CLASSIFICATION AND CONTEXTUAL ENHANCEMENT
OF REMOTELY SENSED DATA

DAVID JULIAN BOOTH

Doctor of Philosophy

THE UNIVERSITY OF ASTON IN BIRMINGHAM

October 1989

ASTON UNIVERSITY

CLASSIFICATION AND CONTEXTUAL ENHANCEMENT
OF REMOTELY SENSED DATA

DAVID JULIAN BOOTH. DOCTOR OF PHILOSOPHY, 1989.

SUMMARY

The aims of the project were twofold: 1) To investigate
classification procedures for remotely sensed digital
data, in order to develop modifications to existing
algorithms and propose novel classification procedures;
and 2) To investigate and develop algorithms for
contextual enhancement of classified imagery in order to
increase classification accuracy.

The following classifiers were examined: box, decision
tree, minimum distance, maximum likelihood. In addition to
these the following algorithms were developed during the
course of the research: deviant distance, look up table
and an automated decision tree classifier using expert
systems technology. Clustering techniques for unsupervised
classification were also investigated.

Contextual enhancements investigated were: mode filters,
small area replacement and Wharton's CONAN algorithm.
Additionally methods for noise and edge based
declassification and contextual reclassification, non-
probabilistic relaxation and relaxation based on Markov
chain theory were developed.

The advantages of per-field classifiers and Geographical
Information Systems were investigated.

The conclusions presented suggest suitable combinations of
classifier and contextual enhancement, given user
accuracy requirements and time constraints. These were
then tested for validity using a different data set. A
brief examination of the utility of the recommended
contextual algorithms for reducing the effects of data
noise was also carried out.

Key Phrases: Remote Sensing
            Digital Image Processing
            Image Classification
            Contextual Enhancement

ACKNOWLEDGEMENTS

CONTENTS

4

7

10

11

12

13

14

List of Tables

List of Figures

CHAPTER 1

1.1 INTRODUCTION

'Effort should be concentrated on improvement of image processing techniques, in particular methods of contextual classification and the integration of other georeferenced data sets, to make maximum use of available imagery.'

(Legg, 1988, report to United Kingdom National Remote Sensing Centre on applications of remotely sensed data in the United Kingdom)

1.2 THE PROBLEM

Classification of remotely sensed images is part of the wider topic of pattern recognition and artificial intelligence. An image classification algorithm is essentially a kind of specialised expert system, fed examples (training data), which then uses this information to induce rules which are used to classify all the image pixels. Different classification algorithms are thus analogous to the different production rules used by expert systems.

The classifiers used for remotely sensed data do, however, differ in two important aspects from these production rules:

1.They must be capable of handling vast amounts of data quickly and efficiently.

19

2.They must be able to cope with the high noise levels inherent in such data (in this context 'noise' is defined in section 2.7.1).

Because of this, remote sensing classifiers have concentrated on parametric techniques at individual pixel level, usually assuming that the frequency distributions of class spectral responses are normal so that the spectral response of a class can be modeled using a combination of mean, variance and covariance.

Of the algorithms used here, the majority make the assumption that a class' spectral response can be adequately represented by a combination of mean digital values (class mean vector), class standard deviations in each spectral band of data and class covariances, although not all of the algorithms use all these statistics. A summary of which statistics are required by which algorithm is given in table 8.4. An attempt has also been made to apply nonparametric methods to the data (look up table and decision tree classifiers). In theory, the lack of assumptions about the data should lead to higher performance classifiers; however, in practice the large data storage requirements, the characteristics of remotely sensed data (section 2.7), and the requirements for many training samples to accurately represent each class, currently restrict the usefulness of these classifiers.

The characteristics of remotely sensed data result in incorrect classifications, at class boundaries (due to the mixed spectral response of these pixels, sometimes called mixels) and randomly throughout the image (due to random or systematic noise, see section 2.7). Assuming a reasonable classification accuracy, these areas will be small relative to those of correct class. The process of contextual reclassification makes use of this assumption to reassign pixels to more appropriate classes.

In its simplest form, a contextual reclassifier can be represented by a mode filter passed over the classified image (a mode filter returns the most commonly returning value over a small area as the central value of that area). More sophisticated techniques involve the identification and declassification of 'small areas' (small areas relative to the overall pattern and containing anomalous classes), or the identification of noise and boundary pixels from the original imagery by means of thresholded high-pass filters and the subsequent declassification and reclassification of pixels identified as such.

Another area of contextual enhancement is relaxation labeling, where information about autocorrelation is used to reclassify all image pixels. In this study, non-probabilistic relaxation methods are examined, using rank order, minimum distance rules and Markov chain theory to enhance classified imagery.

21

An important side effect of contextual enhancement is the generalisation of classified data to a form more suitable for map-type output. The noise and 'mixel' removal process may also remove small, correctly classified, areas. This can be viewed as a process akin to map generalisation, necessary to make maps useful and legible. The generalised output better enables the user to view overall trends in the data without the distractions of fine detail (which will frequently be actually erroneous because of noise).

The development of geographical information systems and their associated digital map data will, in the near future, provide users of remotely sensed data with the chance to classify images on a per-segment basis, with the segments defined using the geographical information system. An initial investigation into the potential benefits of this is presented here.

Also presented is a brief study of the relative contributions of noise, edges and mixels to classification error, using the characteristics of multispectral data to enable identification and separation of true edges from noise elements.

1.3 STUDY AREAS

Four sets of data were used in this study: three Landsat Thematic mapper sub-images (see section 1.4), using various band combinations, and one Landsat MSS sub-image.

1.3.1 PEAK DISTRICT STUDY AREA

The study area covers the Derbyshire Peak District to the north and east of Buxton. The area is a square 25.6 by 25.6 km (1024 by 1024 25m square pixels), with Ordnance Survey grid co-ordinate 400000 370000 as the south west corner.

The area has many differing land cover types from urban areas (Buxton, Chapel-en-le-Frith) and industry (cement works and quarries) to agricultural land, open moorland and reservoirs.

The geology and soils of the area also vary widely. In the north the area predominantly consists of a series of sandstone escarpments, interspersed with shales, allowing valley formation. In the south, this gives way to limestone, forming an upland plateau, dissected occasionally by deep gorges. In the extreme north-west the area includes part of the Cheshire plain.

Terrain is also variable, from areas of high relief in the sandstone area, to the flatter limestone areas in the south.

Overall, the area provides an excellent variety of conditions suitable for testing classification and contextual enhancement algorithms. The only cover type which is lacking is intensive cropland, based on large fields, as occurs in eastern England and East Anglia; however, in view of the other suitabilities of this area,

23

this omission was felt to be insignificant.

In addition to this main data set, three other sets of imagery were used in order to test the findings of this study.

1.3.2 SALISBURY PLAIN STUDY AREA

A three band Landsat TM image and associated training data was made available by one of the author's colleagues. The data consisted of a band 4,5,7 composite image, covering an area of 512 by 512 pixels, over part of Salisbury plain. Ground cover types were defined for natural grasslands, farmland, urban areas, woodland and other man-made features.

1.3.3 NORTH WALES STUDY AREA

A three band Landsat TM image covering an area of North Wales and the Cheshire Plain (bands 4,5 and 7) was made available to the author. Broad ground-cover classes covering urban areas, salt marsh, grassland etc. were defined (see section 10.12).

1.3.4 YEMEN STUDY AREA

A four band landsat MSS image was made available covering a small area of Yemen. Training data was provided by a colleague with considerable experience of the geology of the area.

1.4 SATELLITE SENSORS

1.4.1 LANDSAT MSS

The multispectral scanner carried on board Landsats 1-5 provided data in four spectral bands, corresponding to visible green, red, near and mid infra red. Spatial resolution was 79 by 56 metres, usually corrected to a nominal 80m during processing. Landsat MSS data has been available since the launch of Landsat 1 in 1972. For further information the reader is directed to the textbooks detailed in section 3.2.1.

1.4.2 LANDSAT THEMATIC MAPPER

The thematic mapper sensor was carried on NASA's Landsat 4 and 5 spacecraft. It represents the 'second generation' of sensors, incorporating many alterations and improvements to the multi spectral scanner (MSS) carried on board Landsats 1-3, and also Landsats 4 and 5.

Landsat 4 was launched on 16th. July 1982, Landsat 5 on 1st. March 1984. Both had a near polar sun-synchronous orbit, with a repeat cycle of 16 days.

Thematic mapper scenes cover, in 7 spectral bands, areas 185 by 185 km on the ground, with 5.4% overlap and 7.3% side overlap at the equator, increasing at higher latitudes:

band	wavelength (um)	resolution (m)	notes
1	0.45-0.52	30	blue
2	0.52-0.60	30	green
3	0.63-0.69	30	red
4	0.76-0.90	30	near infra red
5	1.55-1.75	30	near infra red
6	10.4-12.5	120	thermal infra red
7	2.08-2.35	30	near infra red

Imaging is achieved using a scanning mirror, generating 16 lines per sweep (4 for the thermal band). The data is quantised to 8 bits (0-255) reflecting the high radiometric sensitivity of the instrument, for a remote sensing imaging device.

1.5 DIGITAL IMAGE PROCESSING

The images used in this study are digital representations of physical images. Each digital image consists of a series of 'bands' representing different parts of the electromagnetic spectrum. Each band contains an image of surface radiances (or brightness values) for its particular part of the electromagnetic spectrum. These images consist of a regular grid of picture elements (pixels), each represented by a

digital number, the value of which is correlated to
the brightness value of the pixel's corresponding
area on the ground.

The diagram below shows this more clearly:

original   digital image representing this
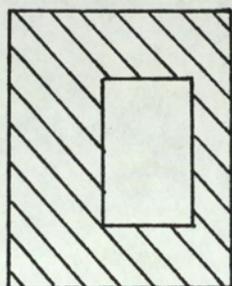
                        8 8 8 8 8 8

                        8 8 1 1 1 8

                        8 8 1 1 1 8

                        8 8 1 1 1 8

                        8 8 8 8 8 8

Conversion of images into digital format enables computer
processing, since the images now consist of numbers, which
can be mathematically manipulated. This is the essence of
digital image processing.

Remotely sensed images are commonly available in digital
format, and digital image processing represents an
important tool to aid their interpretation.

CHAPTER 2

AIMS, OBJECTIVES, DEFINITIONS

The aims and objectives of this study are summarized below:

i.To investigate methods of classification of remotely sensed imagery and contextual methods of increasing the accuracies of these classifications.

ii.From these investigations, to provide a series of recommended classification procedures, given: a) user requirements, and b) the nature of the data to be classified.

2.1 CLASSIFICATION

In terms of classification, the aims of this study were:
a) to investigate currently available classification algorithms for multispectral remotely sensed data and assess the relative benefits of the above algorithms in terms of computational speed and the accuracy of the classifications produced;
b) to develop further classification algorithms, and also to assess these as described above, and
c) to investigate newer algorithms suggested recently in the literature of remote sensing and the technology of expert systems , again assessing these using the criteria of computational speed and accuracy.

d) To use the data from the above analyses to suggest suitable algorithms for use in commonly occurring remote sensing problems.

2.2 CONTEXTUAL ENHANCEMENT

In terms of contextual enhancement, the aims were to investigate and develop algorithms to:

a) produce higher classification accuracies from the output of the classifiers used in section 2.1;

b) generalise imagery to make it more suitable for presentation as thematic maps.

c) compare the algorithms in terms of accuracy increases and computer time expended in producing these increases.

d) investigate any relationship between initial classifier and subsequent contextual enhancement algorithm and recommend suitable pairings of these to be used under differing circumstances.

2.3 CLUSTERING

Under the heading of clustering the aims were to assess the usefulness of unsupervised classification algorithms for image classification and suggest suitable situations for their use.

2.4 PER-SEGMENT CLASSIFICATION

Under the heading of per-segment classification the segment identifier examined was the field boundary and the aim was to examine potential benefits to be obtained from classification on a per-field basis, rather than the currently prevailing per-pixel methods.

2.5 EFFECT OF NOISE AND EDGES ON CLASSIFICATION

Under the heading of noise and edge effects the aim was to assess the relative contributions of noise and edge pixels (pixels whose spectral response is a mixture of those for differing cover types) to error rates in classification of remotely sensed imagery.

2.6 FUTURE STUDY

A further consideration was made of how current and future development in Information Technology may affect approaches to image classification in land and water studies. Here the aim was to discuss and identify areas for future study, notably in the fields of artificial intelligence and geographical information systems, as applied to classification of remotely sensed data.

2.7 SOME DEFINITIONS

2.7.1 NOISE

Noise, in terms of classification of satellite imagery can be defined broadly as any characteristic of a pixel or

image area which causes an incorrect classification to take place.

More specifically, the methods described and developed in this thesis are concerned with the removal of the effects of two types of noise:

The effects of noise on imagery are discussed by Ehlers (1985), who presents methods for estimating image signal to noise ratios in images containing unknown noise levels.

2.7.1.1 RANDOM NOISE

Random noise is defined here as anomalous pixel values caused by, for example, transmission errors between satellite and receiving station or magnetic tape defects. Such noise can commonly be seen as part of remotely sensed data as individual pixels, or groups of pixels which 'clash' with their surroundings when displayed.

2.7.1.2 SYSTEMATIC NOISE

A common occurrence of this type of noise is Landsat MSS 'sixth line banding'. This type of noise is defined as that caused by the physical properties of the sensing device, thus for scanning systems such as Landsat MSS and TM banding occurs, whilst for push-broom sensors, such as the scanner on board the SPOT satellite, vertical striping occurs. These effects are due to the arrays of multiple sensors containing transducers which are not quite identical. For example, in the case of MSS, slight differences in the calibration of the six sensors making

31

up the array for one waveband of data will result in a repeated discrepancy of digital values every six lines of the image.

2.7.2 NOISE, ACCURACY AND GENERALISATION

There is a complex relationship between noise, classification accuracy and generalisation, which can be broadly defined in terms of scale: at large scales it is important to show high-frequency variation, whereas at small scales it is often desirable to suppress this high frequency information in order to show a more general picture, which is more suitable for interpretation of significance.

2.7.3 CONTEXT AND TEXTURE

Context is defined here as the relationship between the class of a pixel and the classes of its neighbouring pixels, over a region of an image. this differs from the broader definition of context which includes all information, of whatever type, relevant to the classification of a pixel.

Texture, as applied to remotely sensed data, has a similar definition to the first given above for context, but here the relationship is between the digital values, not the class labels. Context could, perhaps, be seen as a special case of texture. Some contextual reclassification algorithms are similar to the calculations for texture measures, for example, neighbouring grey level

dependencies (Golton, 1987) are similar in context to the

Markov relationships described in section 10.3.4.

CHAPTER 3

LITERATURE REVIEW

3.1 INTRODUCTION

Image classification is covered in many subject areas
which are often seen as separate. Restricting study to
within the remote sensing literature will limit the
information available. Only too often the remote sensing
community duplicates work done elsewhere, or discovers new
techniques long after they have become commonplace tools
in other fields, such as artificial intelligence, pattern
recognition, robot vision and medical physics.

This can often be justified by the nature of remotely
sensed data. It is very different from the data used in
other fields, and introduces many problems unique to
itself which hinder successful implementation of
algorithms developed elsewhere. Essentially, remotely
sensed data is high-volume, low-precision (even in the
case of so-called 'high radiometric precision' sensors,
such as Landsat Thematic Mapper, whose data is quantised
to 256 grey levels - eight bit precision), prone to noise
and poor radiometric resolution, whereas data used in
other fields is either: high-precision, low-volume and
consequently better suited to processing using
conventional algorithms and computer architecture; or, for
example in the case of robot vision, the problems are
addressed by developing special processing architecture.

34

Another feature is that most vision work has been concerned with single sensor (black/white) imagery.

3.2 TEXTBOOKS

3.2.1 REMOTE SENSING

There are many general textbooks covering remote sensing in varying degrees of detail and complexity. Good general overviews are provided at a basic level by Curran (1985), Barrett and Curtis (1982) and Drury (1987), and in more detail by Mather (1987), Swain and Davis (1978) and Lillesand and Kiefer (1979). For a thorough coverage of all aspects of remote sensing, the reader is referred to

the second edition of the Manual of Remote Sensing (Vol.1, eds: Simonett and Ulaby, 1983. Vol.2, eds: Estes and Thorley, 1983).

There are also a number of textbooks covering specific applications of remote sensing. Notable amongst these is Drury (1987) covering geological remote sensing.

3.2.2 IMAGE PROCESSING

There are also several textbooks, aimed at the remote sensing community, dealing with image processing. An excellent overview is provided by Schowengerdt (1983). Mather (1987) also contains much detail on image processing, for remote sensing applications.

Golton (1987) provides a good review of digital image processing algorithms, covering non remote sensing topics as well as those more closely related to this field.

Duda and Hart (1973) provide an excellent synopsis of image classification algorithms, not limited to the statistical pattern recognition techniques covered by the remote sensing texts, whilst cluster analysis is covered by Anderberg (1973).

The two papers by Jupp, Strahler and Woodcock (1988, 1989) provide a good mathematical and theoretical background to structures within digital images.

3.3 GEOMETRIC AND RADIOMETRIC CORRECTION

In order to relate easily remotely sensed data to maps, and enable classifier training operations, geometric correction of the imagery is necessary. The usual methods of estimating transformation equations by least squares fitting of polynomials using ground control points are covered by the textbooks (Mather (1987), Swain and Davis (1978)).

Choosing suitable ground control points is vital to producing accurately registered images. Davison (1986) details the ground control point selection criteria used by the UK National Remote Sensing Centre. The paper also

describes a technique for automatically locating control

points on imagery, where multiple images must be corregistered.

From time to time alternatives to the conventional polynomial models are suggested in the literature, for example, piecewise correction (Swain and Davis, 1978). Recently Goshtasby (1988) has suggested the use of spline surfaces, enabling corresponding points to be registered exactly. The technique, however, is extremely sensitive to incorrect ground control point registration, since it fits a surface honouring all control points; for this reason, least squares fitting to global polynomials may still be of more use with remotely sensed data, since this technique is more tolerant of incorrect control point location.

Bergeson, Batson and Kieffer (1985) present results showing that Landsat TM data is entirely suitable for accurate mapping at scales of up to 1:100,000, but that geometric distortion is unacceptable at scales of 1:24,000 and larger. Colvocoresses (1986) also describes the use of Landsat TM for 1:100,000 scale mapping.

Radiometric correction of imagery is a problem limited to remotely sensed data. It is necessary either when:

a) absolute physical values (e.g. bi-directional reflectance of a surface) are to be calculated from the image pixel values,

or

37

b) Multitemporal and/or multisensor data are to be directly compared with each other.

Occasionally, there may also be a requirement for radiometric correction where there is a problem of within-scene variation, caused either by sensor characteristics, or variations in atmospheric effects across the scene.

Steven and Rollin (1986) provide a good overview of the needs for such correction, and describe a technique for correcting imagery taken by several airborne scanners (Daedalus airborne thematic mapper, in this case).

Forster (1984) covers band ratioing techniques compared to methods for calculating reflectances using Landsat Multi-Spectral Scanner data with particular reference to urban data.

Robinove (1982) details the corrections necessary to convert Landsat Multi-Spectral Scanner digital values to their equivalent physical values. The paper covers the sensors carried on board Landsat 1,2 and 3, however, the calculations are similar for the more recent sensors, only the constants (provided in the paper for each instrument) requiring alteration.

A comprehensive review of the effects of radiometric variation on classification is provided by Duggin (1985) who lists the following factors as affecting the final brightness value recorded by a sensor:

i) Passage of radiation to target through the atmosphere.

ii) Reflection at the target.

iii) Passage from target to detector through the atmosphere.

iv) The effect of the detector itself.

v) The effects of unresolved cloud and haze.

vi) The effects of random and systematic errors in the recording of radiance levels.

The paper also notes that discrimination of a target by a classifier may be better for some sun-target-sensor geometries than for others.

A thorough review of the radiometric correction techniques used by the Canada Centre for Remote Sensing (CCRS) is provided by Ahern et al (1987), ranging from simple destriping algorithms to absolute radiometric calibration of the data.

Royer et al (1987) detail techniques for rescaling satellite data to radiometrically match the data from other satellite sensors.

A review of the major types of radiometric correction as applied to satellite data in the visible and infra-red

parts of the electromagnetic spectrum is provided by Tiellet (1986) who lists the following:

radiometric calibration and de-striping

radiometric correction of Lambertian scenes

atmospheric considerations

topographic considerations

view-angle considerations

reflectance models

as important for radiometric correction of remotely sensed data.

The paper concludes that the type of radiometric correction to be performed on a data set must be appropriate to the study being performed.

3.3.1 TOPOGRAPHIC EFFECTS

Topography can play an important role in affecting the spectral response of cover types:

Hall-Konyves (1987) examines the effect of sun angle on terrain with slopes of less than 15 degrees using Landsat MSS and TM data. Significant effects are reported only for pasture land, whilst agricultural land and forestry are reported to show minimal topographic effects.

The variation of sun and look angle effects on spectral response over different spectral regions and wavebands is described by Stohr and West (1985). They report that the

shorter wavelengths are most affected by sun-target-sensor geometry.

3.4 ACCURACY ASSESSMENT

Assessing the accuracy of thematic maps (ie classified imagery in this case) is an area of the literature fraught with contradiction. Almost every assumption made about the data, in order for the statistical tests to be valid, seems to be broken by remotely sensed data. For this reason, a variety of accuracy assessment methods have been proposed.

Techniques for accuracy assessment are also used in photointerpretation. Congalton and Mead (1983) propose using discrete multivariate analysis of error matrices. Although the paper is based on designing a sampling strategy to produce desired map accuracies within stipulated confidence limits, the techniques can be reversed to estimate accuracies and confidence limits for given sampling fractions.

Hord (1976) proposes a method for calculating a single 'figure of merit' to assess the accuracy of a classification, using a statistical sampling procedure. The method determines the accuracy, based on the following criteria:

1. Classification accuracy
2. Boundary line placement
3. Control point placement

41

Hay (1979) introduces the concept of stratified sampling to test land use map accuracy. Five problems are identified in assessment of accuracy:

1. What proportions of all decisions are correct?

2. What proportion of the allocation to a category is correct?

3. What proportion of the true category is correctly allocated?

4. Is a category overestimated or underestimated?

5. Are the errors randomly distributed?

The minimum sample size required to answer these questions is given as 50 points (pixels). Questions 1-4 are addressed using binomial theory, whilst question 5 is answered using Poisson frequencies.

Aronoff (1982) addresses classification accuracy from a user's approach, emphasising the probability of rejecting an accurate map against the probability of accepting an inaccurate one (termed producer's and consumer's risk respectively). The paper also stresses the expense to users of time consuming checking procedures, and suggests the use of quick and efficient accuracy assessment strategies.

Card (1982) compares simple random sampling, using known map category marginal frequencies to improve estimates of thematic map accuracy (ie relative areas of each map category).

42

Ginevan (1979) points out shortcomings in statistical procedures (proposed prior to 1979), and suggests the use of acceptance sampling procedures, combined with the binomial probability density function to develop a 'sound statistical methodology for map accuracy validation'.

Aronoff (1985) proposes the use of a minimum accuracy value, defined as the 'lowest expected accuracy of a thematic map, given an observed accuracy test result and the user-selected consumer risk', and provides tables to enable selection of suitable accuracy tests at various confidence levels and consumer risks.

A coefficient of agreement is suggested as a measure of classification accuracy by Rosenfeld and Fitzpatrick-Lins (1986). This is derived from a confusion matrix, but takes non-diagonal elements into account, as well as those along accuracy assessment methods detailed in the remote sensing literature.

Maxwell (1976) covers the application of multivariate system analysis to assessment of classification accuracy.

Accuracy assessment techniques can, in theory, be used to attempt correction of misclassification. Chrisman (1981) introduces Tenenbein's double sampling method for correcting errors in probabilities, in order to adjust area estimates, calculated from classified data, to represent more accurately ground conditions.

Bradbury and MacDonald (1986) contrast per-point with per-field sampling frames, describing the use of arbitrarily defined classes to enable testing of algorithms.

Harrison, Dunn and White (1989) examine the effects of small sample size and site and the relative performances of stratified, random and systematic sampling on the accuracy assessment of classified Landsat Thematic Mapper imagery. They conclude that:

i) A large number of small sample sites gives a more representative accuracy figure.

ii) Systematic sampling is to be recommended because of its operational advantages.

iii) Any gain in efficiency will vary with cover type and sampling efficiency.

Performance analysis of various classification algorithms to determine their utility for large-area forest mapping is detailed by Yool et al (1986). Differences in accuracies between different classification algorithms were attributed to variations in the sensitivities of the algorithms to spectral variations caused by: background reflectance, differential illumination and spatial pattern by species.

The paper discusses some of the problems of bias involved in the selection of training and test data, and concludes that the 'results emphasise the complexity between land-cover regime, remotely sensed data and the algorithms used

to process these data'.

Spatial autocorrelation analysis is used by Congalton
(1988a) to aid design of optimum sampling schemes for
accuracy assessment. Whilst in another paper (Congalton
1988b) the author compares simple random sampling,
stratified sampling, cluster sampling, systematic sampling
and stratified systematic unaligned sampling using
autocorrelation analysis, finding random sampling to
usually provide the best estimate of classification
accuracy.

3.5 CLASSIFICATION

Techniques for classification of remotely sensed digital
data are well established, and covered thoroughly in the
literature. Recommended texts are: Mather (1987) and Duda
and Hart (1973).

Occasionally new classification algorithms have been put
forward, or brought to the attention of the remote sensing
community. Examples include: the binary decision tree
(Belward and DeHoyos, 1987); the spectral shape classifier
(Pendock, 1987); linear discriminant analysis (Tom and
Miller, 1984); temporal trend analysis (Engvall, Tubbs and
Holmes, 1977) and a spectral knowledge based approach
(Wharton, 1987).

Another aspect of recent research into classification
techniques involves improving existing algorithms -
usually by developing faster versions. Examples of this

include: Mather (1985) who shows how the maximum likelihood algorithm may be made more efficient (and therefore faster), and Bryant (1989) who shows how the nearest neighbour (otherwise known as minimum distance) algorithm can be speeded up.

Several texts are devoted to clustering and cluster analysis: recommended is the book by Anderberg (1973). Occasionally this subject is dealt with in the remote sensing literature; for example Holley and Parker (1973), which demonstrates how clustering algorithms can be adapted for use with remotely sensed data.

A further area covered by published papers is that of comparing algorithms in terms of 'usefulness' for classifying remotely sensed data. Common themes are both speed (computational efficiency) and accuracy and these are covered by Mather (1985), Belward and DeHoyos (1987), Wasrud and Lulla (1985) and Booth and Oldfield (1989). Gong and Howarth (1990) describe a Mahalanobis distance classifier.

Saraf and Cracknell (1989) describe the use of linear discriminant and profile analysis on airborne thematic mapper data to distinguish areas of geochemical anomaly by geobotanical effects. The paper also discusses the use of linear discriminant and profile analysis to identify spectral bands providing best discrimination of vegetation types.

A fast maximum likelihood classifier, implemented in hardware, is described by Settle and Briggs (1987).

Clustering in order to segment clouds is described by Seddon and Hunt (1985), who recommend the use of principal components analysis prior to clustering. An iterative minimum-distance based algorithm was employed as the basis of the clustering. The AMOEBA program, with particular emphasis on clustering, is described by Bryant (1990).

Two papers by Wang Ru-Ye (1986a and 1986b) describe different approaches to automated decision tree design, one using a splitting algorithm to define the clusters, the other a merging algorithm.

Csillag (1986) describes a comparison between a supervised maximum likelihood and a change-vector based classifier, based on separability measures as a measure of accuracy. The paper also describes the use of ISODATA for clustering - an iterative nearest neighbour algorithm (Fukunga, 1972), however this paper reports little correlation between the separability of classes, as measured by separability indices, and classifier accuracy.

An attempt at providing a conceptual framework to describe multidimensional feature space and the relative positions of different ground cover classes is made by the 'Tasseled Cap' analogy (Kauth and Thomas, 1976).

Piecewise, linear binary classification rules (decision trees) are proposed for classifying remotely sensed data

47

by Lee (1985), who cites their use as an inexpensive classifier for multitemporal data sets.

The use of clustering to partition images of clouds is discussed by Pairman and Kittler (1986), who found minimum distance clustering to be ineffective for this purpose. The minimum distance (nearest neighbour) algorithm is presented in modified form to account for 'generalised Gaussian clusters' and population differences between clusters. The authors state that:

'Problems occur (with the nearest neighbour algorithm) where the assumptions of clusters with approximately equal variances and populations are not satisfied by the data.'

Two modifications to the nearest neighbour rule are suggested:

i) taking each cluster's covariance into account (see notes on the 'Deviant Distance' classifier in section 4.3.2)

ii) population weighting of the distance measure.

A methodology for evaluating the performance of classifiers is given by Hudson (1987), who describes the use of contingency tables to evaluate the effectiveness of different classification algorithms on landsat MSS data for mapping forest types. The paper concludes that supervised classifiers were found to be superior to unsupervised types, with the maximum likelihood algorithm found to be best. Seasonal variation of vegetation was

48

found to have more effect on classification accuracy than the algorithm used to perform the classification.

The nearest neighbour algorithm was found to be a better classifier than the maximum likelihood algorithm for classification of agricultural Landsat TM data by Ince (1987), who cites the nonparametric nature of the nearest neighbour rule as being better suited to classification of this kind of data.

The use of fuzzy logic for image segmentation is discussed by Cannon et al (1986), who describe a nonparametric clustering algorithm, based on the fuzzy c-means algorithm. This involves two stages:

i)clustering using fuzzy c-means
ii)merging these clusters using a similarity measure

3.5.1 GROUND DATA

In order to train supervised classifiers, and to test the accuracy of all classification algorithms it is necessary to collect some form of reference data.

Curran and Williamson (1985) detail the problems of obtaining accurate ground data for remote sensing, concluding that the ground data used in a remote sensing study will often be less accurate than the remotely sensed data. Sampling schemes are reviewed, and the following criteria are proposed to help in the choice of an

appropriate sampling scheme:

i) The number of ground resolution elements to be sampled.

ii) The number of samples to be taken per ground resolution element.

iii) The area of samples within a ground resolution element.

iv) Any processing of samples taken.

v) The personnel available in the field to carry out the sampling.

The paper concludes that:

i) The spatial variability of the terrain must determine the appropriate sampling strategy.

ii) Sensor errors can contribute greatly to data inaccuracy.

iii) Unwanted variability in ground data can be attributed to the spatial variability of the terrain.

