the values of x_j .

The data used for training are representative of the basic components and can either be extracted from the image (*image training data*), or be created artificially in the laboratory (*reference data*) [73]. The latter can also be field spectroscopy measurements.

The *image training data* can model the image accurately, but they may not be valid for a different image and they were found to vary with time (e.g. seasonal changes in vegetation) [73]. On the other hand the reference training data are more general, but may not model the image very accurately. In the later case we should evaluate how well our model approximates the image using some fitting measures.

In mixture analysis, not only the source of training data has to be decided but also their type. As training data, point data or sets of data can be used. Use of sets of data for training is not very popular although the quality of the results of this approach depends on the balance of constituent components in the mixture. The point data tend to produce more inaccurate (biased) results for extreme population ratios in the mixtures. McCloy and Hall [48] used sets of data from which they calculated the means and used these means for training.

The most popular technique though, is to use points for training which are called *endmembers*. Each endmember spectrum is a point in the space defined by the multi-spectral bands, as is each observation. All possible legitimate mixtures, those with non-negative proportions fill the volume enclosed by the generalised polyhedron whose vertices are points corresponding to the endmember spectra [70, 73]. If image training data are used then an endmember can either be the value of a pixel known to contain a single class, or the mean of a homogeneous region solely composed of one class with more than one pixels. Of course the latter incorporates some of the variability of the terrain and can be considered as an improvement to pixelwise image endmembers.

As mentioned before, the training data should be representative of the basic components. Unfortunately, some times data for the individual components are not available. A much more likely situation is that we have detailed field knowledge of proportional ground cover for a number of pixels. In this case it is still possible to derive the attributes of the basic components using various techniques.

The problem described above can be considered as a multivariate calibration problem. In order to solve such a problem, according to Brown [14], there are three methods to use: the Lwin-Maritz estimator which assumes a non-linear combination of the pure components. This did not perform well in Brown's experiments. The "classical" method, which is equivalent to assuming a linear model for the generation of the composite signal, using training data to find the constants of this model, and then using that model to predict proportions of other pixels. The "inverse method", that looks for the best model to describe the connection of the pure components and then uses regression methods to identify these components.

The prediction of pixel proportions is the unmixing. There have been proposed two main methods for the unmixing: the Least Squares Error (LSE) method, and the Maximum Likelihood Estimation (MLE) method.

2.3.1.1 Least Squares Error (LSE) method

The mixing model can be written in matrix form as:

$$\mathbf{Y} = \mathbf{P}\mathbf{X} + \mathbf{E} \quad (2.1)$$

Where

$\mathbf{Y}$ represents the spectral intensity of a set of mixed pixels. It is a $K \times I$ matrix with I the total number of mixed pixels and K the number of bands used.

$\mathbf{X}$ is a $J \times I$ matrix which represents the proportions of the endmembers in the given mixed pixels. J is the number of the endmembers of the linear model.

$\mathbf{P}$ is a $K \times J$ matrix which represents the spectral intensities of the endmembers.

$\mathbf{E}$ is a $K \times I$ matrix and represents the error between the mixture model and the data.

From the theory of linear equations we know that in general a solution to (2.1), if one exists, will be unique if the number of endmembers is equal to the number of bands, or else $J = K$, and if $J > K$ there will be an infinity of exact solutions. Finally, when $J < K$ there may well not be an exact solution. The excessive number of solutions when $J > K$ does not allow the definition of any sort of best solution from among the infinity of all possibilities. So, we consider only the cases when we have fewer endmembers than bands.

The number of bands considered here may not be equal to the number of the satellite bands. For example, if we are dealing with the six reflectance bands of Landsat TM data and find that the fifth and sixth principal components of the data contain nothing but noise, then the true dimensionality of the data is four and not six.

In the case that $J < K$ a solution to the problem can be obtained by regressing calibration data Y on X . The most common method for *regression* is the least squares error method which yields the following solution to the unmixing problem:

$$\begin{aligned} (\mathbf{P}^T \mathbf{P}) \hat{\mathbf{X}} &= \mathbf{P}^T \mathbf{Y} \Rightarrow \\ \hat{\mathbf{X}} &= (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{Y} \end{aligned} \quad (2.2)$$

Here, we use $\hat{\mathbf{X}}$ to represent the estimate of $\mathbf{X}$ and we have assumed that there is no residual error in the equations. If $(\mathbf{P}^T \mathbf{P})$ is nearly singular, it cannot be inverted in a meaningful way. It means that one or more endmember spectra are too similar to some linear combination of some of the other endmember spectra in $\mathbf{P}$ [56, 70]. In this case a method called Q-R decomposition can be used [44, 25]. The Q-R decomposition method is based on the decomposition of matrix $\mathbf{P}$ into two matrices $\mathbf{Q}$ and $\mathbf{R}$ such that:

$$\mathbf{P} = \mathbf{Q}\mathbf{R}$$

where $\mathbf{Q}$ satisfies

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix with non-zero diagonal elements, and $\mathbf{R}$ is upper triangular with diagonal elements $r_{jj} = 1$. The above definition can be substituted to (2.2) to give the following system:

$$\mathbf{R} \hat{\mathbf{X}} = \mathbf{D}^{-1} \mathbf{Q}^T \mathbf{Y}$$

We can also impose the constraint that all the proportions add to 1 and they can take values from 0 to 1. In this case we have constrained regression and therefore

constrained least squares method can be used. Shimabukuro [71] gave numerical solutions to a mixture problem of 3 components. Schanzer [69] showed that instead of using the constrained least squares method we can use *unconstrained stepwise regression* by eliminating components according to t-statistics.

The classical method described above is especially useful in applications that use satellite data with many spectral bands like the Landsat TM images, in order to allow us to perform the regression procedure.

2.3.1.2 Maximum Likelihood Estimation (MLE) method

In this section we are going to use similar notation to the one we used in the previous section. So, K is the number of bands, J is the number of the endmembers and I is the number of mixed pixels we consider. Furthermore, $\mathbf{P}$ is the matrix with the endmember reflectances, $\mathbf{Y}$ is the matrix with the reflectances of the mixed pixels and $\mathbf{X}$ is the matrix with the proportions of the mixed pixels. This time we will also refer to single rows or single columns of the above matrices which will now be identified by a small letter suffix (e.g $\mathbf{P}_k = (p_{k1}, p_{k2}, \dots, p_{kJ})$, with p_{kj} being an element of matrix $\mathbf{P}$).

In this approach we assume that an endmember is the reflectance of a pixel that contains a single type of object say n_j of them. Each individual object has a reflectance signature which can be represented by a K -dimensional normal distribution. We also assume that there is no correlation between the signals received from the different bands of the sensor. So for a specific band k , a given object of class j will have mean a_{kj} and will exhibit statistical fluctuations due to sensor noise, characterised by a variance b_{kj} . An agglomeration of n_j such objects will have n_j times stronger reflectance, i.e $n_j a_{kj} \equiv p_{kj}$, where p_{kj} is an element of matrix $\mathbf{P}$, with variance n_j times the variance of one individual object, i.e. $n_j b_{kj} \equiv \sigma^2_{kj}$. If we also assume that the endmembers are independent of each other, then the reflectance of a given mixed pixel i with fractional coverage x_{ji} of this class will only contain $x_{ji} n_j$

of such objects and taking into consideration all the individual objects of all types it will contain, we can deduce that its spectral signature can also be represented by a normal distribution with mean y_{ki} in band k given by:

$$y_{ki} = \sum_{j=1}^J p_{kj} x_{ji} \quad (2.3)$$

and variance

$$s_{ki} = \sum_{j=1}^J \sigma_{kj}^2 x_{ji} \quad (2.4)$$

If we now consider the whole set of mixed pixels in a given band k we set:

$$\mathbf{Y}_k \equiv (y_{k1}, y_{k2}, \dots, y_{kI})$$

From equation (2.3) we have $\mathbf{Y}_k = \mathbf{P}_k \mathbf{X}$, so the row vector $\mathbf{Y}_k$ is also normally distributed with mean $\mathbf{P}_k \mathbf{X}$ and covariance matrix:

$$\mathbf{S}_k = \text{diag}(s_{k1}, s_{k2}, \dots, s_{kI})$$

Therefore, the conditional probability $\mathcal{P}(\mathbf{Y}_k|\mathbf{X})$ of a set of mixed pixel values $\mathbf{Y}_k$ in band k with given mixing proportions expressed by $\mathbf{X}$ is given by:

$$\mathcal{P}(\mathbf{Y}_k|\mathbf{X}) = (2\pi)^{-K/2} |\mathbf{S}_k|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y}_k - \mathbf{P}_k \mathbf{X}) \mathbf{S}_k^{-1} (\mathbf{Y}_k - \mathbf{P}_k \mathbf{X})^T\right\}$$

where $|\mathbf{S}_k|$ is the determinant of the covariance matrix.

Then the joint probability $\mathcal{P}$ for all observation vectors $\mathbf{Y}_k$ can be calculated by taking the product of the probability of a certain set of observations to arise in each

band

$$\mathcal{P}(\mathbf{Y}) = \prod_{k=1}^K \mathcal{P}(\mathbf{Y}_k | \mathbf{X}) \mathcal{P}(\mathbf{X})$$

where $\mathcal{P}(\mathbf{X})$ is the prior probability of a certain set of mixing proportions. The idea is to choose the mixing proportions so that this joint probability is maximal. This results in a non-linear optimization problem, which is equivalent to minimizing the log-likelihood function, $\mathcal{Q} = -\ln(\mathcal{P}(\mathbf{Y}))$ [2, 29]:

$$\mathcal{Q} = K \ln(\mathcal{P}(\mathbf{X})) - \frac{1}{2} \sum_{k=1}^K \ln |\mathbf{S}_k| - \frac{1}{2} \sum_{k=1}^K (\mathbf{Y}_k - \mathbf{P}_k \mathbf{X}) \mathbf{S}_k^{-1} (\mathbf{Y}_k - \mathbf{P}_k \mathbf{X})^T$$

Often the analyst has no useful information about $\mathcal{P}(\mathbf{X})$, so equal prior probabilities are assumed and the above equation can be simplified to:

$$\mathcal{Q} = -\frac{1}{2} \sum_{k=1}^K \left(\ln |\mathbf{S}_k| + (\mathbf{Y}_k - \mathbf{P}_k \mathbf{X}) \mathbf{S}_k^{-1} (\mathbf{Y}_k - \mathbf{P}_k \mathbf{X})^T \right)$$

This general case can be solved by using a brute force method, which is based on minimizing $\mathcal{Q}$ over a vector $\mathbf{X}$, such that each component of this vector is a positive integral multiple of $1/r$ for some fixed positive integer r and such that all the components add to 1. An $\mathbf{X}$ which minimizes $\mathcal{Q}$ is taken as an approximation of the estimate. Estimates are not unique in the general case, but if the covariance matrices are "small" enough, then estimates are "almost" unique. Even for a small number of classes this method is not very practical, so usually further assumptions are made.

If we now assume equal variances for all endmembers ($\sigma_{kj}^2 = \sigma_k^2 \Rightarrow \mathbf{S}_{ki} = \sigma_k^2 \sum_{j=1}^J x_{ji} = \sigma_k^2 \Rightarrow \mathbf{S}_{ki} = \mathbf{S}_k \Rightarrow \mathbf{S}_k = s_k \cdot \text{diag}(1, 1, \dots, 1)$), the term $\ln |\mathbf{S}_k|$ can not contribute to the discrimination and therefore it can be omitted. The second part of the summation is the Mahalanobis distance of observation vector $\mathbf{Y}_k$ from mean vector $\mathbf{P}_k \mathbf{X}$. So the Maximum Likelihood Estimate classifier can be simplified

to a Mahalanobis distance classifier. If instead of the assumptions of uncorrelated bands with equal variances, we assume that all endmembers have equal covariance matrices, the covariance matrix of the mixture S_k will be equal to the same matrix and it will again be known.

The Mahalanobis distance classifier may also be weighted using some fuzzy membership functions. These fuzzy membership functions lie on a scale between 0 and 1 and sum to 1 for each case. They are to some extent, similar to the *a posteriori* probabilities of class membership and their magnitude will be related to the proportion of a particular class within a pixel [22]. The fuzzy membership functions can be generated from a supervised version of the fuzzy C-means algorithm [5].

2.3.1.3 Comparison of the LSE and MLE methods

The LSE method is more popular since it is quite simple from the mathematical point of view and can easily be expanded so as to deal with many endmembers. The MLE is a computationally expensive and a relatively slow procedure. Much effort has been put into improving its computational efficiency [20, 47, 63].

Furthermore MLE can perform poorly when it is used to classify high-dimensional data as it is often difficult to get enough training samples to provide reliable estimates of the second order statistics of each class. These statistics, expressed by class covariance matrices, are used in MLE and distinguish it from algorithms using only first order (mean) statistics [63], and therefore it is considered to be a more robust estimator than the LSE.

Finally, MLE relies on a Gaussian probability distribution of the spectral signature of the training data, which often exhibits a non-Gaussian distribution [78].

2.3.2 Unmixing without using prior knowledge

So far in the linear mixing model (2.1) we assumed that the matrix $\mathbf{P}$, which represents the spectral intensities of the endmembers, is known or at least can be derived from the available training data. However, in problems where neither matrix $\mathbf{P}$ nor the proportion matrix $\mathbf{X}$ are known, other techniques can be employed to solve the problem of simultaneous retrieval of subpixel proportions and signatures of the endmembers from the mixture data $\mathbf{Y}$.

A major consideration in this case is the problem of *dimensionality reduction* of matrix $\mathbf{P}$, since we start with no assumptions concerning its dimensions. So it is necessary to define the basic components that correspond to endmembers. The number of these components will be indicative of the different cover types that can be identified.

Factor analysis can be used to identify the basic components [30]. This method involves hypothesizing a minimum number of common factors necessary to reproduce the observed correlations. In absence of any knowledge this means that one has to start with one common factor. This hypothesis is evaluated by applying some criterion to determine whether the discrepancy between the assumed model and the data is trivial. If it is not, a model with one more common factor is estimated and the criterion is applied again. This is continued until the discrepancy is judged to be attributable to sampling error. The problem of factor analysis is that it is not capable of guaranteeing positive factor loadings or common factors. Some work has been done to impose the required constraints when only two components are assumed [42].

Principal component analysis is sometimes used prior to some factor analytic procedures to determine the dimensionality of the common factor space [30, 15]. Principal component analysis is a statistical technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated variables

that represents most of the information in the original set of variables. According to the principal components analysis from matrix $\mathbf{P}$ an eigenvector matrix with the corresponding eigenvalues is derived. The eigenvalues, which are extracted in order of importance, are used to determine the number of physically significant reflecting components present in the data set. The stepwise factor analysis begins then, using only the eigenvector corresponding to the largest eigenvalue.

When the endmembers are selected, the least squares error method described in the previous section can be used to derive the corresponding proportions. In order to provide a measure of how much of the spectral variability is explained by the derived endmembers, the RMS error image is computed. In particular, for each pixel of the image the corresponding proportions are computed and then these proportions are used to reconstruct the values of the image pixels. According to a popular fitting algorithm [3] the maximum *Root-Mean-Square* ($RMS = \sqrt{E(\mathbf{Y} - \hat{\mathbf{Y}})^2}$) error between the image ($\mathbf{Y}$) and that predicted by linear combinations of the endmembers ($\hat{\mathbf{Y}}$) should be less than the noise level of the image data due to the error of the sensor. Furthermore, the RMS error image shows the spatial distribution of the error. Another way of assessing the fitness of the model rather than by rms, is by using *confidence intervals* for the predicted image [14].

Even more band residuals, that is band by band differences between the digital number values modelled by the reference endmembers and the calibrated image data, can be used to increase the performance of modelling complex natural surfaces, since they may be used as an indication of the presence of other cover types [65]. On the other hand, adding endmembers may lead to unstable proportion derivation due to progressive lowering of the spectral contrast between the endmembers.

Another consideration concerning the mixture model is the estimation of a threshold of reliable identification of a component given the errors involved in the procedure [68]. There are two methods for this *continuum threshold analysis* and *residual threshold analysis*. According to the first method a mixed spectrum is re-

peatedly combined with noise and unmixed to estimate the fractions of the endmember spectra. This way a confidence interval is calculated which allows the estimation of the minimum detectable spectrum fraction. In the second method a similar procedure is followed, but this time the target is not included in the reconstruction of the mixed spectrum and is therefore estimated as a residual. The second method is used when the target has low abundance or its spectrum is similar to the other components.

Some times there is some prior information available that helps to decide on which endmembers to use. For example in [37, 78], all possible endmembers were identified from the knowledge of a study region and a library of their spectra was created, appropriate for the study of the specific region. These endmembers represent the main elements expected in the Landsat data. In order to extract the most important components constituting the bulk of the spectral variability throughout the data set a technique known as Spectral Angle Mapping (SAM) was used [37, 78]. SAM calculates the spectral similarity between a test (or pixel) reflectance spectrum and a reference (or laboratory) reflectance spectrum expressed in terms of the average angle between the two spectra treated as vectors in a space with dimensionality equal to the number of spectral bands in the image. The outcome of the SAM for each pixel is an angular difference measured in radians (ranging from zero to $\pi/2$), which is a qualitative estimate of the presence or absence of a surface component.

2.3.3 Geometric models

A variation of the mixing models is to model geometrically the individual plants on the ground and make use of the distribution of the parameters of the geometric models for the unmixing. These are the *geometric* models. These models conceptualize clumps of vegetation as solid three dimensional geometric elements superimposed on a flat soil background. The distribution of the elements themselves can be regular,

as in the case of row crops, or statistical, as for natural vegetated landscapes.

Geometric models have been successfully used to describe much of the variability of semi-vegetated landscapes by altering the shape and density of the geometric canopy elements. Otterman [51] modelled forests and desert vegetation as thin vertical cylinders of random height and spacing. This model assumes that the objects are small and numerous. Therefore, it is not directly applicable to the open forest canopies in which the objects (trees) can occur in low densities and are relatively large. To address these problems, Li and Strahler modelled conifer forests as randomly located cones of similar shape and random height, assuming constant tree and soil background reflectances [41]. Several authors have focused on the problem of fitting stochastic models to spatial point patterns of natural vegetation. Li and Strahler assumed a homogeneous Poisson distribution of conifer tree locations.

Many semi-vegetated landscapes are characterized not only by variations in plant size and density but also by variations in vegetation and soil background. Variations in reflectance are brought about by changes in topography and in the physical properties of the plant (e.g. leaf area, amount of chlorophyll etc.) and soil background (e.g. surface roughness, soil moisture etc.). One method of incorporating the soil and vegetation reflectance variability into geometric models is by treating the reflectance terms as random variables. That is, to be characterized by their mean, variance and, in some areas, spatial covariance [32, 31].

The geometric model provides more equations for the elements of the scene, apart from the basic mixing model equation. These equations help to solve for the unknown parameters of the mixing model. The geometric models can be useful when images from satellites with few spectral bands are used. In these cases we don't have enough uncorrelated mixing equations to regress in order to solve for the unknown parameters of the mixing model. On the other hand, the geometric equations may introduce further unknowns, and in general the inversion procedure, or else the unmixing, is not a trivial procedure as when linear regression is used.

The development of geometric-optical canopy models has contributed greatly to our understanding of the relationship between physical and spectral properties of various vegetation types. However, a lot of semiarid vegetation occurs in heterogeneous terrain that varies in topography, illumination, soil depth and vegetation composition. Explicit scene modelling over such land surfaces may require separate parameter sets for different land classes. Given the difficulty of parameterizing and inverting canopy models, it is important to understand the trade-offs between increased predictive skill, versus model complexity in moving from simple correlative models to more physically based approaches [23].

2.4 Problems with the classical mixture model

Some times the performance of the mixing models is not within acceptable limits. Pure performance of the mixing model can be mainly attributed to deviations from linearity and to inadequency of the training data to model properly the scene.

So far we have discussed linear mixing models, but in some applications this linearity assumption does not hold. In some cases the deviation from linearity is small and we still can assume linearity [50]. Meanwhile non-linear spectral mixing can be linearized by converting reflectance to single scattering albedo. Single scattering albedo can then be linearly mixed to predict non linear mixtures [68].

The validity of this assumption varies with the measured material and with the spectral band. For example, NIR radiation is strongly scattered by green vegetation resulting in a non-linear response as the scattered radiation interacts with soils and other materials [30]. However, violations of the linear assumption may not have serious consequences in situations where non-linearly reflected radiation contributes only a small amount to the total reflectance, e.g. in semiarid areas where vegetation is sparse [74].

Other researchers observed that no linear combination of spectra can fully rep-

resent the measured composite scene spectra and developed methods to handle non linear models. Borel used the bidirectional reflectance distribution function which is given by adding the upwards and the downwards radiosity of the leaves with the radiosity of the ground and results to a non linear equation [6]. More complicated surface structures increase the complexity of the model and of course the unmixing becomes even more difficult.

Furthermore, multi-spectral thermal infrared images may be modelled as mixtures of temperatures instead of reflectances. A linear mixture model is valid for temperatures and it is found to have smaller deviations from linearity than the reflectance mixture model [61]. Even when the conventional reflectance mixing model is used, the temperature of the components should be taken into consideration, since temperature differences between the components may cause error to the proportion estimates. When the components of interest have different temperatures, the proportions of the colder components would be underestimated if temperature is not incorporated in the mixture model. The temperature factor though, increases the complexity of the model and when more than two components are concerned, it is better to assume that all the pixels within the scene have the same temperature [26].

Two other parameters that can influence the performance of the estimates are the shade and the spectral variability of the modelled components. We can distinguish two types of shade, one that is attributed to the terrain morphology (e.g slopes) and one to vegetation (e.g. tree crowns cast shadow to the nearby vegetation) [3]. The first type of shade can be estimated using the appropriate elevation models and forms a shade image that should be subtracted from the original image. The second type of shade can be estimated using geometrical information of the vegetation present [41].

Furthermore, each endmember is used to describe the spectral contribution of an individual component in an image that may, in fact, be spectrally variable within

the scene. This variability can be a source of uncertainty in determining the proportions [68]. Most of the methods discussed above are basically concerned with the classification of individual pixels. The application examined in this thesis though, is concerned with the estimation of proportions present in small (less than 100 pixels) sets of mixed pixels. Dealing with sets of pixels, as opposed to individual pixels offers a major advantage which we can exploit to overcome the problem of modelling the variability within the endmembers.

At first the idea of considering the variance of a mixed pixel as a random function may not appear particularly original. Indeed, in [29, 2], where Maximum Likelihood Estimation was used for the calculation of the mixing proportions, the spectral signatures of the mixed pixels were considered also as random variables. However, those authors assumed that the spectral signature of a pixel is the agglomeration of the spectral signatures of individual objects and they modelled the variance of these individual objects. We actually model the variance of individual pixels. At first sight, what they did seems more correct. After all, pixels do not really have a physical existence. They are digital imaging artifacts. However, it is doubtful that individual objects can actually be identified and modelled. For example, in imaging the rainforest, what is an object? In imaging sparse vegetation, is a shrub with its shadow one object, or are they two objects? How many shadows make up a "pure" class pixel? How can one identify individual objects in class "soil"? In our case, by considering the whole pixel as a random variable, theoretically we can include in modelling its distribution.

On the other hand, when whole sets of mixed pixels are examined, another problem may arise such as the occurrence of outliers. As is known, the natural world highly lacks regularity and is full of human activities, therefore even adjacent pixels could have totally different spectral properties. There is high probability that the data sample extracted from a remotely sensed image will contain some outliers. This problem has not been dealt with for subpixel estimates. When whole pixels

need to be allocated to a given class, a weighted MLE method is proposed to be able to handle small amounts of outliers, especially if a threshold to the maximum acceptable Mahalanobis distance is set [80]. In this thesis we propose the use of robust statistics to improve the classification accuracy for subpixel estimates.

2.5 Visualization

The methods described so far do not allow direct visual interpretation of the procedure in case of multidimensional data. Several studies have been carried out on graphical methods in order to display remotely sensed data and classify them. Such an approach is the nPDF (*n-Dimensional Probability Density Function*) method [16]. This method projects the spectral multidimensional distributions onto a 2-D space defined by a set of selected corners in the data hyper-space. According to this approach the transformed spectral distributions are calculated by determining the hyper-spectral distance of each pixel from the selected corners. We can test different sets of corners to identify those that give the best discrimination. Using training data from pure classes we can locate the region of these classes in the 2-D space. These regions can then be used as a lookup table for the classification procedure.

The technique is user-interactive and can be used to assist in data visualization, image enhancement, and both supervised and unsupervised classification. Cetin and his colleagues [16] have used the nPDF method, with success, to analyze AVIRIS (Airborne Visible/Infrared Imaging Spectrometer), TIMS (Thermal Infrared Multispectral Spectrometer), and TM (Landsat Thematic Mapper) multispectral data.

Furthermore, it is not only important to visualize the classification procedure but also to be able to give a meaningful visual interpretation of the classification results (e.g. proportions derived using spectral mixture analysis). This will allow the observation of temporal changes in the fractions, which can be useful in monitoring, since it assists the identification of some interesting trends. In our case this means

that we should map the acquired proportions to the existing vegetation communities. When a three component mixture is considered, this mapping can be done using ternary diagrams [67].

Figure 2.1: *Ternary classification diagram.*

The *ternary diagram* is a triangle (e.g. ABC) with the property that the sum of the perpendicular distances from any point within the triangle to the three sides equals the altitude of the triangle. We can therefore let the altitude represent 100 percent composition and the distances to the three sides the percentages or proportions of the three components. Each apex of the triangle represents one of the pure components. The distance of any point such as K from the base AB represents the percentage of C in the mixture at K, the distance to the base AC the percentage of B, and that from the base CB the percentage of A. Any point on the side of the triangle represents a binary mixture (*Figure 2.1*).

2.6 Conclusions

In this chapter we have discussed the potentials and limitations of the techniques used for estimating the canopy composition from the measured spectral reflectance. From all the described approaches by far the most popular is that of the linear mixing models. However, it has often been criticised basically for its inadequency to capture nonlinear mixing due to backscatter, and because it can only be used to unmix at most as many components as there are bands. Nonlinear mixing models have started to be presented and although they may be more powerful than the linear ones in modelling far more complex phenomena, they lack the simplicity of the linear models.

Chapter 3

Mixed pixel classification using higher order moments

3.1 Introduction

In this chapter we present a novel method for mixed pixel classification where the classification of groups of pixels is achieved taking into consideration the higher order moments of the distributions of the pure and the mixed classes. The use of higher order moments allows better approximation of the scene and at the same time increases the number of available equations. Thus more classes than available bands can be identified. The method is demonstrated using simulated data and is also applied to real Landsat TM images for which ground data are available.

We assume that all pixels in the set of mixed pixels we have to classify have the *same* mixing proportions. Any variability from one pixel to another is attributed to the intraclass variability of the pure classes assumed. Thus our model does not assume any error in the reflectance values of the mixed pixels and only tries to model the intraclass variability.

In the next section we describe our method in detail and derive the formulae we use. In section 3.3 we demonstrate the method using some simulated data.

In section 3.4 we discuss the problem of determining the distributions of the pure classes, and in section 3.5 we apply the method to some real data. Finally we present our conclusions in section 3.6.

3.2 The proposed method

There is a close relationship between the concept of distribution means and the endmembers concept. But the approach that uses endmembers does not incorporate statistical fluctuations due to noise and intrinsic class variability characterised by the covariance matrix. It assumes a common covariance matrix for all distributions, equal to the noise covariance matrix which is therefore eliminated from the process. Our approach incorporates the within class variances since it makes use of the second order moments. Further, according to the classical method the number of equations used for the estimations is, in the best case, equal to the number of satellite bands, a fact that imposes a constraint to the possible number of endmembers that can be identified. This criticism could possibly be answered with the hyperspectral data where hundreds of bands are available. It is doubtful, however, that these bands contain independent information. For example, it is commonly accepted that the six reflective Landsat TM bands are correlated and they only span a three dimensional spectral feature space [17], so the number of different endmembers that the classical method allows is quite limited. In our approach we increase the number of equations used for the estimation of the proportions. Therefore we do not have such constraints and we can allow the use of more pure classes, which can help us to model an area more accurately.

We worked with Landsat TM multispectral images which have 7 bands. The thermal TM band (TM6) was excluded from the methodology not only because of its sparse resolution (120m) but also because other investigators have shown that, for identification of surface types, thermal information is not readily associated with

that in the reflective part of the spectrum, which in turn may lead to spurious classification [59].

Our aim is to determine, with reference to a set of training sites, the class composition of a test site using its observed spectral response. This class composition is represented by the class proportions. In this study we will consider the case of four pure classes.

For training we are going to use sets of pixels extracted from the remotely sensed image itself, so we are going to use image training data. The sets of pixels used for training belong to small regions on the image that correspond to areas on the ground for which the coverage proportions have been estimated by ground inspection.

In order to accommodate the random subpixel fluctuations in plant and soil properties yet keep the number of model parameters to a minimum, we can represent semi-vegetated landscapes using a stochastic reflectance model. Treating the properties as random variables provides a flexible means of characterising the scene without having to prescribe an inordinate number of detailed vegetation and soil parameters. The training and test sites therefore could be represented by certain distributions. In that respect a test site should have a distribution that is the mixture of the distributions of the reference classes.

Given that we conceptualize our model as a mixture of distributions and not a mixture of point reflectances we propose to calculate its first and second order moments. This eventually leads to two sets of mixture equations that have to be satisfied. These equations can be solved using the *Least Squares Error* (LSE) method.

The pixel value in a spectral band i is given by the linear combination of the spectral responses of each component within the pixel, so the model can be expressed as:

$$w_i = ax_i + by_i + cz_i + dv_i \quad (3.1)$$

where w_i is the known spectral reflectance of a mixed pixel, x_i , y_i , z_i and v_i are the known spectral reflectances of the four possible cover components within the mixed pixel and a , b , c and d are the proportions to be estimated, for each component contained in the mixed pixel.

When the terms of the reflectance model are considered random variables, then the moments of w_i can be expressed in terms of the moments of the individual variables x_i , y_i , z_i and v_i . Such expansions can be achieved by applying fundamental properties of random functions without prescribing the probability density functions of the variates. Also, instead of dealing with individual bands, we shall assume that $\underline{x}$, $\underline{y}$, $\underline{z}$, $\underline{v}$, and $\underline{w}$ are n dimensional vectors, where n is the number of spectral bands used.

The probability density function f_w of $\underline{w}$ is given by:

$$f_{\underline{w}} = \frac{1}{d^n} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\underline{x}, \underline{y}, \underline{z}, \frac{\underline{w} - a\underline{x} - b\underline{y} - c\underline{z}}{d}) d\underline{x} d\underline{y} d\underline{z} \quad (3.2)$$

where $f(\underline{x}, \underline{y}, \underline{z}, \underline{v})$ is the joint probability density function of variables $\underline{x}$, $\underline{y}$, $\underline{z}$ and $\underline{v}$. Since these variables are the reflectances of the four different pure classes, they can be considered as independent and their joint probability can be replaced by the product of the probability density functions of each random variable separately, $f_{\underline{x}}(\underline{x})$, $f_{\underline{y}}(\underline{y})$, $f_{\underline{z}}(\underline{z})$ and $f_{\underline{v}}(\underline{v})$.

Then it is easy to show that

$$\overline{w_i} \equiv \int_{-\infty}^{+\infty} w_i f_{\underline{w}}(\underline{w}) d\underline{w} = a\overline{x_i} + b\overline{y_i} + c\overline{z_i} + d\overline{v_i} \quad (3.3)$$

where $\overline{x_i}$, $\overline{y_i}$, $\overline{z_i}$ and $\overline{v_i}$ are the mean values of the corresponding pure distributions in band i .

Similarly, the elements of the covariance matrices of the four variables can be

shown to obey the following equation:

$$cov_{w_i w_j} \equiv \int_{-\infty}^{+\infty} (w_i - \bar{w}_i)(w_j - \bar{w}_j) f_{\underline{w}}(\underline{w}) d\underline{w} = a^2 cov_{x_i x_j} + b^2 cov_{y_i y_j} + c^2 cov_{z_i z_j} + d^2 cov_{v_i v_j} \quad (3.4)$$

For $n = 6$ equation (3.4) represents a set of 21 equations, while (3.3) represents a set of 6 more equations. So we have 27 equations in total with four unknowns a , b , c and d . We can combine the two sets of equations together and use the *Least Square Error* method to estimate the unknowns. The squared errors for (3.4) and (3.3) are given by:

$$\begin{aligned} \sum_{i=1}^n (\bar{w}_i - a\bar{x}_i - b\bar{y}_i - c\bar{z}_i - d\bar{v}_i)^2 &\equiv \sum_{i=1}^n e_i^2 \\ \sum_{i=1}^n \sum_{j=i}^n (cov_{w_i w_j} - a^2 cov_{x_i x_j} - b^2 cov_{y_i y_j} - c^2 cov_{z_i z_j} - d^2 cov_{v_i v_j})^2 &\equiv \sum_{i=1}^n \sum_{j=i}^n e_{ij}^2 \end{aligned}$$

In order to estimate the total error we calculate the weighted sum of the above expressions. We use weights because the parameters in the two sets of equations do not have the same dimensionality and because the accuracy with which they can be estimated from the data is not the same. Since we are working with sampled distributions we use the standard error for estimating the accuracy with which the mean and the elements of the covariance matrices can be computed. The standard error for the mean is given by $\frac{\sigma_i}{\sqrt{N}}$ where σ_i is the standard deviation for band i , as calculated from the data and N is the number of samples used to represent the given distribution. The standard error for the co-variance is given by $\frac{\sqrt{2cov_{ij}}}{\sqrt{N}}$ where cov_{ij} is the co-variance between bands i and j , as estimated from the sample points [55].

The weighing parameters are set so that the equations with the smaller error should contribute more to the sum, so we weigh each set with the inverse of the corresponding standard error. We also assume that the distributions of the component classes are free of errors therefore, the standard error is calculated only for the mixed

distribution. This is not really true, as the parameters of the distributions of the pure classes are also computed from the image data. However these distributions are represented by a larger number of samples than the mixed class distribution and one may assume that by far the most unreliable term in the equations relating the moments of the various distributions is the term concerning the mixed distribution. Then the total error is given by:

$$E_{total} = \sum_{i=1}^n \left(\frac{N}{var_{w_i}} (\bar{w}_i - a\bar{x}_i - b\bar{y}_i - c\bar{z}_i - d\bar{v}_i)^2 \right) \quad (3.5)$$

$$+ \sum_{i=1}^n \sum_{j=i}^n \left(\frac{N}{2(cov_{w_{ij}})^2} (cov_{w_i w_j} - a^2 cov_{x_i x_j} - b^2 cov_{y_i y_j} - c^2 cov_{z_i z_j} - d^2 cov_{v_i v_j})^2 \right)$$

When we use the Least Squares Error (LSE) method to estimate the proportions we can impose some constraints to the possible solutions and therefore we can use the *Constrained Least Squares* method instead. In our case we have two constraints, namely that the sum of the proportions for any component should be 1 ($a+b+c+d=1$), and the proportion values must be non negative ($0 \leq a \leq 1.0$, $0 \leq b \leq 1.0$, $0 \leq c \leq 1.0$, $0 \leq d \leq 1.0$). We can solve the problem using both the constraints by examining all possible combinations of a , b , c and d to find those that give the minimum square error. This exhaustive search may not be very fast, however it always yields a solution as it finds the best possible solution within the range of acceptable values. One does not need to go to high accuracy when performing the exhaustive search as percentage coverages with accuracy of $\pm 1\%$ are more than adequate.

3.3 Testing of the method with simulated data

The mixed model presented in the previous section was at first assessed using simulated data to represent the pure and the mixed classes. To evaluate this approach one has to create the simulated data very carefully, in order to approximate as much

as possible the data found in real applications.

Since our model conceptualises the data as distributions, the first step to our simulation is to create these distributions. Four six band distributions were artificially created to represent the four pure classes e.g. classes X, Y, Z and V. The means and covariance matrices of each of the simulated distributions were chosen to be the same as those computed from real test sites on a remotely sensed image, that were known to represent pure classes. For example from a site with very dominant soil component we extracted the mean and the covariance matrix in order to create the points of the X distribution. Next, a mixture distribution was created, from the four pure distributions with known mixing proportions as follows: The mean and the covariance matrix of the mixed distribution were derived from the means and the covariances of the pure distributions using the formulae mentioned in the previous section. Using this mean and covariance matrix we created a set of random points to represent the mixed distribution.

We represent each distribution by a set of points and we try to estimate the proportions of the classes in the set of mixed pixels. A set of distributions created this way is shown in *Figure 3.1*, where only two bands are plotted.

The parameters that can vary in this simulation are: the size of the sets used to represent the pure classes, the size of the generated mixed set and the mixing proportions used to generate the mixed set. We examine the effect of the size to the results in order to identify the minimum requirements of our method for reliable statistics for the pure and the mixed sets, and therefore to assess its applicability. Different combinations of proportions are also examined in order to check how the position of the mixed class inside the hyper-volume determined by the pure classes will affect the performance of the classification.

Furthermore, the performance of our model that incorporates the first and the second order moments, is compared to the performance of a model that uses only the means. The latter approach is more similar to the classic approach of mixing

Figure 3.1: A set of simulated distributions to represent the pure classes, projected onto the Band3 - Band4 plane.

models.

All the results presented hereafter are obtained using exhaustive search, with accuracy of $\frac{1}{100}$. The errors presented are the percentage errors of the computed proportions over the corresponding true proportion.

3.3.1 Variation of the size of the distributions

In this experiment we fixed the mixing proportions to $a = 0.1$, $b = 0.2$, $c = 0.3$ and $d = 0.4$.

First we varied the number of samples in the pure classes. All pure classes had the same number of points. The mixed class had 200 points, so as to be considered well defined.

We repeated the experiment for different sizes 100 times, with different randomly

chosen points to represent each set every time, and we calculated the average error and the range of this error in the estimation of each mixing variable. Each experiment was performed twice, once using the means only algorithm, and once using the proposed algorithm. In *Figure 3.2* we plot the errors in a , b , c and d obtained from the two algorithms and for various sizes of the sets that represent the pure classes.

It can be seen that although for many sets of sample points the error of the proposed method was worse than the error of the means only method (notice the overlapping of the error bars), on average, the proposed method performs better than the means only method, particularly, when more than 70 samples were used to represent the pure classes, as in that case the higher order statistics used are more reliable.

Then we varied the number of points in the mixed class. The pure classes used in this experiment had 200 samples each. The samples used were those for which an average error was obtained in the experiment were the size of the mixed samples was varied. Again the experiment was repeated 100 times with different sets each time and for each sample size the average error was computed and the two extreme errors observed were recorded. *Figure 3.3* shows the mean, the standard deviation and the extreme errors in the estimation of a , b , c and d . As it can be seen in these figures the proposed method performed on average better than the means only method, although sets of samples can be found for which the means only method performs better.