Buttner, Hajos and Korandi (1989) suggest that improvements in classification accuracy can be brought about by improving the quality of training data. The paper describes various methods of processing training data prior to classification. These techniques are now beginning to be applied within commercial image processing systems (ERDAS press release on version 7.4 of this package, 1990).

3.6 SEGMENTATION

Image segmentation has recently become an area of intense activity within the field of remote sensing. Allan (1986) suggests that classification accuracy could be increased, whilst at the same time the data required could be reduced if image segmentation could be carried out before classification, either by using existing boundary maps or from the imagery itself.

Saarikoski (1988) provides a good overview of segmentation techniques currently being investigated. Technical detail of specific techniques is given by: Cross, Mason and Dury (1988) (split and merge); Hyde, Fullwood and Corrall (1985) (region extraction, border placement and information-integrating techniques) and DiZenzo (1983) (an overview of advances in image segmentation up to 1983).

A common theme, stressed particularly by Saarikoski (1988), is that of the non-trivial nature of the task of accurately segmenting the often noisy remotely sensed data. Much of this paper deals with the possible increases in classification accuracy to be derived from segmented images, whilst covering research into achieving this segmentation practically. Allan (1986) and Booth, Chidley and Collins (1989) look beyond the solution of this problem to the practicalities of per-field classification. The use of other spatial data, geographical information systems and digital maps are also proposed to circumvent

51

the segmentation problem.

3.7 ARTIFICIAL INTELLIGENCE

The remote sensing community has long been excited by the prospect of artificial intelligence improving consistency of interpretation and freeing operators from routine and mundane tasks. Much has been written on potential uses of artificial intelligence: Smith (1984) suggests uses in geographical problem solving; Lesk (1986) discusses the implications for database handling and Lenat (1986) covers this field from the viewpoint of computer software.

Fiegenbaum (1988) details numerous applications of expert systems in industry and commerce, providing a highly readable account of state-of-the-art expert systems and their uses.

Peacegood (1985) describes a prototype expert system to assess the likelihood of the presence of an aquifer in a remotely sensed scene, based around the proprietary expert system shell Sage. It is proposed that such expert systems could guide inexperienced data users in image interpretation, whilst use of such systems to emulate human-like reasoning for image interpretation is suggested. The use of proprietary expert system shells is advocated in order that: 'the task of writing one (be) eased considerably'.

Schreier and Lavkulich (1979) describe what, with hindsight, would now be looked upon as an Expert or

Geographical Information System, combining data from Landsat imagery, aerial photographs, topographic maps, laboratory results and ground survey reports in order to classify off-road trafficability (the ability of terrain to support vehicle movements) in Northern Canada. This early system exhibited the following important features:

1. Combination of spatially related data sets, from different sources.

2. The use of information available only for small areas within the whole study area to maximise reliability where possible, whilst still enabling decisions to be made from smaller data for other areas.

3. The identification of data sets providing maximum information by use of factor analysis.

Wang and Newkirk (1988) describe the use of a combination of image processing and artificial intelligence techniques for highway network extraction. Low-level processing was carried out using per-pixel techniques, whilst interpretation was carried out using an expert system. Using a knowledge base and inference engine an accuracy of 87.7% was achieved for classification of highways in rural areas on Landsat TM imagery.

The use of artificial intelligence in the form of 'Knowledge-Based Aerial Image Understanding Systems' is discussed by Matsuyama (1987), who introduces a model for expert systems for image processing. The paper concludes

that expert systems must not only 'understand' image processing techniques, but that they must also 'understand' the objects which they are analysing in order to work effectively.

A knowledge-based system for automatic interpretation of aerial images of suburban scenes is described by Nicolin and Gabler (1987), with emphasis placed on knowledge representation and control strategy.

The application of artificial intelligence to geographical information systems, and in particular their user interfaces, methods of spatial data representation and the utilisation of spatial data is reviewed by McKeown (1987)

The use of a-priori knowledge as an aid to interpretation of radar data is described by McGuinness (1988).

Two expert systems, developed by the Canada Center for Remote Sensing, are described by Goodenough et al (1987). These are 'Analyst Advisor' and 'Map Image Congruency Evaluation', based on the expert system shell RESHELL, and written in Prolog. The former system is designed to advise users of remotely sensed data on appropriate image processing techniques, whilst the latter approaches the techniques of photointerpretation. The paper describes the representation of knowledge as a series of rules and frames and discusses methods of interfacing Prolog with the Fortran image processing routines in use at the centre. The paper concludes that: 'remote sensing is a

valid application area for the technology of knowledge-based systems'.

For a fuller description of the application of Artificial Intelligence techniques in the field of remote sensing and land-use planning, the reader is directed to the work by Pooley (1988), which was carried out contemporaneously with this work.

3.8 CONTEXTUAL RECLASSIFICATION AND SPECTRAL MIXING

Contextual reclassification is defined as the process whereby a pixel's class is modified according to its neighbours. The simplest form of contextual reclassification is the use of a mode filter (which replaces the centre pixel of a template area with the most frequently occurring pixel label within this template) on a classified image (Rothery, 1982, Townshend, 1986). In this case small areas tend to be replaced by the regionally dominant class, effectively removing the pixels which have been incorrectly classified due to noise and/or edge effects in the original image. This tendency to remove small areas results in some correctly classified pixels (those representing small areas on the ground) also being removed; a process akin to cartographic generalisation.

A similar, but more specifically targeted, contextual reclassification algorithm is that of small area replacement (Letts, 1979). In this method, contiguous areas containing less than a specified threshold number of

pixels are 'declassified', and subsequently reclassified according to their neighbouring pixels (those still classified). This reclassification can be on the basis of 'nearest classified neighbour', or 'modal class of n nearest classified neighbours'.

A variation on this technique is to try to identify noise and boundary pixels on the original imagery. This is achieved by first passing a suitable edge detector over the image (eg Roberts, Sobell or Laplacian high pass filters), thresholding this to separate the image into edges/non-edges, declassifying the pixels identified as edges, then applying one of the above reclassification algorithms to the declassified image. The thresholding may be done by trial and error, or some automatic solution can be applied (see sections 5.2.2 to 5.2.2.2.4).

Another technique is that of Relaxation Labeling, where the image is reclassified according to the probabilities of a pixel and its neighbours belonging to each class (DiZenzo et al, 1987a, 1987b, Mohn et al, 1987, Smith et al, 1981, Kittler and Illingworth, 1985). This probabilistic relaxation is an extension of the maximum likelihood classification algorithm to cover a neighbourhood, rather than single pixels. Because the algorithm is extremely slow to implement for large neighbourhoods, attention has focused on iterative implementation over small neighbourhoods, best results being achieved after several iterations.

To try to overcome the time problems caused by probabilistic relaxation, non-probabilistic methods have been investigated, for example modifying the minimum distance classifier in a similar manner to the maximum likelihood algorithm above, but classifying according to distances to means, rather than probabilities. Another alternative is to use ranked classes for each pixel, since rank order of classes can be calculated rather faster than the actual probabilities using an equation derived from the maximum likelihood probability estimate (Mather, 1985).

The image can be contextually reclassified using Markov chain theory to model the spatial autocorrelation of the class map (Kittler and Foglein, 1984). This is possible using the image itself, however, research suggests that it is more effective to estimate the transition probability matrix from field data, rather than the 'noisy' classified image (see sections 5.2.1.4.3.1.2, 8.5.2.1, 8.5.2.3 and 10.3.4). This method also has the added advantage of estimating the distance beyond which a pixel has no influence on its neighbours (the order of the chain). This distance could be used to set the size of mode filter window to be used, since the process of Markov relaxation is rather slow.

The autocorrelation distance could also be estimated by constructing a semivariogram for each class (see fig.3.1).
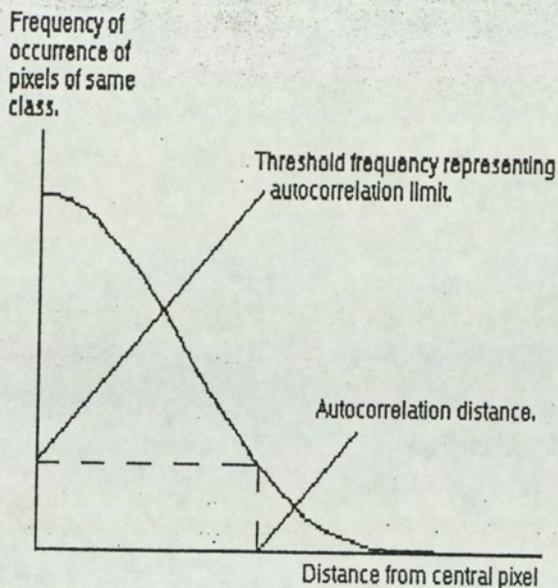
Fig.3.1 Semivariogram estimation of class autocorrelation distance.

Notes:

1. Because distances can only be measured in whole pixels, the semivariogram becomes a bar chart.

2. Beyond the autocorrelation distance, the frequency of occurrence of similarly classified pixels to the central pixel will not be zero, but at a level approximating the overall frequency of occurrence for the class over the whole image.

The principles behind the use of transition probabilities for contextual classification of clouds are described by Kittler and Pairman (1985), who applied these to AVHRR data. The paper states that:

'The Bayes classifier (maximum likelihood) exhibited a problem common to all simple pixel classifiers, that of yielding noisy labeling. The effect of using...the

transition probability classifier has been to clean up noisy boundaries between classes'.

It should be noted, however, that a significant difference exists between this technique and the Markov relaxation technique (section 5.2.4.3), since Kittler and Pairman's algorithm is applied before classification, rather than after, as is the case with Markov relaxation.

A contextual reclassifier, bearing some resemblance to textural classifiers, has been suggested by Wharton (1982). The so-called CONAN algorithm works by first creating a new 'frequency of occurrence' image, the bands of which have values according to each class' frequency of occurrence over an n by n window. The classifier is then 'retrained', using the same areas, on this new image, and the classification rerun on this. The result is an extremely efficient and relatively fast contextual reclassification.

Another contextual reclassification algorithm, and the most important in terms of integration with geographical information systems, is the per-field classifier. This has long been considered desirable (Allan, 1986), but the non-trivial task of extracting regions from remotely sensed imagery has precluded its widespread use. The development of digital mapping and geographical information systems now enables field boundaries to be extracted from maps, and classification to take place on a field by field basis. This classification can either be on a modal basis

(the field's modal class is returned to every pixel within the field), or on a statistical basis (the overall most probable class is assigned to every pixel in the field).

The increases in the amount of data to be processed in recent years led Chiou (1984) to examine the use of simple descriptors for classification of hyperdimensional remotely sensed spectral data. The results of this reveal contextual properties of vegetation, soils and crops. The simple descriptor is also proposed as a method of displaying multidimensional data in pseudocolour using RGB colour space.

Thomas (1980) derives a distance-decaying contextual influence model using analogy with gravitational attraction. This 'proximity function' was found useful in removing noise from classified data.

Rhode (1978) discusses the use of image stratification to allow two differing classes, with similar spectral characteristics, but occurring in different contexts, to be correctly classified.

Merchant (1984) proposes a strategy for image classification which 'seeks to emulate important aspects of visual image interpretation' (notably context). The paper also proposes the use of spatial post-processing for map generalisation.

Gurney and Townshend (1983) introduce the use of contextual information into classification. They

categorise contextual procedures into those applied to raw
or classified data, those applied to individual pixels or
'objects' (groups of similarly classified pixels), and
those working on the form of spatial relationships between
pixels. These methods are compared with existing 'per-
point' classifiers.

Some of the effects of mixed pixels on classification are
discussed by Birnie (1986), who cites the evidence of a
comparison of ground radiometry with Landsat MSS data to
suggest that a mixing of spectral classes within a pixel
is responsible for alteration of the pixel's expected
spectral response.

An attempt to isolate the effects of mixed pixels from the
classification of spectrally pure pixels is presented by
Irons et al (1985) in a paper discussing potential
differences between the Landsat MSS and TM sensors. A 21%
decrease in classification accuracy was reported when
mixed pixels were included in the classification. The
paper concludes with the following points:

i) there is a need to develop new approaches to
classification, which are able to take advantage of the
high spatial resolution of Landsat TM.

ii) the effect of spatial resolution on accuracy depends
on the classification algorithm used: the maximum
likelihood classifier being cited as unable to exploit the
benefit of TM's 30m resolution over the 80m resolution of
MSS data.

iii) increased resolution 'clarifies shapes, sharpens boundaries and alters the textural appearance of classes', thus facilitating manual photointerpretation.

The paper also refers to the use of texture analysis for image classification (see section 3.10).

Another paper concerned with the effects of spectral mixing and ground cover on classification is that by Batista, Hixon and Bauer (1985), in which the most important variable affecting classification of Landsat MSS data was found to be field size: a broad relationship between field size and accuracy (larger fields, higher accuracy), being found to exist.

The use of probabilistic relaxation (see section 5.2.4.1) to improve the output from maximum likelihood classifiers is described by Harris (1985). The iterative technique used was found to initially boost classification accuracy, but with increasing numbers of iterations the classification accuracy dropped. The paper also notes the computational expense involved in the use of iterative techniques.

Pre-classification filtering is described by Atkinson, Cushine and Townshend (1985) who report increases in classification accuracy with Landsat TM data when a mean or median filter is passed over the imagery prior to classification. The paper also notes that the filter size, relative to the spatial extent of classes is important.

62

The use of contextual measures, derived from classified imagery is discussed by Murphy (1985) who used a 3 by 3 operator to generate two measures:

i) The number of classes within the patch
ii) The frequency of occurrence of classes within the patch

These measures were then used as extra bands for further multispectral processing.

3.9 LAND COVER MAPPING IN UPLAND AREAS

The use of satellite data for land cover mapping is well established. Botkin et al. (1984) describe the techniques applied to vegetation monitoring, providing a useful introduction to the field for non-specialists in remote sensing. The agricultural applications of remote sensing are discussed thoroughly by Taylor (1985), whilst Forster (1985) covers the problems peculiar to the application of remote sensing to the monitoring of urban areas.

Morton (1986) and Weaver (1987) detail the use of remotely sensed data for mapping upland vegetation in the United Kingdom, the former using Landsat Multi-Spectral Scanner data, the latter Daedalus Scanner (Airborne Thematic Mapper) data. Both authors provide valuable insight into the problems of using such data for this purpose.

The theoretical advantages of Landsat Thematic Mapper data over Multi-Spectral Scanner data are discussed thoroughly

by Toll (1985), who also provides a good description of the Thematic Mapper sensor.

Forest mapping is dealt with by Malila (1980) who suggests the application of change vector analysis to this problem.

The problems of vegetation surfaces being continua of changing relative occurrences of vegetation type are described by Wood and Foody (1989), who suggest mapping of vegetation as a series of probability surfaces for each vegetation class, rather than as discrete blocks of single classes.

3.10 TEXTURE

Golton (1987) defines texture analysis as the 'study of the relative amplitudes of a number of adjacent pixels in order to obtain information on the type of surface on view'. Two uses of texture analysis are given: classification and edge definition. The paper also mentions human vision modeling, and notes that most texture analysis has concentrated on single-band, rather than multispectral data. The following texture measures are described:

Edge density measure - a measure of average grey-scale gradient over a small image patch.

Structural measures - describing detailed shapes of small regions: only really useful for two-tone images.

Difference measures - can be calculated from a histogram

of the magnitude of differences between pixels, separated by a chosen distance from each other for an image region.

Maximum/minimum measures - directional measures, based on the number of local extrema present in an image patch.

Run-length measures - the number of linearly consecutive pixels having the same digital value, averaged over an image region, measured in different directions.

Co-occurrence measures - an estimate of the probability of one grey level occurring at a distance of d from another grey level, calculated from a matrix of co-occurrences of grey levels (this is similar to Markov relaxation, described in section 5.2.4.3). Golton (1987) states that 14 different texture measures have been derived from co-occurrence matrices, but that some textures which can be readily segmented by eye cannot be differentiated using these measures.

Neighbouring grey level measures - calculated from a tally matrix for the number of times each pixel in a given area has n pixels of the same value at a given distance. This is a directionally independent form of run-length encoding. Several texture measures have been derived from these matrices.

Frequency spectrum measures - using two dimensional power spectra over small areas (Golton (1987) uses 8 by 8 and 32 by 32 pixel areas), usually Fourier transforms are used, but Golton (1987) states that 'Walsh and Slant transforms

65

have been equally useful in land use classification'.

Rank correlation measure - based on pattern matching tests. Pixel values for each image area are sorted by digital number into rank order, to form a vector which is then compared to reference pattern rank vectors for class assignment.

The paper by Haralick, Shanmugan and Dinstein (1973) describes the use of texture for image classification. They describe human vision as being made up of spectral, textural and contextual features, but that 'when a small-area patch has a wide variation of features of discrete gray tone, the dominant property of that area is texture'. In order to quantify texture the spatial grey level dependency matrix is suggested, calculated for a 'resolution cell' (ie a patch over which texture is to be measured). This matrix can then be used to calculate 14 different texture measures. When used for classification of Landsat MSS imagery an accuracy of 83.5% was reported, compared to 74-77% accuracy when using tonal features alone. The paper notes that there is a trade-off between resolution cell size and processing time, and that cell size should be determined by autocorrelation of data (see sections 3.8 and 5.2.3.1)

A simple example of the use of textural information is provided by Gong and Howarth (1990) who describe the use of a Laplacian filter derived image as an extra band for multispectral classification.

Wang and He (1990) describe a statistical method of texture analysis, in which images are broken down into a series of 'texture units' which are used to describe the image by means of its 'texture spectrum', which can then be used for classification or filtering.

Thomas (1981) describes the use of subtractive box filtering, after histogram equalisation, to enhance image texture for subsequent visual interpretation. The paper describes the main role of textural information as an aid to visual image interpretation.

The use of Fourier transforms to classify periodic textures is described by Excell et al (1989). Spatial frequency distributions were used to distinguish between dust clouds and desert, but were found to be poor at distinguishing between clouds and fields.

Majumdar and Bhattacharya (1988) describe the application of the Haar transform for extraction of linear and anomalous patterns within images. The paper states that the transform performs well at this task and is computationally inexpensive.

Gordon and Phillipson (1986) describe a texture enhancement technique to help in the separation of orchard and forest areas using satellite data. The technique involved:

i) passing a 3 by 3 pixel absolute difference filter (ie sum of differences, ignoring sign, between central pixel and its eight neighbours) over Landsat Thematic Mapper bands 3 and 4.

ii) ratioing of the output images from i) to reduce the effects of edges.

iii) separating the ratioed image by density slicing into orchard and non-orchard areas.

The binary image produced by iii) was then used as a selective mask to adjust prior probabilities used by a supervised maximum likelihood classifier, increasing the classifier's accuracy.

The use of edge-density measure for land use mapping from simulated Landsat TM data is discussed by Hlavka (1987), who used a 31 by 31 pixel moving window to calculate these measures over the imagery. The inclusion of texture measures in automated classification procedures was found to produce a worthwhile improvement in classification accuracy. Hlavka (1987) calculated the edge densities using the following steps:

i)edge enhancement
ii)thresholding to produce a binary image
iii)convolution filtering to produce local edge density measures.

Use of the Fourier transform for texture analysis is discussed by Stromberg and Farr (1986). The technique was

68

found to be successful when applied to both synthesised imagery and synthetic aperture radar imagery used for geological interpretation.

Textural processing of remotely sensed data will not be further covered by this thesis. The interested reader is directed to the work of Oldfield (1987), which was carried out in parallel to the work presented here.

3.11 GEOGRAPHICAL INFORMATION SYSTEMS

A pilot geographical information system, linking non-satellite and remotely sensed data, developed by the MaCaulay institute of soil research, the National Remote Sensing Centre and Grampian Regional Council is described by Young (1986). The paper reports that the study was a success, but that the user interface for geographical information systems must be improved before they can become an operational tool.

A review of developments in geographical information at the Natural Environment Research Council's Thematic Information Service (NUTIS) is provided by Jackson and Mason (1986). Stress is placed on data management and efficient data query. The paper describes knowledge based image segmentation (see chapter 7, section 7.3), and mentions the use of an expert system for map and graphic output.

In discussing Intelligent Knowledge Based Systems (IKBS) and Integrated Geographic Information Systems (IGIS)

Jackson and Mason state:

'These developments in turn should open up new opportunities in the field of remote sensing by improving the accuracy of image classification through the use of IGIS supported contextual and IKBS based classification procedures.'

Singh and Dwivedi (1986) describe the combination of information from several different data sources: Landsat MSS interpretations, lithological, topographical and 'other collateral data' to produce a 1:250,000 scale soil map for an area of northern India. they report an accuracy figure of 93.3% for soil landscape boundary definition.

The use of ancillary data - a digital terrain model - to improve classification accuracy is described by Jones, Settle and Wyatt (1988). Slope and aspect information from the model were used to stratify the image prior to classification, however the paper does not comment on the effects of this stratification on classification accuracy.

A 'Relational Image-Based GIS' is described by Qiming Zhou (1989). This comprises two data sets: a spatial, image-based database and an aspatial data base. The paper discusses the problems of integrating these data sets.

CHAPTER 4

CLASSIFICATION

4.1 INTRODUCTION

Classification is the process whereby digital images are split into areas of different classes on the basis of the characteristics of their component pixels. Commonly, classification techniques are used to 'recognise' the spectral characteristics (spectral signatures) of these classes, and assign each pixel to the class which its own spectral characteristics most closely resemble.

The techniques can be divided into supervised and unsupervised. For supervised techniques, the operator is required to define 'training areas' (areas of known class), which are used to provide the necessary statistics for classification. In this case the significance of each class is known (ie what it represents on the ground), so the resultant classification can be used immediately as a thematic map.

Unsupervised classification attempts to use the structure of the data in order to divide the image into classes. Most techniques are a form of spectral clustering. These methods have the advantage of enabling classification of an image without the need for time consuming training however, after classification each class must be correlated with a class on the ground. Unsupervised methods have the theoretical advantage that they are

71

capable of extracting all distinguishable classes from the imagery. They do, however, have practical problems, not least being their large memory requirements and their slow operation in the case of iterative procedures.

4.2 TRAINING FOR SUPERVISED CLASSIFICATION

Supervised classifiers require statistics, generated from example data, to classify imagery. These are usually calculated from training data (areas of known class defined on the image).

Such areas should ideally contain a large number of pixels (greater than 50, Hay, 1979), and be representative of the class' spectral response throughout the image. For this reason, it may be desirable to define more than one training area per class, in different parts of the image. Alternatively, widely separated pixels could be chosen at random, and their ground classes determined by field visits. This method, although statistically valid, cannot be recommended due to time constraints; for example, 50 pixels (the bare minimum) for 10 classes requires at least 500 site inspections, not accounting for the problems associated with finding sufficient pixels to cover uncommon classes (see section 3.4 for other references to training data).

4.3 CLASSIFICATION ALGORITHMS

The algorithms described here are all 'per-pixel' classifiers, relying on the spectral characteristics of

72

individual pixels in isolation to identify the correct class. Other classifiers exist, those listed being the ones used for this study.

4.3.1 BOX CLASSIFIER

The box, or parallelepiped, classifier is the simplest classification algorithm commonly applied to multispectral digital image data. The algorithm divides feature space (the hyperdimensional space defined when each band ('feature') of an image is used as an orthogonal graph axis for plotting the positions of each pixel) into a series of 'boxes' whose upper and lower limits are set using training statistics for each spectral band. The classifier is thus a series of IF...THEN rules (sometimes termed 'production rules' in the artificial intelligence literature), for example:

IF band 1 value is greater than a and less than b
AND band 2 value is greater than c and less than d
THEN the pixel is in class x

Band 1

This area classified
as class A

a   b   Band 2
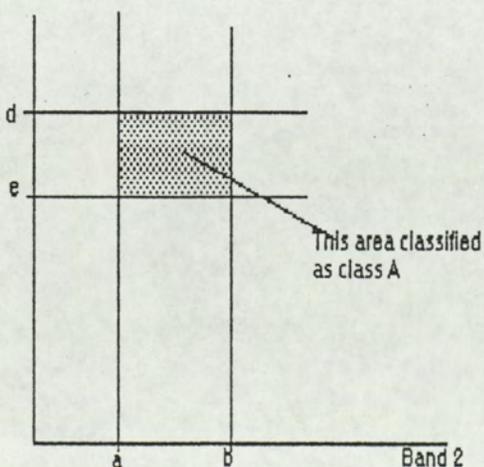
Fig.4.1: to illustrate the box classifier

73

In practice the limits are often set in terms of standard deviations about the training area mean values in order to exclude outlying data points.

Although the algorithm is fast, it suffers from two major drawbacks:

1.Class clusters are rarely box-shaped within feature space.

2.Overlap can occur between adjacent class boxes leading to ambiguous class values.

The complexity of the box shape can be increased, attempting to avoid these problems. Essentially, the algorithm splits the class boxes into smaller boxes, as in Fig.4.2, attempting to avoid overlap by more closely modeling the shape of the clusters.
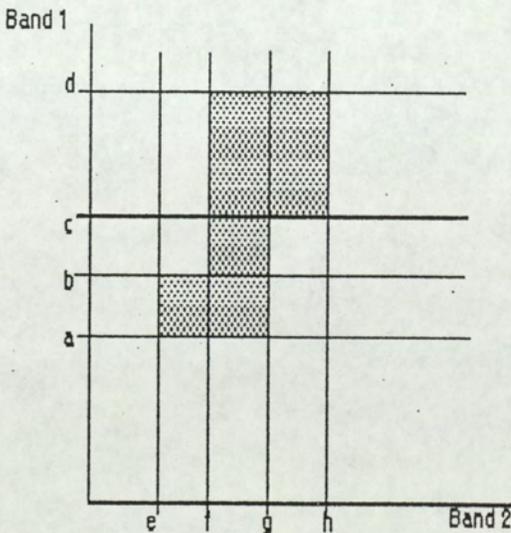
Fig.4.2: to illustrate an improved box classifier
In this case the limits for class x are best summarised in a table:

Table 4.1: class limits for the modified box classifier

Class limits

Band 1   Band 2

Box	Band 1 Lower	Band 1 Upper	Band 2 Lower	Band 2 Upper
1	a	b	e	g
2	b	d	f	g
3	c	d	g	h

An important aspect of this type of classifier is that it can be implemented in real time for three bands using hardware look up tables.

4.3.2 MINIMUM DISTANCE CLASSIFIER

This algorithm assigns pixels to the class to which they are closest in feature space. Each class is assigned a mean vector from training data, consisting of the mean digital values of the training data in each band. The distance of any pixel from this mean vector is commonly calculated in one of two ways - either using Pythagoras' theorem, to give the Euclidean distance, or using the 'round the block' method. Fig.4.3 shows these for the two dimensional case.
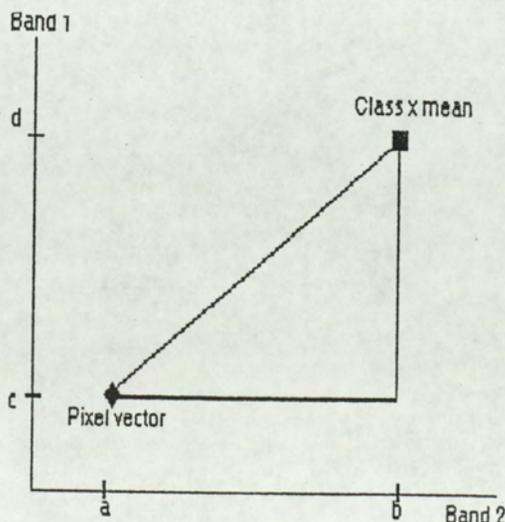
Fig.4.3: to illustrate the minimum distance classifier

The distances from the pixel to class x's mean vector are calculated by:

Round the block: $(|b-a|+|d-c|)$
Euclidean: $((b-a)^2 + (d-c)^2)^{0.5}$

(where $|\ |$ indicates that the sign is ignored)

The pixel to mean vector distance is calculated for each possible class, then the pixel is assigned to the class for which the shortest distance has been calculated.

A problem with this algorithm is its inability to take class probability densities into account. The maximum likelihood algorithm (see section 4.3.5) uses the class probability density function to model changes in probability in feature space, however, this is time consuming. A simpler approximation can be achieved by dividing distances through by the associated class standard deviations, hence the title: 'deviant distance' algorithm. The previous equations thus become:

$$\frac{(|b-a|)}{s_1} + \frac{(|d-c|)}{s_2}$$

and

$$\{\frac{(b-a)^2}{s_1} + \frac{(d-c)^2}{s_2}\}^{0.5}$$

where s_1 and s_2 are the standard deviations of classes 1 and 2 respectively, obtained from training data.

For reduction of computational expense, the square root need not be taken, since this will not affect the rank order of the calculated distances.

This results in distances rescaled for each individual class, roughly approximating the probability density functions for the classes. In other words, distances for class n are rescaled according to the variation of class n in each spectral band, whilst distances for class m are rescaled according to the variation of class m in each spectral band.

4.3.3 DECISION TREE CLASSIFIER

The tendency towards higher spatial and spectral resolution of remotely sensed data has led to the investigation of classification algorithms based on the expert systems approach. Such algorithms rely on the idea that only limited band combinations are necessary to discriminate between one class and its 'background'. The decision tree classifier is an attempt to produce an accurate classification whilst minimising the amount of data which must be examined to classify each pixel.

The first stage is to produce co-incident spectral plots for all classes in all bands from training data. These plots are then used to identify which bands may be used to discriminate between which classes.
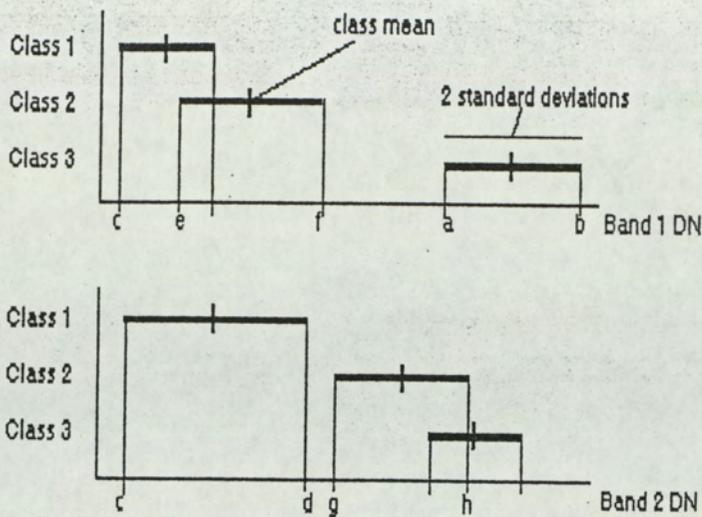
Fig.4.4: co-incident spectral plots for the decision tree classifier

Thus, if a pixel's value lies between a and b on band 1, it can be assigned to class 1, whereas if its value lies between c and d in band 2, it can be assigned to class 3. With greater complexity, if the pixel's value lies between c and f in band 1, then it can either belong to class 2 or 3, but not to class 1. If it lies between g and h in band 2, then it can either belong to class 1 or 2, but not to class 3. The full tree then becomes:

IF value is greater than a and less than b in band 1
THEN the pixel is assigned to class 1
otherwise
IF value is greater than c and less than d in band 2
THEN the pixel is assigned to class 3
otherwise
IF value is greater than e and less than f in band 1
AND value is greater than g and less than h in band 2
THEN the pixel is assigned to class 2
otherwise
the pixel remains unclassified

4.3.3.1 AUTOMATION OF TREE DEFINITION

The most important barrier to the widespread use of the decision tree classifier is the time and effort which must be expended in deriving the tree from class statistics and coincident spectral plots. Moreover, different operators are likely to produce different trees, depending on their own preferences and experience (eg minimising the number of bands to be used, maximising the theoretical accuracy and so on). Different trees will give different classifications, therefore the decision tree algorithm will suffer from inconsistency.