In order to create the above figures we assumed that 200 points were enough to represent the mixed class in the first experiment and the pure classes in the second experiment. But it turned out that even though 200 points give quite good results the effect of undersampling is still significant. In order to eliminate the effect of undersampling in the fixed distributions, we increased the number of points used to represent the mixed class in the first experiment and the pure classes in the second experiment to 10,000 points. The results of this experiment are shown in *Figure*

Figure 3.2: Relative mean error of proportion estimation versus the size of the pure classes data sets. The mixed data sample size is 200 and the set of proportions tested is 0.1, 0.2, 0.3 and 0.4. One line represents the performance of a model that uses only the means and the other of a model that uses the means and the variances as well. The error bars represent the standard deviation of the errors for each method over 100 experiments performed for the same set of parameters but with different data sets. (a) Error for $a=0.1$, (b) Error for $b=0.2$, (c) Error for $c=0.3$, (d) Error for $d=0.4$. The top two curves follow the extreme errors that were observed in each set of 100 experiments, for each method.

Figure 3.3: Relative mean error of proportion estimation versus the size of the mixed data set. The pure class data sets were represented by 200 samples and the set of proportions tested is 0.1 0.2 0.3 0.4. The error bars represent the standard deviation of the errors for each method observed over 100 experiments. (a) Error for $a=0.1$, (b) Error for $b=0.2$, (c) Error for $c=0.3$, (d) Error for $d=0.4$. The top two curves follow the extreme errors that were observed in each set of 100 experiments, for each method.

3.4 and 3.5. As we can see from this experiment the means only model and the proposed model have comparable performances.

To test the significance of these results, we run the *t-test* for the results of *Figure 3.2*. Each time the null hypothesis we tested for was: "Are the two error distributions the same?". For each mean error value shown in these figures we used the associated standard deviation of the distribution of errors and the fact that each distribution was formed by running 100 experiments. Then we used this formula to compute the *t test*:

$$t = \frac{\bar{x} - \bar{y}}{s_D} \quad \text{with} \quad s_D = \left(\frac{\text{var}(x) + \text{var}(y) - 2\text{cov}(x, y)}{N} \right)^{\frac{1}{2}}$$

Finally the probability with which the null hypothesis is true was computed. The results are shown in *Table 3.1*. From these results we see that in most cases the difference in the two distributions is significant, with the proposed method most of the times performing better.

Size of data set	a	b	c	d
30	90%	4%	4%	19 %
50	24%	2%	3%	21%
70	44%	0%	0%	4%
90	90%	8%	1%	31%
110	0%	0%	0%	0%
130	1%	0%	0%	0%
150	0%	0%	0%	0%
200	20%	0%	0%	0%
250	0%	0%	0%	0%

Table 3.1: *Estimated significance of the results shown in Figure 3.2.*

3.3.2 Testing different proportion combinations for the mixed class

In this experiment we used the constraint that the proportion of each class was non zero, so as to have always four classes present in the mixed class. The proportions tested were incremented by 0.10 every time. So the proportion for a class could take

Figure 3.4: Relative mean error of proportion estimation versus the size of the pure classes data sets. The mixed data sample size is 10,000 and the set of proportions tested is 0.1, 0.2, 0.3 and 0.4. One line represents the performance of a model that uses only the means and the other of a model that uses the means and the variances as well. The error bar represent the standard deviation of the errors for each method over 100 experiments performed for the same set of parameters but with different data sets. (a) Error for $a=0.1$, (b) Error for $b=0.2$, (c) Error for $c=0.3$, (d) Error for $d=0.4$. The top two curves follow the extreme errors that were observed in each set of 100 experiments, for each method.

Figure 3.5: *Relative mean error of proportion estimation versus the size of the mixed data set. The pure class data sets were represented by 10,000 samples and the set of proportions tested is 0.1 0.2 0.3 0.4. The error bars represent the standard deviation of the errors for each method observed over 100 experiments. (a) Error for $a=0.1$, (b) Error for $b=0.2$, (c) Error for $c=0.3$, (d) Error for $d=0.4$. The top two curves follow the extreme errors that were observed in each set of 100 experiments, for each method.*

one of the following possible values 0.10, 0.20, 0.30, 0.40, 0.50, 0.60 and 0.70. All possible combinations for the above proportions were tested. For each combination the experiment was repeated 100 times with different randomly chosen points to represent the mixed set every time. We used 200 points to represent the pure distributions and 200 points to represent the mixed distribution.

In order to demonstrate the results of this experiment, we selected one pure class, e.g. class X . Then for every possible value of its proportion a , we computed the maximum error in estimation of this proportion for all the possible combinations of the proportions of the other classes (b , c and d). This error is plotted in *Figure 3.6*.

Figure 3.6: *Error in proportion estimation when different combinations of components a , b , c , d are examined.*

Both models had comparable performance. We can see that small proportions are more likely to give rise to larger errors.

3.3.3 Performance in the presence of noise

Given the fact that all remotely sensed data contain some degree of noise due to instrumentation, we tested experimentally how stable our method is to noise. We added Gaussian noise to the data and performed 1000 runs for each level of noise. The noise we applied has a Gaussian distribution with zero mean and diagonal covariance matrix, with variance given by $\sigma^2 = (\text{noise} \times \mu)^2$, where μ is the minimum component of the six components of the four mean vectors of the original data sets, and *noise* is the percentage of noise we want to apply.

The results of this experiment are presented in Table 3.2 for different levels of noise. For this experiment we used pure distributions of 200 points, mixed distribution of 150 points and the mixture was 10%, 20%, 30% and 40%. The numbers given are the mean estimated error and in brackets the standard deviation of this error computed over the 1000 runs for each noise level.

Noise	Means+ Var				Means			
	a	b	c	d	a	b	c	d
0%	0	5	3.3	5.0	0	10.0	3.3	2.5
1%	1.0 (3.0)	6.8 (2.4)	3.1 (1.5)	4.6 (2.1)	0.0 (0.0)	10.4 (1.4)	3.6 (2.0)	1.8 (0.5)
2%	1.3 (3.3)	7.2 (2.8)	3.5 (2.4)	4.9 (2.6)	0.2 (1.4)	11.0 (2.5)	4.7 (3.0)	1.7 (1.3)
5%	4.7 (6.2)	9.7 (13.1)	10.9 (8.0)	7.2 (5.2)	4.3 (5.3)	11.1 (6.0)	8.63 (6.3)	4.2 (3.9)

Table 3.2: *Effect of noise when applied to the mixed distribution. The error at noise level 0% is the error due to subsampling.*

As we can see from Table 3.2 our method can tolerate reasonable levels of noise and yields comparable results to the means only model. This result is pretty good as the added noise is of zero mean and thus does not influence the means but influences significantly the covariances distorting them towards rounder distributions. (If the variance of the noise in the above experiments is compared to the variances of the class distributions, the noise is of the order of 100%).

So far we examined the effect of white noise to the points of the mixed distribution. Another source of discrepancy between the estimated proportions and the real ones may stem from a small systematic error in the observed reflectances. This

systematic error may be attributed to changes of the background reflectance (soil reflectance) or changes of the illumination.

To test how stable is our method to this sort of disturbance we shifted every pixel of the mixed distribution by the same vector. Every element of this vector is a fraction of the corresponding minimum value of the mean vectors of the pure classes. In order to represent the mixed distribution we used 200 points, so we expected that even when no shift was applied we would have some small errors that arise from the fact that the distributions were underrepresented. Such systematic errors affect only the means and leave the variances unaffected. Thus, as expected, the estimation errors remain constant when only the variances were used for the proportion estimation and they increase less dramatically when the means and the variances model was used in comparison to the means only model. The errors for various levels of added biased noise (measured as percentage at the minimum mean component) can be seen in *Table 3.3*.

Biased Error	Means+Var				Means				Var			
	a	b	c	d	a	b	c	d	a	b	c	d
0%	10.0	10.0	10.0	5.0	0.0	5.0	3.3	0.0	30.0	10.0	6.6	7.5
1%	20.0	15.0	13.3	7.5	20.0	20.0	20.0	10.0	30.0	10.0	3.3	5.0
2%	30.0	20.0	13.3	7.5	40.0	40.0	46.6	25.0	30.0	10.0	6.6	7.5
5%	50.0	30.0	10.0	5.0	80.0	80.0	93.3	50.0	30.0	10.0	3.3	5.0

Table 3.3: *Effect of biased error in the mixed distribution of 200 points. The numbers given are the errors with which the true proportions ($a = 10\%$, $b = 20\%$, $c = 30\%$ and $d = 40\%$) were estimated by the various methods.*

Another experiment that we performed was to use plenty of pixels to represent the distributions so as to have zero error when no shift was applied. Then we shifted the mean of the mixed distribution by a given vector, calculated as before. The results of this experiment can be seen in *Table 3.4*. From this table it is obvious that the method that takes into consideration the variances is much more stable than the method that utilizes only the means of the distributions.

Biased Error	Means+Var				Means				Var			
	a	b	c	d	a	b	c	d	a	b	c	d
0%	0	0	0	0	0	0	0	0	0	0	0	0
1%	20	10	7	5	30	40	50	25	0	0	0	0
2%	30	20	10	5	60	70	83	42	0	0	0	0
5%	60	50	17	2	110	100	97	50	0	0	0	0

Table 3.4: *Effect of biased error in a very well represented mixed distribution. The numbers given are the errors with which the true proportions ($a = 10\%$, $b = 20\%$, $c = 30\%$ and $d = 40\%$) were estimated by the various methods.*

3.4 What if pure class sites are not available?

As we have seen so far, for pure classes we wish to use sets of pixels representative of the pure classes extracted from the remotely sensed image itself. However sometimes, especially if the terrain tends to vary at smaller scales than the size of the test sites, it is difficult to find homogeneous test sites that belong solely to a given pure class.

A solution to this problem is to derive the attributes of the pure classes, or else the mean vectors and the covariance matrices, from test sites for which ground measurements are available [56, 28]. According to our model, for a given site k we have the following equation for the means in a given band i ($i = 1, \dots, n$).

$$\overline{w}_{ki} = a_k \overline{x}_i + b_k \overline{y}_i + c_k \overline{z}_i + d_k \overline{v}_i$$

If we now consider m sites $k = 1, \dots, m$ the above equation can be written in a matrix form as follows:

$$Y = PX$$

where Y is a $m \times n$ matrix that contains the mean vectors (n -Dimensional) for all m sites, P is a $m \times 4$ matrix which contains the quadruples of proportion estimates for the four pure classes within each site, and X is a $4 \times n$ matrix which contains the mean vectors for the four pure classes. In this case we know the matrices Y and P and we want to estimate matrix X . To do so we used *Singular Value Decomposition*

(SVD) to solve the above overdetermined system. According to this method we calculate a matrix X that yields an estimator $\hat{Y}$ for which some appropriate measure of the error $Y - \hat{Y}$ is suitably small.

In order to improve the results of this method, given that the accuracy of the ground measurements may not be the same for all the sites involved in the analysis, we can drop from the analysis those sites for which the residual after the fitting is large, or in other words to remove the univariate outliers. In our case univariate outliers are sites with extreme values in at least one of the bands [75].

Thus we choose a threshold value for the residual and we discard the sites that have residuals above this threshold and repeat the analysis. We measure the residual for each element y_{ki} of matrix Y as $100 \times \frac{|y_{ki} - \hat{y}_{ki}|}{y_{ki}}$, where $\hat{y}_{ki}$ is the estimated value of y_{ki} . If this error is larger than 40 we discard the corresponding site.

With the above procedure we can calculate the mean vectors of the four pure classes. A similar method can be used to derive the covariance matrices of the pure classes. According to our model, for the covariance between bands i, j ($i, j = 1, \dots, n$) of a given site k we have:

$$cov_{w_{ki}w_{kj}} = a_k^2 cov_{x_i x_j} + b_k^2 cov_{y_i y_j} + c_k^2 cov_{z_i z_j} + d_k^2 cov_{v_i v_j}$$

Again if we consider m sites $k = 1, \dots, m$ the above equation can be written in a similar matrix form as follows:

$$\tilde{Y} = \tilde{P}\tilde{X}$$

Where $\tilde{Y}$ is a $m \times \frac{n(n-1)}{2}$ matrix, where each line of this matrix contains the elements of the covariance matrix of the corresponding site. $\tilde{P}$ is a $m \times 4$ matrix which contains the quadruples of the squares of the proportion estimates for the four pure classes within each site and $\tilde{X}$ is a $4 \times n^2$ matrix which contains the elements

of the covariance matrices for the four pure classes. We solve this system with a method similar to the one used for the means.

3.5 Application to real data

Since the simulation results showed that our model performs well, we then tested it with a real application. The aim was to decide on the type of vegetation in an area located close to Athens, the capital of Greece in the province of Attica.

A Landsat TM image that covers the whole area in question was used. This image was corrected by the Institute for Digital Image Processing (DIBAG) at Joanneum Research. The image was first radiometrically calibrated and converted to a reflectance image. This was then atmospherically corrected using the LOWTRAN 7 radiative transfer model. The effects of topography were also corrected using the *C correction* method as it was the adjacency effect using a 10×10 kernel of neighbouring pixels. Finally the image was geocoded using the nearest neighbour resampling method with accuracy of 1 pixel.

The primary vegetation in this study area is composed of conifers, mainly *aleppo pine* (*Pinus halepensis*) with average height 90cm and a variety of shrub species such as *maquis* and *phrygana* with average height 75cm . The surface geology is characterized by limestone and metamorphic rocks. The topography of the area is complex due to excessive relief diversity. The slopes tend to be medium to steep.

Four test areas have been selected because there were forest fires in each of these areas within the last ten years. The test areas are located north (test area of Varnavas), north east (test area of Penteli), west (test area of Pateras) and southeast of Athens (test area of Lavrio). The training site data used in this work were collected using visual field inspection with an estimated accuracy of 5%. They have been collected by the Institute of Mediterranean Forest Ecosystem - National Agricultural Research Foundation (NARF) of Greece for evaluating the potential

forest vegetation regeneration, risk of erosion, and risk of desertification.

The vegetation cover is categorized in four main classes: *bare soil*, *aleppo pine*, *maquis* and *phrygana*. For training we used different sites for which ground data were available. These sites are at least $150 \times 150m^2$ and therefore corresponded to more than 30 pixels in the TM image.

Since we have no regions solely composed of one class, we derive the attributes of the real pure classes (mean vectors and covariance matrices) from sites that we know their composition, using regression techniques. By taking into consideration all pixels in the data sets when computing these covariance matrices, it is possible to underestimate the off diagonal components of the matrix due to the spatial correlation of adjacent pixels. Ideally, we should use a "colouring" scheme like the one proposed in [4] to avoid this error. However, in our case, the sites we have data for consist of very few pixels and we cannot afford the luxury of dividing them into subsets by "colouring"; the statistics would simply become too unreliable. At first we tried to use in the analysis all 39 sites for which ground measurements were available. After the fitting was performed we realized that quite a lot of points were placed far from the fit, as we can see in *Figure 3.7*. In this figure we plotted the largest residual of a site with respect to the least square error solution obtained from all 39 sites. We see that there are very large errors indicating lack of consistency in the data.

To overcome this problem we chose to analyse sites from Pateras separately from the other sites, because we noticed that they exhibited different characteristics. So the linear regression analysis to identify the pure classes was performed two times, once for Pateras and once for the other areas, using all the available sites for each area. In the second case four test sites B-3, B-4, B-5 and PD1-4 were excluded from the analysis, according to their residual error after regression was performed.

When the pure classes were estimated our model was used to identify the composition of the given sites. At first the performance of our model was tested using the

Figure 3.7: *Highest measured error for the elements of the mean vector for each site.*

39 training sites. Two criteria were used to evaluate the obtained results. According to the first criterion a classification result is considered a “hit” if the dominant class is identified correctly, otherwise we have a “miss”. The second criterion is more strict, a classification result is considered a “hit” if the dominant class is identified correctly with accuracy $\pm 15\%$. Moreover two approaches were evaluated, one that made use only of the mean spectral values of the burned areas and one that uses the means and covariances as well.

With the means only model according to the first criterion 30 sites out of the 39 were classified correctly and according to the second criterion 21 sites out of the 39 were classified correctly. Using the mean and covariances model, according to the first criterion 30 sites were classified correctly, while according to the second criterion we had 23 “hits”. The detailed results obtained for these sites are shown in *Tables 3.5 and 3.6*.

In these tables S stands for soil, AP for Aleppo Pine, M for maquis and P for Phrygana. The numbers are percentages of coverage by the corresponding class.

No	Name	Ground				Mean				Mean+Var			
		S	AP	M	P	S	AP	M	P	S	AP	M	P
1	L-1	20	30	0	50	11	37	0	52	4	25	0	71
2	L-2	40	25	5	30	17	1	27	55	9	0	16	75
3	L-3	15	30	0	55	0	23	0	77	12	37	13	38
4	L-4	35	20	10	35	21	5	2	72	31	20	9	40
5	L-5	55	15	10	20	57	20	0	23	48	13	0	39
6	L-6	10	70	0	20	0	67	0	33	9	44	21	26
7	PD1-1	15	55	0	30	19	67	8	6	19	59	15	7
8	PD1-2	40	20	0	40	55	30	0	15	40	14	4	42
9	PD1-3	25	25	0	50	54	29	2	15	33	15	9	43
10	PD1-4	15	45	30	10	25	0	0	75	7	55	18	20
11	PD1-5	10	30	0	60	19	0	31	50	7	60	16	17
12	PD1-6	20	60	0	20	29	39	32	0	5	57	19	19
13	PD1-7	20	35	0	45	31	46	6	17	30	30	18	22
14	PD2-1	0	75	0	25	0	75	0	25	7	48	21	24
15	PD2-2	5	55	5	35	4	21	31	44	8	53	18	21
16	PD2-3	0	70	5	25	1	61	0	38	0	60	0	40
17	PD2-4	0	60	0	40	18	79	0	3	20	80	0	0
18	B-1	15	0	65	20	26	0	53	21	30	0	64	6
19	B-2	10	0	60	30	7	17	61	15	7	43	25	25
20	B-3	30	0	40	30	0	0	56	44	0	36	33	31
21	B-4	5	35	0	60	0	87	0	13	0	52	32	16
22	B-5	0	0	65	35	0	62	38	0	0	55	30	15
23	B-6	5	0	55	40	3	14	48	35	3	13	48	36

Table 3.5: Comparison of the mixed model results with the ground data for Lavrio, Penteli and Varnavas training sites.

Under the heading “Mean” we give the results obtained by using the means of the distributions only in the mixing model. Under the heading “Means+Var” we give the results obtained by using both the means and the covariance matrices of the distributions in the mixing model.

At the second stage of the evaluation of our method, we tested our model using 14 sites (the *test sites*) that they have not been used for the derivation of the means or covariances. According to the first criterion the means only model classified correctly 8 sites, while according to the second criterion it classified correctly 6 sites. The model that used the means and the covariances had 8 “hits” in accordance to the first criterion and 7 “hits” for the second criterion. The detailed results obtained for these sites are shown in Table 3.7. Note that they were obtained using for each site as pure classes those obtained from the corresponding training set. That is for

No	Name	Ground				Mean				Mean+ Var			
		S	AP	M	P	S	AP	M	P	S	AP	M	P
1	P-1	30	30	0	40	42	39	19	0	44	30	26	0
2	P-2	55	10	15	20	50	40	8	2	50	34	7	9
3	P-3	30	30	0	40	68	24	0	8	52	2	0	46
4	P-4	20	35	0	45	18	25	26	31	18	29	24	29
5	P-5	0	65	10	25	0	82	16	2	0	39	32	29
6	P-6	55	10	25	10	37	16	18	29	40	14	21	25
7	P-7	20	40	10	30	9	20	36	35	9	40	29	22
8	P-8	5	30	55	10	0	35	64	1	0	21	65	14
9	P-9	10	15	60	15	13	0	55	32	24	0	70	6
10	P-10	10	15	65	10	3	34	49	14	6	19	58	17
11	P-11	30	10	55	5	46	0	53	1	29	14	31	26
12	P-12	20	15	55	10	19	10	58	13	13	20	47	20
13	P-13	15	60	0	25	30	24	2	44	30	30	1	39
14	P-14	45	25	0	30	47	0	6	47	49	2	11	38
15	P-15	10	60	20	10	11	76	10	3	6	94	0	0
16	P-16	20	40	30	10	16	44	29	11	17	40	30	13

Table 3.6: *Comparison of the mixed model results to the ground data for Pateras.*

all the Pateras sites we used one set of pure class signatures and for all other sites another set.