Inconsistency has been highlighted by Fiegenbaum (1988) as a key problem which artificial intelligence can attack and remove. Artificial intelligence tools, in the shape of expert systems, are already available to address the problems of decision tree design.

The inferencing capabilities of some expert systems are ideal for designing decision trees. Data can be fed into these systems, which then infer the decision rules. In the case of decision tree definition, the data comes from the training areas, and the inferred rules are the 'nodes' of the tree.

It should be noted at this point that not all expert systems are suitable for the purpose of designing a decision tree. The following criteria must be borne in mind:

1.The expert system should have a suitable 'inference engine' for the purpose. It must be able to use integer variables (ie the digital values in each band), and to infer trends in the data, rather than simply building up a look up table from the training data.

2.The expert system must be able to handle 'clashes' caused by noisy data, whilst still inferring useful rules. (There is considerable potential for duplication of spectral signatures between pixels in training areas of different classes.)

3.The expert system must be able to handle sufficiently large amounts of data. In this case, a typical requirement would be: 10 classes, 6 bands (variables), and 200 pixels (cases) per training area: 12,000 8-bit integers.

4.3.4 LOOK UP TABLE AND SPECTRAL SHAPE CLASSIFIERS

A common response to classification problems in the field of artificial intelligence is to store, for each possible combination of variables, the appropriate class to which to assign an object.

For small, noise-free data sets this is an attractive proposition, since each case (combination of values) can be examined and a suitable classification decision made. The next time this combination is encountered, the classifier only has to look up the previous decision. Such a classifier is thus capable of 'learning' to identify classes.

Applying this to remote sensing reveals two major problems: 'the curse of dimensionality' (Duda and Hart, 1973) and problems of noise inherent in the large, low-precision data sets generated by remote sensing devices.

The first problem partially solves itself, since remotely sensed images, despite potentially containing many thousands of different pixel vectors (for Thematic Mapper: 280 million million possible band/value combinations, excluding the thermal band), generally contain only a few thousand different vectors (Mather, 1985). The actual storage requirements for the look up table can thus be reduced to manageable proportions using techniques such as hash coding.

It should be noted, however, that the increases in data volumes for remotely sensed data in recent years have reduced the usefulness of such techniques.

The problem of incorrect pixel values (described in section 2.7.1) is not so easily addressed. Imagery can be smoothed prior to classification in an attempt to reduce this, however, genuine high frequency variation may also be suppressed.

4.3.4.1 SPECTRAL SHAPE CLASSIFIER

The spectral shape classifier (Pendock, 1988) is, in fact, a data compression technique, enabling multi-band data

sets to be quickly classified by look up table or density
slicing.

The algorithm assigns a unique number to each spectral
response shape, producing just one band of data from

Digital value
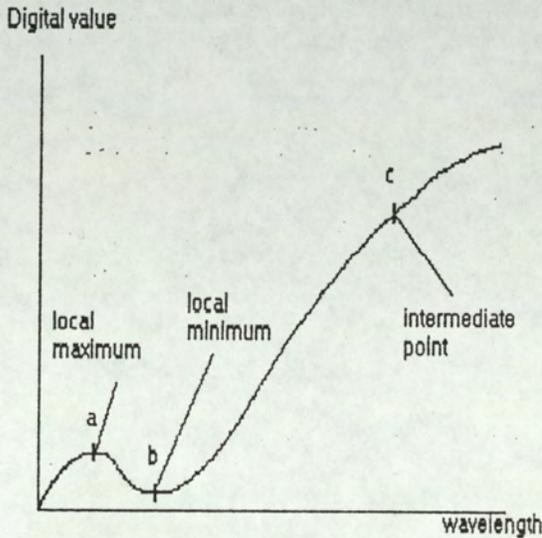
Fig.4.5: to illustrate the spectral shape algorithm

For any point on the spectral response curve shown in
Fig.4.5, the point can be described as a local maximum,
local minimum or an intermediate point. Thus, A is a local
maximum, B a local minimum and C an intermediate
point. This can be approximated in the discrete band case
as:

1.Local Maximum: if band value is greater than the values
of its two adjacent spectral bands.

2.Local Minimum: if band value is less than the values of
its two adjacent spectral bands.

3.Intermediate point: all bands not fulfilling 1 or 2
above.

If the values 0,1 or 2 are used to represent each state, a 'ternary' (base 3) number can be generated for each pixel's spectral shape. In the diagram above, this might be:

A=0
B=1
C=2

giving ternary 12, which can be converted to decimal as 5.

4.3.4.2 OVERALL REFLECTANCE

The problem of storing a large look up table can be reduced to within manageable proportions by compressing the data into fewer bands. The Spectral Shape algorithm (section 4.3.4.1) attempts this, as would principal components analysis followed by discarding less significant components (see section 4.5). Another method of doing this is to produce an 'overall brightness' image by summing the values of a pixel in each band to produce just one band. This single band is then ideal for look up table classification, which, being a simple density slice, can be implemented in hardware.

4.3.5 MAXIMUM LIKELIHOOD CLASSIFIER

None of the previously described classification algorithms is capable of taking the probability density function of a class in feature space into account.

Assuming that a class' pixels are normally distributed in
feature space, the maximum likelihood classifier estimates
the probability density function for each class from
training data. Pixels are then assigned to the class to
which they are most likely to belong. This can improve
upon the previously described classification algorithms,
as described below:

Frequency

class mean values (assuming normal
distribution)

a

b

d c

digital value.
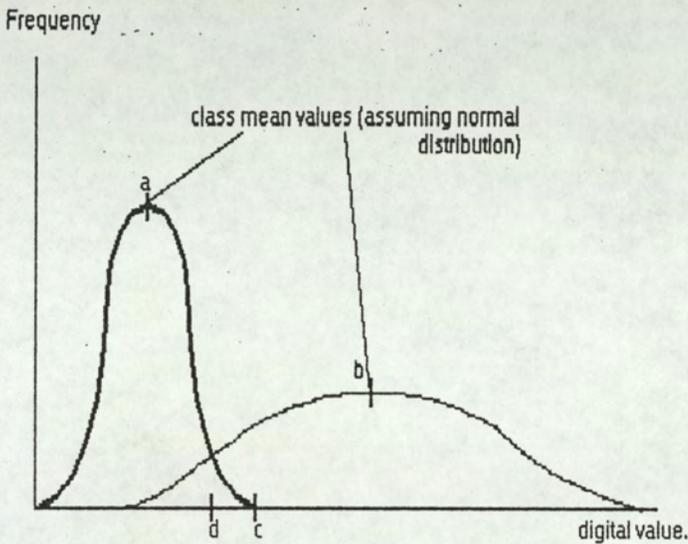
Fig.4.6: to illustrate the maximum likelihood classifier

In Fig.4.6, A and B are the class means for class 1 and 2
respectively. A minimum distance classifier will place the
decision boundary equidistant from the means at C. As can
be seen, this will incorrectly assign class 2 pixels to
class 1.

The maximum likelihood algorithm, on the other hand,
calculates its decision boundary to be at D, using the

probability density functions of the two classes. This
position of the boundary gives the classification with the
minimum error.

The algorithm is unfortunately susceptible to data noise and, more importantly, distributions other than normal. For example, skewed and bimodal distributions occur frequently in remotely sensed data. This is, however, a problem common to all parametric classifiers which have to make assumptions about the data to which they are applied.

4.3.5.1 IMPLEMENTATION

The probability density functions for each class can be estimated from the class variance-covariance matrix, which can itself be estimated from training data.

These form the basis of the discriminant functions, which are used to assign pixels to classes. These functions take the form:

$$gi(x)=P(x/wi)P(wi)$$

where:

$gi(x)$=the probability that pixel x belongs to class i.
$P(x/wi)$=the probability of x belonging to class i, calculated using the probability density function.
$P(wi)$=the a-priori probability of any pixel belonging to class i.

$P(x/wi)$ can be estimated by:

$$\hat{p}(x/wi)=(1/[\{2pi\}^{0.5}]\hat{s}i)exp(-0.5\{[x-\hat{u}i]^2\}/[\hat{s}i]^2)$$

where:

$\hat{s}i$=estimated variance of class i

85

ûi=estimated mean of class i

p̂(x/wi)=estimated probability of pixel x belonging to class i

expanding this into n dimensions gives:

$$\hat{p}(x/wi)=(1/[\{2pi\}^{\{n/2\}}]|E|i^{0.5})$$

$$exp(-0.5\{trans[x-ui]*[E]^{-1}*[x-ui]\})$$

where:

E=variance-covariance matrix

x=pixel vector

ui=mean vector for class i

trans(n)=transpose of vector n

Thus, to calculate class probabilities, the variance-covariance matrix is first estimated, inverted and its determinant calculated. The only part of the equation which must be repetitively calculated for each pixel is:

$$exp(-0.5\{trans[x-ui]*[E]^{-1}*[x-ui]\})$$

the rest of the equation being constant for any given variance-covariance matrix.

The calculations can be further reduced, since all that is actually required to assign pixels to classes is a set of values in the same rank order as the probabilities. Such techniques are detailed by Mather (1985, 1987).

4.4 CLUSTERING

Clustering is generally used in remote sensing as an unsupervised classification technique. Most 'unsupervised' classifiers require several parameters to be supplied by the operator, such as: number of clusters (classes); threshold distances and probabilities, before classification.

In order to make a reasonable guess at the values of these parameters, the operator must have a detailed knowledge of the scene to be classified. For this reason, the term 'semi-supervised' might be more appropriate.

At best, an unsupervised classifier will extract all 'spectral' classes (those classes which can be distinguished by their image spectral response), which must be matched with their corresponding ground classes. In practice, the time required to produce such an optimum classification often precludes the use of clustering techniques in remote sensing.

4.4.1 TWO PASS CLUSTERING

A fast method of clustering, involving only two passes through the data, is implemented by generating an n-dimensional histogram on the first pass through the data (n is the number of spectral bands in the image). Cluster centres are then identified as local maxima within the histogram. On the second pass each image pixel is assigned

87

to a cluster according to a proximity function (usually nearest neighbour).

As data dimensionality increases the storage capacity required by the program increases, until coarsening the radiometric resolution (for example reducing 8 bits of precision (256 grey levels) to 7 bits (127 grey levels) by dividing all values by 2) or hash coding must be considered to enable implementation of the algorithm.

4.4.2 ITERATIVE CLUSTERING

A less memory intensive clustering algorithm relies on the assumption that, given an appropriate number of clusters, repetitive classification, recalculating the cluster centres after each pass through the data, will eventually result in cluster membership converging to stationarity. This is the essence of Forgy and Jancey's algorithm, detailed in Anderberg (1973), and the commonly used ISODATA program for unsupervised classification of remotely sensed data.

In practice, this involves making a first guess at cluster centres, or taking random pixels from the image as 'seeds', classifying the image (by nearest neighbour rule), recalculating the cluster mean vectors using the new cluster membership and reclassification of the image using these new class vectors. Convergence can be checked for by comparing subsequent classifications. In an ideal case this would occur when no pixels change class between

iterations, however in practice, a threshold must be set to allow for noise within the data.

4.5 DIMENSION REDUCTION TECHNIQUES

Duda and Hart (1973) describe 'the curse of dimensionality', whereby increases in the amount (number of bands) of data used in a classification result in disproportionately large increases in the computational expense of the classification.

The paper by Fusco and Trevese (1985) discusses the correlation between adjacent spectral bands in the context of using data from such bands to reconstruct missing data within satellite images. If, as suggested in this paper, adjacent spectral bands are correlated then by careful band selection an image containing a reduced number of spectral bands can be used for classification without too great a drop in accuracy.

Computational expense can, therefore, be reduced by lowering the number of bands to be processed. Techniques for such data reduction fall into two categories: discarding the least useful bands from the data set, or processing the data to produce a smaller data set containing as much of the original information as possible.

An examination of stepwise discriminant functions for band selection is provided by Labovitz (1985). The levels of separability achieved using such discriminant functions

was found to be sub-optimal. The paper also examines the effect of spacing in training samples, finding that autocorrelation of training samples often leads to overestimation of accuracy.

4.5.1 A BRIEF REVIEW OF DIMENSION REDUCTION TECHNIQUES

Canas and Barnett (1985) report on the use of principal components analysis to retain as much information in three new bands generated from all four bands of landsat MSS data as possible, whilst reducing subsequent processing times by approximately one quarter. Figures quoted indicate that 98% of the total variance of the original four data bands was retained in the first three principal components (although variance and information content cannot always be equated). They describe principal components analysis as: 'a systematic means of reducing the dimensionality of multichannel image data', and recommend the application of this technique to other satellite image data in addition to the landsat MSS data used.

Singh and Harrison (1985) also applied principal components analysis to Landsat MSS data, but they recommend that the components are standardised , since this was found to increase signal-to-noise ratio and image enhancement. They noted, however, that: 'whether standardisation is desirable is, in the ultimate analysis, to be decided on non-statistical grounds'.

The application of principal components analysis to multitemporal classification is described by Jiaju (1988), who cites the removal of calibration requirements for multitemporal data as an advantage of the technique which enables more effective classification of multitemporal data.

The technique of discarding redundant information is covered in the review article by Thomas et al. (1987), which addresses the problem of: 'which channels contain the best information to separate the classes of interest to the user?'.

The paper attempts to solve this problem by examination of class statistical profiles, which are used to derive a number of multi-channel separability indices, giving some indication of the probable effectiveness of different band combinations for class separations (see section 4.3.3). The separability indices recommended in the paper are summarised in table 4.2.

Table 4.2 Separability indices used by Thomas et al (1987).

Separability Index	Optimum conditions for use
Divergence	classes tending towards homogeneity limited number of bands in use
Transformed Divergence	as for divergence, but more suitable where many bands are in use
Bhattacharyya Distance	less homogeneous classes, limited number of bands in use
Transformed Bhattacharyya Distance	as Bhattacharyya distance, but more suitable where many bands are in use

Michaelis (1988) describes the use of band covariance matrices to determine whether Landsat Thematic Mapper data contains redundant information. The paper also describes the use of principal components analysis to reduce the number of bands of data to be processed.

Schrier and Lavkulich (1979) used factor analysis to identify data components containing maximum information content, describing the use of this technique on a data set containing many varied data types (see section 3.7).

A separability measure termed 'ellipsoid volume' is described by Sheffield (1985) who applies this to the task of choosing the optimum three bands from a multi-band data set for image classification. The paper includes a Basic program to calculate the measure.

4.5.2 DIVERGENCE AND SEPARABILITY MEASURES

Thomas et al (1987) describe a separability index as a measure which attempts: 'to define the relative feasibility of (a) pixel being a member of class A whilst also being a member of class B, and vice-versa'.

Divergence is described as the integral of the likelihood ratio for a pixel belonging to one class as opposed to any other class. In other words, if a graph were to be drawn showing the difference between the likelihood of a pixel belonging to its most probable class (calculated by the maximum likelihood rule, see section 4.3.5) and the probability of it belonging to any other class for all

possible pixel values, and the area under the curve
calculated, the result would be the divergence of that
particular band from which the values were taken (see
fig.4.7).

Probability of pixel
belonging to class n.

The area under the curve
is the divergence for class
n in the band used.

Digital value of pixel.
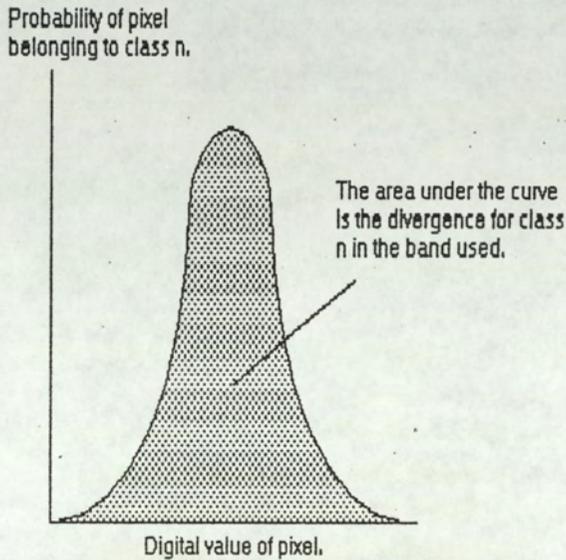
Fig.4.7 An interpretation of divergence.

Thomas et al (1987) suggest improving the divergence
calculation to enable easier identification of the point
at which sufficient separability of classes has been
achieved to give a satisfactory classification. This is
termed the transformed divergence, and is achieved by
introducing a saturating form to the expression for
divergence. For broader class probability distributions
the Bhattacharyya distance is recommended as an index of
class separability (Thomas et al (1987) eqn.32), and again
a saturating form is introduced into the equation to give
a transformed Bhattacharyya distance.

A comparison of four separability measures is given by

Mausel, Kramer and Lee (1990): transformed divergence, Jeffreys-Matusita distance, Bhattacharya distance and divergence. They tested the band combinations suggested by these indices against all possible four channel combinations of their data set, finding transformed divergence and Jeffreys-Matusita distance to be the best indicators of class separability, and therefore accuracy.

4.5.3 PRACTICAL APPLICATIONS

As discussed in section 4.3.3, some rationale for discarding redundant information, whilst still retaining reasonable class separability is desirable. This is especially true in situations where computational expense must be minimised.

Section 4.3.3 describes band selection by examination of co-incident spectral plots to assess class separability. The class separability indices discussed in section 4.5 may provide a more theoretically sound basis for band selection.

It must be remembered, however, that discarding data is carried out in many cases with the objective of reducing computational expense. If the method used to determine which bands of data to retain and which to discard is itself computationally expensive, then the user of the data might be better off retaining all bands, giving maximum separability, or referring to established 'traditions' within the remote sensing community to decide on which bands to use for a particular task.

4.5.3.1 CALCULATION OF DIVERGENCE

Thomas et al (1987) give an expression for divergence as follows:

$$D_{ij} = 1/2 \ \text{tr} \ (V_i - V_j)(V_i^{-1} - V_j^{-1})$$
$$+ 1/2 \ \text{tr} \ (V_i^{-1} + V_j^{-1})(U_i - U_j)(U_i - U_j)^T$$

where:

D_{ij} is the divergence of class i from class j across all channels.

V_i is the variance-covariance matrix for class i across all channels

V_j is the variance-covariance matrix for class j across all channels

U_i is the mean vector for class i across all channels

U_j is the mean vector for class j across all channels

tr indicates the trace of a matrix (sum of the leading diagonal elements).

From this equation, it can be seen that for any band combination, all that is required to calculate divergence for each class pair is the variance-covariance matrix for that pair and the mean vectors of the two classes.

4.5.3.2 EXAMPLE OF THE USE OF DIVERGENCE

Thomas et al (1987) describe two applications of divergence:

i) to identify class pairs which are poorly separated by any band combination - usually with a view to using

additional bands to improve classification accuracies.

ii) to recommend a reduced number of bands for classification, whilst maintaining a high level of accuracy, in order to reduce computer overheads.

The following example describes the calculation of divergence for two 3-band data sets, presents the results of these calculations, and discusses the practical implications of these results.

4.5.3.2.1 DATA SETS AND CALCULATION

Two of the data sets used elsewhere in this study were selected for divergence calculation. Unfortunately these consisted of only three bands of data each, and so the use of divergence to determine the theoretical optimum three-band combination for visual analysis from a multi-band data set cannot be discussed here. For full explanation of this technique the reader is directed to the paper by Thomas et al (1987), which describes such a study.

The data sets were: Landsat TM of an area of North Wales and Landsat MSS of an area of Yemen, both of which are described in section 8.1.1. The mean vectors for each class and the variance-covariance matrices were those calculated by the maximum likelihood classification algorithm described in section 4.3.5).

Divergences for each class pair were then calculated using the equation given in section 4.5.2.1. These are set out in table 4.3 below.

Table 4.3 inter-class divergences calculated for two 3-band data sets

North Wales

Class j

Class i

	1	2	3	4	5	6	7
2	5006.5890						
3	2949.2900	252.5379					
4	3032.3280	94.5350	105.2883				
5	19705.5900	1476.6310	2706.3410	2643.0790			
6	11682.5000	1091.0140	832.5231	1382.8270	1657.8410		
7	17838.8900	1087.9340	2803.4910	2279.1420	822.3004	3198.0220	
8	11861.2100	432.1631	896.2335	917.3776	216.0645	635.4537	793.8749

Yemen

Class j

class i

	1	2	3	4	5	6	7
2	50.0232						
3	61.2144	247.4882					
4	1024.1350	635.2363	1865.3920				
5	627.7104	373.5067	1241.3630	351.8843			
6	385.9738	266.9448	807.5260	804.3631	116.4684		
7	561.8875	518.7795	917.9494	1348.6990	340.9630	84.2851	
8	780.1611	900.3976	994.1033	2475.3750	960.7924	417.3784	139.7164

4.5.3.2.2 DISCUSSION

The results of maximum likelihood classifications of these data sets are summarised in the confusion matrices in table 4.4 below.

Table 4.4 Confusion matrices for maximum likelihood classifications.

North Wales data set

Classified as ------>

	U	1	2	3	4	5	6	7	8	
1	0.4038	1.	0	1.	0.	0.	0.	0.		4040.
2	0.	0.	359.	0.	5.	0.	0.	6.	0.	370.
3	0.	0.	1.	185.	1.	0.	20.	0.	2.	209.
4	0.	0.	57.	2.	108.	0.	0.	0.	3.	170.
5	0.	0.	34.	0.	0.	5.	0.	0.	15.	54.
6	0.	0.	0.	8.	0.	0.	89.	0.	0.	97.
7	0.	0.	25.	4.	0.	0.	0.	4.	2.	35.
8	0.	0.	64.	2.	7.	0.	0.	0.	447.	520.

0.4038. 541. 201. 122. 5. 109. 10. 469.

Normalised accuracy: 95.27%

Yemen data set

Classified as ------>

	U	1	2	3	4	5	6	7	8	
1	0.	11.	0.	531.	712.	1594.	818.	374.	0	.4040.
2	0.	0.	0.	31.	63.	36.	163.	77.	0	. 370.
3	0.	12.	2.	3.	116.	22.	26.	28.	0	. 209.
4	0.	24.	4.	67.	18.	57.	0.	0.	0	. 170.
5	0.	0.	0.	2.	4.	47.	0.	1.	0	. 54.
6	0.	0.	0.	0.	11.	3.	22.	61.	0	. 97.
7	0.	0.	0.	0.	1.	1.	10.	23.	0	. 35.
8	0.	9.	0.	6.	155.	25.	192.	133.	0	. 520.

0. 56. 6. 640.1080.1785.1231. 697. 0.

Normalised accuracy: 2.26%

For the North Wales data set the applications of divergence can be seen clearly: classes which can be discriminated easily using the three bands available have high divergence values (for example class 1 and classes 5,6,7 and 8) these correlate with extremely low occurrences of misclassification within these class pairs (shown in the contingency table); classes which may require the addition of further bands to aid discrimination have low divergence values (for example

98

classes 2 and 4 and classes 2 and 8), these values corresponding to low values in the contingency table.

For the Yemen data the picture is less clear. There is some correlation between low divergence values and inter class mis-classification, but this relationship is largely obscured by the poor classification accuracy.

4.5.3.2.3 CONCLUSIONS

The results presented above lead to the conclusion that, where a data set is suited to multispectral classification (as is the case with the North wales set) then divergence can provide a useful preview of the probable results of a maximum likelihood classification. Divergence can thus be used to aid in optimum band selection without the necessity of performing numerous classifications of the data.

For data sets which are unsuited to multispectral classification, such as the Yemen data set above, divergence can be of little use and alternative methods of classification should be examined.

4.6 EFFECTS OF TERRAIN ON REMOTELY SENSED DATA

Since the data used in this study consists of pixel brightness values from surfaces illuminated by the sun, the effects of sun-surface-sensor geometry and the interactions between surfaces and incoming radiation at different angles and wavelengths are important.

Although no attempts were made to compensate for the above interactions within the test data sets, the reader should be aware that variations in spectral response due to terrain represent a potential source of classification error. The following notes refer the interested reader to two papers on this subject.

Pinter et al (1987) examine the effects of terrain, illumination angle and sensor view angle on the reflectance characteristics of wheat, using a field radiometer. They conclude that these factors greatly influence the reflectance characteristics of this land-cover type. This variation in reflectance characteristics was found to still be present even after band ratioing the data.

Kay and Barnsley (1989) examine the effects of sensor view angle on multispectral classification accuracy. Although concerned with changes in sensor attitude, their results would also seem to be valid for terrain effects (since both sensor angle and terrain affect the sensor-surface geometry), although tall, vertical vegetation on a slope will undoubtedly have a different spectral response to similar vegetation when viewed from an angle. They conclude that: 'the effect (of differing sensor angle) makes the consistent classification of land cover types...extremely difficult'.

CHAPTER 5

CONTEXTUAL RECLASSIFICATION

5.1 INTRODUCTION

This chapter describes the various methods which have been
used in contextual enhancement of classified remotely
sensed data. A critical appraisal of these methods is
given and proposals are made for improved techniques.

There are many reasons why a classified image may be
imperfect or incorrect. Two of the most commonly occurring
reasons are spectral inseparability of ground classes and
the introduction of 'noise' into the image by any
combination of sensor characteristics, atmospheric effects
and data transmission errors (see section 2.7.1).

Spectral inseparability of ground classes must often be
accepted as an unfortunate fact. Even if the classes
are theoretically separable, given specific growth stages
and season, the likelihood of being able to obtain a
suitable image of the area at a suitable date is low.
Legg (1988), in a study of the suitability of Landsat
Multi-Spectral Scanner data for agricultural monitoring in
the United Kingdom, reports a likely figure of one
suitable image per three years for any part of the United
Kingdom. In many cases, just one image is insufficient to
accurately identify all classes, and multitemporal data
must be used (Belward and DeHoyos, 1987). The chances of

obtaining this sort of coverage will be even lower than those for a single scene.

The noise problem is easier to address by image processing. In cases of systematic noise (for example, Multi-Spectral Scanner 6th line banding), statistical routines can be used effectively. In the case of random noise the broad group of techniques known as contextual processing can be employed.

Context, as applied to remotely sensed digital imagery, is simply the relationship between single pixels and their neighbouring pixels. Intuitively, it can be seen that certain classes are unlikely to occur next to or within areas of other classes, for example deciduous woodland is unlikely to occur within areas of upland moor. Also, small areas differing from their surroundings may be the result of incorrect classification, due to noise within the data, rather than their presence on the ground, or their inclusion on a finished map may be inappropriate to the scale of the map (see sections 5.1.1 and 2.7.2).

5.1.1 ACCURACY AND SCALE

Accuracy of classification is not easily defined. For land cover classification, an accurate classification at one scale might show only two classes, for example 'sea' and 'land'. At a slightly larger scale this might be subdivided to give 'sea','farmland' and 'urban area'. Increasing scale would increase the number of classes

102

required for an 'accurate' classification. At the largest scales, for example, a field containing a crop could even be subdivided into 'hedges','crop', 'weeds' and 'bare soil'.

This problem of defining what is 'accurate' is further compounded by the requirements of the user of the final classification. The terms 'producers- ' and 'users-accuracy' have been mentioned by Aronoff (1982, section 3.4 of this thesis). For the author's purposes an 'accurate' classification is defined as a classification which closely resembles the requirements of its user, at the scale and detail level required.

5.1.1.1 GENERALISATION

The paper by Bryukhanov (1985) provides an important insight into the meaning of the term 'generalisation', when applied to satellite imagery:

'The term "generalisation" was introduced at about the time of the first (satellite images), to describe a number of specific properties of images communicated from space. Conceptually the term denotes scale generalisation achieved through a process that naturally brings together small units of geologic content to produce larger units.

'The generalisation process is one that diminishes spatial ground resolution while generally expanding (satellite image) area coverage.'

When applied to land cover classification, this implies

103

the amalgamation of small areas of differing class replacing these with larger areas of homogeneous class. For example, the merging of pixels classified as 'road', 'grass' and 'hedge' to form a larger area with the class name 'agricultural grassland'.

Bryukhanov (1985) goes on to describe the levels of generalisation achieved for different spatial and radiometric resolutions of satellite data. A complex relationship between map scale, viewing distance and data resolution is also developed, which can then be used to calculate the optimum level of detail to be included in a map.

5.2 CONTEXTUAL ENHANCEMENT ALGORITHMS

5.2.1 MODE FILTERING

Of the many smoothing filters available for use on remotely sensed data, the mode filter is the only type which can be successfully applied to classified imagery (the median filter is sometimes quoted as acceptable, but will tend to be biased towards classes which have labels (values on the image) lying in the middle of the range of classes, since it returns the middle class value when these are arranged in numerical order).

A classified image consists of an array of pixels, whose values represent particular classes, thus adding up these class numbers and taking their average for an n by n patch of pixels will give a meaningless result. A modal filter,

however, will return the class value of the most frequently occurring class within the patch. It can thus be used to smooth imagery after classification.

5.2.2 SMALL AREA REPLACEMENT

In theory, noise and anomalous classifications will occur in small groups of pixels, or single pixels, differing from the surrounding overall pattern. An intuitively simple method of noise removal is to identify homogeneous regions containing less than a predetermined number of pixels (commonly 8), then reclassify these areas according to their context.

Unfortunately, a true small area identification algorithm is extremely expensive computationally (Oldfield, 1988), since for each pixel a search must be made for all of its similarly classified neighbours, then a search for all of the similarly classified neighbours of these, and so on, making sure that no pixel is counted more than once, until either the threshold number of similarly classified pixels is reached or no further similarly classified neighbours are found. If the latter is the case, then all the similarly classified neighbours found up to this point can be declassified.

A good approximation, however, is achieved by applying a threshold to a square patch: if the patch contains less than the threshold number of pixels of similar class to the central pixel, then the central pixel is declassified.

105

5.2.2.1 EDGE AND NOISE DETECTION

Edges appear in remotely sensed imagery where adjacent pixels differ from each other to a greater extent than is common in the image. Such edges usually occur at the boundaries between ground cover types. In many cases, the actual boundary on the ground will occur within an image pixel. This will result in a 'mixel' - a pixel whose spectral response is a mixture of those for different cover types. In these cases, the resultant spectral response may cause the classification algorithm to assign the pixel to an incorrect class. For this reason, it is desirable to identify edge pixels and to reclassify these according to their context.

In the case of noise, random or systematic anomalies exist in the image. These can often be identified as single pixels or small groups of pixels differing from the 'background'. Edge detection algorithms can also be used to identify noise for this reason. Unfortunately, in some cases, these 'noise' pixels may actually represent true ground conditions. Without site inspection, it is impossible to identify such areas, so a compromise must be reached where such areas are 'generalised' out, to produce a classification approximating to the overall picture at whatever scale the end product is required.

Ehrich (1977) states that:

'Much of our knowledge of edge detection comes from

artificial intelligence investigations in which image conditions are not particularly severe'

going on to define boundaries as edges between areas of relatively uniform texture. This paper provides a good review of edge detection literature in the fields of: perception theory; biomedical image analysis and terrain analysis. The paper notes that 'serious noise levels' often cause difficulties and describes edges within Landsat images as falling into two categories:

i) Short and curved (difficult to find)
ii) Long and of low curvature or straight (easier to find, when noise levels are low)

The paper also describes edge detection using matched filters, allowing line growing from segments detected by this method.

The relationship between scale and edges is also discussed: 'A large variety of types (of lineament) may exist at several different scales.'

A knowledge-based approach for lineament extraction is described by Parikh (1986), using low-level operators (filters) to detect edges, followed by interpretation by high-level decision processes, which are able to set off further low-level processing. The paper also refers to the two types of edges discussed by Ehrich (1983), and also mentions the possibility of:

i) Combination of image data with ancillary data, such as

geological or topographic maps.

ii) Collection and codification of geological expertise in a form suitable for automatic processing.

The paper mentions the computational effort involved in such approaches, proposing the use of a massively parallel processor, and suggests the use of 'minimum cost-path algorithms' for achieving specific objectives.

A five step convolution procedure to produce directionally enhanced images is described by Moore (1983), as a method of reducing 'controversy' when edges are defined by human operators. The methodology advocated consists of:

i) Generating a low spatial frequency image with an averaging function.

ii) Extraction of directional data by means of a convolution filter.

iii) Smoothing the directional data using an average or tangent function.

iv) Further smoothing the data by extracting directional trends in the tails of the image histograms.

v) Adding the enhanced directional trends to the image.

Frequency domain filtering, instead of convolution filtering described in section 5.2.2.1.1, is described in the paper by Duggin, Rowntree and Odell (1988).

5.2.2.1.1 HIGH PASS FILTERING

A high pass (edge detection) filter is a filter which

enhances high frequencies in an image, boosting the high frequency information in the image relative to the low frequency information.

For a 3 by 3 pixel filter kernel, a filter leaving the image
unchanged is:

```
0 0 0              0  0  0
0 1 0      or      0 9/9 0  (1)
0 0 0              0  0  0
```

and a filter allowing the low frequency information to pass is given by:

```
        1 1 1            1/9 1/9 1/9
1/9 *   1 1 1     or     1/9 1/9 1/9  (2)
        1 1 1            1/9 1/9 1/9
```

A high pass filter can be created by subtracting (2) from (1): in other words removing the low frequency information from the image to give:

```
        -1 -1 -1          -1/9 -1/9 -1/9
1/9 *   -1  8 -1   or     -1/9  8/9 -1/9
        -1 -1 -1          -1/9 -1/9 -1/9
```

Any filter which accentuates the difference between a pixel and its neighbours will enhance the high frequency components of an image. A good example of such a filter, based on theory, is the Laplacian edge detector:

```
 0 -1  0
-1  4 -1
 0 -1  0
```

This has been found to be an excellent filter to use for 'sharpening' images.

109

The filters so far described are capable of enhancing or detecting high frequency information (detection = filter alone, enhancement = filter result added back to the original image), but can provide no information on the 'direction' of the detected edges. To estimate edge direction, orthogonal filter pairs have been used, such as the template edge detectors shown below:

Roberts

```
0 -1  -1  0
1  0   0  1
```

Sobel

```
 1  2  1    -1  0  1
 0  0  0    -2  0  2
-1 -2 -1    -1  0  1
```

Prewitt

```
 1  1  1    -1  1  1
 1 -2  1    -1 -2  1
-1 -1 -1    -1  1  1
```

Each filter pair is used to find the magnitude of an edge in two orthogonal directions. The direction of the edge vector can then be resolved as shown in Fig.5.1.
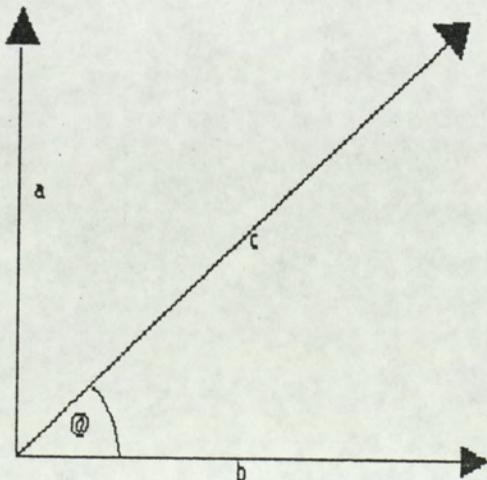
Fig.5.1 Calculation of vector direction

110

The combined magnitude, c, is given by:

$$c=(a^2 + b^2)^{0.5} \text{ (Pythagoras)}$$

and the direction angle, @, by:

$$@=arctan(a/b)$$

In the case of the Prewitt and Sobel filters these values can be returned to the position of the central pixel, however, the Roberts filter presents the problem that none of the four pixels can be described as 'central'. In this case an arbitrary pixel position must be chosen from the four possibilities available. In this study, the top left position was chosen.

Estimation of edge direction can be taken a step further by using the technique of template matching (Golton, 1988), where a series of templates (usually 8 at 0, 45, 90... degrees to each other) are applied to the image to ascertain the direction of edges.

5.2.2.1.2 AUTOMATIC THRESHOLDING OF EDGE DETECTOR OUTPUT

The automatic threshold generator applied to the data uses an automatic threshold selection method based upon a simple image model to choose an appropriate threshold for dividing the image into object and background.

In this case, object refers to the edges, and background the rest of the image.

5.2.2.1.3 DIFFERENTIATION OF NOISE AND EDGES

In theory, random noise, affecting individual bands separately, can be differentiated from true edges, which will occur at the same positions in each band, due to correlation between adjacent spectral bands.

If an edge detector is passed over all n bands of an image and the results thresholded to give edge and non-edge pixels, then corresponding pixel values in each band (0 or 1) are summed over the n bands, the result will be an image showing the likelihood of pixels being edge, noise or neither.

True edge pixels will tend to exist in all bands (in the case of close spectral bands, showing high inter-band correlation (see section 4.5 and Fusco and Trevese, 1985), as is the case with Landsat Thematic Mapper bands 1-5 and 7, used here), whereas noise will tend only to occur in single bands. On summation, the closer the result for a pixel to n, the more likely it is to be a true edge pixel, whilst pixels with non-zero values smaller than n are more likely to represent the locations of random noise. A value of zero will represent the location of a non-edge, non-noise pixel.

5.2.2.2 RECLASSIFICATION

To complete the contextual processing sequence, once small areas and noise pixels have been declassified, using any of the previously mentioned algorithms, they can be reclassified according to their context. This implies a reclassification scheme based on neighbouring, classified pixels.

5.2.2.2.1 NEAREST NEIGHBOUR

The simplest reclassification technique searches from an unclassified pixel to discover its nearest classified neighbour. The pixel is then assigned to the same class as this. An efficient search algorithm can be used which stores a list of co-ordinates, relative to the unclassified pixel, to search, in order of increasing distance. The search is stopped as soon as a classified pixel is encountered.

5.2.2.2.2 MODE OF NEAREST N NEIGHBOURS

An extrapolation of the above technique finds the nearest n classified neighbours of an unclassified pixel and assigns the modal class of these to the unclassified pixel. This technique is less susceptible to any noise remaining in the declassified image.

5.2.2.2.3 MODAL CLASS OF N BY N PATCH

This is effectively a selective mode filter. When an unclassified pixel is encountered the algorithm returns the modal class of an n by n pixel patch centred on the unclassified pixel. The algorithm is fast, but care must be taken to choose an appropriate patch size: too small and some areas will remain unclassified; too big and the image will be overgeneralised, and the time required by the algorithm will increase drastically.

5.2.2.2.4 EFFICIENT SEARCHING FOR NEAREST NEIGHBOURS

For efficient implementation of the nearest neighbour reclassification algorithms, a fast method for finding a pixel's nearest classified neighbours is necessary.

The simplest method to implement by computer, that of calculating the distance to the target pixel from every other classified image pixel, then searching through these to find the shortest distance, is impossibly slow to implement over large images. For example, a 1024 by 1024 pixel image, with, say 10% unclassified pixels will require about 922*922 separate distance calculations for each of 102 unclassified pixels: 86,708,568 calculations.

Clearly, the search must be reduced. One method is to limit the search to a (relatively) small patch, centered on the target pixel, and search as before. For, say, a 101 by 101 pixel patch, with on average 10% of the pixels within the image unclassified, as before, the number of

calculations becomes 91*91 pixels searched for each of the 102 unclassified pixels: 844,662 calculations, and much more feasible.

The method does, however, have two drawbacks:

1.Some pixels may remain unclassified in images containing few classified pixels: increasing the patch size to cope with this increases the number of calculations dramatically.

2.Many pixels could probably be classified by pixels much closer than half the size of the patch - in this case a great many calculations are 'wasted' examining pixels much further away than the nearest neighbour.

The technique used here to keep calculations to a minimum is to carry out an ordered search, outwards from the central pixel, until a classified neighbour is encountered. The search is then terminated.

In this case, only the absolute minimum number of pixels is examined.

The problem of designing an ordered search algorithm remains, however. This has been overcome in this case by:

1.Calculating the distances of all pixels in a large patch from the centre of the patch, and storing these in an array, together with the co-ordinates of the pixels (relative to the central pixel).

2.Array elements are then sorted, according to distance order, using a Shell sort algorithm (Press, 1986). The distance-ordered co-ordinates are then written to a file.

3.Any algorithm requiring ordered searching can now read in this file of search co-ordinates, and examine the pixels indicated by the list in order, until a classified neighbour is encountered. When this occurs, the algorithm can move on to the next unclassified pixel and restart the search from the 'top' of the list.

The author has not encountered any reference to this type of ordered search algorithm in the literature.

5.2.3 WHARTON'S CONAN ALGORITHM

This algorithm (CONtextual ANalysis), developed by Wharton (1982) falls somewhere between context and texture analysis. It requires a classified image as input, but reassigns pixels to new classes according to a texture-like measure of class frequency calculated for an n by n pixel patch.

The algorithm can be split into two parts: first, a mode filter type of operator is passed over the image, but, instead of returning the modal class, the class frequency histograms are saved, forming a temporary image of nclass bands, where nclass is the number of classes. This temporary frequency image is used as the subject of another classifier, in this case minimum distance, which is trained using the same areas as before. The whole image

116

is then reclassified by this, classifying pixels according to their context, rather than their spectral properties.

5.2.3.1 PATCH SIZE AND STORAGE REQUIREMENTS

Experiments elsewhere in this study, using mode filters and Markov relaxation, suggest that the autocorrelation function for classified imagery of this type decays to zero at a distance somewhere between six and ten pixels from a central pixel. The implication of this for Wharton's CONAN algorithm, is to suggest that the ideal patch size should lie between 13 by 13 pixels (approximate maximum distance of 6 pixels) and 21 by 21 pixels (approximate maximum distance of 10 pixels). This leads, however, to serious speed implications (the larger the patch, the slower the algorithm), therefore 9 by 9 and 15 by 15 pixel patches were tried. The former size represents a compromise to increase speed, whilst the latter is the largest size enabling the frequency values to be stored as BYTE (8 bit unsigned integer) variables, saving memory space (15*15=225, 8 bit integers can cover 0-255, therefore values up to 255 can be stored).

An alternative method of estimating autocorrelation distances, by semivariogram estimation, is discussed in section 3.8.

5.2.4 RELAXATION

Relaxation is a technique whereby classified imagery can be reclassified according to contextual information obtained from the original classification. Relaxation is often applied iteratively, but it must be realised that the technique can only be successful when it is applied to a reasonably accurate classification, since the technique assumes that the majority of image pixels are correctly classified, and can therefore be used as the basis of classification modification.

5.2.4.1 PROBABILISTIC RELAXATION

Essentially this involves calculating the probability of membership of each class for each image pixel, then assigning class labels according to the most probable class within an n by n pixel patch centered on the pixel to be classified.

A maximum likelihood classifier can be modified to produce these results, but various short-cuts to enhance the speed of the algorithm (Mather, 1985) must be sacrificed in order to calculate actual probabilities. The algorithm thus becomes extremely slow to implement.

5.2.4.2 RELAXATION BASED ON THE MINIMUM DISTANCE CLASSIFIER

Probabilistic relaxation techniques require a classification algorithm which actually calculates class probabilities. As noted earlier, the maximum likelihood

118

algorithm, in its practical form (as used in this study and described by Mather (1987)), does not actually calculate true probabilities; merely a series of values which have the same rank order as the probabilities. True probabilistic relaxation is, therefore, extremely computationally expensive: there is scope for the development of relaxation algorithms which do not deal in probabilities, but use other, less expensively calculated, statistics.

The computationally inexpensive minimum distance algorithm is presented here as a possible solution to this problem, in two forms: distance based and rank based.

5.2.4.2.1 DISTANCE BASED

For each pixel, the feature-space distances from each class mean centre are calculated, and summed over a 3 by 3 pixel patch (ie over the pixel's 8 nearest neighbours). The pixel is then assigned to the class which has the smallest summed distance associated with it. This provides a relaxation across the patch.

5.2.4.2.2 RANK BASED

For this, the ranks of distances for each class are calculated on the basis of nearest=1, furthest=nclass, where nclass is the number of classes, for all pixels in the patch (3 by 3 was used in this case, as above). They are then summed, and the class with the lowest summed rank is assigned to the central pixel. This technique could

119

equally easily be applied to the maximum likelihood algorithm, in the form used here, since this also is capable of modification to give ranked classes.

5.2.4.3 MARKOV RELAXATION

First order Markov chains represent a method for modeling the effects of autocorrelation. Essentially, from a tally matrix of adjacencies, obtained either in the field, or from the image itself, the probabilities of a pixel of class (a) occurring at a distance d pixels from a pixel of class (b) can be calculated. In this way, relaxation can be refined to include a distance weighting function for classification modification.

The order of the chain can also be used to set boundaries for filtering operations. For an image with Markov adjacency order n, then any reclassification or contextual operations can be confined to within a 2n+1 by 2n+1 square patch (ideally a circular patch, diameter 2n+1), since pixels a greater distance than this apart will be uncorrelated.

5.2.4.3.1 IMPLEMENTATION

Figure 5.2 summarises the steps necessary to complete this
algorithm.

```
┌─────────────────────────────────────────────────────────────┐
│ Estimation of tally matrix                                    │
└─────────────────────────────────────────────────────────────┘
┌─────────────────────────────────────────────────────────────┐
│ Normalise tally matrix to give transition                     │
│ probabilities                                                 │
└─────────────────────────────────────────────────────────────┘
┌─────────────────────────────────────────────────────────────┐
│ Multiply transition probability matrix by itself              │
│ until convergence is reached, thus calculating the            │
│ probabilities of each class occurring at any distance         │
│ from any other class                                          │
└─────────────────────────────────────────────────────────────┘
┌─────────────────────────────────────────────────────────────┐
│ For every pixel in the image calculate and sum the            │
│ probabilities of the occurrence of each class at this         │
│ location                                                      │
└─────────────────────────────────────────────────────────────┘
┌─────────────────────────────────────────────────────────────┐
│ Reclassify the pixel according to the most probable           │
│ class                                                         │
└─────────────────────────────────────────────────────────────┘
```

Fig.5.2 Markov relaxation.

5.2.4.3.1.1 TALLY MATRIX

The first stage of the process is to construct an
adjacency tally matrix, either from the classified image,
or from ground data. Entries are made for the neighbours
of each pixel as in Fig.5.3 below:

```
              Pixel class
              1    2    3    4
           1
Neighbour  2
class      3
           4
```

Fig.5.3 format of tally matrix

5.2.4.3.1.2 TRANSITION PROBABILITY MATRIX

The tally matrix is next converted to show the probabilities of neighbours occurring in each class pair by dividing each element of each row by the total for that row. In this case the matrix is symmetrical about its leading diagonal, since the probabilities of class n occurring next to class m and vice-versa are equal. In other cases, notably in modeling changes of state over time (time series analysis), this may not be the case (for example, the probability of 'lunch' occurring after 'breakfast' is much higher than that for 'breakfast' occurring after 'lunch').

5.2.4.3.2 CALCULATION OF SUBSEQUENT TRANSITION PROBABILITIES

The probability of a pixel of class m having a neighbour of class n at a distance of x pixels away from the pixel can be calculated by multiplying the transition probability matrix by itself x times (in other words raising it to the power of x). The elements of this matrix then represent the probabilities that a pixel of class n will occur at a distance x from a pixel of class m.

5.2.4.3.3 CONVERGENCE

Eventually, a distance will be reached when the probabilities of the pixel belonging to any class must no longer differ significantly from each other, in other words, the central pixel ceases to affect the

122

probabilities of class membership of its neighbours
beyond this distance. This is the limit of autocorrelation
for this class and may be used to determine the
theoretical optimum size of contextual filters.

5.2.4.3.4 AUTOCORRELATION MODELING

The transition probability matrices up to the convergence
of probabilities can be used to model the autocorrelation
function for a classified image, and therefore as part of
a relaxation scheme to contextually reclassify the image.

CHAPTER 6

USE OF AN EXPERT SYSTEM TO AID MULTISPECTRAL
CLASSIFICATION

As discussed in sections 3.7 and 6.4, expert systems
with induction and inferencing capabilities can be used
as suitable tools for defining decision trees for
classification, and for classifying multispectral remotely
sensed data.

An example is Super Expert, a proprietary system for IBM
PCs and compatibles, from Intelligent Terminals Ltd.
This expert system has an induction engine, based on
Quinlan's ID3 algorithm (Intelligent Terminals Ltd. 1986),
which will accept example data on an integer scale.

Any 'supervised classification' procedure can be described
as an induction engine, in that examples are presented to
the 'induction' engine which goes on to produce a set of
rules to be used in classification. It is the mathematics
of the rule induction mechanism which changes from one
method to another. One advantage of the ID3 type of
algorithm, or other pattern matching algorithms, is that
they make few assumptions about the nature of the data and
therefore data on various scales including numeric and
categorical can be mixed.

This chapter contains an example of the use of Super
Expert for multispectral classification.

6.1 USE OF SUPER EXPERT

(Please see chapter 8 for a description of the data sets used).

6.1.1 PEAK DISTRICT DATA SET

Super Expert requires example data from each class in order to infer the decision rules. In this case, this was obtained from the training data used by the other classifiers. Super Expert was set up to accept the attributes shown in table 6.1:

Table 6.1: attributes used by Super Expert

integer band1	integer band2	integer band3	integer band4	integer band5	integer band6	decision Class
23	24	3	1	15	3	"class8"
127	110	123	78	96	54	"class3"
2	1	4	23	34	6	"class5"

(Note: 'band6' refers to TM band 7, since TM band 6 data was not available.)

where 'band1-band6' are the digital values in each of the 6 Thematic Mapper spectral bands used, and 'Class' can be any of the classes defined by the training data (see section 8.4).

The example data can be stored in an ASCII file, with the suffix .TXT, containing records in the following format:

b1,b2,b3,b4,b5,b7,"class descriptor"

(where b1-b5 and b7 are the digital values in each of the 6 TM bands used)

A short Fortran program was written to extract training data in this format from the image files stored on the Vax cluster. Because Super Expert is not designed explicitly for use with large data sets such as those used in remote sensing, the number of examples had to be restricted to 40 per class. This data was then transferred to the PC via Kermit and loaded into Super Expert.

Super Expert requires an ASCII text file, consisting of records in the format given in section 6.1. This corresponds to the Fortran-77 format:

Im,',',Im,',',...',"'An,'"'

where:

m is the number of digits required for each integer value
n is the number of characters used for the class descriptor

A text file where each record has this format can be directly read into Super Expert, but only if the file suffix is '.TXT'. This file is imported into Super Expert.

A decision rule was then induced by Super Expert (by typing !) from these examples and is shown in Fig.6.1.

Fig.6.1 Decision rule from 40 examples per class.

```
band4
    <  64 : band1
        <  74 : band5
            <  50 : band2
                <  22 : class1
                >= 22 : class4
            >= 50 : band4
                <  45 : band6
                    <  23 : band1
                        <  68 : class4
                        >= 68 : class5
                    >= 23 : class7
                >= 45 : band3
                    <  32 : class5
                    >= 32 : band2
                        <  29 : class7
                        >= 29 : class8
        >= 74 : band5
            <  77 : band1
                <  76 : band4
                    <  41 : class3
                    >= 41 : band4
                        <  46 : class8
                        >= 46 : class3
                >= 76 : class3
            >= 77 : band1
                <  91 : band2
                    <  30 : class7
                    >= 30 : class8
                >= 91 : class3
    >= 64 : band1
        <  79 : band5
            <  86 : class10
            >= 86 : class9
        >= 79 : band1
            <  104 : class6
            >= 104 : band2
                <  58 : class3
                >= 58 : class2
```

Explanation of this diagram, showing the tree in the same
format as Super Expert's output, can be obtained by the
following interpretation of the first six lines:

If the pixel's value in band 4 is less than 64 and the
pixel's value in band 1 is less than 74 and the pixel's
value in band 5 is less than 50, then, if the pixel's
value in band 2 is less than 22, the pixel is assigned to

127

class 1, otherwise, if the pixel's value in band 2 is greater than or equal to 22, the pixel is assigned to class 4.

A rule was also induced using 25 examples per class, which is omitted here for clarity. Both rules were tested as described below.

These rules were then encoded in Fortran on the Vax 8650 cluster as a series of IF...THEN...ELSE statements, using the decision tree shell discussed elsewhere. The resultant classifiers were then applied to the study area data and normalised accuracies calculated. Contingency tables and accuracies for these classifications, generated using the verification data used elsewhere in this study (see section 8.4), are given in tables 6.1, 6.2 and 6.3.

In addition, the accuracy figure for the manually derived decision tree classifier, described elsewhere in this study, is also given to provide a comparison of automatically and manually derived decision rules.

6.1.2 NORTH WALES DATA SET

A similar procedure was carried out using the North Wales Thematic Mapper data. This time only three bands were used (TM bands 4,5 and 7), and only eight classes were defined. A manually derived decision tree was not constructed.

6.1.3 YEMEN DATA SET

The three band (MSS bands 1,2 and 4, or 3,5 and 7,
depending on notation) Landsat MSS data for Yemen was also
used, again with eight classes defined. A manual tree was
not derived for this data set.

6.1.3.1 SUN WORKSTATIONS: DATA FILES

The program mentioned in section 6.1.1 was also used on
the Sun workstation where the data sets were situated,
however, the unix command 'unix2dos' had to be run on the
text files before they were suitable for use by Super
Expert on the PC. This program adds suitable DOS end-of-
record markers to Sun created files.

6.1.4 RESULTS

Summarised results are presented in table 6.2. Contingency
tables are presented in the appendix.

Table 6.2 SUMMARY OF ACCURACIES AND TIMINGS

Table 6.2a PEAK DISTRICT DATA SET

rule derived from	normalised accuracy %	time (sec) to derive tree	
25 examples	65.13	560	
40 examples	66.63	1160	
co-incident spectral plots	58.23	-	(manually defined tree approximate time: 1 hour)

Table 6.2b NORTH WALES DATA SET

	Normalised accuracy (%)
25 examples	92.58
40 examples	87.17

Table 6.2c YEMEN DATA SET

	Normalised accuracy (%)
25 examples	21.15
40 examples	41.56

6.2 DISCUSSION

The advantages of the expert systems approach to the design of decision trees are demonstrated when the classification accuracies for the manually derived tree are compared with those produced by the trees generated by Super Expert. For the Peak district data set the automatically generated trees outperformed the manually derived tree by approaching 8%, using less training data.

The time involved is also shorter for the expert systems approach, since the operator no longer has to grapple with statistics and co-incident spectral plots. Another advantage of this is the consistency which automatic tree design brings to decision tree classification.

The reduction in the amount of training data required to produce a satisfactory classification when using an expert system may well be a useful factor in remote, inaccessible areas, where field data is expensive to gather, but it must be remembered that currently available micro-computer based expert systems, such as Super Expert, cannot handle the large amounts of training data routinely used in remote sensing.

For the North Wales data set very high accuracies were produced by the decision trees. These accuracies were produced by relatively simple tree structures (35 and 43 nodes for the 25 example tree and the 40 example tree respectively), however, for the Yemen data complex trees were produced (81 nodes, plus one clash and 100 nodes, plus two clashes for the 25 example tree and the 40 example tree respectively). These complex trees gave low accuracy figures.

This is probably a result of the spectral structure of the classes with which the expert system is presented: if a class has a uniform spectral response, then the tree definition algorithm is well suited to differentiating that class from other classes. If, on the other hand, there is little similarity between a class' pixels, a complex tree will result.

This version of Super Expert is not an ideal tool for classifying remotely sensed data: as mentioned above, large data volumes cannot be handled, and the tree derived had to be recoded for use on a more powerful processor. This experiment does, however, show the potential of an expert system, properly designed for use with remotely sensed data, for rapid and consistent design of decision trees.

With the increasing desire to include data from sources other than remote sensing in classification schemes, expert systems may provide a useful form of classifier.

Super Expert allows 'don't care' values to be inserted into the examples before rules are inferred. This could prove extremely useful in cases where spectral information must be overridden by other data, for example in classification of areas where suitable map data is available: if an area is classified adequately on the map, then spectral information need not be consulted.

It should be noted that the trees induced using 40 examples per class showed similar structures to those induced using 25 examples, those from 40 examples merely exhibiting higher levels of IF THEN nesting, reflecting the greater variations provided by more examples.

6.3 CONCLUSIONS

The use of an expert system is preferable to manual methods of decision tree definition where classes are spectrally homogeneous, since accuracy is improved, preparation time is reduced and consistency is achieved.

In situations where class spectral responses are of a less homogenous nature (for example, the Yemen data set), complex trees of little merit are derived.

This version of Super Expert is not an ideal expert system for remote sensing purposes, but points the way towards similar systems able to handle the high data volumes necessary for efficient use with remotely sensed data.

6.4 SOME NOTES ON THE APPLICABILITY OF ARTIFICIAL INTELLIGENCE TECHNIQUES TO MULTISPECTRAL CLASSIFICATION

6.4.1 CLASSIFICATION AS RELATED TO PATTERN RECOGNITION AND EXPERT SYSTEMS

Image classification may, at first, seem an ideal area for the application of expert systems. Training data is used to derive a series of rules which, when applied to the imagery, allow it to be subdivided into different classes. Moreover, the use of expert systems allows information from sources other than remotely sensed data to be integrated into the classification procedure.

The nature of remotely sensed data, however, may require the user to think again: expert systems derive their rules on the basis that all information supplied is true; they are generally geared to working with small, controlled data sets. The low precision, high volume nature of remotely sensed data implies that some of the information sent to the expert system will be untrue: a case in point being the effects of noise in imagery, leading to training areas which contain unrepresentative pixels.

This 'all information is true' approach is implemented by the look up table classifier (although clashes between classes are taken into account). The poor performance of this algorithm highlights the effects of applying noise sensitive techniques to inherently noisy data.

A more useful approach to the use of expert systems for remote sensing is to bring in knowledge based rules after classification. The contextual algorithms used here are, in fact, examples of this, relying on the knowledge that remotely sensed data is noisy, but has high levels of spatial autocorrelation.

6.4.2 GEOGRAPHICAL INFORMATION SYSTEMS

The increasing availability of computerised spatial data (for example the Soil Survey of England and Wales' Land Information System (LANDIS), Ordnance Survey digital maps and the forthcoming 1991 UK census) will lead to great potential for the use of expert systems. Unreliable, but comprehensive, remotely sensed data can be augmented by other, more reliable (but restricted in terms of coverage), data sets, to which expert systems are much better suited. Remote sensing will thus become only one input to a fully integrated Spatial Information System (see chapter 7 for further notes on GIS).

CHAPTER 7

GEOGRAPHICAL INFORMATION SYSTEMS AND IMAGE SEGMENTATION

7.1 THE IMPACT OF GEOGRAPHICAL INFORMATION SYSTEMS ON IMAGE CLASSIFICATION

The development of Geographical Information Systems, enabling rapid processing of spatial data has great potential in the field of land cover classification. In the near future, most sources of data complementary to that obtained by remote sensing will become available in GIS-compatible format (for example: digital maps and spatially related databases), which will be either directly compatible with remotely sensed data (raster GIS), or easily converted to suitable format (vector GIS: vector to raster conversion being relatively easy, compared to raster to vector).

This information will remove a great deal of frustration from users of remotely sensed data: regions of interest can be quickly extracted from digital maps and relevant ancillary data can instantly be called upon from spatial databases (for example: soils; relief; aspect; map category), avoiding much tedious digitisation work which would have to be carried out at present.

The use of GIS will also allow contextual classification to use a broader definition of 'context', as opposed to the currently applied limited definition of 'characteristics of immediately adjacent pixels', for

example, the paper by Drayton et al (1989) describes the use of population data, from UK census records, combined with remotely sensed data to increase the accuracy of urban classifications using per-pixel algorithms.

7.2 EXAMPLE CASE STUDY: NOTTINGHAMSHIRE

A brief example will enable the reader to envisage some of the savings which GIS can bring to the field of operational remote sensing.