It can be seen that both models perform quite well overall, in the sense that they produced results that agree with the ground data. In conclusion we may say that both methods of mixed pixel classification achieved classification accuracy 57% (8 sites) when applied to the 14 test sites in the sense that the correct dominant class was identified. The model that utilises the covariance matrices had better performance in identifying the coverage of the dominant class with error in the range $\pm 15\%$.

The best results for the test sites were obtained when training sites and test sites were situated near by. This is something one should expect since some of the classes, especially bare soil, vary in reflectance signature with rangeland type, so it is clearly desirable to compare the results of a stratified analysis where each rangeland type is treated separately. In order to have a visual interpretation of the separability of the pure classes obtained this way, we plotted the derived mean signatures of the pure classes in the six bands, *Figure 3.8*.

No	Name	Ground				Mean				Mean+Var			
		S	AP	M	P	S	AP	M	P	S	AP	M	P
1	BAR-T1	5	0	50	45	0	27	53	20	0	49	32	19
2	BAR-T2	0	0	85	15	0	36	38	26	0	0	74	26
3	LAV-T1	15	65	0	20	73	27	0	0	49	6	0	45
4	LAV-T2	10	50	0	40	61	39	0	0	32	26	4	38
5	PEN-T1	35	35	5	25	37	63	0	0	29	42	8	21
6	PEN-T2	30	50	5	15	22	26	36	16	18	42	16	24
7	PEN-T3	60	10	0	25	63	0	29	8	63	0	34	3
8	PEN-T4	5	55	10	30	0	59	0	41	0	64	3	33
9	PAT-T1	25	15	20	40	11	0	67	22	13	15	62	10
10	PAT-T2	15	10	55	20	10	8	54	28	14	16	56	14
11	PAT-T3	30	0	50	20	0	24	54	22	0	27	50	23
12	PAT-T4	10	75	10	5	6	32	39	23	4	28	34	34
13	PAT-T5	70	20	0	10	30	29	0	41	30	31	0	39
14	PAT-T6	70	0	20	10	59	0	17	24	50	11	4	35

Table 3.7: Comparison of the mixed model results to the ground data for testing sites.

3.6 Discussion and conclusions

It is clear from the above experiments and results that the proposed approach is promising. Its sensitivity and accuracy have been demonstrated with the simulated data. There are various reasons for which the discrepancies when applied to real data might have arisen: First, there is some time difference between the time the image was taken (September 1993) and the time the ground data were collected (February 1993). Second, a major problem when analysing a set with as small number of pixels as the sets used here, is to locate in the image the exact position where the field data had been collected. So there is a great need for accuracy of geometric registration. As spatial variation of vegetation is high, the selection of pixels that fit the measured ground elements is critical: a shift of one pixel yields significantly different training sets, regression coefficients and cover estimates. Finally, the ground data are only estimates by people who visited the sites and it is possible that these estimates are themselves erroneous.

There seem to be a systematic difference between the images from the Pateras training area and all other training areas in terms of overall brightness. This is

Figure 3.8: *Derived means of the four pure classes. (a): 23 sites from Lavrio, Penteli and Varnavas. (b): 16 sites from Pateras.*

inexplicable as the data for all sites came from the same TM image which was radiometrically corrected. We can attribute the discrepancy to either of two factors: The radiometric corrections were performed locally and not on the full image and there may be some discrepancy between the corrections applied to individual sites. Alternatively, it is possible that the bare soil that is present in the Pateras site is of different type from that in the other sites and this causes a distortion in the spectral signature of all classes due to nonlinear reflectivity effects. An alternative method using accurate radiometric corrections and reference endmembers taken from a spectral library or field spectrometry could be more relevant. Clearly, these points have to be further investigated and the source of any discrepancy which arises identified. As a conclusion, we consider that the proposed approach has a lot of potential and it certainly deserves further investigation.

The simulation proved that the proposed model is at least as reliable as the model that makes use of the means only, when different sizes were used to represent the pure and the mixed distributions and for any proportion combination. In the presence

of low levels of noise our model proved quite reliable and when the distribution distortion was due to a systematic error, our model was by far more stable than the means only model.

One of the strong points of the proposed method is that it allows the introduction of more pure classes to the model, since it increases the number of equations used to estimate the proportions. Also it produces more stable results when a biased error is present to the mixed distribution that may be attributed to changes of illumination. Finally, the "pure" classes do not need to be entirely distinct and non-overlapping as the equations can cope with overlap. The only case the method cannot deal with is when the mean of the mixed class is not enclosed by the square defined by the means of the four "pure" classes. The obvious conclusion in that case is that the site in question cannot just be represented as a mixture of the four classes considered, and further "pure" classes have to be introduced into the system.

Chapter 4

Mixed pixel classification using Robust Statistics

4.1 Introduction

In the previous chapter we have proposed an algorithm for the unmixing of sets of mixed pixels, where the value of each pixel is considered as a random variable and higher order moments of the distributions of these random variables were taken into account. There may be cases, however, where the set of pixels that has to be unmixed may contain several outlier pixels. Further, the regions which one is interested in monitoring may only consist of a handful of pixels in the TM data. Then the calculation of statistical descriptors of their distributions may not be particularly reliable. Thus, a more robust way of solving the problem is needed.

In this chapter we present such a robust method for mixed pixel classification where the *Hough transform* and the *Trimmed means* methods are used to classify small sets of pixels. We compare the performance of these methods with the *Least Squares Error* method presented in the previous chapter and we show that in the presence of outliers the trimmed means method is much more reliable than the traditional Least Squares Error method. The method is exhaustively tested using

simulated data and it is also applied to real Landsat TM image for which ground data are available.

In the next section we discuss the Hough Transform in general. In section 4.3 we describe our method in detail and derive the parameters used. In section 4.4 we demonstrate the robustness of the method with simulated data. In section 4.5 we discuss the problem of determining the reflectances of the pure classes from sets of mixed pixels, and in section 4.6 we apply the method to some real data. Finally we present our conclusions in section 4.7.

4.2 Brief overview of the Hough Transform

The *Hough Transform*, HT, converts a difficult global detection problem in image space into a more easily solved local peak detection problem in a parameter space [36, 39]. The key ideas of the method can be illustrated by considering sets of collinear points in an image. A set of points which lie on a straight line can be defined by a function f , such that:

$$f((a, b), (x, y)) = y - ax - b = 0 \quad (4.1)$$

where a and b are two parameters, the slope and intercept, which characterise the line. Equation (4.1) maps each value of the parameter combination (a, b) to a set of image points. The mapping is one to many from the space of possible parameter values to the space of image points. The HT uses the idea of equation (4.1) which can be viewed as defining a one to many mapping from image points to a set of possible parameter values. This corresponds to calculating the parameters of all straight lines which belong to the set that pass through a given image point (x, y) :

$$f((x, y), (a, b)) = y - xa - b = 0 \quad (4.2)$$

In the case of a straight line, each image point (x, y) defines a straight line in the (a, b) parameter space. Points that are collinear in image space all intersect at a common point in parameter space and the coordinates of this parameter point characterises the straight line connecting the image points. The HT identifies these points of intersection in parameter space. Determination of the point of intersection in parameter space is usually considerably easier than detecting extended point patterns in image space.

4.3 The proposed method

In the linear mixing model adopted here, it is assumed that the pixel value in any spectral band is given by the linear combination of the spectral responses of each component within the pixel, so the model can be expressed as:

$$w = ax + by + cz \quad (4.3)$$

where w is the known spectral reflectance of a mixed pixel, x , y and z are the known spectral reflectances of the three possible cover components within the mixed pixel, and a , b and c are the proportions of the components contained in the mixed pixel that have to be estimated.

The linear mixing equation above, is actually the equation of a plane in luminance space where we measure one type of luminance along each axis. What we are interested in identifying are the parameters a , b and c of this plane. The method usually used for this purpose is that of least squares fitting. It is well known, however, that the method of least squares is particularly sensitive to outliers. What we propose in this chapter is the use of Hough transform to identify the best values of a , b and c . Hough transform is known to be a robust technique which can tolerate large amounts of outliers and still produce good results. In its most commonly used

form it is used to identify straight lines in images [36], but more generally Hough transform can be thought of as a transformation into the parametric domain where we seek to identify sets of real data that indicate the same values of the parameters for the parametric surface they define.

In our case this surface is a plane defined in the 3D domain (x, y, z) , which is parameterised by different values of (a, b) . Since the values of these parameters have to sum up to 1, we can eliminate the third one in terms of the other two and equation (4.3) becomes:

$$w - z = (x - z)a + (y - z)b \quad (4.4)$$

Thus, our method consists of the definition of an accumulator 2D array defined in the parametric (a, b) domain. For each quadruple (x, y, z, w) we have a different line defined in the (a, b) domain. This line intersects various cells of the accumulator array the occupancy number of which is incremented by 1. When all possible quadruples of the data have been considered, the highest peak in the accumulator array defines the best values of the mixing parameters a and b .

In practice, instead of using the one to many mapping implied above we can choose two quadruples at a time. Thus, instead of computing one parametric curve for each quadruple, from the two selected quadruples (x_1, y_1, z_1, w_1) , (x_2, y_2, z_2, w_2) a set of a and b values is calculated. Then only the corresponding cell in the accumulator array is incremented.

Usually, the data contain two or more bands, so the two quadruples can be selected from different bands. For instance, with two-band data, the first quadruple concerns the luminance values of the four pixels chosen in the first band and the second one consists of the luminances of the same pixels in the second band. The use of sets of quadruple values from separate bands sharpens the peaks in the accumulator array since usually separate bands do not yield similar parametric lines,

whereas quadruples from the same band do. As a result, the intersections of the lines are more accurate.

The next step in our procedure is to estimate the appropriate bin size in the accumulator space for the two parameters a and b . In our applications we do not really need to know the values of a and b to better than two significant figures accuracy, so the size of our accumulator array will not be larger than 100×100 , but generally it will be smaller. The problem with a high resolution array is not only that it is computationally expensive but that it is some times unrealistic to expect to obtain results with such a high accuracy.

In (4.4) x , y and z are the reflectances of pure classes. Due to intraclass variability, however, each of these variables can be thought of as a random variable distributed according to some distribution, which given enough data, can be modelled parameterically. As the reflectances of the pure classes that enter into equation (4.4) are drawn from these distributions, it is obvious that w is expected to have its own variability, and that even if we have an exact value for it, parameters a and b cannot possibly be estimated with accuracy higher than the accuracy dictated by the intraclass variability of x , y and z .

We can better explain that by realizing that a given reflectance w of a certain mixed pixel can be created by more than one combinations of values x , y and z , all of which could be legitimate reflectances of the corresponding pure classes. Therefore, the values of a and b we find must reflect this uncertainty. Thus, the bin sizes we have to use in the accumulator array have to reflect the uncertainties in x , y and z .

When we solve (4.4) for a , or b we get:

$$\begin{aligned} a &= \frac{w - z - b(y - z)}{x - z} \\ b &= \frac{w - z - a(x - z)}{y - z} \end{aligned}$$

The worse case error for a and b can then be computed by:

$$\begin{aligned}\Delta a &= \left| \frac{\partial a}{\partial x} \right| \Delta x + \left| \frac{\partial a}{\partial y} \right| \Delta y + \left| \frac{\partial a}{\partial z} \right| \Delta z \\ \Delta b &= \left| \frac{\partial b}{\partial x} \right| \Delta x + \left| \frac{\partial b}{\partial y} \right| \Delta y + \left| \frac{\partial b}{\partial z} \right| \Delta z\end{aligned}$$

Thus the error for a and b turns out to be:

$$\begin{aligned}\Delta a &= \left| \frac{a}{x-z} \right| \Delta x + \left| \frac{b}{x-z} \right| \Delta y + \left| \frac{b+a-1}{x-z} \right| \Delta z \\ \Delta b &= \left| \frac{a}{y-z} \right| \Delta x + \left| \frac{b}{y-z} \right| \Delta y + \left| \frac{b+a-1}{y-z} \right| \Delta z\end{aligned}$$

Note that a and b are non-negative numbers, while $a+b \leq 1$. If for simplicity we consider that x , y and z vary within the same range, i.e. if we assume for the moment that $\Delta x \sim \Delta y \sim \Delta z \equiv \sigma$, we can see that the uncertainty in a and b is:

$$\begin{aligned}\Delta a &\sim \frac{\sigma}{|x-z|} \\ \Delta b &\sim \frac{\sigma}{|y-z|}\end{aligned}\tag{4.5}$$

The above assumption is clearly an oversimplification as there is no reason to expect that the reflectances of each pure class due to intraclass variability vary within the same range of values. We may say that for a conservative estimate of the error in the parameters we shall take all these ranges of the reflectances of the pure classes to be equal to the largest one.

For multiband data we shall have different ranges of the reflectances of the pure classes for each band. As we want to combine the data from as many bands as possible in order to estimate the mixing proportions, we must have accumulator arrays that are compatible with each other. For this reason we use the minimum value calculated from equations (4.5) to decide the size of the accumulator cells.

The Hough transform picks the peak in this accumulator array and thus specifies the sought set of parameter values as the values of the parameters at the centre of the picked cell. We shall refer to this method as the Standard Hough Transform (SHT).

An alternative approach is to set up an accumulator array which has the finest resolution we desire, say 100 bins along each of the two axes. When we finish with the accumulation process using this fine resolution we identify the peak in the accumulator space and around this peak we place a window with size calculated from (4.5) as mentioned above. Within this window we calculate the mean values of a and b . We shall refer to this algorithm as the Trimmed Means (TM) approach.

A variation of this method is to place similar size windows all over the accumulator space, calculate the mean within each window and then select the maximum. Note that the only difference between this method and that of SHT is that SHT picks the mode of the accumulator array, while the last method calculates the trimmed mean inside each cell before the maximum is picked. This last method was found to be less accurate than the TM method described earlier, so in the experiments that follow only the SHT and the TM methods will be compared. The LSE method will also be used as a benchmark.

4.4 Testing of the methods with simulated data

The methods presented in the previous section were at first assessed using simulated data to represent the pure and the mixed classes. To evaluate this approach one has to create the simulated data carefully, in order to approximate as much as possible the data found in real applications. The data can be conceptualised as distributions, so the first step to our simulation is to create these distributions. Since we do not know what distribution will model this type of data most accurately we tested two types of distributions: normal and uniform.

Three such distributions were artificially created to represent the three “pure” classes e.g. classes X, Y and Z. The mean and the spread of the simulated distributions were chosen to be the same as those computed from real test sites on a remotely sensed image, that were known to represent almost pure classes. Next, a mixture distribution was created, from the three pure distributions with known mixing proportions, as described in chapter 3.

4.4.1 Time and accuracy concerns

The Hough Transform is known to require heavy computations and therefore we give an estimate of the CPU times of the Hough methods presented in this chapter. The times presented here are expressed in seconds and were estimated using an Alpha workstation. The main parameter that may influence the execution time is the number of points used to represent the various classes.

Figure 4.1: *CPU times estimated using an Alpha workstation versus the number of points representing each class.*

As shown in *Figure 4.1* the computation time increases exponentially with the

number of points that represent each class. For sets with less than 100 points the computation time remains reasonably small.

Since the experiments with the simulated data mentioned below were carried out using only 30 samples to represent each class, in order to assess the effect of subsampling the underlying distributions, we estimated the error for the same proportion (30%) using different datasets derived from the same distributions. The estimation errors are shown in *Figure 4.2*.

Figure 4.2: *Influence of the choice of datasets.*

As we can see from *Figure 4.2* the estimation error depends on the selected datasets, that is why we will present our results in the form of an average error over 100 experiments in order to reduce the effect of subsampling.

4.4.2 Normal distributions

We represent each distribution by a set of points and we try to estimate the proportions of the classes in the set of mixed pixels. Some sets used to represent the “pure” classes are plotted in *Figure 4.3*. Each set consists of about 30 samples. Only two

bands were used for all the experiments described below.

Figure 4.3: *Some example sets used to represent the three “pure” classes.*

The performance of TM and the SHT is compared with that of the least square error solution (LSE), presented in the previous chapter, which was used as a benchmark. The results are presented in terms of the percentage errors of the computed proportions over the corresponding true proportion.

4.4.2.1 Testing different proportion combinations for the mixed class

To test the behaviour of the algorithms for different mixture proportions, we ran the following experiments: We created a series of datasets where one class is represented by some fixed proportion, and the two other classes take all possible combinations of proportions so that the sum of the three is always 100%. (For simplicity we restricted ourselves to proportions that are integer multiples of 10.)

We ran then the algorithm as many times as we had combinations of proportions and picked from all the runs the maximum error in estimating the value of the

proportion that was kept fixed. Further, to eliminate the error of the representation of each class by 30 points only, we repeated this experiment 100 times for different datasets drawn from the same distributions. Then we calculated the average error over these 100 runs for both algorithms and plotted it versus the value of the proportion that was kept fixed each time (and to which the calculated error was referring).

The results are shown in *Figure 4.4*. All algorithms have higher errors for small proportions, as it was expected, due to the natural intraclass range of possible values which may result to negative or exactly 0 proportion values (which our programs reject) and to positive but near 0 values (which our programs accept) when incompatible or near incompatible quadruples are picked from the data. The TM performs better than the SHT for small and large proportions since it smooths the effect of underestimation or overestimation, which as we mentioned before is more likely to happen in the extreme cases. For the middle proportions that are in general better defined the loss of resolution takes its toll and the TM performs slightly worse than the SHT.

As expected the LSE method is the best. However, the use of robust statistics is not expected to outperform LSE in general. It is expected to outperform LSE in the presence of outliers.

4.4.2.2 Presence of outliers

In this experiment some outliers were added to the mixed set. The sets used to represent the “pure” classes were the same for all the experiments, described below and are plotted in *Figure 4.5*. The three pure classes were generated using the same covariance matrix. Each set consisted of about 30 samples. The proportions used to create the mixed sets were $a = 30\%$, $b = 60\%$ and $c = 10\%$. Only two bands were used for the proportion estimation.

The parameters that vary in the following experiments are: the type of outliers,

Figure 4.4: Average maximum relative errors in estimating different proportions for *TM*, *SHT* and *LSE*.

the number of outliers and the proportions of the outliers added to the mixed set. We can distinguish two main types of outliers. The ones that follow a certain pattern e.g. a cluster (*coherent outliers*) and those that have random proportions (*random outliers*). The influence of the outliers depends on their distance from the mixed distribution in the luminance space and therefore various distances have been examined to demonstrate how they affect the obtained results.

a. Coherent Outliers

The outliers of this type tend to create clusters in an arbitrary distance from the mixed distribution in the luminance space. Such outliers are shown in *Figure 4.5*.

Distant outliers may be present in the mixed set due to the existence of another class that we have no data to describe. At first we considered outliers placed at different distances from the mixed set. The unit used to express these distances was the average standard deviation of the mixed distribution. The standard deviation of the cluster of outliers was half the average standard deviation of the mixed set.

Figure 4.5: *Outliers placed at a given distance (3, 6, and 9 times the standard deviation) from the mean of the mixed set.*

(The spreading of the mixed set is really described by a covariance matrix and not a standard deviation. However, an average standard deviation can be defined by taking the average of the square roots of the diagonal elements of the covariance matrix.)