7.2.1 INTRODUCTION

Much of Nottinghamshire's water supply is drawn from boreholes into a large sandstone aquifer to the north of the city of Nottingham. There is currently some concern over the nitrate levels in the water, raised by application of fertilisers to the agricultural land overlying the aquifer which are washed into the aquifer by rainfall.

Figures have been calculated to give average nitrate contribution per unit area for specific crops grown in the area, however, a complete ground survey of crops over the whole area would be too costly to undertake. For this reason the use of Landsat Thematic Mapper data was examined.

136

7.2.2 METHOD

The necessary stages of the study are shown in table 7.1 below, and the benefits which could be realised by use of GIS are noted.

Table 7.1 Comparison of GIS with conventional methodology

OPERATION OF GIS	METHOD USED	POTENTIAL BENEFITS
classification	supervised maximum likelihood	removal of non-agricultural areas prior to classification. per-field classification
isolation of aquifer	frame grab of map, followed by manual boundary digitisation	use of available digital geology map: faster with more accurate registration of data sets.
calculation of areas under each crop	custom written routine, or by hardcopy and planimeter	area calculation routine available
multiplication of areas by nitrate potential	by hand or spreadsheet	routine available
production of final map	ink-jet colour plot	labeling can be added from existing data. Better classification gives superior product.

7.2.3 CONCLUSIONS

As can be seen, the use of GIS in this project would have:

1. Significantly decreased the time spent on trivial tasks.

2. Increased the accuracy of the end products.

3. Improved the quality and presentation of any hardcopy output.

7.3 IMAGE SEGMENTATION AND PER-SEGMENT CLASSIFIERS

The contextual classification algorithms discussed previously have concentrated on noise removal and small area replacement. A further type of contextual classifier relies on the properties of remotely sensed imagery. Because of the autocorrelation between adjacent pixels, and because real ground conditions usually result in homogeneous areas of similar class, then if these areas can be identified and a classification based on the whole area applied, the classification accuracy may be increased.

Barnsley et al (1989) describe the use of map data for segmentation of images prior to classification, enabling the definition of different prior probabilities to a maximum likelihood classifier, according to image segment. The paper reports improvements in urban classification accuracy from the use of this technique.

The use of map data would be particularly useful in agricultural areas, where the fields provide a framework

for segmentation and subsequent classification.

7.3.1 MANUAL SEGMENTATION

Despite the necessity of manually defining segments on the imagery involved, this is the simplest method of dividing imagery into meaningful, homogeneous segments prior to classification. Using this method it is also possible to use information other than that from remote sensing. In the UK, for example, the Ordnance Survey 1:25,000 series of maps show urban areas, woodland, water features and field boundaries, which could all be used to reduce classification error, or to avoid the need for

classification altogether in some cases (particularly water and urban areas, which can be defined from the map alone). Unfortunately, owing to the infrequency of revision of the field boundaries shown on these maps, their utility may be questionable.

Field boundaries provide an ideal framework for image segmentation. If an image is partitioned into fields, and classified on a per-field, rather than a per-pixel basis, then classification accuracy can be increased, since the final classification will be more robust in terms of its noise rejection abilities (Allan, 1986).

If maps with field boundaries are unavailable, then it may still be possible to partition the image on this basis. It is possible to identify individual fields on Landsat

imagery by eye.

This still leaves the problem of actually providing a
'map' of segments which can be used by the classification
algorithms. In the future this may not be a problem,
since Ordnance Survey digital maps should become more
generally available, from which fields could be extracted.
At present, however, the boundaries must be added to
either imagery or digitised maps. This can be simply, if
slowly, done using interactive image/graphics
software, with a minimum requirement that the package
to be used allows:

1.Boundary definition, and

2.Region filling within these boundaries.

For this task, software used for interactive definition of
training areas may be suitable.

7.3.2 AUTOMATIC SEGMENTATION

Automatic image segmentation is possible, although at high
computational expense. Methods used can be split into two
categories: those which segment the imagery on the basis
of boundaries/edges found by techniques such as high-pass
filtering; and those which work on the principle of
combining similar neighbouring pixels to form homogeneous
regions. Of this latter type, a subset relies on comparing
the texture of adjacent areas, whilst another subset
amalgamates pixels on the basis of spectral similarity.
Basically these latter techniques are similar to

classification algorithms, but with added spatial constraints.

7.3.2.1 AUTOMATIC SEGMENTATION BASED ON EDGE-DETECTION AND REGION GROWING

The following method of automatic image regionalisation was tried by the author during this study, but little success was achieved.

7.3.2.1.1 METHOD

The method consisted of the following stages:

i) Edge detection, using a Laplacian detector (section 5.2.2.1.1) - the results of which were thresholded to 'edge' and 'non-edge' pixels (section 5.2.2.1.2).

ii) Calculation of the distance of all 'non-edge' pixels from their nearest 'edge', using the ordered search technique detailed in section 5.2.2.2.4.

iii) Identification of local maxima within the distance image produced by step ii.

iv) Region growing, using the local maxima as seeds for outward growth, until each image pixel was assigned to a region.

7.3.2.1.2 DISCUSSION

The technique was not found to produce satisfactory image regionalisation. The author suggests the following reasons

for this lack of success:

i) The edges defined by the use of a thresholded filter as the first stage of the process may have been unsuited to the task of defining boundaries between homogeneous regions.

ii) Outward growth of the regions from local maxima did not take region size into account - a simple decision, based on minimum distance to region centres (local maxima) was used.

CHAPTER 8

MATERIALS AND METHODS

8.1 IMAGE PROCESSING

Image processing for the Peak District scene was carried out using a combination of an IBM PC-AT based image processing system (the ITS-30) and a cluster of two DEC Vax 8650 minicomputers. Tasks requiring visual interpretation of the imagery were carried out on the PC based system, whilst the rest of the image processing was done on the Vax cluster.

Transfer of data between the two systems was via the software communications package Kermit; a time-consuming procedure, although the availability of an Ethernet Local Area Network will remove this bottleneck in the near future.
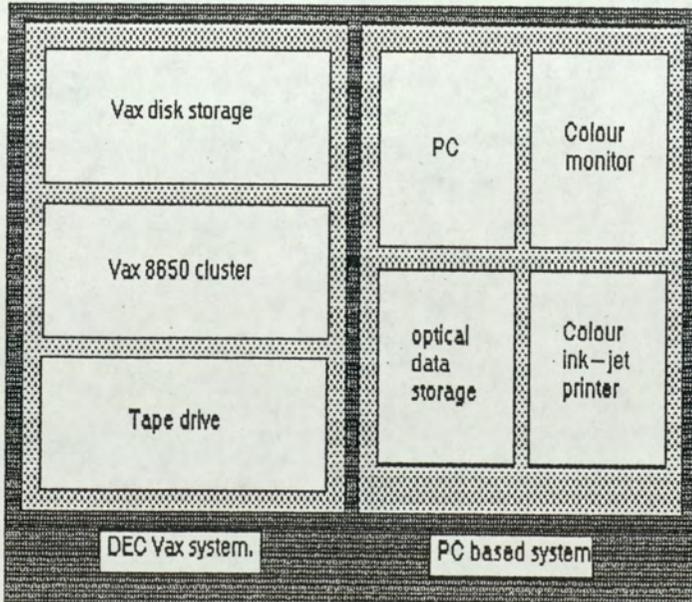
Fig.8.1: Overview of the image processing facilities used

Table 8.1 overleaf shows the breakdown of how each system was used.

Table 8.1: Breakdown of system use

system	process
IBM PC	defining training areas display of False Colour Composite and classified images defining Ground Control Points for geometric correction production of hard copy
Vax Cluster	geometric correction classification contextual enhancements accuracy estimation image manipulation (subsampling, extraction) statistical routines

For the North Wales and Yemen data sets, image processing was carried out using ERDAS software, running under UNIX on a Sun 386i workstation.

8.1.1 SOFTWARE

The IBM PC based software consists of commercially available software, written in C. In addition to this, some routines were specially written to enable files from the Vax to be displayed by the system, and for system files to be written to the PC hard disk in a form suitable for transfer to the Vax.

All image processing software on the Vax cluster was written by the author in Fortran-77 (using DEC's compiler) specifically for the project. Part of this software was converted to run under UNIX on the Sun 386i for use on the North Wales and Yemen data by the author. Subroutines for

Shell sort and matrix manipulation were used from the following sources:

Numerical Recipes, Press,W.H. (1986) (Shell sort)
Statistics and Data Analysis in Geology, Davis,J.C. (1973)
(matrix operations)

8.1.2 DATA

Four data sets were used (see section 1.3). Imagery of the UK Derbyshire Peak District was used to test all the algorithms, whilst the other three data sets (North wales, salisbury and Yemen) were used to further test selected algorithms after initial testing using the Peak District data. These latter data sets were supplied to the author by colleagues, therefore their exact content, for example the criteria used to define training data, was less well known than for the Peak District data. For this reason it is felt that greatest emphasis should be placed on the results for the Peak District study area, with the results for the other data sets serving as a useful cross-reference.

The author would like to express thanks to colleagues at Aston University and the Robertson Group plc for allowing the use of the three additional data sets in this study.

8.1.3 FIELDWORK

It was necessary to undertake much fieldwork to obtain accurate training and verification data for the Peak

District data set. This was carried out over a period of six days during March 1988.

Visits were made to suitable sites throughout the area which were photographed, enabling accurate location of suitable training areas, and the annotation of these on to Ordnance Survey 1:25,000 scale maps of the area. It was also necessary to return to some areas to enable accurate recording of along-transect cover types for markov relaxation, which was also achieved by the photographic method outlined above (see section 5.2.4.3.2).

8.2 GEOMETRIC CORRECTION - PEAK DISTRICT

The raw image data was read on to the Vax cluster as six files of approximately 4,000 by 3,000 pixels each. From these a rough area 1,500 by 1,500 pixels was extracted from each band, centred on the study area. This subscene was used to provide the data for geometric correction, producing a 1024 by 1024 pixel extract with 25 by 25 metre pixels covering the study area.

Correction was carried out using a least squares fit to a polynomial of the form:

$$x=a+bX+cY+dXY \quad (1) \quad y=e+fX+gY+hXY \quad (2)$$

where:

a-h are coefficients

(x,y) are the old image pixel co-ordinates

(X,Y) are the new image pixel co-ordinates

146

The equation coefficients, ground control points and the resultant errors are shown in table 8.2.

Table 8.2: Parameters for geometric correction for Peak District image

EQUATION COEFFICIENTS.

X	Y
-88.24609	273.8955
0.8136215	-0.1979294
0.1978970	0.8099709
-4.5262277E-06	-4.8801303E-07

POINT	REAL X	REAL Y	PREDICTED X	PREDICTED Y	ERROR
1	94.0000	845.0000	94.1648	844.7307	0.3158
2	115.0000	904.0000	115.6681	904.7001	0.9677
3	147.0000	761.0000	146.6721	760.6548	0.4761
4	243.0000	1031.0000	242.4660	1030.6981	0.6134
5	962.0000	660.0000	961.9591	660.6709	0.6721
6	1009.0000	726.0000	1009.0871	725.3714	0.6346
7	149.0000	469.0000	148.8365	469.2608	0.3078
8	495.0000	242.0000	494.0186	241.9474	0.9828
9	522.0000	322.0000	523.4440	321.3442	1.5860
10	634.0000	443.0000	633.6382	443.5960	0.6972

RMS ERROR FOR WHOLE SET IS: 0.8096696 (approximately 0.81)

The image was resampled using the nearest neighbour algorithm to preserve radiometric fidelity. Pixels were resampled from the original 30 by 30 metres to 25 by 25 metres, since a 25 metre pixel is equivalent to a 1 millimetre square on a 1:25,000 scale map, and this enables rapid conversion from pixel to map co-ordinates and vice-versa.

The ten points used were chosen from an initial set of 16 points, being the largest subset of points providing a root mean square error of less than one pixel.

Point selection was carried out by computing the transform equations for the set of points, calculating the RMS error

(see section 8.2.1), and removing the point with the highest individual error (if the overall RMS error was greater than 1.0), then re-computing the transform equations, using the new, smaller, set of points. This process was continued until an RMS error of less than 1.0 was achieved.

The use of this order of transformation, giving rotation, translation and scaling, was felt to be adequate for the data sets.

8.2.1 RMS ERROR

The RMS (Root Mean Square) error is calculated as the square root of the sum of the squared deviations of X and Y (ie the difference between actual and predicted X,Y co-ordinates).

The Salisbury data set was supplied corrected to the Ordnance Survey's national grid, whilst the Yemen and North wales data were not geometrically corrected.

8.3 RADIOMETRIC CORRECTION

Radiometric correction was not attempted on the imagery used. Since no physical parameters were required from the data, and only one scene was used, the original data was felt to be satisfactory.

Extension of the classification to adjacent scenes, however, would almost certainly require some form of radiometric calibration between scenes: in the simplest

case, dark pixel subtraction (Booth, 1987) would be adequate.

8.4 TRAINING DATA AND ACCURACY ASSESSMENT

8.4.1 DEFINITION OF TRAINING AREAS

The geometrically corrected 1024 by 1024 pixel scene for the Peak District was split into 4 512 by 512 pixel quadrants, to enable display at full resolution on the PC based image processing system. Training areas for each class were defined interactively, using the 'image classification' software package available with this system, on each quadrant. Files containing these areas were then transferred to the Vax for subsequent processing.

With the intention of splitting these areas into two sets (training and verification areas), at least 200 pixels were chosen for each class (Hay (1979) recommends at least 50 pixels per class for testing classification accuracy).

For the other data sets suitable training and verification areas were supplied with the images.

8.4.2 TRAINING AREAS AND CLASSES - PEAK DISTRICT

Ten classes were chosen as representative of the ground cover in the study area. These are given in table 8.3:

Table 8.3 Classes chosen to represent the Peak District study area

Class number	Description	Notes
1	water	mostly reservoirs
2	quarry	limestone or sand
3	urban	towns/villages
4	coniferous woodland	
5	broad-leaf woodland	
6	bracken	
7	black moor (heather)	
8	white moor (grasses)	
9	Agriculture (pasture)	
10	Agriculture (better pasture)	

It should be noted that these arbitrary classes were not meant to provide an accurate classification of the area at any specific level of detail, but to enable the comparison of different classification algorithms on a common basis.

Homogeneous areas for each class were identified from a combination of Ordnance Survey map data (at scales of 1:25,000 and 1:50,000), oblique terrestrial photography and field notes made during visits to the area. This information was transferred to the Ordnance Survey 1:25,000 maps of the area to enable rapid identification of areas suitable for training and verification.

8.4.3 SPLITTING

The four quadrants were recombined to form a 1024 by 1024 pixel image on the Vax cluster. The areas of known class were then split into two equal sets: one for training the classifiers; the other for testing their accuracy. This was achieved by splitting the image into two, taking alternate pixels, in a 'check' pattern to form the two images, and filling in the remainder with zeros (see

```
   original         training      verification

 1  2  3  4        1    3           2     4
 5  6  7  8           6    8     5     7
 9 10 11 12        9    11           10    12
13 14 15 16           14   16    13    15
17 18 19 20        17   19           18    20
```

Fig.8.2 Splitting of areas of known class into training and verification areas

8.4.3.1 NOTE ON ACCURACY IMPLICATIONS

The autocorrelation of the imagery will lead to the two sets of data comprising pixels with very similar vectors, despite being made up of entirely different pixels. Two observations have been made about this situation (Swain and Davis, 1978):

1.Use of the same (or very similar) areas to test accuracy as those used to train the classifier will result in accuracy estimates which are higher than those occurring in reality.

2.If the accuracy estimates from training data are significantly different than those from verification data, then the choice of training data is poor.

These two conflicting observations make it difficult to justify any accuracy assessment technique as being 'superior' to another. Fortunately in this case, only relative accuracy figures are required. It should, however, be noted that all accuracy estimates given should only be used to compare different algorithms, and should not be taken as truly representative of the performance of

the individual classifiers.

8.4.4 OTHER DATA SETS

The training and verification data was supplied with the data sets. The data consisted of separate training and verification data, for representative cover types for these areas.

8.4.5 DERIVATION OF TRAINING AND VERIFICATION STATISTICS

Training areas were defined for 10 classes, representative of cover types in the Peak District, using the ITS-30 PC based image processing system. These were then saved as image files and sent via the Kermit communications software package to Aston University's DEC Vax 8650 cluster where all subsequent processing took place.

8.4.5.1 SPLITTING OF TRAINING AND VERIFICATION AREAS

Testing classification accuracy using training data is not considered to provide an accurate estimate of accuracy (Swain and Davis, 1978). To provide a more accurate estimate, Duda and Hart (1973) suggest the process of 'leaving one out', where one pixel is removed from the training data prior to classification, then its actual class is compared with the class to which it has been assigned. This is carried out for each training area pixel in turn, eventually arriving at a figure for correct and incorrect assignments, from which percentage accuracy can be calculated. Unfortunately, in the case of remotely

sensed data, where there may be several thousand training pixels, the time required to carry out such an accuracy assessment reduces this method to being of theoretical interest only. In the case of per-pixel classifiers, the calculations could be reduced to only those pixels contained within the training sets, as set out below:

i) Extract a single pixel from training data.

ii) Calculate class statistics for classification using remaining training pixels.

iii) Assign the single pixel to a class according to the classification rules derived from the statistics.

iv) Check the single pixel's assigned class against its true class (from training data).

v) Update accuracy figure using result of this test (accuracy = number of correct pixels/total number of pixels tested).

vi) Return to step i, and continue until each pixel in the training set has been tested.

In the case of contextual algorithms, the neighbouring classified pixels are used to reclassify pixels, therefore the above method could only be applied if the whole image were to be classified for each combination of extracted pixel and remaining training data, since, in theory the class of every image pixel is related to the class of every other image pixel.

A more practical, yet less satisfactory, solution is to have two separate sets of ground data: one used to train the classifier; the other to assess its accuracy by use of a contingency table (sometimes called a confusion matrix). Duda and Hart (1973) point out that the use of this method will, in fact, lower the theoretically achievable classifier accuracy, since it is deprived of half of the potential training data, but this must be accepted if time constraints preclude the 'leaving one out method'.

Having defined the training areas for this study, a method of splitting them into two separate sets of training and verification areas was necessary for subsequent accuracy assessment. This was achieved by taking every other pixel from the full training areas image for actual classifier training, whilst the remaining pixels were used to assess the classification accuracy.

8.4.6 CONTINGENCY TABLES

The contingency table is a device for summarising the accuracy of a classification. In simplest form, it consists of a tally table of actual class (from ground data) against assigned class (the class to which the classification algorithm has assigned the pixel). The matrix is easier to interpret if the class order is the same on both axes, since the leading diagonal values relative to the other entries quickly give a 'feel' of the classification's accuracy:

154

```
              classified as
                U    1   2   3   4
            1   -   56   -   2   -
ground  2   3   27 102   -   -
class   3  29   5   3  97   1
            4  10   2   -   -   8
```

Where U represents any unclassified pixels.

The column for unclassified pixels is optional, since in many cases classifiers are designed to assign all pixels to a class, leaving no image pixels unclassified.

The higher the values on the leading diagonal relative to the values in the other cells, the more accurate the classifier. The simplest method of calculating the classification's accuracy is to sum the leading diagonal elements to give the total number of pixels correctly classified, then express this as a percentage of the sum of all the matrix elements. This is the normalised classification accuracy.

This only gives an accuracy figure for the whole matrix. If individual class accuracies are required then the matrix must be normalised on a per-column or a per-row basis. Here each row or column element is expressed as a fraction of the relevant row or column total. Thus, expressed by rows, the matrix now shows proportions of ground data correctly or incorrectly classified, whereas, expressed by column, the matrix shows the proportions of image pixels assigned to correct or incorrect classes.

If areal estimates are required, these figures can be applied as correction factors to the class pixel totals

for the image (Chrisman, 1981).

The figures quoted throughout this study are overall normalised accuracies.

8.4.7 CONFIDENCE LIMITS FOR ACCURACY ESTIMATES FROM CONTINGENCY TABLES

The normalised accuracy figure derived earlier represents the proportion of samples correctly allocated to a class:

proportion correct = c/n

c = number of pixels correctly classified in area
n = total number of pixels in area

This is also known as the sampling fraction, and confidence limits for this can be derived from statistical tables, given n. A suitable table is included in Pearson and Hartley (1966).

A problem with the tables is that the values of n for which they are drawn up tend to be relatively small. The tables in Pearson and Hartley (1966) cover n=8 to n=1,000. Unfortunately for this study the value of n is nearer 8,000 when overall classification accuracies are calculated.

Another method of estimating these confidence limits is to use the binomial standard error formula:

$$SE = (P.Q/N)^{0.5}$$

SE = standard error (%)

P = percentage of classifications correct

Q = percentage of classifications incorrect

N = number of pixels in sample

If confidence limits must be estimated for any of the accuracies quoted in this study, then use of Pearson and Hartley's table 41, with a value of 1,000 for n is recommended for the following reasons:

1.the resultant confidence limits will be reasonably narrow (+/-<5% in most cases), but will provide a 'safety factor' to guard against inaccurate results.

2.the nature of the verification areas, close to the training areas may lead to over high estimates of classification accuracy.

Alternatively, Snedecor and Cochran's formula, given by Baines (1988) can be used. This calculates the 95% confidence interval and is given as:

$$\hat{P}+/- = 1.96(P*Q/N)^{0.5} + 50/N$$

where:

$\hat{P}$+/- is the percentage range of accuracies at 95% confidence.

P is normalised accuracy (%)

Q is given by 100-P

N is the number of samples

157

8.5 NOTES ON IMPLEMENTATION OF THE ALGORITHMS

8.5.1 DERIVATION OF CLASS STATISTICS FROM TRAINING DATA

Table 8.4 below lists the classification algorithms together with the necessary statistics required to be calculated from training data to use the algorithms.

Table 8.4 Statistics required by classifiers

Algorithm	Class means	Class standard deviation	Variance-covariance matrix	Coincident spectral plots
M L	yes	no	yes	no
D D	yes	yes	no	no
M D	yes	no	no	no
Box	yes	yes	no	no
D T	yes	yes	no	yes
LUT	no	no	no	no

M L = Maximum Likelihood

D D = Deviant Distance

M D = Minimum Distance

D T = Decision Tree

LUT = Look Up Table

8.5.1.1 NOTES

1.Co-incident spectral plots cannot be drawn without class means and standard deviations.

2.The look up table classifier is nonparametric, requiring no assumptions to be made about the shapes of the class frequency distributions, therefore no statistics are required.

3.The Box classifier could be modified as 'min-max'

requiring class minima and maxima rather than means and standard deviations.

8.5.2 MARKOV RELAXATION

8.5.2.1 DERIVATION OF MARKOV TRANSITION PROBABILITY MATRIX FROM MAP AND TRAINING DATA

Estimating the transition probabilities from one class to another using an existing classification is prone to the effects of noise. Ideally, to minimise this, the transition probabilities should be obtained direct from field data. This raises the problems of designing a sampling scheme which is both quick and accurate in terms of the transition probabilities estimated. Two methods were investigated for this study: stratified random sampling and along-transect sampling. Of these, transect sampling was selected as the most practical method, for the following reasons:

1. Stratified random sampling requires detailed knowledge of a large area, or repeated visits to pixel locations chosen by the sampling.

2. For small, relatively unimportant, classes (for example bracken) a great deal of time can be wasted attempting to find sufficient points within these classes.

3. It was felt that the field data was insufficiently accurate to be used for stratified random sampling (which requires sub-pixel locational accuracy).

8.5.2.2 ALONG TRANSECT SAMPLING

This sampling method, either along random or carefully selected 'representative' transects, provides a quick method for generating neighbour tally matrices from which estimates of transition probability matrices can be calculated.

The first stage of the process is to define a transect across the study area. It is important to try to avoid introducing any directional bias into the data at this stage - ideally some random element should be used to choose transects: in this case north-south and east-west Ordnance Survey 1:25,000 map grid lines were randomly selected in this case as being convenient and relatively unbiased in their positioning. Run lengths of each cover type are then recorded in the sequence in which they occur (in this case, using 1:25,000 maps, annotated with field data, the lengths were recorded in millimetres, since one millimetre at 1:25,000 scale represents 25 metres on the ground, which is the resampled pixel size of the imagery).

For example:

10(2) 25(7) 91(8)...

represents 10 mm of class 2, followed by 25 mm of class 7, followed by 91 mm of class 8. From this a tally matrix of neighbours can be made up, as shown in Fig.8.3.

```
              class
              2  ... 7   8
          2   10 ... 1   0
          .   .  ... .   .
class     .   .  ... .   .
          7   -  ...25   1
          8   -  ... -  91
```

Fig.8.3 Tally matrix

(Note: the matrix is symmetrical about the leading
diagonal, therefore only half has been shown for clarity.)

From this tally matrix, the data can be normalised to give
the transition probability matrix, as explained in section
5.2.4.3.1.2.

8.5.2.3 DERIVATION OF MARKOV TRANSITION PROBABILITY MATRIX
FROM IMAGE DATA

In some cases it may not be possible to obtain enough
field data to estimate the transition probability matrix.
In this case, in order to use the Markov relaxation
technique, the matrix must be estimated from the
classified imagery. In doing this, it is assumed that the
noise levels (incorrect classifications) are small
relative to the correct classifications. If this is the
case, then a tally matrix derived from neighbour
statistics for pixels in the classified image will
approximate the true transition probability matrix for the
data after normalisation.

For this study, the tally matrix was created by examining
the 'right hand' neighbour of each image pixel (with the
exception of pixel 1024 in each line, whose right hand
neighbour was not present). The relevant cell of the tally
matrix was incremented by 1 for each pixel, then, after

161

this was completed, the two 'halves' (either side of the matrix) were summed, reflected element by reflected element, to produce a symmetrical matrix, since, in this case class n on the right of class m was considered to be the same as class m on the right of class n.

The matrix was then normalised.

The 'quality' of the estimated transition probability matrix from this method will depend on the quality of the classification used to produce the matrix. Given that the tally table produced from a 1024 by 1024 pixel image will contain 1023*1024 entries, then there will be scope to remove pixels whose class is dubious before generating the tally table, hopefully further reducing the error.

8.5.2.4 TALLY MATRIX FROM FIELD DATA

The tally matrix, estimated from the map and field data transects, for the class adjacency statistics of the area to be used by the Markov enhancement algorithm is given in Fig.8.4. Because the matrix is reflected along the principal diagonal, it is only necessary to show half the matrix.

```
                class
            1   2   3   4   5   6   7   8   9  10

        1   26  0   0   1   0   1   0   1   0   0
        2       69  0   0   2   0   0   1   0   4
        3           94  0   1   0   0   0   2   6
        4               145 2   1   0   8   1   0
 class  5                   94  2   0   12  11  14
        6                       30  1   2   1   0
        7                           120 6   0   0
        8                               511 8   3
        9                                   270 2
        10                                      587
```

Fig.8.4 Tally matrix derived from field data

8.5.3 SIMULATION OF PER-FIELD CLASSIFIER

In order to assess the value of developing per-field classifiers, a simulation study was devised.

The projected classifier would assign all pixels within a field to their modal class. To simulate this, training data was used to provide a realistic selection of classified pixels from each class. These pixels were then divided into groups of 100, each group representing a 'field'. This was done for 100 fields for each of the 10 classes. The modal class of each field was obtained, and compared with the correct class. Accuracy was assessed as:

1.Number of correctly classified fields/total number of fields

2.Number of correctly classified pixels/total number of pixels

where 1 represents the per-field classifier's performance, and 2 the performance of an ordinary per-pixel classifier, given the same data set.

Training areas were used to define the appropriate mixture of correct and mis-classified pixels for each class. These were identified using both sets of training and verification areas used to train and test the per-pixel classification algorithms.

For each training area (of each class), the classes to which each pixel was assigned by the classifiers were read into an array, noting the totals for each area. These were then split into 'fields' of 100 pixels (starting at array element 1 each time the number of pixels chosen exceeded the number of array entries). The classifier accuracy and the per-field accuracies were calculated as above.

8.5.4 THRESHOLDED HIGH-PASS FILTERS

8.5.4.1 CHOICE OF IMAGE BAND FOR HIGH-PASS FILTERING

An arbitrary decision was made to use the band with the largest dynamic range for the filtering operations. Results are presented in table 8.5 showing the minimum and maximum values for each band on the image. the dynamic range is also given.

Table 8.5 Dynamic range of spectral bands

Thematic Mapper band	minimum value	maximum value	dynamic range
1	61	87	26
2	21	41	20
3	18	48	30
4	30	111	81
5	27	116	89
7	5	53	48

164

8.6 ALTERATIONS TO EXISTING ALGORITHMS AND NEW ALGORITHMS

Many of the algorithms used for this study were developed
specifically for the project, whilst others represent
novel modifications to techniques already in general use
by the remote sensing community. The list below attempts
to summarise the algorithms developed and implemented in
code by the author during the course of this research
and briefly detail the differences (where
appropriate) from conventional algorithms in use.

8.6.1 LOOK UP TABLE CLASSIFIER

Because of the often high dimensionality of remotely
sensed data it is rarely possible to store a sufficiently
large look up table for classification of these data sets,
although the DEC Vax used for this study is able to store
8 byte integers (Fortran INTEGER*8, 64 bits). Mather
(1985) employs hash coding to enable a look up table
to be stored as part of the maximum likelihood
algorithm. The use of look up tables has not, however,
been reported for classification of remotely sensed data
directly from spectral information.

8.6.2 DECISION TREE CLASSIFIER

Belward and deHoyos (1987) detail the use of binary
decision trees for classification of remotely sensed data.
This study differs from their research in two respects:
firstly, the manually defined tree was not limited to
binary decisions, and secondly, in the application of

expert systems technology to tree definition, which has not previously been reported.

8.6.3 DEVIANT DISTANCE CLASSIFIER

This represents a simple modification to the more usual minimum distance algorithm.

8.6.4 ITERATIVE NEAREST NEIGHBOUR CLUSTERING

The ISODATA algorithm represents a frequently used example of this type of classifier applied to remotely sensed data.

8.6.5 SMALL AREA REPLACEMENT BY THRESHOLDED EDGE DETECTORS

The use of edge detection filters to define 'small areas' for reclassification by their context has not been reported, workers generally preferring to identify such areas on classified imagery by means of pixel counting algorithms.

8.6.6 NON-PROBABILISTIC RELAXATION

Work on relaxation labeling to date has concentrated on extension of the maximum likelihood algorithm to cover local areas. Presented here is an algorithm based on the minimum distance classifier which either uses locally summed distances or ranks to relax the classification.

8.6.7 MARKOV RELAXATION

The Markov chain model for relaxation of classified imagery has been developed specifically for this research.

8.6.8 WHARTON'S CONAN ALGORITHM

The classification procedure used here differs from that used by Wharton (1982). In this case a simple minimum distance algorithm is applied, representing a significant decrease in computational expense.

8.7 APPLICABILITY OF CONTEXTUAL ENHANCEMENTS TO OTHER AREAS

In order to assess the generality of the conclusions of this study, those algorithms found to be most effective on the Peak District data were applied to a different data set. The details of this test are set out below.

8.7.1 AREA AND IMAGERY

The test area is part of the United Kingdom Salisbury plain, consisting of small towns, agricultural areas, woodland and natural grasslands. Landsat Thematic Mapper data for 1985 was available for this area, with a series of training areas defined as part of a separate research project within the Remote Sensing Unit. I am grateful to Mr. John McGuire for making this data available.

8.7.2 IMAGE PROCESSING

The data consisted of a geometrically corrected 512 by 512 pixel image, stored on the optical disk system of the PC-based ITS-30 image processor. In order to keep file transfer between the Vax cluster and this system to a minimum, the classification of the data was carried out using the ITS-30. The following classifiers were applied, using bands 3,4 and 5 of the data:

Box classifier
Minimum distance classifier
Maximum likelihood classifier

The classified images, and training data were then transferred to the Vax cluster, where overall accuracies were calculated and the contextual enhancements, detailed in table 8.6 were carried out.

Table 8.6 Classifiers and contextual enhancements

classifier	contextual enhancement
box	9 by 9 mode filter 9 by 9 CONAN small area replacement (8,5)
minimum distance	9 by 9 mode filter 9 by 9 CONAN small area replacement (8,5)
maximum likelihood	9 by 9 mode filter 9 by 9 CONAN Markov relaxation (tally matrix from image)

Notes: (8,5) indicates declassification of areas of 8 pixels or less, followed by reclassification by mode of 5 nearest classified neighbours.

168

8.8 ALGORITHM IMPLEMENTATION

The algorithms used in this study were coded by the author in Fortran-77, either from published descriptions of the algorithms or from the author's own ideas, with the exception of the use of published routines for such operations as matrix manipulation and data sorting (see section 8.1.1) and some commercial software (see section 8.1.1).

The reasons for coding the algorithms were as follows:

i) At the start of the project the availability of proprietary image processing software could not be guaranteed.

ii) It was envisaged that some coding of algorithms would be necessary, even if full use were to be made of existing image processing software. To this end it was felt that experience in coding the simpler image processing algorithms would be of great value during the latter stages of the project.

iii) Certain computationally expensive algorithms would place too great demands on existing image processing software and hardware. Thus the use of the University's central computing facilities was desirable.

Whilst not attempting to describe all the algorithms in detail, or to list the source code of each program, this section is provided to describe some of the features common to these programs.

8.8.1 DATA STRUCTURES

Satellite images consist of large numerical arrays. As supplied on magnetic tape, and as stored by computer these images are made up of a series of records, each of which usually represents one line of image data (in some cases these records are 'blocked' to contain several lines of data, or the data structure is different, for example band interleaved by pixel or pixel pair data, however, the author will confine this discussion to band sequential, unblocked data in an attempt to avoid unnecessary complication). All the data used in this study was either supplied in this format, or converted to it, before processing.

This file structure lends itself in the case of 8-bit or 7-bit (stored in the lower 7 bits of 8-bit integers) data to reading on a line by line basis by programs such as the one below:

```
      PROGRAM EXAMPLE
C
C An example of how to read an image one line at a time C
from Fortran
C
      BYTE IN(1024),OUT(1024)
C
      OPEN(10,FILE='INFILE.DAT',STATUS='OLD',FORM=
   &  'UNFORMATTED')
      OPEN(11,FILE='OUTFILE.DAT',STATUS='NEW',FORM=
   &  'UNFORMATTED')
C
      DO I=1,100
      READ(10)IN
      DO J=1,1024
      OUT(J)=IN(J)
      ENDDO
      WRITE(11)OUT
      ENDDO
C
      STOP
      END
```

This program will read the first 100 lines from a file called INFILE.DAT, which has a line length of 1024 pixels per line. It will then copy the input array IN byte by byte into another array OUT which is then written to the file OUTPUT.DAT.

The BYTE variable type can also be replaced by INTEGER*1 or LOGICAL*1 if this type is not recognised by the compiler in use.

Unfortunately, some unix systems do not store their data files as discrete records. In this case, the easiest way to access individual pixels within files from Fortran is to open the image files as direct access read, with a record length of one byte, then combine the individual bytes into line records within the program. Despite its slowness, the author was forced to modify his programs in this way when faced with data processing on the Sun workstation for the Yemen and North Wales data sets.

Individual pixels can be accessed within byte arrays as individual array elements, however care must be taken in the case of eight bit data, which is often interpreted as signed seven bit data by the compiler, causing values greater than 127 to become negative. A statement similar to:

        IF(I.LT.0)I=I+256

will solve this problem, where I is the integer value of the pixel byte. A similar statement can be used to reverse

171

this:

```
IF(I.GT.127)I=I-256
```

For per pixel and per line operators, the above method of reading an image one line at a time is sufficient, however for some applications (for example box filters) it becomes desirable to hold several lines of an image in the computer's memory at once. This can be achieved by declaring n BYTE arrays (one for each of n lines, where n is the linear dimension of an n by n pixel box) then:

i) Replacing the contents of each line's array by the contents of the next line's array and:

ii) Reading the next line of the image into the nth array.

This results in a moving strip of n lines stepping through the image (a special case is necessary at the top of the image to ensure that all line arrays are always full).

Some algorithms (for example, those requiring spatial searching) require that the whole image is held in memory. This can be achieved by use of n by m byte arrays (n=No. of lines, m=No.of pixels), filled one line at a time from file. This method was not used universally because of the restrictions on image size introduced by having to hold such large arrays.

172

8.8.2 SUBROUTINE LIBRARIES

Two subroutine libraries were available for possible inclusion in the programs: the NAG subroutine library, and the SPIDER routines. Of these, the former consists largely of mathematical functions, for example matrix manipulation, which could be taken from other readily available sources (see section 8.8.1), whilst the latter is a specific image processing library. The SPIDER algorithms were not used, however, for two reasons:

i)The algorithms are almost exclusively single-band operators and therefore unsuited to multispectral processing.

ii)The time factor involved in using even simple subroutines from the library (since most routines call further routines, which call further routines, and so on) often results in it being faster to write an image processing program from scratch.

Had the emphasis of this thesis been on texture, rather than context, the SPIDER routines would have been invaluable. the interested reader is directed to Oldfield (1987) for further notes on their application to textural image processing.

CHAPTER 9

RESULTS

9.1 PEAK DISTRICT DATA SET

The results of the experiments are a series of accuracy/time pairs. For each algorithm, or variation of the algorithm, the classification accuracy was estimated, using the verification pixels. The Vax 8650 CPU time used to produce each classified image was noted (this is the 'charged CPU time' figure from the log file created by running the program as a non-interactive batch job).

9.2 CLASSIFIERS

Results are presented for the following classification algorithms, using the same training data, with accuracies calculated using the same verification data:

Maximum Likelihood
Deviant Distance
Minimum Distance
Box
Decision Tree
Look Up Table

In the case of the box classifier, limits of 2 standard deviations on either side of the class mean reflectance in each band were set. Beyond these limits the pixels remained unclassified.

174

The decision tree classifier

was designed to use only three out of the six available bands (Thematic Mapper bands 1, 4 and 5). These bands were selected because, on inspection of the co-incident spectral plots, they appeared to be capable of distinguishing between all the classes in the training data. Alternative, and less subjective, methods of band selection (for example, the use of divergence) are discussed in section 4.5.

Results are given for two versions of the look up table classifier. For the zero thresholded version, all table entries were used in classification. For the thresholded (threshold set to 10) version, any table entries containing less than the threshold value were deleted before classification.

Since all the classifiers require some kind of statistics to be generated from the training data before they can be run, this time is included in the CPU time figures.

The look up table classifier was applied to the overall brightness image as well as to the spectral shape image.

9.3 CONTEXTUAL ENHANCEMENTS

The classifiers listed above provide a useful variety of classification accuracies to test the effectiveness of the contextual enhancement algorithms used here. For this reason, each algorithm was tested on each classified image, and accuracy/time pairs are again listed.

175

The following contextual algorithms were applied to the
classified imagery:

9.3.1 MODE FILTERS

3 by 3 pixel patch

5 by 5 pixel patch

7 by 7 pixel patch

9 by 9 pixel patch

11 by 11 pixel patch

13 by 13 pixel patch

9.3.2 SMALL AREA REPLACEMENT

areas containing 1 or less pixels declassified, followed
by reclassification by:

 nearest classified along line neighbour

 modal class of 9 by 9 pixel patch

 nearest classified neighbour

 modal class of nearest five classified neighbours

areas containing 4 or less pixels declassified, followed
by reclassification by:

 nearest classified along line neighbour

 modal class of 9 by 9 pixel patch

 nearest classified neighbour

 modal class of nearest five classified neighbours

areas containing 8 or less pixels declassified, followed

by reclassification by:

 nearest classified along line neighbour

 modal class of 9 by 9 pixel patch

 nearest classified neighbour

 modal class of nearest five classified neighbours

9.3.3 THRESHOLDED HIGH PASS FILTERS

filters run on band 5 of the unclassified image:

 3 by 3 pixel Prewitt filter, automatically thresholded.
Pixels set to 1 declassified. Reclassified by:

 nearest classified neighbour

 modal class of nearest five classified neighbours

 3 by 3 pixel Prewitt filter, automatically thresholded.
Pixels set to 1 and any pixels within a radius of 2 pixels
of these declassified. Reclassified by:

 nearest classified neighbour

 modal class of nearest five classified neighbours

 2 by 2 pixel Roberts filter, automatically thresholded.
Pixels set to 1 declassified. Reclassified by:

 nearest classified neighbour

 modal class of nearest five classified neighbours

 2 by 2 pixel Roberts filter, automatically thresholded.

Pixels set to 1 and any pixels within a radius of 2 pixels
of these declassified. Reclassified by:

nearest classified neighbour

modal class of nearest five classified neighbours

filters run on the classified image:

3 by 3 pixel Prewitt filter, automatically thresholded.
Pixels set to 1 declassified. Reclassified by:

nearest classified neighbour

modal class of nearest five classified neighbours

3 by 3 pixel Prewitt filter, automatically thresholded.
Pixels set to 1 and any pixels within a radius of 2 pixels
of these declassified. Reclassified by:

nearest classified neighbour

modal class of nearest five classified neighbours

2 by 2 pixel Roberts filter, automatically thresholded.
Pixels set to 1 declassified. Reclassified by:

nearest classified neighbour

modal class of nearest five classified neighbours

2 by 2 pixel Roberts filter, automatically thresholded.
Pixels set to 1 and any pixels within a radius of 2 pixels
of these declassified. Reclassified by:

nearest classified neighbour

modal class of nearest five classified neighbours

178

9.3.4 MARKOV RELAXATION

relaxation using tally matrix estimated from classified image relaxation using tally matrix estimated from ground data

9.3.5 WHARTON'S CONAN ALGORITHM
based on 9 by 9 pixel patch
based on 15 by 15 pixel patch

9.3.6 NON-PROBABILISTIC RELAXATION

The two non-probabilistic relaxation algorithms were modified versions of the minimum distance classifier. Results are presented here for:

relaxation based on summed pixel vector-class mean distances over a 3 by 3 pixel patch

relaxation based on summed class ranks over a 3 by 3 pixel patch

In addition to these experiments, the Markov relaxation algorithm was applied to the maximum likelihood image for a second iteration, as recommended by proponents of probabilistic relaxation techniques. The results were not, however, considered sufficiently successful to continue the experiment to cover the other classified images.

179

9.4 ASSESSMENT OF RELATIVE CONTRIBUTIONS OF NOISE AND EDGES TO CLASSIFICATION ERRORS

Applying the theory described earlier for separating noise and edges, the classified images were reclassified according to the following rules:

noise pixels removed; reclassification by nearest classified neighbour

edge pixels removed; reclassification by nearest classified neighbour

noise and edge pixels removed; reclassification by nearest classified neighbour

The accuracies after these steps are given for each classifier.

9.5 CLUSTERING

9.5.1 HISTOGRAM CLUSTERING

The histogram clustering algorithm, described earlier, was applied to the six bands of the geometrically corrected Thematic Mapper image. Training data for each class was then used to determine which class each cluster could be most closely associated with (the class containing the highest proportion of a cluster's pixels). The clustered image was then passed through a look up table, reassigning each cluster to its associated class. Normalised

accuracy was calculated as for the supervised techniques on the resultant image.

9.5.2 ITERATIVE CLUSTERING

The iterative, nearest neighbour, clustering algorithm, with 10 clusters, was applied to the output of the minimum distance classifier to investigate possible improvements by 'fine-tuning' the classification. Results are presented after 1, 2 and 3 iterations.

9.6 SIMULATED PER-FIELD CLASSIFIER

A per-field classifier, with fields accurately defined by geographical information systems techniques, was simulated using training and classified data from all the classifiers. The results of this are presented in a 'before and after' format, all accuracies having been calculated directly from the data, without contingency tables.

9.7 NOTES ON CONFIDENCE LIMITS FOR ACCURACY ESTIMATES

Many different methods have been proposed for assigning confidence limits to the accuracy estimates obtained from contingency tables (see literature review and chapter 8).

Calculations are presented here for the 95% confidence limits on the lowest unenhanced classification and the highest contextually enhanced accuracies achieved, using Snedecor and Cochran's formula, given by Baines (1988).

9.7.1 95% CONFIDENCE LIMITS FOR LOWEST UNENHANCED CLASSIFICATION ACCURACY

Snedecor and Cochran's formula is given as:

$$\hat{P}+/- = 1.96(P*Q/N)^{0.5} + 50/N$$

where:

$\hat{P}+/-$ is the percentage range of accuracies at 95% confidence.

P is normalised accuracy (%)

Q is given by 100-P

N is the number of samples

For the decision tree classifier (accuracy 58.23%):

$$\hat{P}+/- = 1.96*(58.23*41.77/8084)^{0.5} + 50/8084$$

$$= 1.08 \text{ (to 2 dec. pl.)}$$

Therefore, at the 95% confidence interval, the accuracy lies between 57.15 and 59.31%.

9.7.2 95% CONFIDENCE LIMITS ON HIGHEST CONTEXTUALLY ENHANCED ACCURACY

For the maximum likelihood classifier and Markov relaxation, with transition probabilities estimated from field data (accuracy 96.39%):

$$\hat{P}+/- = 1.96*(96.39*3.61/8084)^{0.5} + 50/8084$$

$$= 0.09 \text{ (to 2 dec. pl.)}$$

Therefore, at the 95% confidence interval, the accuracy lies between 96.30 and 96.48%.

Table 9.1: Summary - greatest reductions in error

classifier	accuracy %	algorithm	accuracy %	reduction in error %
maximum likelihood	92.22	Markov relaxation	96.39	53.60
deviant distance	87.06	small area replacement	92.55	42.43
minimum distance	80.07	Markov relaxation	90.50	52.33
box	66.12	small area replacement	91.66	75.38
decision tree	58.23	Markov relaxation	82.47	58.03
look up table	59.81	Markov relaxation	60.98	0.03

9.8 RESULTS FOR OTHER AREAS

In order to further test the more successful algorithms listed above, they were applied to three other data sets. The results for these data sets are summarised in table 9.2 below and full results are presented in appendix A.

Table 9.2 Summary results for other areas.

9.2a Salisbury data set

classifier and enhancement	accuracy	% decrease in error due to enhancement
box	38.69	-
9 by 9 mode	36.96	(-2.82)
9 by 9 CONAN	37.95	(-1.21)
SAR (8,5)	37.35	(-2.20)
minimum distance	93.99	-
9 by 9 mode	99.51	91.85
9 by 9 CONAN	99.68	94.68
SAR(8,5)	99.89	98.17
maximum likelihood	96.94	-
9 by 9 mode	99.40	80.39
9 by 9 CONAN	99.58	86.27
Markov relaxation	79.90	(-556.86)

9.2b North Wales data set

classifier	accuracy (%)	best contextual enhancement	accuracy (%)	reduction in error (%)
maximum likelihood	95.27	9*9 CONAN	98.27	63.42
minimum distance	95.45	9*9 CONAN	97.32	41.10
deviant distance	94.04	-	-	-

9.2c Yemen data set

classifier	accuracy (%)	best contextual enhancement	accuracy (%)	reduction in error (%)
maximum likelihood	54.38	SAR (8,1)	70.06	34.37
minimum distance	37.54	9*9 mode	45.13	12.15
deviant distance	42.21	-	-	-

CHAPTER 10

DISCUSSION

10.1 CLASSIFICATION

10.1.1 ACCURACY

Of the algorithms investigated, the maximum likelihood
algorithm is undoubtedly capable of producing the most
accurate results. This suggests that the assumptions
necessary in the application of Bayesian and Gaussian
probability theory (and particularly that of normally
distributed deviations of reflectances about the class
mean reflectance) are, at least approximately, correct
when applied to the samples of remotely sensed data
used.

The potential problems of multi-modal classes are
outweighed by the algorithm's ability to accurately
determine the classes of marginal pixels.

Second in terms of accuracy came the two classifiers based
on the minimum distance rule.

The drawback of the minimum distance algorithm is that it
is unable to take into account differences in the 'spread'
of class distributions - it assumes that a pixel at
feature space distance d from two or more class means is
equally likely to belong to each class, whereas the class
distributions may be such that the true class membership
likelihoods at this distance are very different.
Attempting to enable such comparisons by dividing

185

distances by their associated class standard deviations in each band before calculating the final feature space distances (the 'deviant distance' algorithm) does not truly enable modeling of class probability density functions, as does the maximum likelihood algorithm. The results presented here, however, do show a significant improvement over the basic minimum distance algorithm. The results of such classifications (minimum distance and deviant distance) would certainly be of use for quick

assessments of the effectiveness of training areas before more rigorous (and time consuming) classification using the maximum likelihood rule.

The box classifier is probably the most commonly available classification algorithm on image processing systems. The results for this algorithm show it to be reasonably effective, producing accurate classifications when the box boundaries are set to produce a small box (+/-1 or 2 standard deviations from the class means). This accuracy is at the expense of high levels of unclassified pixels, lying outside the boxes in feature space. As will be shown later, this tendency to leave pixels unclassified instead of assigning them to incorrect classes is a virtue when contextual post-processing is to be carried out. It is therefore concluded that the box classifier alone is unlikely to produce classifications of sufficient accuracy to merit its use, but may be useful as either a 'building block' for subsequent contextual processing or as a 'quick

look' facility to assess the potential effectiveness of a maximum likelihood classification using the same training areas.

The decision tree classifier has been put forward as the answer to the problems of processing large, multi-temporal data sets efficiently (Belward and DeHoyos, 1987). The algorithm also has a great deal in common with expert systems. It does, however, require much time to be spent determining suitable decision boundaries before the algorithm can be implemented, and the accuracies achieved here using single-date imagery are not really high enough to merit its use. In the UK, Legg (1988) has shown that the possibility of obtaining multitemporal imagery for crop classification and monitoring, to which the properties of the decision tree classifier are ideally suited, is small. Unless expert systems technology is used to determine tree structure, avoiding time and effort on behalf of the operator, the decision tree classifier cannot be recommended for use with remotely sensed data.

It is difficult to present a definitive conclusion for the look up table classifier. On one hand, to enable its use currently on multidimensional data sets, so many compromises must be made in order to be able to store and search the table (discarding bands, coarsening radiometric resolution) that any benefits from the algorithm's lack of parametric assumptions about the data are hidden. On the other hand, when an effective solution is encountered (the

187

overall brightness image) the classifier can be most effective (75% accuracy was achieved using this image). The other problem with this algorithm is that it 'learns' from the training data: if a table entry does not occur in this data, then any similar pixels in the image cannot be classified without resort to another algorithm. The more table entries possible, the more training data will be required to cover these entries.

To some extent, the appropriateness of a classification algorithm depends on the task in hand. For land-cover mapping, the accuracy must be as high as possible, for all classes. Some applications, however, may only require accurate mapping of one class type, which may easily be differentiated by one of the faster algorithms. In this case all the algorithms were able to accurately classify the water areas: for a task such as reservoir monitoring, it may be cheaper to divide the image only into water and non-water classes, using a fast algorithm.

10.1.2 TIMING

There is a strong correlation between the accuracy produced by a classifier and the time it takes to produce the classification. The most accurate classifier, the maximum likelihood algorithm, is also the slowest. The relationship is not linear: in order to gain an extra few percent accuracy a disproportionately long time must be spent on image processing. Thus, any study using classified remotely sensed data must be examined from the

188

standpoint of the desirable classification accuracy, weighed against budgetary implications. For 'one off' classifications, where obtaining, correcting and other processing of imagery may be time consuming, the additional processing time required by the maximum likelihood algorithm may only represent a small proportion of the total time to be spent on the project. The use of the maximum likelihood algorithm can therefore be justified more easily than for routine classification of, say, meteorological satellite data. In this case regular commitment of computer resources to time consuming classification algorithms may be undesirable, and slightly poorer classification accuracies can be accepted.

The look up table classifier is particularly susceptible to variations in the information content of the input data. For the spectral shape image, accuracies were relatively poor (59%), whilst for the overall brightness image the accuracy was increased to 73%. Clearly trouble must be taken over feature selection to optimise the results from this algorithm, which has the great advantage over the other algorithms studied of being nonparametric. A listing of the look up table mapping for the overall brightness image might provide useful clues for multimodal classes, and subsequent splitting of training areas.

10.2 CONTEXTUAL ENHANCEMENT

The advantages of subsequent enhancement of classified imagery can be split into two sections: the improvements

189

in classification accuracy to be gained and the useful generalisation to produce more 'map like' end products.

The latter is difficult to quantify, so comments will be restricted to subjective impressions from visual analysis of 'before and after' images.

10.2.1 GENERALISATION

A classified image, no matter how accurate the algorithm used to produce it, never looks like a map. Boundaries are ill-defined and the image has a 'speckled' appearance due to mis-classified pixels within areas which a map would show as homogeneous.

Simple 'contextual smoothing' can be achieved by application of a mode filter. In this study, sizes of patch from 7 by 7 to 11 by 11 pixels were found to be most effective, and initial use of a 9 by 9 patch is recommended.

10.2.2 ACCURACY IMPROVEMENT

The results show that contextual post-processing algorithms can significantly improve the accuracies of classified images. Not all algorithms were successful in this respect, and again there would seem to be a disproportionate increase in processor time needed to produce slightly greater enhancements than those produced by simpler algorithms.

Another important feature of these algorithms is their need for good classifications in the first place. Most rely on the assumption that the classified pixels used to contextually reclassify the image have been correctly classified. For this reason classifiers such as the tightly constrained box algorithm, where 'difficult' pixels are often left unclassified often enable larger increases in accuracy to be obtained. The opposite often occurs in the cases of algorithms which assign all pixels to a class on the basis of nearest (minimum distance) or most likely (maximum likelihood) class. In these cases a significant number of incorrectly classified pixels is often used in the reclassification process, leading to further erroneous classification.

The following section deals more specifically with the results obtained from the individual algorithms, or categories of algorithms where they can conveniently be grouped.

10.3 CONTEXTUAL ALGORITHMS

10.3.1 MODE FILTERS

With the exception of their application to the spectral shape and look up table classified image the application of mode filters consistently raised the accuracy of the classified images, consuming relatively low amounts of CPU time in the process. The choice of patch size has an effect on the improvements obtained: a gradual increase in improvement is obtained with increasing patch size, until

a peak is reached, after which accuracy drops again slowly. In this study, this peak patch size was found to lie between 7 by 7 and 11 by 11 pixels. Since the algorithm requires relatively little time to run it is suggested that a suitable strategy is to start at a 7 by 7 pixel filter and incrementally increase the size of window by 2 (the smallest increment possible) until a slight decrease in accuracy is obtained. The image produced by the previous filter should then be used.

10.3.2 SMALL AREA REPLACEMENT

The thresholded declassification algorithms and associated reclassifiers are susceptible to the quality of their input (classified imagery). This quality is not necessarily reflected in the classification accuracy. In this case it is the box classified image which benefits most from the application of the technique (91% accuracy achieved with areas of up to 8 pixels declassified and subsequent nearest neighbour reclassification).

Of the three reclassification algorithms investigated, the use of a 9 by 9 pixel selective mode filter unsurprisingly produced poor results where large unclassified areas were present, owing to its inability to find classified pixels in these areas. Surprisingly there seems to be little difference between the fast nearest along line neighbour search, and the true ordered search nearest neighbour algorithms, and taking the mode of the nearest 5

neighbours has a barely perceptible effect. Theoretically a true outward search for neighbours must be more robust, however the results presented here suggest that it may be possible to save time by using the along-line search method.

Generally the improvements in accuracy increase with increasing threshold - the larger the areas declassified, the better the reclassified accuracy. The algorithm presented here is incapable of being extended beyond 9 pixel areas, since it uses a 3 by 3 pixel window. True small area identifiers require the use of extremely time consuming 'structured walk' algorithms (Oldfield, 1988) and are not investigated here.

10.3.3 THRESHOLDED HIGH PASS FILTERS

10.3.3.1 ON UNCLASSIFIED IMAGERY

The declassification of imagery based on noise and edges detected by high pass filters on the raw imagery produces similar results to the small area replacement algorithms discussed above. There seems little to choose between the two filters used here (Roberts and Prewitt), despite their differing window sizes.

Increasing the number of declassified pixels by including all pixels within a threshold radius (2 pixels in this case) from the edge or noise pixels in the declassification scheme has an adverse effect on classification accuracy. This consistently reduced the

193

final accuracy figures.

The time required to run the algorithms is similar to the small area replacement algorithms discussed earlier.

No experimentation was carried out with threshold levels: the automatic algorithm was used throughout. It was felt that additional time spent experimenting with different threshold levels would be impractical in operational situations.

10.3.3.2 ON CLASSIFIED IMAGERY

Applying the same techniques to classified imagery produced similar (although slightly higher) accuracies to those above. Experience from the previous experiments led to not trying either different filters or distance thresholding in this case.

Again it was with the box classified imagery that the greatest improvements in accuracy were achieved, implying that this algorithm is able to accurately classify all but the most 'difficult' of pixels, which can subsequently be classified by contextual techniques.

10.3.4 MARKOV RELAXATION

The two methods of estimating the class transition probability matrix: from field data and from the image were both investigated. By far the most successful implementation of the algorithm results from use of the field data estimated transition probability matrix. In the

case of the maximum likelihood classified image, use of a transition probability matrix estimated from the classified image itself actually resulted in a 17% drop in accuracy. This aside, and with the exception of the poorly performing spectral shape and look up table classifier, the algorithm produced higher accuracies than those for the classifiers alone with both estimates of transition probability matrix.

Alone among the contextual algorithms mentioned so far, the Markov relaxation technique was able to better the performance of a 9 by 9 modal filter on the maximum likelihood image. This was achieved, however, at the expense of a great deal of extra processing time.

Once again, the results for the box classified image show a great deal of improvement. For this reason it was decided to try an iterative implementation of the algorithm on this image.

For the transition probability matrix estimated from the image, a slight improvement in accuracy (less than 0.5%) was achieved on the second iteration, followed by a slight drop in accuracy (again less than 0.5%) on the third iteration, however, by the second iteration there were no pixels in the verification area classified as class 5 (broad leaf woodland), and class 6 (bracken) had similarly disappeared on the third iteration. The consistency of the accuracy figures shows that the classification of some classes was being improved upon considerably, at the

expense of other classes.

When the technique was applied using the transition probability matrix estimated from the field and map data the results were similar (0.5% accuracy increase, with the loss of class 5).

Because of the effectiveness of the technique, it is felt that the algorithm is useful despite its slowness. In cases where the maximum accuracy is desired from a classification it would seem to be appropriate to choose a maximum likelihood classifier, followed by Markov relaxation, with transition probabilities estimated from field data.

Despite the accuracy achieved with the box classified image, it is still possible to achieve similar, or higher, accuracies in less time, using a maximum likelihood classifier, followed by a mode filter.

10.3.5 WHARTON'S CONAN ALGORITHM

This algorithm, borrowing from textural processing, presents several differences to those algorithms discussed above:

1. It can be implemented relatively quickly.
2. It does not, in theory, rely on the accuracy of the previous classification.
3. It does not deal with adjacency or proximity functions for reclassification.

Point 2 requires further explanation: the algorithm uses the training areas used by the preceding classifier to estimate the relative proportions (numbers of pixels in an n by n pixel patch) of each class occurring for each true class over an n by n pixel patch. Thus it is theoretically possible that a classified image where all pixels of one class have been misclassified as other classes in constant proportion can be reclassified: the reclassification being such that all n by n pixel patches containing this proportion of the two classes will have their central pixels assigned to the correct class. This theory is borne out to some extent in practice. The decision tree classified image's accuracy was increased by 22% when the algorithm was applied over a 15 by 15 pixel patch (58.2% before to 80.6% after), however the classification accuracy of the spectral shape and look up table classified image was decreased by application of the algorithm.

The algorithm was tested using 9 by 9 and 15 by 15 pixel patches, results seeming to indicate that poorer initial classification accuracies could be improved more by use of the larger patch size, whilst the better classifications responded more favourably with the 9 by 9 pixel patch. For the maximum likelihood classified image the results obtained were similar to the best achieved using mode filters, in approximately the same times. Choice between the two algorithms here would have to depend on such 'unmeasurables' as operator preference and the 'look' of

the end product. In many cases this choice may be academic since commercial digital image processing systems are unlikely to have implementations of the CONAN algorithm available.

With the exception of the box classified image and the look up table and spectral shape image, the algorithm performed similarly to the Markov relaxation algorithm, accuracies being slightly lower (of the order of 1-2%), which, given the considerable time savings available from the use of CONAN, may be acceptable in some circumstances.

10.4 RELATIVE CONTRIBUTION OF NOISE AND EDGES TO CLASSIFICATION ERRORS

This experiment consistently identified the pixels classed as 'noise' to have the most detrimental effect on classification, greater accuracies being achieved by reclassification of these pixels than either reclassification of 'edge' pixels or of both edge and noise pixels. This suggests that many of the edge pixels identified were classified correctly and were important during reclassification to act as correctly classified neighbours.

Only the box and decision tree classified images benefited from the experiment, the rest of the images had their accuracies reduced. This method of edge and noise detection cannot, therefore, be recommended, despite its theoretical elegance.