Different numbers of outliers were also examined. Each experiment was repeated 100 times, each time a different set to represent the “mixed” class was drawn. Then the average and the standard deviation of the errors in estimating the proportions were calculated.

The results obtained from this experiment are shown in *Table 4.1* for TM, SHT and the LSE method. The true proportions of the mixed distribution were $a = 30\%$, $b = 60\%$ and $c = 10\%$. Note that as expected, the LSE method is the most sensitive to outliers, while the two robust methods are quite resilient to their presence, with the TM giving the best estimate, especially for proportions $b = 60\%$ and $c = 10\%$. The error in the estimation of proportion c is the largest one, which was expected since c is small.

Dist	Out	TM			SHT			LSE		
		a	b	c	a	b	c	a	b	c
3	10%	20 (10)	7 (7)	66 (50)	15 (13)	10 (8)	87 (35)	30 (18)	9 (6)	110 (61)
-	20%	20 (9)	6 (6)	61 (51)	51 (34)	15 (11)	10 (6)	87 (41)	16 (8)	176 (121)
-	30%	21 (9)	6 (5)	64 (49)	15 (11)	9 (6)	89 (33)	63 (34)	25 (13)	222 (152)
-	40%	17 (11)	6 (5)	56 (52)	18 (14)	8 (6)	86 (48)	82 (29)	35 (16)	263 (215)
6	10%	19 (9)	6 (5)	65 (50)	17 (11)	8 (6)	90 (34)	47 (33)	16 (8)	152 (103)
-	20%	18 (8)	6 (5)	66 (49)	14 (11)	8 (5)	84 (39)	78 (30)	34 (15)	235 (194)
-	30%	19 (11)	6 (5)	65 (57)	18 (14)	8 (6)	93 (48)	88 (38)	44 (24)	195 (211)
-	40%	19 (9)	7 (6)	62 (55)	16 (12)	10 (6)	93 (43)	101 (40)	57 (26)	233 (249)
9	10%	17 (8)	6 (5)	62 (45)	15 (9)	8 (5)	80 (35)	65 (30)	24 (12)	184 (143)
-	20%	20 (9)	6 (5)	67 (54)	16 (12)	9 (6)	86 (41)	83 (36)	45 (27)	275 (270)
-	30%	19 (10)	6 (5)	57 (51)	15 (11)	9 (6)	89 (36)	110 (44)	62 (29)	260 (284)
-	40%	18 (12)	7 (7)	69 (66)	20 (12)	9 (7)	97 (55)	126 (60)	71 (28)	236 (277)

Table 4.1: *Effect of coherent outliers present in the mixed distribution. The numbers in the table are the average errors over 100 experiments conducted. Inside parentheses is the standard deviation of these errors.*

If the cluster of outliers is within the convex hull defined by the “pure” class sets we may be able to identify the two mixture compositions. In order to achieve this, we must examine the second significant peak in the Hough space as well. The separability of the two peaks in the Hough space depends on how many points each of the mixture sets has, and how similar the two mixtures are. *Figure 4.6* shows an example of a Hough space with two peaks clearly separated.

b. Random Outliers

This type of outliers does not form a second mixture distribution but they can be scattered anywhere on the feature plane. In this experiment, for a certain number of random outliers, a number of mixed sets were generated and tested. The average and the standard deviation of the errors in proportion estimation based on 100 experiments are shown in *Table 4.2*.

Out	TM			SHT			LSE		
	a	b	c	a	b	c	a	b	c
10%	21 (12)	6 (10)	69 (85)	18 (16)	9 (12)	93 (88)	31 (26)	11 (7)	122 (91)
20%	19 (9)	6 (4)	59 (43)	15 (10)	8 (6)	86 (24)	38 (30)	15 (10)	133 (102)
30%	19 (13)	6 (7)	56 (52)	18 (16)	9 (8)	91 (47)	46 (33)	17 (13)	160 (128)
40%	24 (29)	9 (16)	73 (87)	22 (33)	12 (17)	102 (88)	56 (33)	20 (14)	181 (154)

Table 4.2: *Effect of random outliers present in the mixed distribution. The numbers in the table are the average relative errors over 100 experiments conducted, and inside parentheses is the standard deviation of these errors.*

Figure 4.6: *The mixed set contains points from two mixtures: half of the points belong to a mixture with composition 30-60-10 and the other half to a mixture with composition 30-10-60. The two clusters indicate the two mixtures.*

The Hough methods, and TM in particular performed well, while the LSE method is more significantly affected by the outliers, although not as much as by the presence of the coherent outliers, because the effect of random outliers tends to average out. Thus, the more the outliers are and the more uniformly distributed about the mixed distribution they are, the better the LSE method will perform because the mean of the mixture distribution will not be affected by their presence. However, such an improvement in the performance of the LSE method is clearly artificial.

4.4.3 Uniform distributions

In order to create the sets to represent the pure and the mixed classes in the previous experiments, we assumed that the data follow the normal distribution. After performing a normality test on the real data we noticed that the data distributions differ from the normal distribution. So we repeated the above experiments, but this time the simulated data sets were created using uniform distributions. The set up for the experiments was very similar to the previous one. Each set consisted of about 30 samples. The proportions used to create the mixed sets were $a = 30\%$, $b = 60\%$ and $c = 10\%$. Again only two bands were used for the proportion estimation.

First we tested the effect of coherent outliers present in the mixed distribution. The results of this experiment are shown in Table 4.3. Then we tested the effect of random outliers. The results can be seen in Table 4.4.

Dist	Out	TM			SHT			LSE		
		a	b	c	a	b	c	a	b	c
3	10%	6 (5)	7 (3)	53 (17)	16 (12)	6 (3)	49 (31)	34 (20)	8 (4)	112 (62)
-	20%	13 (32)	10 (15)	53 (19)	23 (38)	9 (16)	59 (29)	56 (34)	15 (8)	173 (118)
-	30%	15 (37)	11 (17)	51 (19)	26 (42)	10 (18)	62 (29)	62 (33)	25 (13)	181 (138)
-	40%	7 (6)	6 (3)	47 (17)	22 (15)	5 (4)	65 (29)	75 (32)	33 (16)	190 (169)
6	10%	5 (4)	8 (3)	51 (16)	20 (14)	6 (3)	51 (31)	59 (34)	17 (9)	201 (121)
-	20%	6 (5)	7 (3)	51 (17)	18 (13)	6 (4)	53 (32)	76 (34)	32 (15)	177 (155)
-	30%	7 (5)	7 (3)	52 (19)	19 (14)	6 (4)	59 (28)	84 (32)	55 (21)	286 (252)
-	40%	7 (6)	7 (3)	50 (18)	19 (14)	6 (4)	58 (31)	109 (45)	63 (26)	240 (273)
9	10%	5 (5)	8 (2)	56 (17)	16 (13)	7 (4)	46 (31)	66 (33)	23 (12)	207 (141)
-	20%	6 (4)	7 (3)	52 (17)	19 (14)	5 (4)	55 (31)	93 (41)	52 (24)	278 (267)
-	30%	6 (6)	8 (3)	53 (19)	19 (14)	6 (4)	55 (32)	107 (53)	66 (26)	255 (285)
-	40%	7 (5)	7 (3)	51 (20)	22 (15)	6 (4)	62 (31)	112 (51)	66 (27)	247 (281)

Table 4.3: Effect of coherent outliers present in the mixed uniform distribution. The numbers in the table are the average errors over 100 experiments conducted. Inside parentheses is the standard deviation of these errors.

Out	TM			SHT			LSE		
	a	b	c	a	b	c	a	b	c
10%	6 (5)	7 (3)	56 (16)	15 (11)	6 (4)	48 (31)	39 (31)	18 (12)	128 (108)
20%	6 (9)	8 (5)	55 (19)	18 (16)	6 (7)	54 (31)	49 (33)	20 (16)	180 (141)
30%	5 (5)	8 (3)	51 (18)	19 (14)	6 (4)	57 (31)	61 (35)	25 (19)	178 (168)
40%	11 (18)	10 (14)	68 (108)	23 (22)	10 (16)	78 (115)	56 (34)	25 (17)	190 (150)

Table 4.4: Effect of random outliers in the mixed uniform distribution. Mixed set composition $a = 30\%$, $b = 60\%$, $c = 10\%$.

Form Tables 4.3 and 4.4 we can see that when uniform distributions were used to

represent the classes, the robust methods give, in general, lower errors than the errors obtained when normal distributions were used to represent the classes. The LSE method has a similar performance in both cases. We may attribute this difference to the fact that when we derived the expected accuracy of our estimation the error analysis we performed (see equation (4.5)) was based on the worst case scenario and complete ignorance of the distributions of the pure classes, which is equivalent to uniform distributions, was assumed.

Given that Gaussian and uniform distributions are probably the two extreme ends of the distributions one expects to have to deal with in reality, we may conclude that the use of robust statistics is expected to make significant difference in all practical situations when outliers are present.

4.5 What if pure classes are not available?

Ideally, for pure classes we would like to use sets of pixels representative of the pure classes extracted from the remotely sensed image itself. However sometimes, especially if the terrain tends to vary at smaller scales than the size of the test sites, as in our case, it is difficult to find homogeneous test sites that belong solely to a given pure class. Since we have no regions solely composed of one class we have to derive the spectral characteristics of the real pure classes from mixed sites for which we know their composition [56, 28].

A number of different methods to derive the pure classes were tried. At first the attributes of the pure classes (mean vectors, covariance matrices) were derived using the Least Squares Error method as described in chapter 3. Then simulated sets of 30 points were created to represent the “pure classes” assuming that they follow the normal distribution. These sets were used to classify the available sites.

Another solution to this problem is to use Hough Transform to derive the pure

classes. According to our model we have:

$$w = ax + by + cz \quad (4.6)$$

We can make use of the Hough transform to identify the best values for x , y and z . We can consider that equation (4.6) is an equation of a plane in the 3D space (a , b , c), which is parameterised by different values of w (which is known) and we are interested in identifying the luminances x , y , z (which are unknown). In this case we have a 3D accumulator array defined in the parametric (x , y , z) domain. We can then use the luminance values w of the training sites, with the estimated (by ground inspection) values of their mixture parameters, to identify values of (x , y , z) which can be considered to be the means of the corresponding pure classes. Clearly we must perform a different Hough transform for each band. Since x , y and z are luminances, they can take integer values in the range 0 to 255, so we have a 3-D accumulator array $256 \times 256 \times 256$. We select three samples (i.e. three different w 's) from three different sites (each with different but known proportions) at a time, and we solve a system of three equations similar to equation (4.6) for the values of x , y and z . Then only the corresponding cell in the accumulator array is incremented.

After the training part of the classification is concluded, the derived values for x , y and z are used to classify any mixed set. Therefore, we need to know the intraclass variability in x , y and z in order to create simulated data that represent the "pure classes". Ideally, the covariance matrix of each "pure class" should be computed. When x , y , z and w are treated as random variables, the corresponding elements of their covariance matrices are related via equations similar to equation (4.6), which are linear in a^2 , b^2 and c^2 , as described in chapter 3. A separate Hough transform should be performed for each element of the covariance matrix.

This method, however, proved to be too expensive computationally and too unreliable. Instead, a much simpler method was used for the calculation of the

intraclass variability which was based on the steepness of the peak in the Hough space used to calculate the mean luminances of the pure classes. Let us assume that for a derived triplet (x_0, y_0, z_0) we have a peak value f_{x_0, y_0, z_0} in the Hough space. Then at the point $(x_0 + \sigma_x, y_0, z_0)$ we have:

$$\frac{f_{(x_0 + \sigma_x, y_0, z_0)}}{f_{(x_0, y_0, z_0)}} = e^{-\frac{1}{2}}$$

From the above we can derive σ_x and in a similar way σ_y and σ_z . These standard deviations and the values of x , y and z , calculated before, are used both to create simulated sets of 30 points, assuming that they follow the normal distribution, and to classify any mixed set with unknown proportions. Both robust methods (SHT and TM) were used for the derivation procedure.

This derivation procedure assumes that the ground data are completely error free, which is quite a strict assumption. Since one can not expect to avoid small errors in ground data altogether, we test how much these errors influence the resultant "pure classes" to get an indication of how accurate results we can expect when the aforementioned classification methods are applied to real data. In order to test the stability of the derivation procedure we used again simulated data.

We decided on three sets to represent the pure classes (the ones used in previous simulations) and from these we created four mixed sets with known proportions (we need at least three mixed sets). The proportions used for the mixed sets were: set 1 (30% – 60% – 10%), set 2 (50% – 10% – 40%), set 3 (70% – 20% – 10%), and set 4 (40% – 20% – 40%). These proportions were carefully selected to avoid degeneracy of the equations we have to solve, i.e. so that the convex hull defined by the means of the sets of the pure classes has a non zero volume. Then we used the derivation procedures mentioned above to derive the statistics of the original pure classes. When no error was introduced, the pure classes were derived with high accuracy. Then we introduced some error in one of the mixed sets: set 1 was treated

as having proportions (30% – 50% – 20%) instead of (30% – 60% – 10%). When this slightly erroneous values were used during the derivation of the pure class statistics, one of the derived “pure classes” was quite misplaced while another was slightly moved, as we can see in *Figure 4.7*. This test proved that even small errors in the ground data can affect the recovery of the statistics of the “pure classes”. This of course has a great impact to the classification performance.

Figure 4.7: *The means of the pure classes derived from mixed sets of known composition when slightly erroneous mixing proportions for one of the sets were used in the derivation.*

Thus, we can conclude from this study that it is possible to solve the inverse problem, i.e. to derive the statistics of the pure classes from the available data of mixed classes with known (by ground inspection) proportions, but the values of these known proportions have to be known with quite high accuracy (perhaps better than 10%). As expected, the LSE method gives slightly better results than the robust methods since no outliers are involved.

4.6 Application to real data

Since the simulation results showed that our model performs well, we then tested it with some real data. The test area Pateras has been selected for this purpose. The primary vegetation in this study area, as already mentioned, is composed of conifers, mainly *aleppo pine* and a variety of shrub species. So the vegetation cover is categorised in three main classes: *bare soil*, *aleppo pine* and *other vegetation*.

Since we had no regions solely composed of one class, we derived the statistics of the pure classes using the methods described in the previous section. 16 training sites were used for this purpose. The algorithms were then tested on 6 sites which had not been used for training and for which the composition was known from ground inspection as well ("test sites"). Two criteria were used to evaluate the obtained results. According to the first criterion a classification result is considered a "hit" if the dominant class is identified correctly, otherwise we have a "miss". The second criterion is more strict, a classification result is considered a "hit" if the dominant class is identified correctly with accuracy $\pm 15\%$. The results of the robust methods were compared with the results obtained by the LSE method.

The classification results can be seen in Table 4.5. The top line of the table refers to the method used to derive the "pure classes", the second line refers to the method used to classify the sites, while the third line refers to the criterion used to decide on a "hit" or "miss". The numbers presented in this table corresponds to the number of "hits" according to the relevant criterion.

No	TM						SHT						LSE					
	TM		SHT		LSE		TM		SHT		LSE		TM		SHT		LSE	
	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd
16	11	8	11	8	11	9	11	8	11	9	11	10	10	8	10	5	12	8
6	4	4	4	4	4	4	5	4	5	3	4	4	4	3	4	2	3	2

Table 4.5: Classification results obtained using TM, SHT and LSE when the end-members were derived with different methods.

As we see in Table 4.5 the classification performance of all methods, is reasonably consistent. This can be attributed to the fact that the classification procedure is

greatly dependent on the sets used to represent the “pure classes”; different pure sets yield different classification results which explains the inconsistencies in performance of the same method when different “pure” sets were used. The derivation procedures used assume that the ground data were completely error free, as mentioned before, which may not be the case here.

4.7 Discussion and conclusions

A novel method of mixed pixel classification, which is based on the *Hough transform* and a local *trimmed mean* smoothing has been presented. The simulation results showed that the Hough transformed method can tolerate a large amount of outliers and still retain an acceptable performance. So the Hough method seems more attractive in terms of performance, but the price that one has to pay is the increase in computational complexity.

The method was applied to a problem involving a mixture of three components. The estimation of the mixing proportions involves quadruples of pixels, each drawn from the mixed and three pure class populations was respectively. Each quadruple then votes for a particular set of mixing parameters.

The problem of exponential explosion of the number of quadruples one can potentially generate also has to be considered. Indeed, if each one of the distributions that represents a pure class and the mixed distribution consists of 30 points, we have to consider 30^4 possible combinations which is about 10^6 combinations. This is really the limiting factor in our approach: It is not feasible to use it for large data sets or for many “pure” classes. However, the method is not really meant for large data sets as it is only introduced for the case that the datasets are not sufficiently large to allow reliable statistics to be extracted from them.

All methods when applied to real data achieve reasonably good performance. One possible source of inconsistency may be attributed to the error introduced

during the “pure class” derivation procedure resulted from ground data inaccuracies. These may have been caused due to the way ground data were collected which was done by field inspection.

Chapter 5

Mixed pixel classification using the Randomized Hough Transform

5.1 Introduction

It has been proposed in the previous chapter that the process of unmixing can be performed with the help of Hough transform as this approach can cope with the presence of outliers as well. In this chapter we propose the use of the *Randomized Hough Transform* algorithm for the determination of the proportions of pure classes present in sets of mixed pixels for large datasets (for which the deterministic Hough is prohibitively slow).

The most important drawback of the Standard Hough Transform is that it performs an exhaustive search of the possible combinations of the data samples and it becomes prohibitively expensive for large sets of data. The Randomized Hough on the other hand, samples the data set and thus runs much faster making it appropriate for large data sets as well. The question is whether the sampling we are required to perform to make the problem computationally tractable introduces unacceptable errors in estimating the mixing proportions. That is why it had to be tested and compared with the Hough that performs exhaustive search. In particular we

compare the results of the Standard Hough Transform (SHT) method proposed in chapter 4 with the results obtained from the Randomized Hough Transform (RHT) proposed in [79, 34].

In section 5.2 a brief description of the Randomized Hough transform [79, 34] is given. Then the performance of SHT and RHT is compared using simulated data in section 5.3 and some real data for which the ground truth is known in section 5.4. Finally we present some conclusions in section 5.5.

5.2 The Randomized Hough Transform (RHT)

The basic idea of the Randomized Hough Transform [79, 34] is that instead of analysing the parameter space at the end of the vote accumulation process, we analyse it dynamically during the process of accumulation and look for signs of forming peaks. The data sets are randomly sampled to choose a pair of quadruples at a time. The random sampling is continued until an evident candidate for a global maximum or a peak is detected in the accumulator array.

There are several ways to detect a peak. For instance, either the number of votes in the cell must reach a fixed threshold t to be considered as the maximum, or the accumulator array is studied threshold by threshold ($t = 1, 2, \dots$). In the latter approach, with each threshold t , a global maximum is detected and the random sampling is stopped when the same global maximum has been detected to be a winner at separate threshold levels t_m times. If the number of the threshold levels needed is known or is determined beforehand, the method is called *the constant threshold method* and if the number of the threshold levels is automatically adjusted according to the input data, the method is called *the variable threshold method*. The variable threshold method detects “clear” maxima sooner but it requires, besides the “normal” accumulator, another accumulator to store the winner of each threshold level. In this chapter we shall use the constant threshold method.

As the accumulation process is terminated early as soon as the stopping criterion is met, the resulting accumulator space is sparse. Consequently the accumulator array can have the form of a dynamic structure like a tree. Moreover, the reflectances of pixels x , y , z and w can be chosen with or without the replacement of the selected reflectances. In our experiments, they are chosen with the replacement which means that the content of a class, i.e. the set of reflectances, will remain unchanged during the sampling.