10.5 NON-PROBABILISTIC RELAXATION

The two algorithms investigated, based on the minimum distance classifier, produced similar results to their parent algorithm, but in a considerably longer time. The accuracies were about 1% higher than those for the minimum distance classifier alone (80.1% [minimum distance alone], 81.3% and 81.4% [non-probabilistic relaxation]).

There seems little difference between using ranked classes compared to actual distances, with the ranked class version coming out slightly more accurate (81.4% compared to 81.3%).

10.6 SIMULATION OF PER-FIELD CLASSIFIER

Intuitively, the usefulness of a classifier able to assign whole fields to single classes would seem high, particularly in agricultural areas. Definition of 'fields', by complex algorithms, from the image data is a non-trivial task, and a great deal of effort has been expended towards this goal. In the UK some of the necessary field boundary information has been readily available for many years: the boundaries are shown on Ordnance Survey maps of 1:25,000 scale, although, as noted in section 7.3.1 the utility of this information may not be as high as would at first appear to be the case.

This study has concentrated on the potential benefits to be obtained from per-field classifiers, assuming that a method of field extraction will shortly be

available.

The results of this study are, unfortunately, inconclusive. Whilst the accuracies for three of the classifiers increased to 100% (deviant distance, minimum distance and box), the accuracies of the other classifiers dropped considerably (by 9% for the maximum likelihood and decision tree algorithms, and 50% for the spectral shape and look up table classified image).

In view of the 'quick' nature of this experiment, and the encouraging improvements to some of the classifiers, it is felt that a comprehensive study of per-field classifiers should be undertaken, the scope of which is beyond this study. Some suggestions for this are included in the 'future study' section.

10.7 CLUSTERING

Unsupervised techniques for image classification suffer from the drawback of only being able to extract 'spectral classes' from the imagery, which may not correspond to the 'informational classes' required on the final map. Assigning the spectral classes to informational classes seems unlikely to provide a classification of great accuracy, as the results show.

10.7.1 ITERATIVE NEAREST NEIGHBOUR CLUSTERING

The iterative clustering algorithm provided images of steadily decreasing accuracy with each iteration, when the image classes were mapped to corresponding informational classes. These drops in accuracy were linked to the merging of small, but important, classes (such as water) with their nearest large spectral classes (in this case, water pixels gradually became associated with coniferous woodland). It would thus seem that the biggest failing of the algorithm is its inability to cope with small, or spectrally close classes.

This algorithm is used by ISODATA package, although this was not used here.

10.7.2 HISTOGRAM CLUSTERING

The accuracy produced (after spectral class to informational class mapping) of 77.25% by this algorithm was encouraging. The lack of any class 5 (broad-leaf woodland) again highlights the inability of the clustering algorithms to identify small or spectrally indistinct classes, which a supervised algorithm can, to some extent, be forced to identify.

The relatively fast speed of this algorithm, however, enables its use as a preliminary exploratory measure for use before attempting classification of imagery using time-consuming supervised techniques such as the maximum likelihood algorithm.

10.8 CONTEXTUAL ENHANCEMENT AND QUANTITATIVE ANALYSIS

Despite the increases in calculated accuracy obtained by the use of contextual enhancement routines, images produced by this method are not recommended for further quantitative analysis for the following reasons:

1. The generalisation (ie improvement) is largely subjective and visual. The original classification represents the most accurate image from which to obtain areal estimates.

2. The more processing carried out on the data, the higher the risk of compounding errors produced by earlier processing stages.

3. Accuracy increases calculated may be due to definition of block training areas. The contextual algorithms tend to block up classifications, therefore the calculated accuracies increase.

10.9 SUITABILITY OF ALGORITHMS FOR USE WITH OTHER DATA

The importance of integrating other forms of digital spatial data with remotely sensed data is increasing. Because of the assumptions made about remotely sensed data by many classification algorithms these algorithms may be unsuited to use on such augmented data sets. This section sets out some of the considerations which must be taken into account when using such data sets.

10.9.1 CLASSIFICATION

Additional data sets present three barriers to use of parametric classifiers:

1. There may be incomplete coverage by these data sets
2. Data values may be labels or categories rather than points on an interval scale
3. Resolution may be coarser than the remotely sensed data.

In the case of incomplete data coverage a classifier must be capable of making decisions based on the available data: conventional parametric algorithms (maximum likelihood, minimum distance, box) are unable to function in such cases.

The presence of label data renders the concepts of feature space meaningless, and again conventional parametric classifiers cannot be used.

Coarse resolution of additional data sets presents a more subtle problem: the data is useful, but the poor resolution reduces its locational accuracy. In cases such as this the data would ideally be used to provide background information whilst placing greatest reliance on high resolution data.

These three points suggest that a change of thinking is required if additional data is to be successfully integrated with remotely sensed imagery. As has been seen earlier, the use of expert systems can address these problems: additional decision rules can be developed for

cases where some data is missing, and label data can be accommodated as 'logical' variables.

10.10 APPLICABILITY OF RESULTS TO DIFFERENT AREAS

This study has been limited in its scope in terms of geographical area, limiting the types and patterns of land cover to those occurring within the study area. The following notes briefly cover the use of the algorithms and techniques developed during this study in general terms. It is felt that the overall pattern of results is broadly representative of the techniques and algorithms in general, although individual results should be viewed with caution and in the context of the study as a whole.

10.10.1 SUPERVISED CLASSIFICATION

The effectiveness of any spectrally based classification algorithm is limited by the spectral separability of the classes which the algorithm has to identify. In addition to this, sensor characteristics, particularly spatial and radiometric resolution, also affect the quality of the final classification. If classes are not spectrally distinct, given the limitations of the data in use, then a per-pixel classifier cannot hope to produce accurate results.

10.10.2 CONTEXTUAL ENHANCEMENT

As mentioned above, sensor characteristics often preclude accurate identification of classes at a per-pixel level. Mostly this results in classifications where most of the image pixels have been correctly assigned, but whose accuracies are lower than is desirable. Contextual enhancement of such classifications represents a useful tool for identification and reclassification if the incorrectly assigned pixels.

10.11 TIMING IN THE CONTEXT OF A PROJECT

Figure 10.1 graphically illustrates the diminishing benefits obtained with increased time spent on classification. This does not tell the whole story, however, since the nature of any remote sensing project will affect the practicality of applying certain classifiers. The following notes attempt to clarify this statement.

Fig.10.1 Normalised accuracy vs. processing time

205
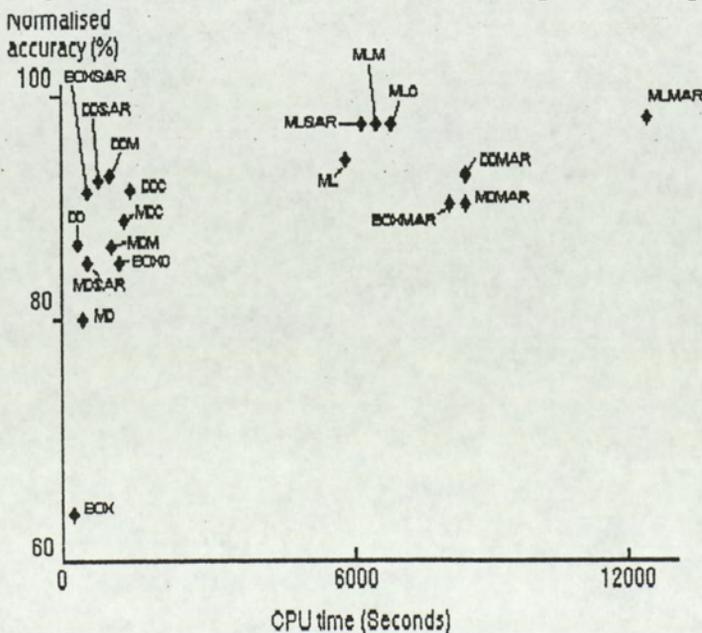
Notes:
```
ML    maximum likelihood classifier
DD    deviant distance classifier
MD    minimum distance classifier
BOX   box classifier
MAR   Markov relaxation
C     Wharton's CONAN algorithm
SAR   small area replacement
M     mode filter
```

10.11.1 DATA ACQUISITION

The time spent in acquiring the data in many remote
sensing projects can be considerable. Field surveys can
take weeks or months, and even the acquisition of the
remotely sensed data from its suppliers may take some
time. In this context, a full day's processing to produce
a Markov enhanced, maximum likelihood classification may
become insignificant. On the other hand, if data is
immediately or routinely available (for example,
meteorological satellite data), such lengthy processing
times may be undesirable.

10.11.2 DATA VOLUME

Another factor affecting the suitability of certain
algorithms is the volume of data which must be processed.
As an example: the United Kingdom is covered by
approximately 40 Landsat Thematic mapper scenes, each of
which contains approximately 25 1024 by 1024 pixel areas
as used for this study. Taking a day for processing each
1024 by 1024 extract (maximum likelihood followed by
Markov relaxation) gives a total processing time in excess
of three years. In cases such as this it is clearly
impractical obtain the highest possible accuracies, and

some compromise between speed and accuracy must be reached.

10.12 EFFECT OF NOISE ON CLASSIFICATION AND CONTEXTUAL ENHANCEMENT ALGORITHMS

In order to assess the robustness of the recommended classifiers and contextual enhancement algorithms when presented with noisy data, an experiment was performed where random noise was added to the image data and classifier performance monitored.

10.12.1 METHOD

Random numbers were added to each image band's data as follows:

(1) 0 or 1 added, bias towards 0, representing only mildly noisy data.

(2) 0,1,2,3 or 4 added, bias towards 0, representing moderately noisy data.

(3) Numbers between 0 and 10 added, representing very noisy data.

The same training areas that were used in the main study were used to generate the statistics used by the classifiers listed in table 10.1 below. These algorithms were then applied to the altered data.

Small area replacement and Markov relaxation were then applied to the resultant classified imagery, the resultant accuracies are given in table 10.1.

Table 10.1 below shows the normalised accuracies produced by the selected algorithms:

Table 10.1 Accuracies of classifiers and contextual enhancements when applied to noisy data

algorithm	normalised accuracy			
	unaltered data	mildly noisy data	moderately noisy data	very noisy data
box	66.12	65.05	62.36	60.74
plus SAR	91.66	91.12	91.87	87.91
deviant distance	87.06	87.04	78.86	85.02
plus SAR	92.55	92.31	86.11	91.90
max like	92.22	91.61	89.83	86.24
plus Markov	96.39	96.52	96.14	94.52

10.12.2 DISCUSSION

10.12.2.1 CONFIDENCE LIMITS

The 95% confidence intervals for these accuracy figures were calculated in chapter 9 as lying between 1% (accuracies around 60%) and 0.1% (accuracies around 95%). To ensure a reasonable safety factor in assessing these results, changes in accuracy of less than 1% will be considered insignificant. This effectively means that the slight increases in accuracy achieved by some of the algorithms when applied to noisy data can be attributed to chance, rather than genuine increases in classifier performance.

10.12.2.2 TRENDS

The overall trend of the results is for the classification accuracies to decrease with increasing noise. The accuracies of the contextually enhanced classifications, however, decrease at a slower rate than those for their associated per-pixel classifiers. The one anomaly to this trend is the results for the deviant distance classifier, whose accuracies actually increased when applied to the very noisy data. It must be assumed, however, that this increase occurred by co-incidence, rather than because of any special properties of the algorithm.

The Markov relaxation algorithm, when applied to the maximum likelihood classified image consistently boosted the classification accuracy: accuracies dropping by only 2% with increasing noise, compared to over 6% for the maximum likelihood classifier alone.

10.12.3 CONCLUSIONS

The experiments show, for this particular data set, that although all classifiers are affected adversely by noisy data, the contextual enhancement algorithms proposed in this study are less susceptible to noise than existing per-pixel classifiers. The use of such contextual algorithms is therefore recommended where data signal to noise ratios may be poor, and the combination of the maximum likelihood classifier with Markov relaxation would seem to be particularly useful in reducing the adverse effects of noise.

10.13 APPLICATION TO OTHER AREAS

10.13.1 SALISBURY DATA SET

Before commenting on the contextual enhancements, it should be noted that the box classification algorithm used in this case differs from that used in the main study. In this case the algorithm is 'min-max', in other words the decision boundaries are set as the lowest and highest values for each class in each band, obtained from training data. Since means and standard deviations are not calculated this is a very fast version of the box classifier. It does, however, have the drawback that anomalous pixels within training areas can adversely affect classification accuracy. In this case the resultant accuracy was extremely low because: 1) the 'urban' training data contained a wide variety of spectral responses, and, 2) this class was assigned a low label number - the algorithm works by checking for class membership in the numerical order of class labels until a positive result; once this has occurred no further classes are checked.

It should also be noted that the accuracies given were calculated from the areas used to train the classifiers. For this reason, they are rather higher than would be expected (Swain and Davis, 1978).

10.13.1.1 BOX CLASSIFIER

Contextual enhancement of a classification of such low accuracy as this cannot be expected to produce significant improvements in accuracy. The results presented here bear this out, with all the enhancements tried producing slight decreases in accuracy.

10.13.1.2 MINIMUM DISTANCE AND MAXIMUM LIKELIHOOD CLASSIFIERS

The results for these classifications agree with the results presented in the main study. Given a reasonably accurate classification, contextual enhancement can significantly decrease any errors present.

It was felt that not enough was known of the study area to enable estimation of transition probabilities by the author. Markov relaxation was, therefore, limited to estimating these from the classified image itself, producing similar results to those presented in the main study: a drop in accuracy.

10.13.2 NORTH WALES DATA SET

The results for the North Wales data set are broadly similar to those achieved with the Peak District data set. These two areas contain a similar range of ground cover types, although large areas of estuarine land and sea exists for the North Wales data set, which is not present in the Peak District scene.

Image processing for this and the Yemen data set was carried out using the commercially available Erdas software, running on a Sun 386i workstation under Unix, up to the classification stage. Post-classification processing was carried out using the same programs as for the Peak District imagery. These programs were modified to run under Unix using Sun 1.1.1 Fortran (a compiler for Fortran-77, see section 8.8.1 for further details).

10.13.3 YEMEN DATA SET

This was the only non-Thematic Mapper data set used in this study, and the most difficult for which to obtain satisfactory training data - this being derived from photointerpretations.

The contextual results resemble those for the classification algorithms which produced lower accuracies when applied to the Peak District data, in that contextual processing was not found to be of benefit.

The poor classification accuracies are the single most likely factor to have caused the poor performance of the contextual algorithms. The reader should therefore be aware that the application of post processing is unlikely to improve poor classifications (see chapter 11).

Image processing for this data set was carried out as detailed in section 10.13.2.

10.14 COSTS

The cost of implementing the algorithms is directly
related to the computer time to execute them. This must,
however, be seen in the context of the continuing decrease
in cost of CPU time in real terms, brought about by
constant improvement in processor power.

Based on a CPU charge of £15 per hour, costs for the
algorithms used in this study are given in table 10.2,
when applied to the Peak District 1024 by 1024 by 6 band
data set.

Table 10.2 Costs of algorithms

algorithm	cost @ £15 per hour CPU time, and £30 per hour operator time		Total cost (£)
	CPU	operator	
box classifier	0.70	15	15.70
minimum distance	1.20	15	16.20
maximum likelihood	24.00	15	39.00
9 by 9 mode filter	1.50	-	1.50
Small area replacement	0.25	-	0.25
CONAN	2.00	-	2.00
Markov relaxation	19.00	30	49.00

CHAPTER 11

CONCLUSIONS

11.1 INTRODUCTION

The aims of this study were set out in chapter two, and were divided into two sections. Section 11.2 will deal with the investigation of classifiers and contextual enhancement (aim i), whilst section 11.3 will cover the recommendations resulting from the study (aim ii).

11.2 INVESTIGATION

In response to aim i (chapter 2) the following classification methods were investigated:

Box classifier

Minimum distance classifier

Deviant distance classifier

Maximum likelihood classifier

Decision tree classifier (manually defined)

Decision tree classifier (expert system defined)

Look up table classifier (based on overall brightness image)

Look up table classifier (based on spectral shape algorithm)

Histogram clustering

Nearest neighbour clustering

All algorithms were tested on Landsat Thematic Mapper imagery of the Derbyshire Peak District, whilst those

214

algorithms producing highest accuracies were additionally tested on three other data sets: Salisbury and North Wales (Landsat Thematic Mapper) and Yemen (Landsat Multi Spectral Scanner).

Also in response to aim i, the following contextual enhancements were investigated:

Mode filters

Small area replacement

Wharton's CONAN algorithm

Markov relaxation

Non-probabilistic relaxation

Per-segment classification

All algorithms, with the exception of per-segment classification, were tested on the Peak District data set (using the output from the classifiers above). The algorithms achieving the highest accuracies were also tested on the three other data sets.

Comparison of the algorithms above was carried out on the basis of the following criteria:

i) classification accuracies achieved

ii) speed of operation (both computer and operator time)

iii) consistency of results where the algorithms were tested on the additional data sets

From these comparisons the recommendations set out in section 11.3 below have been derived.

11.3 RECOMMENDATIONS

Aim ii of the study was to provide a series of recommended classification algorithms for use with remotely sensed data. These were determined according to: a) user requirements (accuracy levels desired, time available) and b) the nature of the data (dimensionality, spectral homogeneity of classes etc.). These recommendations are set out below:

11.3.1 CLASSIFIERS AND CLUSTERING

These algorithms can be split into two types:

1. Quick and moderately accurate algorithms, useful for rapid exploration of data before more computationally expensive classification.

Recommended algorithms in this category are:

i. Histogram clustering

ii. Box classifier

iii. Look up table classifier using the overall brightness image.

iv. Decision tree classifier, designed by expert system.

2. Slower, more accurate classifiers, for operational classification, where the highest accuracies must be achieved.

Recommended algorithms in this category are:

  i.Maximum likelihood classifier

 ii.Deviant distance classifier

The reasons for not recommending the other algorithms examined are as follows:

11.3.1.1 MINIMUM DISTANCE

The deviant distance algorithm represents a simple and computationally inexpensive improvement over this algorithm.

11.3.1.2 MANUALLY DEFINED DECISION TREE

This algorithm was found to be not particularly accurate and the time penalties incurred in designing the tree were prohibitive.

11.3.1.3 LOOK UP TABLE CLASSIFIER ON SPECTRAL SHAPE IMAGE

This combination of algorithms gave poor accuracy figures, although it could become more useful with more spectral bands.

11.3.1.4 ITERATIVE NEAREST NEIGHBOUR CLUSTERING

This algorithm was found to be slow, offering no significant benefits over histogram clustering.

11.3.2 CONTEXTUAL ENHANCEMENTS

The algorithms listed below were found to be most effective at increasing the accuracies of classified images:

1.Markov relaxation, with transition probabilities estimated from field data.

2.Small area replacement, ideally with an area threshold of 8 pixels for declassification.

3.Wharton's CONAN algorithm.

4.Mode filters.

The algorithms respond in slightly different ways to different data. For the best results the following conditions for use should be observed if possible:

11.3.2.1 MARKOV RELAXATION

This algorithm should be used when time is unimportant. It was found to be capable of achieving maximum accuracy with the maximum likelihood algorithm. The algorithm may not be available to an operator, precluding its use.

11.3.2.2 SMALL AREA REPLACEMENT

This algorithm should be used when time limits are important, but high accuracy levels must be achieved. The larger the size of threshold area for declassification, the slower the algorithm.

11.3.2.3 WHARTON'S CONAN ALGORITHM

This is a good algorithm to use when time is short, but it may not be available to an operator.

11.3.2.4 MODE FILTERS

These are the most likely algorithm to be available on a commercial image processing system at present. They are recommended for improving the presentation of the output image, particularly when hard copy is to be made.

11.3.3 RECOMMENDED CLASSIFIER AND CONTEXTUAL ENHANCEMENT PAIRINGS

The following pairings of classifier and contextual enhancement algorithm were found to produce the highest accuracies in this study (table 11.1):

Table 11.1: Pairings of classifier and enhancement algorithm producing highest accuracies

classifier	contextual enhancement
maximum likelihood	Markov relaxation
deviant distance	small area replacement
minimum distance	Markov relaxation
box	small area replacement
decision tree	Markov relaxation
look up table	Markov relaxation

Note: these are results from the specific image and training data used here. Their use as a general indication of suitable algorithm pairings is recommended only with caution. Operators should experiment and use their own experience to determine pairings suitable for use in their own sphere.

219

11.3.4 PAIRINGS NOT RECOMMENDED

The following contextual enhancement algorithms are not recommended:

1.Markov relaxation, with transition probabilities estimated from the classified imagery.

This is susceptible to the noise that the algorithm is attempting to remove.

2.Thresholded edge filters, followed by reclassification of pixels identified as edges.

This is a circuitous and insufficiently effective algorithm. The theory is convincing, but in practice the performance is poor.

3.Non-probabilistic relaxation.

These algorithms performed quite well, however the algorithms recommended above managed to perform equally well in less time.

11.3.5 PER-FIELD CLASSIFIERS

These are recommended for areas where boundary data is readily available. They represent a promising area for future research.

11.3.6 CLASSIFICATION ACCURACY AND COMPUTATIONAL EXPENSE

11.3.6.1 SUMMARY

Table 11.2 below sumarises the recommended classification procedures, determined from this study, for different routine applications of remotely sensed data.

Table 11.2: Recommended classification procedures

application	recommendations
quick look/'exploration' (no ground data)	histogram clustering
as above, but with ground data	box classifier, or decision tree classifier, designed by expert system
rapid land-cover map production	box classifier, followed by small area replacement or a mode filter
high-quality map production	maximum likelihood classifier, followed by Markov relaxation, with transition probabilities estimated in the field
production of classified images for statistical analysis	maximum likelihood alone (contextual enhancement is a form of generalisation reducing the accuracy of aerial estimates etc.)

11.3.7 FUTURE STUDY

Two important areas for future study in the field of classification and enhancement are:

1.The development of techniques integrating remote sensing with geographical information systems.

2.An assessment of the potential of expert systems for classification of remotely sensed data.

Geographical information systems represent a rapidly developing area of spatial information processing. Much data, complimentary to remotely sensed material, is, or will shortly become, available in geographical information system-compatible format (for example: Ordnance Survey maps; the 1991 UK census; the Soil Survey of England and Wales' land information system [LANDIS]). This information, in combination with spectral reflectance data provided by remote sensing, will provide a most powerful tool for terrestrial monitoring in the near future.

With the increase in available information, mentioned above, will come a need for faster, more efficient data processing. Expert systems, able to infer decision rules from example data, are a possible answer to these needs. Such tools would enable fast identification of relevant attributes, followed by rule generation to distinguish different classes. This would provide accurate classifications using the minimum of data required, something a human operator would be unable to match with any degree of consistency.

Expert systems with inference engines capable of handling the large, low-precision data sets provided by remote sensing are a prerequisite for any further work in this field.

11.3.7.1 PER-FIELD CLASSIFIERS

The area of per-field classification, with the advent of geographical information systems and digital map data, is an important one for future study. Topics which should be addressed are:

1. Derivation and extraction of 'fields' from geographical information and digital map data.

2. Investigation of appropriate per-field classifiers; both in terms of the classification algorithms themselves and the algorithms used to average these classifications over individual fields.

3. Investigation of the parts of each field to use for classification: should boundary pixels be avoided; should anomalous areas be removed?

11.4 RELATIVE CONTRIBUTIONS OF NOISE AND EDGES TO CLASSIFICATION ERRORS

The experiments carried out suggest that noise is a more critical factor in impairing the performance of classifiers than pixels with mixed spectral response occurring at edges (class boundaries).

REFERENCE LIST

Ahern,F.J. Brown,R.J. Cihlar,J. Gauthier,R. Murphy,J. Neville,R.A. and Tiellet,P.M. (1987) 'Radiometric Correction of Visible and Infrared Remote Sensing Data at the Canada Centre for Remote Sensing' International Journal of Remote Sensing, Vo.8 No.9 p1349-1376.

Allan,J.A. (1986) 'How Few Data do we Need: Some Radical Thoughts on Renewable Natural Resources Surveys' proc. Symposium on Remote Sensing for Resources Development and Environmental Management, Enschede, Netherlands, August, p901-906.

Anderberg,M.R. (1973) 'Cluster Analysis for Applications' Academic Press, New York.

Aronoff,S. (1982) 'Classification Accuracy: A User Approach' Photogrammetric Engineering and Remote Sensing, Vol.48, No.8, p1299-1307.

Aronoff,S. (1985) 'The Minimum Accuracy Value as an Index of Classification Accuracy' Photogrammetric Engineering and Remote Sensing, Vol.51, No.1, p99-111.

Atkinson,P, Cushine,J.P and Townshend,J.R.G (1985) 'Improving Thematic Mapper Land Cover Classification Using Filtered Data' International Journal of Remote Sensing, Vol.6 No.6 p955-961.

Baines,L.M. (1988) 'The Application of Remote Sensing to the Management of Urban Wildlife habitats' PhD thesis, Aston University.

Barnsley,M.J. Saddler,G.J. and Shepard,J.W. (1989) 'Integrating Remotely Sensed Images and digital Map Data in the Context of Urban Planning' in: Remote Sensing For Operational Applications, technical contents of the 15th. annual conference of the Remote Sensing Society, university of Bristol, 13-15 September. p25-32.

Barrett,E.C. and Curtis,L.F. (1982) 'Introduction to Environmental Remote Sensing' Chapman and Hall, London.

Batista,G.T. Hixson,M.M. and Bauer,M.E. (1985) 'Landsat MSS Crop Classification as a Function of Scene Characteristics' International Journal of Remote Sensing, Vol.6 No.9 p1521-1533.

Belward,A.S. and deHoyos,A. (1987) 'A Comparison of Maximum Likelihood and Decision Tree Classification for Crop Cover Estimation from Multitemporal LANDSAT MSS Data' International Journal of Remote Sensing, Vol.8, No.2, p229-235.

Birnie,R.V. (1986) 'Pixel-Mixing Effects and their Significance to Identifying Snow Condition from Landsat MSS data' International Journal of Remote Sensing, Vol.7 No.7 p845-853.

Booth,D.J. (1987) 'Use of Landsat Thematic Mapper Multitemporal Imagery to Map Bracken Encroachment in the Kinder Scout Area of the Derbyshire Peak District' MSc Thesis, Silsoe College, Silsoe, Bedford. MK45 4DT.

Booth,D.J. and Oldfield,R.B. (1989) 'A Comparison of Classification Algorithms in Terms of Speed and Accuracy After the Application of a Post-Classification Modal Filter' International Journal of Remote Sensing, Vol.10, No.7. p1271-1276.

Booth,D.J. Chidley,T.R.E. and Collins,W.G. (1989) 'Integration of Context Classifiers with GIS' proc. IGARSS'89, Vancouver, Canada, in press.

Borgeson,W.T. batson,R.M. and Kieffer,H.H. (1985) 'Geometric Accuracy of Landsat-4 and Landsat-5 Thematic Mapper Images' Photogrammetric Engineering and Remote Sensing, Vol.51 No.12 p1893-1898.

Botkin,D.B. Estes,J.E. MacDonald,R.M. and Wilson,M.V. (1984) 'Studying the Earth's Vegetation from Space' Bioscience, Vol.34, No.8, p508-514.

Bradbury,P.A. and Macdonald,A. (1986) 'Accuracy Assessment of Land-Cover Classifications' in: proceedings of a remote sensing workshop: Ground Truth for Remote Sensing, Nottingham University, 17th April. p126-132.

Bryant,J. (1989) 'A Fast Classifier for Image Data' Pattern Recognition, Vol.22, No.1, p45-48.

Bryant,J. (1990) 'AMOEBA Clustering Revisited' Photogrammetric Engineering and Remote Sensing Vol.56 No.1 p41-47.

Bryukhanov,B.N. (1985) 'Some Key Parameters of Space Imagery: A Look at Geologic Information Content' Soviet Journal of Remote Sensing Vol.3(4) p576-585.

Buttner,G. Hajos,T and Korandi,M. (1989) 'Improvements to the Effectiveness of Supervised Training Procedures' International Journal of Remote Sensing Vol.10 No.6 p1005-1013.

Canas,A.A.D. and Barnett,M.E. (1985) 'The Generation and Interpretation of False-Colour Composite Principal Component Images' International Journal of Remote Sensing Vol.6 No.6 p867-881.

Cannon,R.L. Dave,J.V. Bezdek,J.C. and Trivedi,M.M. (1986) 'Segmentation of a thematic Mapper Image using the Fuzzy C-Means Clustering Algorithm' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-24 No.3 p400-408.

Card,D.H. (1982) 'Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy' Photogrammetric Engineering and Remote Sensing, Vol.48, No.3, p431-439.

Chiou,W.C. (1984) 'Simple Descriptor for Contextual Classification of Hyperdimensional Remotely Sensed Spectral Data' Applied Optics, Vol.23, No.3, p465-468.

Chrisman,N.R. (1981) 'Beyond Accuracy Assessment: Correction of Misclassification' ASP-ACSM Convention, Technical Papers, 47th Meeting.

Colvocoresses,A.P. (1986) 'Image Mapping With the Thematic Mapper' Photogrammetric Engineering and Remote Sensing, Vol.52 No.9 p1499-1506.

Congalton,R.G. and Mead,R.A. (1983) 'A Quantitative Method to Test for Consistency and Correctness in Photointerpretation' Photogrammetric Engineering and Remote Sensing, Vol.49, No.1, p69-74.

Congalton,R.G. (1988a) 'Using Spatial Autocorrelation Analysis to Explore the Errors in Maps Generated From Remotely Sensed Data' Photogrammetric Engineering and Remote Sensing, Vol.54 No.5 p587-592.

Congalton,R.G. (1988b) 'A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data' Photogrammetric Engineering and Remote Sensing Vol.54 No.5 p593-600.

Cross,A.M. Mason,D.C. and Dury,S.J. (1988) 'Segmentation of Remotely-Sensed Images by a Split-and-Merge Process' International Journal of Remote Sensing, Vol.9, No.8, p1329-1345.

Csillag,F. (1986) 'Comparison of Some Classification Methods on a Test Site (Kiskore, Hungary): Separability as a Measure of Accuracy' International Journal of Remote Sensing, Vol.7 No.12 p1705-1714.

Curran,P.J. (1985) 'Principles of Remote Sensing' Longman, London.

Curran,P.J. and Williamson,H.D. (1985) 'The Accuracy of Ground Data used in Remote-Sensing Investigations' International Journal of Remote Sensing, Vol.6 No.10 p1637-1651.

Davis,J.C. (1973) 'Statistics and Data Analysis in Geology' 1st. ed. Wiley, Chichester.

Davison,G.J. (1986) 'Ground Control Pointing and Geometric Transformation of Satellite Imagery' International Journal of Remote Sensing, Vol.7, No.1, p65-74.