5.3 Testing of the methods with simulated data

The basic difference between the Randomized Hough and the Standard Hough, is that the latter uses exhaustively all quadruples of pixels available and at the end it identifies the peak in the accumulator array, while the Randomized Hough draws at random quadruples and monitors the accumulator array. Once a peak has been identified in it with value above the preset threshold, the process stops. Thus, not all the data are necessarily used by RHT. The purpose of the experiments presented in this section is to assess the effect of this sampling and termination process and compare it with the exhaustive search. As the Standard Hough performs the exhaustive search, the two algorithms have to be assessed with small data sets for which the exhaustive search is computationally possible. The performance of the RHT algorithm with larger data sets will be investigated with real data in the next section.

Three distributions were artificially created to represent the three "pure" classes, which were assumed normally distributed in each band. The mixture distribution was created from the three pure distributions with known mixing proportions. We represent each distribution by a set of points and we try to estimate the proportions of the classes in the set of mixed pixels. The sets used to represent the "pure" classes are similar to those used in the experiments in the previous chapter as we can see

in *Figure 5.1*. Each set consists of about 30 samples. Only two bands were used for all the experiments described below.

Figure 5.1: *Some example sets used to represent the three "pure" classes.*

The performance of RHT is compared with that of SHT and also with the results of the least square error solution (LSE), described in chapter 3. The results are presented in terms of the relative errors in the computed proportions.

5.3.1 Time and accuracy concerns

The Hough Transform is known to require heavy computations and therefore we give an estimate of the CPU times of the Hough methods presented in this chapter. The times presented here are expressed in seconds and were estimated using a Sun SPARCstation 20. The main parameter that may influence the execution time is the number of points used to represent the various classes.

As shown in *Figure 5.2* the RHT is very fast while the computation time needed by the SHT increases exponentially with the number of points that represent each

Figure 5.2: *CPU times estimated using a Sun SPARCstation 20 versus the number of points representing each class. The RHT allows for greater error (about 5% greater) than the SHT in this case.*

class. Due to the stochastic nature of the RHT, each value of the CPU time given in the figure was the average of 100 experiments repeated with the same data sets but with different initializations of the random number generator that determines the sequence of the chosen quadruples. Since SHT is deterministic, the CPU times given are from single experiments.

Since the experiments with the simulated data mentioned below were carried out using only 30 samples to represent each class, in order to assess the effect of subsampling the underlying distributions, we estimated the error in proportion estimation using different datasets derived from the same distributions. For each set of data, RHT was repeated 100 times and the average errors and their standard deviations (represented by error bars) are shown in *Figure 5.3*.

It is obvious from *Figure 5.3* that the behaviour of RHT is very erratic and it could vary significantly according to the datasets that represent the relevant classes. To improve upon this behaviour, we decided to iterate RHT, i.e. to run it a few

Figure 5.3: *Influence of the choice of datasets. The error bars indicate the standard deviation of the distribution of the observed errors with 100 different initializations of the random number generator, for fixed set of data.*

times for each given set of data, (each time with different initialization of the random number generator) and use as the final result the average of the results. Figure 5.4 shows the average error and its variance computed over 100 different runs using the same dataset, as a function of the number of iterations performed by the algorithm. The result recorded was the average result over these iterations. This was compared with the true value in order to calculate the error. For example for *Number of iterations* = 15, we ran the experiment 1500 times with different initialization of the random number generator each time, and we divided these results into 100 groups of 15 each. The average error of each group was calculated and the mean and the variance of these 100 average errors was computed and plotted in this figure. It is clear that after about 10 iterations the behaviour of the algorithm with respect to its error stabilizes, and it is not worth considering a higher number of iterations.

Figure 5.5 is similar to Figure 5.3, but each run of the RHT algorithm now

Figure 5.4: *Average relative error for proportion estimation, when we vary the number of iterations used. The error bars represent the standard deviation of the distribution of the observed errors with 100 different initializations of the random number generator, for fixed set of data.*

corresponds to 10 iterations over the same datasets but with different initializations of the random number generator. We can see that RHT now behaves better than SHT. This is surprising as SHT is deterministic and it should be better. (Note that the behaviour of SHT is somehow different in *Figure 5.5* from that in *Figure 5.3* because different datasets were chosen for the construction of the two figures.)

To understand this behaviour, we examined closer the type of peak in the accumulator array. It became obvious that the peak is rather flat, due to the intrinsic uncertainty in the data, as discussed in chapter 4, and neither of the two algorithms really picks the right peak. By iterating RHT, however, and taking the average result, we manage most of the time to go closer to the true peak. This is obvious from *Figure 5.6* where we present as a grey image the peak of the accumulator array and mark the positions of the results of the 10 RHT iterations, their average, the SHT result and the true values of the mixture.

Figure 5.5: *Influence of the choice of datasets when RHT is iterated 10 times. The error bars indicate the standard deviation of the distribution of the observed errors with 100 different initializations of the random number generator, for fixed set of data.*

Clearly, 10 iterations slow down the algorithm by a factor of about 10, but this is a small overhead given the speed of the algorithm. It can be seen that if we multiply roughly the CPU times that correspond to RHT by 10 in *Figure 5.2*, the algorithm will still be sufficiently fast and more efficient than SHT for datasets represented by more than about 50 pixels each.

5.3.2 Testing different proportion combinations for the mixed class

As we have seen in the previous chapters, the performance of the algorithms when tested on different mixed sets tends to vary with the mixture proportions. To test this behaviour we repeated a similar experiment to that presented in the previous chapters. According to this experiment we created a series of datasets where one class is represented by some fixed proportion, and the two other classes take all

Figure 5.6: *The area around the peak in the accumulator array. This is a part of the accumulator space with size 32×38*

possible combinations of proportions so that the sum of the three is always 100%.

We ran then the algorithm as many times as we had combinations of proportions and picked for all the runs the maximum error in estimating the value of the proportion that was kept fixed. Further, to eliminate the error of the representation of each class by 30 points only, we repeated this experiment 100 times for different datasets drawn from the same distributions. Then we calculated the average error over these 100 runs for both algorithms and plotted it versus the value of the proportion that was kept fixed each time (and to which the calculated error was referring).

The results are shown in *Figure 5.7* for the non-iterated RHT. As expected, on average it behaves worse than SHT. In *Figure 5.8* we show the results of similar experiments where RHT was iterated 10 times each time. Again as expected, RHT now performs on average better than SHT. All algorithms have high errors for small proportions due to the natural intraclass range of possible values which may result to negative or exactly 0 proportion values (which our programs reject) and to

positive but near 0 values (which our programs accept) when incompatible or near incompatible quadruples are picked from the data.

Figure 5.7: Average maximum relative errors in estimating different proportions for *LSE*, *SHT* and non-iterated *RHT*.

5.3.3 Outliers

In this experiment some outliers were added to the mixed set. The parameters that vary in the following experiments are: the type of outliers, the number of outliers, and the position of outliers added to the mixed set. We examined the influence of coherent and random outliers as we did in chapter 4.

The coherent outliers tend to create clusters in an arbitrary distance from the mixed distribution. At first we consider outliers placed in different distances from the mixed set. The unit used to express these distances is the standard deviation of the mixed distribution. The standard deviation of the cluster of outliers will be half the standard deviation of the mixed set. Different number of outliers are also be examined. The experiment was repeated 100 times, each time a different set to

Figure 5.8: Average maximum relative errors in estimating different proportions for LSE, SHT and iterated RHT.

represent the “mixed” class was drawn. Then the average and the variance of the errors in estimating the proportions were calculated.

The results obtained from this experiment are shown in *Table 5.1* for the iterated 10 times RHT, for SHT and the LSE method. The true proportions of the mixed distribution were $a = 30\%$, $b = 60\%$ and $c = 10\%$. Note that as expected, LSE method is the most unstable to outliers, while the two Hough methods are quite resilient to their presence. The error in estimation of proportion c is the largest one, which was expected since c is small.

Then we tested the influence of random outliers. This type of outliers do not form a second mixture distribution but they can be scattered anywhere on the feature plane. In this experiment, for a certain number of random outliers, a number of mixed sets were generated and tested. The average and the variance of the errors in proportion estimation based on 100 experiments are shown in *Table 5.2*. The Hough methods performed well, while the LSE method is more significantly affected by the outliers, although not as much as by the presence of the coherent outliers, because

Dist	Out	RHT			SHT			LSE		
		a	b	c	a	b	c	a	b	c
3	10%	12 (8)	6 (5)	54 (41)	14 (7)	8 (6)	81 (30)	29 (19)	8 (6)	110 (56)
-	20%	12 (10)	5 (4)	55 (47)	16 (14)	8 (6)	87 (49)	48 (33)	17 (9)	168 (120)
-	30%	17 (27)	9 (16)	69 (110)	18 (31)	11 (16)	89 (80)	61 (34)	25 (11)	186 (146)
-	40%	19 (34)	7 (15)	63 (53)	26 (43)	16 (23)	109 (133)	76 (33)	33 (16)	254 (205)
6	10%	13 (9)	5 (4)	58 (41)	15 (10)	8 (5)	82 (40)	51 (33)	17 (10)	179 (121)
-	20%	12 (11)	6 (8)	61 (72)	16 (14)	9 (8)	87 (37)	81 (30)	34 (14)	223 (202)
-	30%	12 (12)	6 (9)	58 (82)	16 (16)	10 (14)	97 (115)	82 (34)	47 (23)	243 (237)
-	40%	12 (12)	6 (8)	55 (44)	17 (15)	9 (12)	91 (83)	101 (43)	52 (28)	238 (240)
9	10%	13 (10)	5 (4)	55 (48)	18 (12)	8 (6)	94 (41)	67 (32)	23 (13)	197 (149)
-	20%	12 (9)	5 (4)	53 (40)	14 (11)	9 (6)	88 (34)	86 (33)	45 (25)	213 (215)
-	30%	13 (13)	6 (7)	57 (44)	16 (13)	9 (8)	85 (29)	109 (51)	64 (31)	259 (284)
-	40%	13 (10)	6 (5)	63 (55)	18 (14)	9 (10)	94 (85)	126 (63)	73 (27)	261 (290)

Table 5.1: *Effect of coherent outliers present in in the mixed distribution. The numbers in the table are the average errors over 100 experiments conducted. Inside parentheses is the variance of these errors. RHT results were obtained with the iterating version of the algorithm.*

Out	RHT			SHT			LSE		
	a	b	c	a	b	c	a	b	c
10%	13 (12)	5 (9)	60 (89)	17 (17)	9 (14)	99 (117)	29 (22)	8 (6)	102 (60)
20%	13 (11)	6 (7)	57 (70)	16 (10)	8 (6)	88 (38)	41 (31)	14 (10)	134 (108)
30%	14 (10)	6 (5)	62 (48)	17 (11)	8 (5)	89 (27)	44 (30)	16 (12)	144 (115)
40%	18 (28)	7 (12)	54 (40)	19 (24)	10 (12)	91 (33)	47 (34)	17 (12)	131 (93)

Table 5.2: *Effect of random outliers present in the mixed distribution. The numbers in the table are the average relative errors over 100 experiments conducted and inside parentheses is the variance of these errors. RHT results were obtained with the iterating version of the algorithm.*

the effect of random outliers tends to average out.

5.4 Application to laboratory data

The aim of this experiment is to take real images of mixed pixels with known composition and then test our methods on these images. For the image acquisition we used a high precision RGB camera connected to an 8-bit frame grabber [82, 83]. The next step was to decide on the materials that were going to represent the “pure” classes and consequently mixed with known proportions to create the “mixed” classes.

The choice of the materials had to comply with several restrictions. In this case our sensor measures colour so the materials used to represent different classes should be different in colour terms. Since the resolution of the camera is quite high we need

to have very small grains (of the order of $10^{-6}m$) in order to be able to have many of them in the same pixel. Furthermore, there should be an easy and accurate way to measure the proportions in the mixture. Given that the materials are solid, similar densities are desirable in order to be able to use weight measurements to create the mixtures.

For the above reasons we decided to mix coloured chalk, such an image is shown in *Figure 5.9*. We took ordinary bars of chalk and we graded them using a mill so as to create very fine powder. Then we used a sieve in order to use only the chalk grains with diameter less than a certain value ($200\mu m$). As soon as we had enough material of the three colour chalks we created the mixtures by mixing the relevant weights. Then using equal quantities of chalk for each class, we grabbed the images.

Figure 5.9: *An image of the coloured chalks.*

In order to avoid the influence of the spatial variability of the illumination to our results we transformed the RGB space to the chromaticity space and used the chromaticity values to test our models. For this experiment we used blue, red and green chalk. The results we obtained are presented in *Table 5.3*. The RHT results were obtained with 10 iterations of the algorithm. The datasets this time consisted

of thousands of points each and RHT used all the samples to draw quadruples from. SHT, however, could not possibly make use of all the data, so it was run on subsets of 30 pixels only from each image. Inspite of this SHT performed quite well. However, this performance would not have been so good if the mixed distribution created were less uniform, as its representation by 30 points only would not have been adequate.

Real			RHT			SHT			LSE		
Red	Green	Blue	Red	Green	Blue	Red	Green	Blue	Red	Green	Blue
60	10	30	62	9	29	63	10	27	62	9	29
10	20	70	12	21	67	12	22	66	12	21	67
40	20	40	46	10	44	47	7	46	46	10	44

Table 5.3: *Results on real images.*

A second set of real data was created, which consisted of pure classes with colours more similar to each other: the same green, another red and yellow. The pure classes of the two experiments in the chromaticity space are shown in *Figure 5.10*. The values of $G/(R+G+B)$ and $R/(R+G+B)$ are within the range of $0 \dots 1$ and we scale them to $0 \dots 255$. The results of the experiments with the second set of data are shown in *Table 5.4*.

Figure 5.10: *The two colour sets tested.*

Real			RHT			SHT			LSE		
Red	Green	Yellow	Red	Green	Yellow	Red	Green	Yellow	Red	Green	Yellow
70	10	20	70	5	25	68	2	30	70	5	25
30	60	10	37	56	7	40	59	1	38	57	5

Table 5.4: Results on real images.

A third set of experiments with real data was run after we added some outliers in one of the mixture to see how that affects the results. We used red, green and blue to represent the pure classes. In order to introduce the outliers we threw some instant coffee grains on the flat surface of the mixture. The relative position of the outliers to the mixture in the chromaticity space can be seen in *Figure 5.11* and the results obtained in *Table 5.5*.

Figure 5.11: The three pure classes and the mixture with several random outliers.

Real			RHT			SHT			LSE		
Red	Green	Blue	Red	Green	Blue	Red	Green	Blue	Red	Green	Blue
10	20	70	13	19	68	13	18	69	17	18	65

Table 5.5: Results on real image when outliers are present.

All the methods gave reasonable results in the real image tests. In the presence of outliers, however, the SHT and the RHT performed better than the LSE method.

Furthermore, the computation times of the RHT run with 10 iterations, is in the range of 0.27-2.72 seconds, while the SHT. The SHT took about 3 seconds for the 30 pixels, giving a less accurate answer.

5.5 Discussion and conclusions

In this chapter we proposed the use of the Randomized Hough Transform for the unmixing problem when large datasets are used to represent the pure and the mixed classes. With a series of experiments performed with simulated and real laboratory data, we showed that RHT can be made very stable and at the same time be much faster than the deterministic Hough for large data sets.

Its major advantage over the deterministic Hough is its more or less fixed requirements in CPU time irrespective of the number of samples used. All our results were compared with the classical Least Square Error method which was used as a benchmark. In the absence of outliers this method is clearly the best, but the advantage of both Hough methods is their ability to cope with large numbers of outliers in the mixture distribution.

Chapter 6

Mixed pixel classification using the hypothesis testing Hough transform

6.1 Introduction

In this chapter we present a new method for mixed pixel classification where the classification of groups of mixed pixels is achieved by using the Hypothesis Testing Hough Transform. The motivation of this work is that some other estimation methods based on robust statistics, like the Standard Hough Transform, have been criticized that although they can cope with the presence of outliers, they give poor performance in the absence of outliers in comparison to the least squares error method. The method proposed in this chapter is demonstrated using simulated data and proved to perform equally well in the presence and in the absence of outliers. It is also applied to real Landsat TM data.

In chapter 4, it has been proposed the use of robust statistics for the linear unmixing of sets of mixed pixels. In particular, the Standard Hough Transform (SHT) and the Trimmed Means (TM) method were proposed and shown to be able

to deal with the problem of unmixing in the presence of outliers. Both approaches were compared with the Least Squares Error (LSE) solution which was affected significantly by the outliers. However in the absence of outliers, the LSE solution was clearly better than the solutions obtained either by TM or SHT.

As we have seen so far the peak in the accumulator space, for the kind of data we are using, is quite broad with a number of local peaks. As a result the SHT was often confused and therefor the results produced were not as good as those of the LSE. In this chapter, we propose the use of the Hypothesis Testing Hough Transform (HTHT) which relies on the use of a soft voting kernel and which smooths this broad peak enabling the Hough algorithm to find the true maximum.

In the next section we describe our method in detail and discuss the parameters used. In section 6.3 we demonstrate the performance of the method with simulated data. In section 6.4 we apply the method to some real data. Finally we discuss our conclusions in section 6.5.

6.2 The proposed method

In the linear mixing model, it is assumed that the pixel value in any spectral band is given by the linear combination of the spectral responses of each component within the pixel. When we are interested in identifying the proportions of three pure classes, we may write:

$$w = ax + by + cz \quad (6.1)$$

where w is the known spectral reflectance of a mixed pixel, x , y and z are the known spectral reflectances of the three possible cover components within the mixed pixel and a , b and c are the proportions to be estimated for each component contained in the mixed pixel.

To allow for intraclass variability, we assume that we have at our disposal sets of

pixels of pure class coverage as well as a set of mixed pixels, the coverage of which has to be estimated.

The advantage of the Hypothesis Testing Hough transform over the conventional one is that it creates a continuous accumulator *function*, as opposed to a discrete accumulator *array*, which can be sampled as densely as it is necessary and which can be calculated at any point with as much accuracy as it is allowed by the data (i.e. the binning error of the conventional Hough is eliminated) [52]. As our proportions have to sum up to 1, we can eliminate c from (6.1) and our accumulator function becomes 2D. This function is sampled at a set of chosen (a,b) points. For each hypothesized set of (a,b) values and each triplet of (x,y,z) values drawn from the sets of points that represent the pure classes, we can estimate the reflectance vector w_H using:

$$w_H = z + a(x - z) + b(y - z) \quad (6.2)$$

The components of this vector, which are the hypothesized reflectances of a mixed pixel in each of the spectral bands, are then compared in turn with the corresponding components of each of the mixed pixels in the mixed pixel set. The hypothesized values (a,b) are confirmed or rejected by each mixed pixel tested according to a kernel function used, that weighs the difference between the hypothesized reflectances of the mixture and each of the true reflectances. A suitable kernel function as proved in [52] is given by the following equation.

$$P(\xi) = \prod_{i=1}^n Q\left(\frac{\xi_i}{width_i}\right) \quad (6.3)$$

with

$$Q(x) = \begin{cases} 1 - \frac{m}{m-2}x^2 + \frac{2}{m-2}x^m & \text{for } x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $i = 1, \dots, n$ are the bands that are going to be used in the analysis, $\xi_i = w_{H_i} - w_{T_i}$ is the difference between the hypothesized reflectance and the real one, m

is an even integer greater than 2 (in our case we decided to use $m = 4$), and $width_i$ is the width of the voting kernel in band i .

The width of the kernel should reflect the uncertainty with which we can estimate the proportions. We can better explain that by realizing that a given reflectance w of a certain mixed pixel can be created by more than one combinations of values x, y, z , due to the intraclass variability of the pure classes. As we have seen in chapter 4 the uncertainty in a, b is given by the standard error for a and b . If for simplicity we consider that x, y and z vary within the same range, i.e. if we assume that $\Delta x \sim \Delta y \sim \Delta z \equiv \sigma$, the uncertainty in a and b is:

$$\begin{aligned}\Delta a &\sim \frac{\sigma}{|x - z|} \\ \Delta b &\sim \frac{\sigma}{|y - z|}\end{aligned}\tag{6.4}$$

The above assumption is clearly an oversimplification but we may say that for a conservative estimate of the error in the parameters we shall take all these ranges of the reflectances of the pure classes to be equal to the largest one. So a suitable width for the voting kernel that incorporates these parameters would be:

$$width_i = \max\left(\frac{\sigma_i}{|\bar{x}_i - \bar{z}_i|}, \frac{\sigma_i}{|\bar{y}_i - \bar{z}_i|}\right)\tag{6.5}$$

where σ_i is the standard deviation of the most diverse class and $\bar{x}_i, \bar{y}_i, \bar{z}_i$ are the means of the pure classes in band i .

When all possible quadruples of the data have been considered, the set of hypothesized (a, b) values with the highest vote are the best values of the mixing parameters a and b . As we are dealing with a continuous accumulator function, we could use these values to perform a sort of hill climbing and identify the correct a, b values [53]. However, in most cases no such high accuracy is required.

The computational time requirements for this algorithm are quite heavy so a

way to speed up the processes is needed. For this reason we used a hierarchical two stage Hough algorithm [54]. In the first stage we perform a Hough Transform using a coarse grid. The purpose of this pass is to act as focus of attention for the second pass. The idea is to determine which are the approximate proportions that contribute to the mixed set in question. This information is then used to generate a new accumulator space that is much smaller than the previous one and therefore can be sampled with as much accuracy as we want (in our case resolution 1%).

The coarse sampling rate can be selected to be the same as the uncertainty in a and b , calculated from (6.4). If the size of the grid calculated this way is very small, as in cases when we have very distinct pure classes, then the two stage Hough would really become a one stage Hough, since the first pass will dominate the process. In cases that the pure classes are quite similar, the bins in the first stage will be very large, and as a result the second stage will become dominant, tending to a full size, full resolution Hough. In order to avoid such extreme situations we use the values calculated from equations (6.4) as an indication of the appropriate size of the coarse grid, but also we use some safeguards to ensure that the size of the first grid will be cost effective (for example the sampling points in the coarse grid are not allowed to be closer than 10% and further apart than 50%).

At the first level of processing it is crucial to pinpoint the right neighbourhood of proportions, otherwise any further processing will be useless. As the function is grossly subsampled, all detail structure of it is omitted. The kernel width used is roughly of the same size as the sampling distance (given by equation (6.4)) so no further significant smoothing to the function is imposed, apart from the smoothing implied by the undersampling. At the second stage, we require the kernel width to be reasonably large since the sampling of the parameter space is now very dense and lots of small scale details of the accumulator function are picked up. Using a wide kernel is equivalent to low pass filtering the accumulator function before sampling it densely to find the most prominent peak. Thus, we allow pixels to spread their

vote over a large region of the parameter space, defined by the kernel width, which is now 10 times larger than the kernel width used in the first stage.

In order to find how many times larger should the width of the kernel be in the second stage of processing, compared to the width used in the first stage, we performed a series of experiments. Each time we increased the width by 0.1 and calculated the values of the accumulator function along a single line (for $b = \text{const}$ for example), used as an indicator. These values were plotted as a simple cross-section view of the 2D accumulator function. We stopped the process when the accumulator function became smooth enough, so that the dense sampling that was used could pick the single dominant peak. This process can be made automatic by checking whether the cross-section used as an indicator of the appearance of the accumulator function, has a single local maximum, which is also the global maximum.

6.3 Testing of the method with simulated data

Our method is at first assessed using simulated data to represent the pure and the mixed classes. Three distributions were artificially created to represent the three “pure” classes, which were assumed normally distributed in each band. The means and covariance matrices of each of the simulated distributions were chosen to be the same as those computed from real test sites that were known to represent almost “pure classes” in a remotely sensed image. The sets used to represent the “pure” classes were the same for all the experiments, described below and are plotted in *Figure 6.1*. A mixture distribution was created, from the three pure distributions using the following proportions $a = 30\%$, $b = 60\%$ and $c = 10\%$. We represent each distribution by a set of points (30 samples) and we try to estimate the proportions of the classes in the set of mixed pixels. Only two bands were used for all the experiments described below.

The results of the Hypothesis Testing Hough Transform (HTHT) method are

Figure 6.1: *Some example sets used to represent the three “pure” classes.*

compared with the solution obtained by the least squares error (LSE) method and by the Trimmed Mean method (TM) which was shown to outperform the Standard Hough in chapter 4. The results are presented in terms of the relative errors in the computed proportions.

6.3.1 Time and Accuracy Concerns

The computational time needed for the HTHT depends on three parameters: the number of points used to represent the sets, the size of the coarse grid (since in our case the size of the finer grid is always constant with intersample distance equal to 0.01), and the separability of the sets used to represent the “pure” classes.

If we refer to the same “pure” classes and increase the number of samples used to represent the sets, the computational time will increase exponentially. As far as the second factor is concerned, the larger the grid, the faster the first stage of processing for the HTHT will be and the slower the second stage. So the size of

the coarse grid is also a trade-off of computational performance. We generally look for a reasonably fast first stage so as to speed up the second stage that usually requires heavier computations. In the experiments presented below the coarse grid was 39×19 and the kernel had width 0.4 and 0.7 for the two bands in the coarse level and in the fine level the grid resolution was 1, with the kernel width 0.6 and 1.0 in the two bands.

Finally the more similar the pure classes are the more likely it is that more than one proportion combinations will give non-zero voting values and therefore enter the analysis, which will inevitably increase the computational time. In the simulations that follow, the “pure” sets are quite similar and the initial grid is rather coarse, so the computational time will reflect the worst case scenario for the 30 point sets used. The computational time when estimated on an Alpha workstation was approximately 160 seconds.

Since the experiments with the simulated data mentioned below were carried out using only 30 samples to represent each class, in order to assess the effect of undersampling the underlying distributions, we estimated the error for the same proportion ($a = 30\%$) using different datasets derived from the same distributions.

As we can see from *Figure 6.2* the estimation error depends on the selected datasets, that is why we will present our results in the form of an average error over 100 experiments in order to reduce the effect of undersampling the class distributions. We can see from this figure that although the HTHT performs in general better than the TM method, it still does not perform as well as the LSE method. This is due to the way HTHT is applied and also due to the nature of our data.

According to the HTHT, for each quadruple (x_1, y_1, z_1, w_1) we get one estimate (a_1, b_1) from equation (6.2). If we check another quadruple (x_2, y_2, z_2, w_2) then we are going to get another estimate (a_2, b_2) . If the difference of the corresponding points (e.g. $|x_1 - x_2|$) is small in comparison to the intraclass distance (e.g. $|x_1 - y_1|$) then the estimates (a_1, b_1) and (a_2, b_2) will be quite similar. So, if we have

Figure 6.2: *Relative error in estimating a whose true value is 30% versus different sets of samples representing the same mixture.*

small intraclass variances in the pure and the mixed classes in comparison to their interclass difference, we are going to get always an accurate estimate for (a, b) , similar to that obtained by using the LSE method [52]. As the interclass variance increases the performance of the algorithm is expected to deteriorate.

We have discussed in section 6.2 that if we select an appropriate width for the voting kernel we can compensate for the uncertainty in our data. But if the sets we use have quite large intraclass variability in comparison to the interclass difference the width of the voting kernel becomes large and consequently the accumulator space becomes quite flat. Thus the peak the algorithm has to detect becomes very uncertain. In summary, the more compact the data sets that represent the pure classes are, the closer the performance of HTHT is to that of LSE. For example, if there was no intraclass variability at all, HTHT would produce an estimate as good as the LSE method both in the presence and in the absence of outliers.

In the following experiment we are going to test how the intraclass variance affects the accuracy of the obtained results for a given interclass distance. The

means of the sets used to represent the pure classes are selected to be the same as the means of the pure classes in our Landsat TM image, that is they form a triangle in the 2-band scattergram with sides equal to $(19, 8)$, $(7, 3)$ and $(12, 11)$.

In this experiment we check how the variance in the sets used to represent the pure classes will affect the accuracy of the obtained results. We used just one point to represent the mixed class. The proportions used to create this mixed class point were $(a = 30\%, b = 60\%, c = 10\%)$. Then we created sets of 30 points for the three pure classes from Gaussian round distributions with increasing equal variances. We started with a covariance matrix which has its diagonal elements equal to 1 and each time we increased the size of its elements by 1. Each experiment was repeated 100 times and the average error in estimating each proportion was calculated. The results of this experiment can be seen in *Figure 6.3*.

As we can see from *Figure 6.3* for pure classes with small variance the HTHT gives an estimate as accurate as the LSE estimate. As the variance becomes larger the estimate of HTHT becomes less accurate than the LSE estimate. The TM almost in every case gives the worst estimate in comparison to the other two methods. The sets that we are going to use in the following experiments have large variance, as we can see from *Figure 6.1*. So in the absence of outliers, the HTHT is not expected to give as good estimates as the LSE, but it is expected to give better estimates than the TM (which in its turn had given better estimates than the Standard Hough).

To test the behaviour of the algorithms for different mixture proportions, we run the following experiments: We created a series of datasets where one class is represented by some fixed proportion, and the two other classes take all possible combinations of proportions so that the sum of the three is always 100%. (For simplicity we restricted ourselves to proportions that are integer multiples of 10.)

We run then the algorithm as many times as we had combinations of proportions and picked from all the runs the maximum error in estimating the value of the proportion that was kept fixed. Further, to eliminate the error introduced by

Figure 6.3: *Relative average error in proportion estimation when we vary the spread of the sets used to represent the pure classes.*

representing each class by 30 points only, we repeated this experiment 100 times for different datasets drawn from the same distributions. Then we calculated the average error over these 100 runs for both algorithms and plotted it versus the value of the proportion that was kept fixed each time (and to which the calculated error was referring).

The results are shown in *Figure 6.4*. As expected the LSE method gives the best estimate. If we exclude though the extreme cases of very small (10%) or very large proportions (80%), (which anyway are more likely to be underestimated, due to the overestimation of the smallest proportion), the performance of HTHT is quite similar to the performance of the LSE. In the next experiment we will demonstrate that HTHT can keep such high performance standards even when outliers are present.

Figure 6.4: Average maximum relative errors in estimating different proportions for HTHT, TM and LSE.

6.3.2 Outliers

In this experiment some outliers were added to the mixed set. The parameters that vary in the following experiments are: the type of outliers, the number of outliers and the proportions of the outliers added to the mixed set. We examined *coherent* and *random* outliers. The influence of the outliers depends on their distance from the mixed distribution in the luminance space and therefore various distances have been examined to demonstrate how they affect the obtained results. As a unit measure of the distance we used the standard deviation of the mixed set. At first we examined *coherent* outliers.

From Table 6.1 we can see that the HTHT gives better estimate than the TM for $a = 30\%$ and $b = 60\%$. For the estimation of $c = 10\%$ the TM gives on average better result than the HTHT, but since c is quite small and usually generates high error, it is not a good indicator of the method's performance. Furthermore HTHT seems to be less dependent on the dataset selection since it gives in general smaller

Dist	Out	HTHT			TM			LSE		
		a	b	c	a	b	c	a	b	c
3	10%	15 (9)	5 (3)	56 (39)	19 (8)	5 (5)	57 (46)	27 (18)	9 (5)	96 (55)
-	20%	17 (8)	5 (4)	66 (37)	18 (10)	6 (5)	62 (50)	51 (33)	16 (10)	180 (116)
-	30%	17 (9)	5 (4)	68 (42)	19 (10)	7 (5)	66 (57)	68 (32)	23 (13)	208 (145)
-	40%	16 (9)	5 (3)	57 (38)	19 (11)	6 (4)	54 (44)	79 (29)	33 (17)	236 (198)
6	10%	17 (8)	5 (3)	65 (33)	20 (8)	7 (4)	69 (50)	48 (35)	17 (9)	179 (117)
-	20%	16 (9)	5 (4)	55 (40)	18 (10)	6 (5)	61 (45)	72 (32)	32 (16)	205 (178)
-	30%	16 (9)	5 (4)	62 (39)	18 (9)	6 (5)	58 (46)	90 (33)	46 (23)	245 (239)
-	40%	17 (9)	6 (5)	72 (49)	19 (9)	6 (5)	66 (54)	95 (38)	58 (29)	303 (296)
9	10%	16 (8)	5 (4)	66 (40)	20 (9)	6 (4)	64 (51)	61 (34)	24 (12)	181 (136)
-	20%	16 (8)	5 (3)	60 (41)	20 (9)	6 (5)	61 (49)	85 (35)	45 (22)	208 (204)
-	30%	16 (9)	6 (4)	58 (42)	20 (9)	7 (5)	66 (52)	105 (47)	61 (28)	199 (236)
-	40%	17 (9)	6 (5)	65 (38)	20 (9)	7 (5)	50 (45)	111 (63)	64 (31)	232 (254)

Table 6.1: *Effect of coherent outliers present in the mixed distribution. The numbers in the table are the average errors over 100 experiments conducted. Inside parentheses is the standard deviation of these errors.*

error deviations. Finally the robust methods were not affected by the outliers no matter how far from the mixture mean they were placed, while the LSE method seems to be greatly affected by this distance, and its performance deteriorates as the outliers were placed further away.

To visualize better the results of *Table 6.1* we plotted the results in estimation of proportion $a = 30\%$ in *Figure 6.5*, which is more representative, since it is not very big ($b = 60\%$ is in general well estimated) and not very small ($c = 10\%$ gives the highest errors). This figure consists of three plots. The first one corresponds to the case where the outliers were placed 3 standard deviations away from the mean of the mixed set. In the second plot the outliers were placed 6 standard deviations away and finally the third plot describes the case where the outliers were placed 9 standard deviations away. The shaded strips represent the average errors plus or minus 1 standard deviation of these errors.

So far we examined only coherent outliers, now we are going to examine *random* outliers as well. These outliers are randomly chosen at distances in the range between 0 and 12 standard deviations from the mixed set. In this experiment, for a certain number of random outliers, a number of mixed sets were generated and tested. The average and the standard deviation of the errors in proportion estimation based on

Figure 6.5: *Effect of coherent outliers present in the mixed distribution. The strips represent the average errors plus or minus 1 standard deviation of these errors. The black dashed line is the mean error of the TM method minus one standard deviation.*

100 experiments using different “mixed” sets (with the same mixture proportion always), are shown in Table 6.2 and in Figure 6.6.

Outliers	HTHT			TM			LSE		
	a	b	c	a	b	c	a	b	c
10%	16 (8)	5 (4)	61 (39)	19 (8)	6 (5)	64 (45)	33 (25)	11 (8)	115 (89)
20%	15 (9)	5 (3)	55 (40)	21 (20)	6 (10)	67 (46)	39 (32)	14 (9)	143 (104)
30%	15 (8)	6 (4)	62 (40)	19 (12)	6 (7)	64 (46)	49 (32)	16 (12)	162 (139)
40%	15 (8)	6 (4)	59 (39)	19 (9)	7 (6)	64 (55)	56 (36)	19 (15)	186 (153)

Table 6.2: *Effect of random outliers present in the mixed distribution. The numbers in the table are the average relative errors over 100 experiments conducted, and inside parentheses are the variances of these errors.*

Figure 6.6 consists of three plots, the first one corresponds to errors in estimation of proportion $a = 30\%$, the second one to errors in estimation of proportion $b = 60\%$ and the last one to errors in estimation of proportion $c = 10\%$

Figure 6.6: *Effect of random outliers present in the mixed distribution. The strips represent the average errors plus or minus 1 standard deviation of these errors. The black dashed line is the mean error of the TM method minus one standard deviation.*

As we can see from Table 6.2 and Figure 6.6 the robust methods performed well, in particular the HTHT outperformed the TM, even in the estimation of the small c proportion. In this experiment the datasets used to represent the mixed class differ a lot from one another due to the nature of the outliers present, so since HTHT seems to be less dependent on the choice of the dataset, it performs better than the TM even for small proportions. The LSE method, on the other hand, is significantly affected by the outliers, although not as much as by the presence of the coherent outliers, because the effect of random outliers tends to average out.

6.4 Application to real data

Since the simulation results showed that our model performs well, we then tested it with a real application. The test area Pateras was selected for this purpose. As we had no regions solely composed of one class, we derived the attributes of the real pure classes (e.g mean vectors) from sites of known composition, using LSE or TM. For this purpose we used 16 sites for which coverage proportions had been estimated by ground inspection. Two methods were used to derive the attributes of the pure classes. According to the first method, we applied the Least Squares Error method to the mixed model of equation (6.1) and solved for the pure classes mean vectors (x, y, z) . Furthermore the corresponding elements of the covariance matrices of the pure class are related via equations similar to equation (6.1), which are linear in a^2 , b^2 and c^2 , so using LSE these can be estimated as well. The second method utilizes the TM method to solve the mixed model for the unknowns (x, y, z) . The variance of the pure classes is estimated based on the steepness of the peak in the Hough space used to calculate the mean luminances of the pure classes.

At the stage of the evaluation of our method we created 30 simulated points to represent the pure classes using the statistics of these classes computed in the training phase. Then we tested our model with 6 sites that had not been used for the derivation of the pure class attributes. Two criteria were used to evaluate the obtained results. According to the first criterion a classification result is considered a “hit” if the dominant class is identified correctly, otherwise we have a “miss”. The second criterion is more strict, a classification result is considered a “hit” if the dominant class is identified correctly with accuracy $\pm 15\%$.

The classification results can be seen in *Table 6.3*. The top line of the Table refers to the method used to derive the “pure classes”, the second line refers to the method used to classify the sites, while the third line refers to the criterion used to decide on a “hit” or “miss”. The numbers presented in this Table correspond to the

number of “hits” according to the relevant criterion.

No	TM						LSE					
	HTHT		TM		LSE		HTHT		TM		LSE	
	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd
16	11	7	11	8	11	9	10	6	10	8	12	8
6	5	4	4	4	4	4	5	3	4	3	3	2

Table 6.3: *Classification results obtained using HTHT, TM and LSE when the end-members were derived with different methods.*

As we see in Table 6.3 the classification performance of all methods, is reasonably good. The errors can be attributed to the fact that the procedures used to derive the “pure classes” assume that the ground data were completely error free, which may not be the case here.

6.5 Conclusions

In this chapter we proposed the use of the Hypothesis Testing Hough Transform for the unmixing problem. With a series of experiments performed with simulated data we showed that HTHT can be very stable. All our results were compared with the classical Least Squares Error method and Trimmed Means method. The advantage of the HTHT method is its ability to give an accurate estimate both in the presence and in the absence of outliers, comparable to that of the LSE in the absence of outliers.

The problem of combinatorial explosion was dealt with, using a two stage Hypothesis Testing Hough Transform, but still it remains a limiting factor when larger data sets or many “pure” classes are used.

All methods when applied to real data achieve reasonably good performance. One possible source of inconsistency may be attributed to the error introduced during the “pure class” derivation procedure resulted from ground data inaccuracies. In our case the ground data were field inspection estimates and it is doubtful that they had better than 20% accuracy.

Chapter 7

Discussion and conclusions

In this thesis the *mixed pixel classification* problem has been identified as a very important issue in remote sensing. Careful investigation of already existing algorithms has been performed and new efficient algorithms to overcome known limitations have been suggested. Finally the major contributions of this thesis are presented.

7.1 Major contributions of this thesis

The aim of this thesis was to propose algorithms for efficient mixed pixel classification in remote sensing applications. The mixed pixel classification *issues* that we are concerned with in this work are:

1. Model accurately the variability of a class.
2. Increase the number of different classes that can be effectively identified within a mixed set.
3. Eliminate the effect of outliers.
4. Eliminate the effect of outliers and maintain small computational time.
5. Eliminate the effect of outliers and maintain high performance.

In order to handle effectively the aforementioned issues a set of algorithms is proposed in this thesis. These algorithms make the following *assumptions* concerning the data:

1. A full representation of the scene can be done using a *linear mixture model*.
2. Whole regions of homogeneous coverage are going to be classified or else we are interested in identifying the composition of sets of mixed pixels rather than single pixels.
3. We have ground data that can be used to derive the attributes of the “pure” classes.

These assumptions, are considered reasonable as they have already been adopted by many other researchers with success [40, 71, 74]. Thus, the major *contributions* of this thesis can be summarized as follows:

1. A new algorithm for the linear unmixing was proposed that utilizes *higher order moments*. This algorithm is believed to model more accurately the variability of the terrain since it incorporates second order moments and in principle it could contain any moments of as high order as the data allow us to estimate reliably. It also augments the equations entering the linear regression and consequently increases the number of classes that can be identified in the mixed set [7, 8].
2. A new algorithm was proposed based on the *Hough Transform* in order to cope with the possible existence of outliers in the mixed set. It has been shown that it removes very efficiently the influence of outliers. Also a *Trimmed Means*, TM, method was compared to the *Standard Hough Transform*, SHT, and proved even more accurate [9, 10].

3. The *Randomized Hough Transform*, RHT, was proposed in order to reduce the computational time required by the SHT and still maintain the same performance [33, 11].
4. The *Hypothesis Testing Hough Transform*, HTHT, was proposed in order to improve the accuracy of the Hough transform and it was demonstrated to be more accurate than the TM method [12, 13].

Simulated data were created to represent the “pure” and the “mixed” sets. In order to create such data, we assumed that they could be conceptualized as distributions and thus we created these distributions. The performance of the proposed methods was explored, even under extreme circumstances, using simulated data. The performance of the proposed algorithms was compared with that of the *Least Square Error* (LSE) method on simulated and real data. The LSE method was selected since it is the most popular method used to solve the linear equations connecting the means of the “pure” and “mixed” sets. The performances of the various robust methods were compared with each other in order to explore their strong and weak points when applied on simulated and real data. The proposed algorithms were also applied to real Landsat TM data. Further, some images of materials with controlled composition (chalk) were acquired used for testing some of the proposed algorithms. This experiment demonstrates also, the potential of the above methods to *Industrial Inspection* applications.

As we can see from the above, a number of methods has been proposed that solve the *mixed pixel classification* problem, even in extreme situations. These methods proved to be quite accurate when applied to simulated data. However they did not achieve the same level of performance when applied to real Landsat TM data, even though they retain acceptable results. This deterioration in performance is not a worrying factor since the well established *Least Squares* method was even more affected. Unfortunately we did not have much control over the ground data

collection and the image registration and correction processes, moreover there was some time difference between the image acquisition and the ground data collection, which is almost certain to introduce inconsistencies.

Bibliography

- [1] Asrar G., "Theory and applications of optical remote sensing", *John Wiley & Sons*, 1989.
- [2] Arai K., Y. Terayama, Y. Ueda and M. Moriyama, "Adaptive least squares method for estimation of partial cloud coverage within a pixel", *Int. J. Remote Sensing*, vol. 16, no. 12, pp. 2197-2206, 1995.
- [3] Adams J.B., M.O. Smith, P.E. Johnson, "Spectral mixture modelling: A new analysis of rock and soil types at the Viking Lander 1 site", *Journal of geophysical research*, vol. 91, no. B8, pp. 8098-8112, 1986.
- [4] Besag J., J. York and A. Mollie, "Bayesian Image Restoration with Two Applications in Spatial Statistics", *Ann. Inst. Statist. Math.*, vol.43, pp.1-59, 1991.
- [5] Bezdek J.C., R. Ehrlich and W. Full, "The fuzzy c-means clustering algorithm", *Computers and Geosciences*, vol. 10, pp. 191-203, 1984.
- [6] Borel C.C. and S.A. Gerstl, "Nonlinear spectral mixing models for vegetative and soil surfaces", *Remote Sensing of Environment*, pp.403-416, 1994.
- [7] Bosdogianni P., M. Petrou and J. Kittler, "Mixed pixel classification in Remote Sensing". Proceedings of the EUROPTO series, Image and Signal Processing for Remote Sensing, vol.2315, pp.494-505, 1994.

-
- [8] Bosdogianni P., M. Petrou and J. Kittler, "Mixture models with higher order moments", to appear in *IEEE Trans. Geosci. Remote Sensing*, 1996.
 - [9] Bosdogianni P., M. Petrou and J. Kittler, "Robust mixed pixel classification in Remote Sensing", *VIII ISPRS-Congress*, vol. VII, pp.85-90, Vienna, Austria, 1996.
 - [10] Bosdogianni P., M. Petrou and J. Kittler, "Mixed pixel classification with Robust Statistics", submitted to *IEEE Trans. Geoscience and Remote Sensing*, 1996.
 - [11] Bosdogianni P., H. Kälviäinen, M. Petrou and J. Kittler, "Robust unmixing of large sets of mixed pixels", submitted to *Journal of Pattern Recognition*, 1996.
 - [12] Bosdogianni P., M. Petrou and J. Kittler, "Robust mixed pixel classification using the hypothesis testing Hough transform", *IGARSS*, Lincoln, USA, pp.1376-1378, 1996.
 - [13] Bosdogianni P., M. Petrou and J. Kittler, "Classification of sets of mixed pixels with the hypothesis testing Hough transform", submitted to *IEEE Vision, Image and Signal Processing*, 1996.
 - [14] Brown P.J, "Multivariate calibration (with discussion)", *Journal of the Roy. Stat. Soc. Ser. B*, vol. 44, pp. 287-321, 1982.
 - [15] Bryant R.G., "Validated linear mixture modelling of Landsat TM data for mapping evaporate minerals on a playa surface: methods and applications", *International Journal of Remote Sensing*, vol. 17, no. 2, pp. 315-330, 1996.
 - [16] Cetin H., T. Warner, D. Levandowski, "Data classification, visualization and enhancement using n-dimensional probability density functions (nPDF): AVIRIS, TIMS, TM, and geophysical applications", *Photogrammetric Engineering and Remote Sensing*, vol. 59, no. 12, pp. 1755-1764, 1993.

-
- [17] Coppin P.R. and M.E. Bauer, "Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features", *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no.4, pp.918-927, 1994.
- [18] Cross A.M., J.J. Settle, N.A. Drake, R.T.M. Paivinen, "Subpixel measurement of tropical forest cover using AVHRR data", *Int. J. Remote Sensing*, vol. 12, no. 5, pp. 1119-1129, 1991.
- [19] Duncan J. and J. Franklin, "Estimating fractional vegetation cover at the sub-pixel scale in a semiarid region using a statistical mixture model and remotely sensed data", *IGARSS*, 1994.
- [20] Ediriwickrema J., S. Khorram, "Classification accuracy improvement and delineation of mixed pixels using hierarchical image classification", *IGARSS*, Lincoln, USA, pp. 793-796, 1996.
- [21] Farrand W.H., R. B. Singer and E. Merenyi, "Retrieval of apparent surface reflectance from AVIRIS data:a comparison of empirical line radiative transfer and spectral mixture methods", *Remote Sens. of Environ.*, pp. 311-321, 1994.
- [22] Foody D.B.G.M.and D.P.Cox, "Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions", *International Journal of Remote Sensing*, vol. 15, pp. 619-631, 1993.
- [23] Franklin J., "Thematic Mapper analysis of tree cover in semiarid woodlands using a model of canopy shadowing", *Remote Sensing of Environment*, pp. 189-201, 1991.
- [24] Gebbinck M.S.K. and T.E. Schouten, "Decomposition of mixed pixels", *Image and Signal Processing for Remote Sensing II*, SPIE 2579, pp.104-115, Paris, 1995.
- [25] Germmd D. and Ake B., "Numerical methods", Englewood Cliffs N.J., Prentice-Hall, 1974.

-
- [26] Gillespie A.R., "Spectral mixture analysis of multispectral thermal infrared images", *Remote Sensing of Environment*, vol. 42, pp. 137-145, 1992.
- [27] Goel N.S., "Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data, *Remote Sens. Rev.*, pp.1-213, 1988.
- [28] Gong P. , J.R. Miller and M. Spanner, "Forest canopy closure from classification and spectral unmixing to scene components - multisensor evaluation of an open canopy", *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 5, pp.1067-1079, 1994.
- [29] Horwitz H.M., R.F. Nepleka, P.D. Hyde and J.P. Morgenstern, "Estimating the proportions of objects within a single resolution element of a multispectral scanner", *Proceedings of the 7th International Symposium of Remote Sensing of Env. ERIM*, no. 10259-1-x, pp. 1307-1320, 1971.
- [30] Huete A.R., "Separation of Soil-Plant spectral mixtures by factor analysis", *Remote sensing of environment*, pp. 237-251, 1986.
- [31] Jasinski M.F. and P.S. Eagleson, "The structure of Red-Infrared scattergrams of semivegetated landscapes", *IEEE Trans. Geosci. Remote Sensing*, vol. 27, no. 4, pp. 441-451, 1989.
- [32] Jasinski M.F. and P.S. Eagleson, "Estimation of subpixel vegetation cover using red-infrared scattergrams", *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 2, pp. 253-267, 1990.
- [33] Kälviäinen H., P. Bosdogianni, M. Petrou and J. Kittler, "Mixed Pixel Classification with the Randomized Hough Transform", *ICPR*, Vienna, Austria, July, 1996.

-
- [34] Kälviäinen H., P. Hirvonen, L. Xu, and E. Oja, Probabilistic and Non-probabilistic Hough Transforms: Overview and Comparisons, *Image and Vision Computing*, vol. 13, pp. 239-252, 1995.
- [35] Kauth R.J. and G.S. Thomas, "The tasselled cap- A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat", *Proc. Symp. on Machine Processing of Remotely Sensed data*, pp. 4B41-4B51, 1976.
- [36] Illingworth J. and J. Kittler, "A survey of the Hough Transform", *CVGIP*, vol.44, pp.87-116, 1988.
- [37] Kruse F.A., Lefkoff A.B., Broadman J.W., Heidebrecht K.B., Shapiro A.T., Barloon P.J., Goetz A.F.H, "The Spectral Image Processing System (SIPS) interactive visualization and analysis of imaging spectrometer data", *Remote Sensing of environment*, vol.44, pp.145-163, 1993.
- [38] Lawton W.H. and E.A. Sylvestre, "Self modeling curve resolution", *Technometrics*, vol. 13, pp. 617-633, 1971.
- [39] Leavers V.L., "Shape detection in computer vision using the Hough transform", *Springer-Verlag London Limited*, 1992.
- [40] Lenington R.K. and C.T. Sorensen, R.P. Heydorn, "A mixture model approach for estimating crop areas from Landsat data", *Remote Sensing of environment*, vol. 14, pp. 197-206, 1984.
- [41] Li X. and A.H. Strahler, "Geometric-optical modeling of a conifer forest canopy", *IEEE Trans. Geosci. Remote Sensing*, vol. 23, pp. 705-721, 1985.
- [42] Liang S., A.H. Strahler and X. Li, "Simultaneous retrieval of subpixel proportions and signatures from two component mixture", *ACSM ASPRS Fall conversion*, Atlanta, USA, pp. B-108-117, 1991.

-
- [43] Lillesand K., "Remote sensing and image interpretation", *John Wiley & Sons*, 1994.
- [44] Mahar K., Affi M.S., "Linear and correlation analysis for computerized identification of categories in Landsat images", *Int. J. Remote Sensing*, vol. 16, no. 12, pp. 2277-2284, 1995.
- [45] Marsh S.E., P. Switzer, W.S. Kowalik, R.J.P. Lyon, "Resolving the percentage of component terrains within single resolution elements", *Photogrammetric Engineering and Remote Sensing*, no. 8, pp. 1079-1086, 1980.
- [46] Martens H., T. Naes, "Multivariate Calibration", *John Wiley & Sons Ltd.*, 1989.
- [47] Mather P.M., "A computationally-efficient maximum likelihood classifier employing prior probabilities for remotely sensed data", *Int. J. Remote Sensing*, vol. 6, pp. 369-376, 1985.
- [48] McCloy K.R. and K.A. Hall, "Mapping the density of woody vegetative cover using Landsat MSS digital data.", *International Journal of Remote Sensing*, vol. 12, no. 9, pp. 1877-1885, 1991.
- [49] Mertes L.A.K., M.O. Smith and J.B. Adams, "Estimating suspended sediment concentrations in surface waters of the Amazon river wetlands from Landsat images", *Remote Sens. of Environ.*, vol. 43, pp. 281-301, 1993.
- [50] Millington A.C., N.A. Drake, J.R. Townsend, N.A. Quarmby, J.J. Settle and A.J. Reading, "Monitoring salt playa dynamics using Thematic Mapper data", *IEEE Transactions on geoscience and remote sensing*, vol. 27, no. 6, pp. 754-761, 1989.
- [51] Otterman J., "Albedo of a forest modeled as a plane with dense protrusions", *J. Climate Appl. Meteor.*, vol. 22, pp. 297-307, 1984.

-
- [52] Palmer P., M. Petrou and J. Kittler, "A Hough Transform Algorithm with a 2D Hypothesis Testing Kernel", *CVGIP Image Understanding*, Vol. 58, No.2, pp. 221-234, 1993.
- [53] Palmer P., J. Kittler and M. Petrou "An optimising line finder using a hough transform algorithm", in press *CVGIP Image Understanding*, 1996.
- [54] Palmer P., M.Petrou and J.Kittler, "Using focus of attention with the Hough Transform for accurate line parameter estimation", *Pattern Recognition*, vol. 27, No. 9, pp. 1127-1134, 1994.
- [55] Papoulis A., "Probability random variables & stochastic processes", John Wiley & Sons, 1991.
- [56] Pech R.P., A.W.Davis, R.R.Lamacraft and R.D.Craetz, "Calibration of LANDSAT data for sparsely vegetated semi-arid rangelands", *Int. J. Remote Sensing*, vol. 7, no. 12, pp. 1729-1750, 1986.
- [57] Pickup G., V.H. Chewings and D.J. Nelson, "Estimating changes in vegetation cover over time in arid rangelands using Landsat MSS data", *Remote Sens. of Environ.*, vol. 43, pp. 243-263, 1993.
- [58] Peddle D.R., F.G. Hall and E.F. LeDrew, "A comparison of NDVI and spectral mixture analysis for estimating biophysical properties in Boreal forest terrain", *IGARSS*, pp.230-232, 1996.
- [59] Price J.C., "The contribution of thermal data in Landsat multispectral classification", *Photogrammetric Engineering and Remote Sensing*, vol. 47, pp.229-236, 1981.
- [60] Privette J.L., "An efficient strategy for the inversion of bidirectional reflectance models with satellite remote sensing data", *PhD thesis*, University of Colorado, Dept. of Aerospace Engineering Science, 1994.

-
- [61] Rambal S., B.Lacaze and T. Winkel, "Testing an area-weighted model for albedo or surface temperature of mixed pixels in Mediterranean woodlands", *Int. J. Remote Sensing*, vol. 11, no. 8. pp. 1495-1499, 1990.
- [62] Richardson A.J., C.L. Wiegand, "Distinguishing vegetation from soil background information", *Photogrammetric Engineering and Remote Sensing*, vol. 43, no. 12, pp. 1541-1552, 1977.
- [63] Roger R.E., "Sparse inverse covariance matrices and efficient maximum likelihood classification of hyperspectral data", *International Journal of Remote Sensing*, vol. 17, no.3., pp. 589-613, 1996.
- [64] Rokos D. and D. Argialas, "Study of forest vegetation regeneration based upon Landsat TM images analysis: preliminary results".
- [65] Roberts D.A., M.O. Smith, J.B.Adams, D.E. Sabol, A.R. Gillespie and S.C. Willis, "Isolating woody plant material and senescent vegetation from green vegetation in AVIRIS data", *Proc. Airborn Sci. workshop AVIRIS*, JPL Publications 90-54, pp. 42-57, 1990.
- [66] Rosenthal W. and J. Dozier, "Automated mapping of Montane snow cover at subpixel resolution from the Landsat Thematic Mapper", submitted to *Water Resources Research*, 1996.
- [67] Sabol D.E., D.A. Roberts, J.B. Adams and M.O.Smith, "Mapping and monitoring changes in vegetation communities of Jasper Ridge, CA using spectral fractions derived from AVIRIS images", *Proc 4th Annual JPL Airborne Geoscience Workshop*, Wash. DC, 1993.
- [68] Sabol D.E., J.B. Adams and M.O. Smith, "Quantative subpixel spectral detection of targets in multispectral images", *Journal of Geophysical Research*, vol. 97, no. E2, pp. 2659-2672, 1992.

- [69] Schanzer D.L., "Comments on The least-squares mixing models to generate fraction images derived from remote sensing multispectral data", *IEEE Transaction on geoscience and remote sensing*, vol. 31, no.3, 1993.
- [70] Settle J.J. and N.A. Drake, "Linear mixing and the estimation of ground cover proportions", *International Journal of Remote Sensing*, vol. 14, pp. 1159-1177, 1993.
- [71] Shimabukuro Y.E. and J.A. Smith, "The least-squares mixing models to generate fraction images derived from remote sensing multispectral data", *IEEE Trans. Geosci. Remote Sensing*, vol. 29, no. 1, pp. 16-20, 1991.
- [72] Smith M., D.Roberts, J.Hill, W. Mehl, B. Hosgood, J.Verdebout, G.Schmuck, C.Koechler and J.Adams "A new approach to quantifying abundances of materials in multispectral images", *IGARSS*, 1994.
- [73] Smith M.O., J.B.Adams and A.R. Gillespie, "Reference endmembers for spectral mixture analysis", *fifth Australian remote sensing conference*, vol.1, pp.331-340, 1990.
- [74] Smith M.O., S.L. Ustin, J.B. Adams and A.R. Gillespie, "Vegetation in Deserts: I. A regional measure of abundance from multispectral images", *Remote Sensing of Environment*, vol.31, pp.1-26, 1990.
- [75] Tabachnick B.G. and L.S. Fidell "Using Multivariate Statistics", HarperCollins Publishers, 1989.
- [76] Townshend J.R.G. and C.O. Justice, "Information extraction from remotely sensed data. A user view.", *International Journal of Remote Sensing*, vol. 2, pp. 313-329, 1981.

- [77] Tucker C.J., "A comparison of satellite sensor bands for vegetation monitoring", *Photogrammetric engineering and remote Sensing*, vol. 44, no. 11, pp. 1369-1379, 1978.
- [78] Var der Meer F., "Spectral unmixing of Landsat Thematic Mapper data", *Int. J. Remote Sensing*, vol.16, no.16. pp. 3189-3194, 1995.
- [79] Xu L., E. Oja, and P. Kultanen, A New Curve Detection Method: Randomized Hough Transform (RHT), *Pattern Recognition Letters*, vol. 11, pp. 331-338, 1990.
- [80] Zhang D., J.Vandeneede, P. Wambacq and A.Oosterlinck, "Robust classification of remote sensing images", *IGARSS*, pp.1530-1534, 1993.
- [81] Williamson H.D., "Estimating herbage cover with regression and linear mixture models in a semi-arid area", *Remote Sensing from research to operation*, pp.388-397, 1993.
- [82] "Sony 3CCD Color Video Camera, Model DXC930P, Operating Instructions", 1995.
- [83] "Digicolor - True Color Digitizer and Display Device Set, Hardware Reference Manual", Document No. HW355-3.0, Datacube Inc., 1991.