DiZenzo,S. (1983) 'Advances in Image Segmentation' Image and Vision Computing, Vol.1, No.4, p196-210.

DiZenzo,S. Bernstein,R. Degloria,S.D. and Kolsky,H.G. (1987a) 'Gaussian Maximum Likelihood and Contextual Algorithms for Multicrop Classification' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.6

DiZenzo,S. Bernstein,R. Degloria,S.D. and Kolsky,H.G. (1987b) 'Gaussian Maximum Likelihood and Contextual Algorithms for Multicrop Classification Experiments Using Thematic Mapper and Multispectral Scanner Sensor Data' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.6.

Drayton,R.S. Brown,C.M. and Martin,D. (1989) 'Contextual Classification of Urban Areas Using Population Data' in: Remote Sensing For Operational Applications, technical contents of the 15th. annual conference of the Remote Sensing Society, university of Bristol, 13-15 September. p101-105

Drury,S.A. (1987) 'Image Interpretation in Geology' Allen and Unwin, London, 1987.

Duda,R.O. and Hart,P.E. (1973) 'Pattern Classification and Scene Analysis' Wiley Iterscience, New York.

Duggin,M.J. (1985) 'Factors Affecting the Discrimination and Quantification of Terrestrial Features Using Remotely Sensed Radiance' International Journal of Remote Sensing, Vol.6 No.1 p3-27.

Duggin,M.J. Rowntree,R.A. and Odell,A.W. (1988) 'The Application of Spatial Filtering Methods to Urban Feature Analysis using Digital Image Data' International Journal of Remote Sensing Vol.9 No.3 p543-553.

Ehlers,M. (1985) 'The Effects of Image Noise on Digital Correlation Probability' Photogrammetric Engineering and Remote Sensing, Vol.51 No.3 p357-365.

Ehrich,R.W. (1977) 'Detection of Global Edges in Textured Images' IEEE Transactions on Computers, Vol.C-26 No.6 p589-603.

Engvall,J.L. Tubbs,J.D. and Holmes,Q.A. (1977) 'Pattern Recognition of Landsat Data Based on Temporal Trend Analysis' Remote Sensing of Environment' No.6, p303-314.

Estes,J.E. and Thorley,G.A. (eds) (1983) 'Manual of Remote Sensing: vol.2. Theory, Interpretation and Applications' 2nd. ed. American Society of Photogrammetry, Falls Church, Virginia.

Excell,P.S. Campbell,D. Dodd,D. Newby,S. and Stevens,R. (1989) 'Classification of Quasi-Periodic Textures in Thematic mapper Scenes using Fourier Transforms' in: Remote Sensing For Operational Applications, technical contents of the 15th. annual conference of the Remote Sensing Society, university of Bristol, 13-15 September. p123-128.

Fiegenbaum,E. McCorduck,P. and Nii,H.P. (1988) 'The Rise of the Expert Company' Times Books, New York.

Forster,B.C. (1983) 'An Examination of Some Problems and Solutions in Monitoring Urban Areas from Satellite platforms' International Journal of Remote Sensing, Vol.6, No.1, p139-151.

Forster,B.C. (1984) 'Derivation of Atmospheric Correction Procedures for LANDSAT MSS with Particular Reference to Urban Data' International Journal of Remote Sensing, Vol.5, No.5, p799-817.

Fukunga,K. (1972) 'Introduction to Statistical Pattern Recognition' Academic Press, New York.

Fusco,L and Trevese,D (1985) 'On the Reconstruction of Lost Data in Images of More Than One Band' International Journal of Remote Sensing, Vol.6 No.9 p1535-1544.

Ginevan,M.E. (1979) 'Testing Land-Use Map Accuracy: Another Look' Photogrammetric Engineering and Remote Sensing, Vol.45, No.10, p1371-1377.

Golton,E. (1987) 'A Survey of Image Processing Algorithms' RAL-87-105, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon.

Gong,P. and Howarth,P.J. (1990) 'The Use of Structural Information for Improving Land-Cover Classification Accuracies at the Rural-Urban Fringe' Photogrammetric Engineering and Remote Sensing, Vol.56 No.1 p67-73.

Goodenough,D.G. Goldberg,M. Plunkett,G. and Zelek,J. (1987) 'An expert System for Remote Sensing' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.3 p349-367.

Gordon,D.K. and Phillipson,W.R. (1986) 'A Texture-Enhancement Procedure for Separating Orchard From Forest in Thematic Mapper Data' International Journal of Remote Sensing, Vol.7 No.2 p301-304.

Gurney,C.M. and Townshend,J.R.G. (1983) 'The Use of Contextual Information in the Classification of Remotely Sensed Data' Photogrammetric Engineering and Remote Sensing, Vol.49, No.1, p55-64.

Hall-Konyves,K. (1987) 'The Topographic Effect on Landsat Data in Gently Undulating Terrain in Southern Sweden' International Journal of Remote Sensing, Vol.8 No.2 p157-168.

Haralick,R.M. Shanmugam,K. and Dinstein,I (1973) 'Textural Features for Image Classification' IEEE Transactions on Systems, Man and Cybernetics Vol.SMC-3 No.6 p610-621.

Harris,R (1985) 'Contextual Classification Post-Processing of Landsat Data Using a Probabilistic relaxation Model' International Journal of Remote Sensing, Vol.6 No.6 p847-866.

Harrison,A.R. Dunn,R. and White,J.C. (1989) 'The use of Satellite Imagery for Simulating and Testing the Efficiency of Sampling Approaches for the Measurement of Land Use Cover' in: Remote Sensing For Operational Applications, technical contents of the 15th. annual conference of the Remote Sensing Society, university of Bristol, 13-15 September. p187-192.

Hay,A.M. (1979) 'Sampling Designs to Test Land-Use Map Accuracy' Photogrammetric Engineering and Remote Sensing, Vol.45, No.4, p529-533.

Hlavka,C.A. (1987) 'Land-Use Mapping using Edge Density Texture Measures on Thematic Mapper Simulator Data' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.1 p104-108.

Holley,W.A. Parker,H.D.Jnr. and Kan,E.P. (1973) 'The JSC Clustering Program ISOCLS and its Applications' Machine Processing of Remotely Sensed Data, proc. conf. Lafayette, Indiana, Oct. p36-50.

Hord,R.M. and Brooner,W (1976) 'Land-Use Map Accuracy Criteria' Photogrammetric Engineering and Remote Sensing, Vol.42, No.5, p671-677.

Hudson,W.D. (1987) 'Evaluation of Several Classification Schemes for mapping Forest Cover Types in Michigan' International Journal of Remote Sensing Vol.8 No.12 p1785-1796.

Hyde,J. Fullwood,J.A. and Corrall,D.R. (1985) 'An Approach to Knowledge-Driven Segmentation' Image and Vision Computing, Vol.3, No.4, p198-205.

Ince,F. (1987) 'Maximum Likelihood Classification, optimal or Problematic? A Comparison with Nearest Neighbour Classification' International Journal of Remote Sensing Vol.8 No.12 p1829-1838.

Intelligent Terminals Ltd. (1986) 'SuperExpert User Manual' draft version. Intelligent Terminals Ltd. George House, 36 North Hannover Street, Glasgow G1 2AD, Scotland UK.

Irons,J.R. Markham,B.L. Nelson,R.F. Toll,D.L. Williams,D.L. Latty,R.S. and Stauffer,M.L. (1985) 'The Effects of Spatial Resolution on the Classification of Thematic Mapper Data' International Journal of Remote Sensing, Vol.6 No.8 p1385-1403.

Jackson,M.J. and Mason,D.C. (1986) 'The Development of Integrated Geo-Information Systems' International Journal of Remote Sensing, Vol.7 No.6 p723-740.

Jiaju Lu (1988) 'Development of Principal Component Analysis Applied to Multitemporal Landsat TM Data' International Journal of Remote Sensing Vol.9 No.12 p1895-1908.

Jones,A.R. Settle,J.J. and Wyatt,K.B. (1988) 'Use of Digital Terrain Data in the Interpretation of SPOT-1 HRV Multispectral Imagery' International Journal of Remote Sensing Vol.9 No.4 p669-682.

Jupp,D.L.B. Strahler,A.H. and Woodcock,C.E. (1988) 'Autocorrelation and Regularization in Digital Images I. Basic Theory' IEEE Transactions on Geoscience and Remote Sensing, Vol.26 No.4 p463-473.

Jupp,D.L.B. Strahler,A.H. and Woodcock,C.E. (1989) 'Autocorrelation and Regularization in Digital Images II. Simple Image Models' IEEE Transactions on Geoscience and Remote Sensing, Vol.27 No.3 p247-258.

Kay,S.A.W. nd Barnsley,M.J. (1989) 'The Relationship Between Vegetation Canopy Geometry and Image Statisticsin a Multiple View-Angle Data Set' in: Proceedings of the Remote Sensing Society's Annual Conference, Sept.1989, Bristol University. p207-213.

Kauth,R.A. and Thomas,G.S. (1976) 'The Tasseled Cap - a graphic Description of the Spectral - Temporal Development of Agricultural Crops as Seen by Landsat' Proc. Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, p41-53.

Kittler,J and Foglein,J (1984) 'Contextual Classification of Multispectral Pixel Data' Image and Vision Computing, Vol.2 No.1.

Kittler,J and Illingworth,J (1985) 'Relaxation Labeling Algorithms-a Review' Image and Vision Computing, Vol.3 No.4.

Kittler,J. and Pairman,D. (1985) 'Contextual Pattern Recognition Applied to Cloud Detection and Identification' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-23 No.6 p855-863.

Labowitz,M.L. (1986) 'Issues Arising from Sampling Designs and Band Selection in Discriminating Ground Reference Attributes using Remotely Sensed Data' Photogrammetric Engineering and Remote Sensing, Vol.52, No.2 p201-211.

Lee,T. (1985) 'A Low-Cost Classifier for Multitemporal Applications' International Journal of Remote Sensing, Vol.6 No.8 p1405-1417.

Lenat,D.B. (1986) 'Computer Software for Intelligent Systems', Scientific American.

Lesk,M. (1986) 'Computer Software for Information Management' Scientific American.

Lillesand,T.M. and Kiefer,R.W. (1979) 'Remote Sensing and Image Interpretation' Wiley, New York.

Legg,C.A. (1988) 'Operational Remote Sensing in the UK: Problems of Image Acquisition' proceedings IGARSS'88 Remote Sensing: Moving Towards the 21st. Century, Edinburgh, UK, Sept.12-16. p1457-1462.

Letts,P.J. (1979) 'Small Area Replacement in Digital Thematic Maps' Proc. 1979 Machine Processing of Remotely Sensed Data Symposium.

Majumdar,T.J. and Bhattacharya,B.B. (1988) 'Application of the Haar Transform for Extraction of Linear and Anomalous Patterns over Part of Cambay Basin, India' International Journal of Remote Sensing Vol.9 No.12 p1937-1942.

Malila,W.A. (1980) 'Change Vector Analysis: an Approach for Detecting Forest Changes with Landsat' Machine Processing of Remotely Sensed Data Symposium, p326-336.

Mather,P.M. (1985) 'A Computationally-Efficient Maximum Likelihood Classifier Employing Prior Probabilities for Remotely Sensed Data' International Journal of Remote Sensing, Vol.6 No.2.

Mather,P.M. (1987) 'Computer Processing of Remotely Sensed Images: an Introduction' Wiley, Chichester.

Matsuyama,T. (1987) 'Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.3 p305-316.

Mausel,P.W. Kramber,W.J. and Lee,J.K. (1990) 'Optimum Band Selection for Supervised Classification of Multispectral Data' Photogrammetric Engineering and Remote Sensing Vol.56 No.1 p55-60.

Maxwell,E.L. (1976) 'Multivariate System Analysis of Multispectral Imagery' Photogrammetric Engineering and Remote Sensing, Vol.42, No.9, p1173-1186.

McGuinness,R. (1988) 'A Knowledge-Based Colour Graphics Display System for Multi-Parameter Radar Data' International Journal of Remote Sensing Vol.9 No.3 p515-526.

McKeown,D.M. (Jr.) (1987) 'The Role of Artificial Intelligence in the Integration of Remotely Sensed Data with Geographic information systems' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.3 p330-348.

Merchant,J.W. (1984) 'Using Spatial Logic in Classification of Landsat TM Data' proc. 9th. Pecora Symposium on Technologies for Remote Sensing Today and Tomorrow, p378-385.

Michaelis,M. (1988) 'Interpretation of Geometrical Aspects of Thematic Mapper Data' International Journal of Remote Sensing Vol.9 Nos.10-11 p1723-1737.

Mohn,E. Hjort,N.L. and Storvik,G.O. (1987) 'A Simulation Study of Some Contextual Classification Methods for Remotely Sensed Data' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.6.

Moore,G.K. (1983) 'Objective Procedures for Lineament Enhancement and Extraction' Photogrammetric Engineering and Remote Sensing, Vol.49 No.5 p641-647.

Morton,A.J. (1986) 'Moorland Plant Community Recognition Using Landsat MSS Data' Remote Sensing of Environment, Vol.20, p291-298.

Murphy,D.L. (1985) 'Estimating Neighbourhood Variability with a Binary Comparison Matrix' Photogrammetric Engineering and Remote Sensing, Vol.51 No.6 p667-674.

Nicolin,B. and Gabler,R. (1987) 'A Knowledge-Based System for the Analysis of Aerial Images' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.3 p317-329.

Oldfield,R.B. (1988) 'Lithological Mapping of Northwest Argentina with Remote Sensing Data Using Tonal, Textural and Contextual Features' PhD thesis, Aston University.

Pairman,D. and Kittler,J. (1986) 'Clustering Algorithms for use with Images of Clouds' International Journal of Remote Sensing, Vol.7 No.7 p855-866.

Parikh,J.A. (1986) 'Automated Techniques for Extraction of Geological Fracture Patterns' in: Pattern Recognition in Practice, II (eds. Gelsema,E.S. and Kanal,L.N.) Elsevier Science B.V. North-Holland p373-383.

Peacegood,G. (1985) 'Expert Systems in Remote Sensing: a Preliminary Application' in Remote Sensing: Data Acquisition, Management and Applications. Proceedings of the poster sessions at the international conference of the Remote Sensing Society and the Center for Earth Resources Management. University of London 9th-12th Sept. 1985. p273-282.

Pearson,E.S. and Hartley,H.O. (1966) 'Biometrika Tables for Statisticians' Vol.1, 3rd. ed. Cambridge University Press, Cambridge.

Pendock,N. (1987) 'Fast Classification of Image Data With Large Spectral Dimension' South African Journal of Photogrammetry, Remote Sensing and Cartography Vol.14 part 6 p351-358.

Pinter,P.J.Jnr, Zipoli,G. Maracchi,G. and Reginato,R. (1987) 'Influence of Topography and Sensor View Angles on Near Infra-Red/Red Ratio and Greenness Vegetation Indices of Wheat' International Journal of Remote Sensing Vol.8 No.6 p953-957.

Pooley,M.R. (1988) 'Computer Automation of Land Use Data' PhD Thesis, Department of Civil Engineering, Aston University, Birmingham. B4 7ET.

Press,W.H. (1986) 'Numerical Recipes, the Art of Scientific Computing' Cambridge University Press, Cambridge.

Rhode,W.G. (1978) 'Improving Land Cover Classification by Image Stratification of Landsat data' proc. 12th. International Symposium on Remote Sensing of Environment, Manila, Philippines, p729-741.

Robinove,C.J. (1982) 'Computation with Physical Values from Landsat Digital Data' Photogrammetric Engineering and Remote Sensing, Vol.48, No.5, p781-784.

Rosenfeld,G.H. and Fitzpatrick-Lins,K. (1986) 'A Coefficient of Agreement as a Measure of Thematic Classification Accuracy' Photogrammetric Engineering and Remote Sensing, Vol.52, No.2, p223-227.

Rothery,D.A. (1982) 'Supervised Maximum-Likelihood Classification and Post-Classification Filtering Using MSS Imagery for Lithological Mapping in the Oman Ophiolite' proc. International Symposium on Remote Sensing of Environment, Second thematic Conference, remote Sensing for Exploration Geology, Fort Worth, Texas, Dec.6-10,

p417-425.

Royer,A. Charbonneau,L. Brochu,R. Murphy,J.M. and
Tiellet,P.M. (1987) 'Radiometric Comparison of Landsat-5
TM and MSS Sensors' International Journal of Remote
Sensing, Vol.8 No.4 p579-591.

Saarikoski,A. (1988) 'Survey of Segmentation Techniques
for Image Analysis' Photogrammetric Journal of Finland,
Vol.11, No.1, p68-80.

Schowengerdt,R.A. (1983) 'Techniques for Image Processing
and Classification in Remote Sensing' Academic Press, New
York.

Schreier,H and Lavkulich,L.M. (1979) 'A Numerical Approach
to Terrain Analysis for Off-Road Trafficability'
Photogrammetric Engineering and Remote Sensing Vol.45 No.7
p635-642.

Seddon,A.M. and Hunt,G.E. (1985) 'Segmentation of Clouds
Using Cluster Analysis' International Journal of Remote
Sensing, Vol.6 No.5 p717-731.

Settle,J.J. and Briggs,S.A. (1987) 'Fast Maximum
Likelihood Classification of Remotely-Sensed Imagery'
International Journal of Remote Sensing, Vol.8 No.5 p723-
734.

Sharaf,A.K. and Cracknell,A.P. (1989) 'Linear Discriminant
Analysis an Aid in Remote Sensing for Geobotanical
Investigation' International Journal of Remote Sensing,
Vol.10 No.11 p1735-1748.

Sheffield,C. (1985) 'Selecting Band Combinations from
Multispectral Data' Photogrammetric Engineering and Remote
Sensing, Vol.51 No.6 p681-687.

Simonett,D.S. and Ulaby,F.T. (eds) (1983) 'Manual of
Remote Sensing: vol.1. Theory, Instruments and Techniques'
2nd. ed. American Society of Photogrammetry, Falls Church,
Virginia.

Singh,A.N. and Dwivedi,R.S. (1986) 'The Utility of Landsat
Imagery as an Integral Part of the Data Base for Small-
Scale Soil Mapping.' International Journal of Remote
Sensing, Vol.7 No.9 p1099-1108.

Sing,A and Harrison,A (1985) 'Standardized Principal
Components' International Journal of Remote Sensing Vol.6
No.6 p883-896.

Smith,B.W. Siegel,H.J. and Swain,P.H. (1981) 'Contextual
Classification on a CDC Flexible Processor System' Proc.
1981 Machine Processing of Remotely Sensed Data Symposium.

Smith,T.R. (1984) 'Artificial Intelligence and its Applicability to Geographical Problem Solving' The Professional Geographer, Vol.36, No.2, p1147-158.

Steven,M.D. and Rollin,E.M. (1986) 'Estimation of Atmospheric Corrections from Multiple Aircraft Imagery' International Journal of Remote Sensing, Vol.7, No.4, p481-497.

Stohr,C.J. and West,T.R. (1985) 'Terrain and Look Angle Effects Upon Multispectral scanner Response' Photogrammetric Engineering and Remote Sensing, Vol.51, No.2. p229-235.

Stromberg,W.D. and Farr,T.D. (1986) 'A Fourier-based Textural Feature Extraction Procedure' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-24 No.5 p722-731.

Swain,P.H. and Davis,S.M. (1978) 'Remote Sensing: the Quantitative Approach' McGraw-Hill, New York.

Taylor,J.C. (date unknown) 'Agricultural applications of Remote Sensing' Remote Sensing Unit, Silsoe College, Silsoe, Bedford. MK45 4DT.

Thomas,I.L. (1980) 'Spatial Postprocessing of Spectrally Classified Landsat data' Photogrammetric Engineering and Remote Sensing, Vol.46, No.9, p1201-1206.

Thomas,I.L. (1981) 'Textural Enhancement of a Circular Geological Feature' Photogrammetric Engineering and Remote Sensing Vol.47 No.1 p89-91.

Thomas,I.L, Ching,N.P, Benning,V.M, and D'Aguanno,J.A. (1987) 'A Review of Multi-Channel Indices of Class Separability' International Journal of Remote Sensing Vol.6 No.6 p897-901.

Toll,D.L. (1985) 'Effect of Landsat Thematic Mapper sensor Parameters on Land Cover Classification' Remote Sensing of Environment, Vol.17, p129-140.

Tom,C.H. and Miller,L.D. (1984) 'An Automated Land-Use Mapping Comparison of the Bayesian Maximum Likelihood and Linear Discriminant Analysis Algorithms' Photogrammetric Engineering and Remote Sensing, Vol.50, No.2, p193-207.

Townsend,F.E. (1986) 'The Enhancement of Computer Classifications by Logical Smoothing' Photogrammetric Engineering and Remote Sensing Vol.52 No.2 p213-221.

Tiellet,P.M. (1986) 'Image Correction for Radiometric Effects in Remote Sensing' International Journal of Remote Sensing, Vol.7 No.12 p1637-1651.

Wang,F. and Newkirk,R. (1988) 'A Knowledge-Based System for Highway Network Extraction' IEEE Transactions on Geoscience and Remote Sensing, Vol.26 No.5 p525-531.

Wang Li and He,D.C. (1990) 'A New Statistical Approach for Texture Analysis'Photogrammetric Engineering and Remote Sensing Vol.56 No.1 p61-66.

Wang Ru-Ye (1986a) 'An Approach to Tree-Classifier Design Based on Hierarchical Clustering' International Journal of Remote Sensing, Vol.7 No.1 p75-88

Wang Ru-Ye (1986b) 'An Approach to Tree-Classifier Design Based on a Splitting Algorithm' International Journal of Remote Sensing, Vol.7 No.1 p89-104.

Wasrud,J. and Lulla,K. (1985) 'Evaluation of Classification Algorithms' proc. 19th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, Oct.21-25, p433-441.

Weaver,R.E. (1987) 'Spectral Separation of Moorland Vegetation in Airborne Thematic Mapper Data' International Journal of Remote Sensing, Vol.8, No.1, p31-42.

Wharton,S.W. (1982) 'A Context-Based Land-Use Classification Algorithm for High-Resolution Remotely Sensed Data' Journal of Applied Photographic Engineering, Vol.8, No.1. p46-50.

Wharton,S.W. (1987) 'A Spectral Knowledge-Based Approach for Urban Land-Cover Classification' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25, No.3, p272-282.

Wood,T.F. and Foody,G.M. (1989) 'Analysis and Representation of Vegetation Continua from Landsat Thematic Mapper Data for Lowland Heaths' International Journal of Remote Sensing Vol.10 No.1 p181-192.

Yoeli,P. (1982) 'Cartographic Drawing with Computers' Computer Applications, special issue, Vol.8.

Yool,S.R. Star,J.L. Estes,J.L. Botkin,D.B. Eckhardt,D.W. and Davis,F.W. (1986) 'Performance Analysis of Image Processing Algorithms for Classification of Natural Vegetation in the Mountains of Southern California' International Journal of Remote Sensing, Vol.7 No.5 p683-702.

Young,J.A.T. (1985) 'Remote Sensing and an Experimental Geographic Information System for Environmental Monitoring, Resource Planning and management' International Journal of Remote Sensing, Vol.7 No.6 p741-744.

Quiming Zhou (1989) 'A Method for Integrating Remote
Sensing and Geographic Information Systems'
Photogrammetric Engineering and Remote Sensing Vol.55 No.5
p591-596.

APPENDIX

RESULTS

A1 LAYOUT

The results in this appendix are presented as follows:

Best cases (highest clasification accuracies achieved by adjusting parameters) for each classifier-contextual enhancement pairing on Peak District Data set (Tables A1a and b).

Accuracies achieved using the two clustering algorithms described in the thesis using the Peak District data set (Table A2).

Results of the algorithms tested on the Salisbury, North Wales and Yemen data sets (Tables A3, A4 and A5).

Confusion matrices for the results in table A1, A3, A4 and A5 (Table A6).

Table A1a: Summary - best cases for each contextual algorithm

algorithm and parameters		classifier	accuracy % before	after	reduction in error %
mode filter	9	maximum likelihood	92.22	95.76	45.50
	9	deviant distance	87.06	92.42	41.42
	11	minimum distance	80.07	87.64	37.98
	7	box	66.12	71.26	15.17
	11	decision tree	58.23	65.13	16.52
	3	look up table	59.81	59.20	-1.52

Note: parameter denotes size of patch in pixels

algorithm and parameters	classifier	accuracy % before	after	reduction in error %
small area replacement 8,5	maximum likelihood	92.22	95.83	46.40
8,5	deviant distance	87.06	92.55	42.43

238

algorithm and parameters		classifier	accuracy % before	after	reduction in error %
small area replacement	8,5	minimum distance	80.07	86.32	31.36
	8,5	box	66.12	91.66	75.38
	4,5	decision tree	58.23	79.34	50.54
	4,L	look up table	59.81	60.22	1.02

Notes: first parameter indicates maximum size of area declassified in pixels, second indicates number of nearest classified neighbours used in reclassification. L signifies use of nearest along-line classified neighbour.

algorithm and parameters		classifier	accuracy % before	after	reduction in error %
thresholded edge filter		maximum likelihood	92.22	93.20	12.60
		deviant distance	87.06	88.61	11.98
		minimum distance	80.07	83.60	17.71
	*1	box	66.12	89.00	67.53
	*2	decision tree	58.23	79.30	50.44
		look up table	59.81	60.47	1.64

Notes: all Roberts filter + automatic thresholding on classified image, with reclassification by mode of nearest five neighbours, except: *1:Prewitt filter + automatic thresholding on band 5 of original image, with reclassification by mode of nearest five neighbours, and *2:Prewitt filter + automatic thresholding on classified image, with reclassification by mode of nearest five neighbours.

algorithm and parameters		classifier	accuracy % before	after	reduction in error %
Markov relaxation	M	maximum likelihood	92.22	96.39	53.60
	M	deviant distance	87.06	92.10	38.95

algorithm and parameters		classifier	accuracy % before	after	reduction in error %
Markov relaxation	M	minimum distance	80.07	90.50	52.33
	M*2	box	66.12	90.70	72.55
	I9	decision tree	58.23	82.47	58.03
	I6	look up table	59.81	60.98	0.03

Notes: M signifies relaxation based on transition probabilities estimated from map and field data, M*2 indicates two iterations of this, I signifies use of the classified image to estimate transition probabilities, the figure after this indicating the order of the chain derived from these.

algorithm and parameters		classifier	accuracy % before	after	reduction in error %
Wharton's CONAN algorithm	9	maximum likelihood	92.22	95.68	44.47
	15	deviant distance	87.06	91.62	35.24
	15	minimum distance	80.07	89.16	45.61
	15	box	66.12	86.45	60.00
	15	decision tree	58.23	80.59	53.53
	15	look up table	59.81	55.95	-9.60

Notes: the figures indicate the size of the patch used in pixels.

Table A1b: Accuracies and times for all algorithms (Peak District Data Set)

classifier	enhancement	accuracy %	time seconds
maximum likelihood	none	92.22	5670.42
	3 by 3 mode	94.64	5915.39
	5 by 5 mode	95.48	5936.39
	7 by 7 mode	95.62	5968.55
	9 by 9 mode	95.76	6001.89
	11 by 11 mode	95.19	6050.38
	13 by 13 mode	94.53	6121.51
	15 by 15 mode	93.89	6173.05
	SAR 1 plus:		
	NAL neighbour	92.13	5699.52
	mode of patch	92.16	5890.73
	NN	92.16	5733.40
	mode of 5 NN's	92.16	5828.27
	SAR 4 plus:		
	NAL neighbour	92.02	5699.52
	mode of patch	92.83	5890.73
	NN	92.52	5733.40
	mode of 5 NN's	92.59	5828.27
	SAR 8 plus:		
	NAL neighbour	95.45	5699.52
	mode of patch	83.10	5890.73
	NN	95.60	5733.40
	mode of 5 NN's	95.83	5828.27
	Prewitt on B5, auto threshold, declassification plus:		
	NN	92.13	5795.76
	mode of 5 NN's	92.55	5886.63
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	85.39	5820.20
	mode of 5 NN's	84.76	5911.07
	Roberts on B5, auto threshold, declassification plus:		
	NN	92.18	5795.76
	mode of 5 NN's	92.79	5886.63

classifier	enhancement	accuracy %	time seconds
Maximum Likelihood	as above, but declassification of pixels within a radius of 2, plus:		
	NN	86.05	5820.20
	mode of 5 NN's	85.27	5911.07
	Prewitt on classified image auto threshold, declassification plus:		
	NN	92.27	5795.76
	mode of 5 NN's	92.68	5886.63
	Roberts on classified image auto threshold, declassification plus:		
	NN	92.85	5795.76
	mode of 5 NN's	93.20	5886.63
	Markov relaxation		
	image (order 7)	75.05	10285.64
	map	96.39	12351.93
	map (2nd itern.)	95.72	19033.44
	CONAN (9 by 9)	95.68	6130.28
	(15 by 15)	95.11	6330.94
	edge/noise		
	noise by NN	91.54	–
	edges by NN	89.36	–
	both by NN	89.28	–
deviant distance	none	87.06	337.97
	3 by 3 mode	89.50	582.94
	5 by 5 mode	91.11	604.33
	7 by 7 mode	92.00	636.10
	9 by 9 mode	92.42	669.44
	11 by 11 mode	92.28	717.93
	13 by 13 mode	91.55	789.06
	15 by 15 mode	91.03	840.60

classifier	enhancement	accuracy %	time seconds
Deviant Distance	SAR 1 plus:		
	NAL neighbour	87.46	367.07
	mode of patch	87.52	558.28
	NN	87.51	400.95
	mode of 5 NN's	87.49	491.82
	SAR 4 plus:		
	NAL neighbour	87.77	367.07
	mode of patch	88.54	558.28
	NN	87.99	400.95
	mode of 5 NN's	88.16	491.82
	SAR 8 plus:		
	NAL neighbour	92.34	367.07
	mode of patch	76.00	558.28
	NN	91.77	400.95
	mode of 5 NN's	92.55	491.82
	Prewitt on B5, auto threshold, declassification plus:		
	NN	87.20	463.31
	mode of 5 NN's	87.52	554.18
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	83.33	487.75
	mode of 5 NN's	83.29	578.62
	Roberts on B5, auto threshold, declassification plus:		
	NN	87.18	463.31
	mode of 5 NN's	87.75	554.18
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	83.03	487.75
	mode of 5 NN's	80.24	578.62

classifier	enhancement	accuracy %	time seconds
Deviant Distance	Prewitt on classified image auto threshold, declassification plus:		
	NN	87.58	463.31
	mode of 5 NN's	88.29	554.18
	Roberts on classified image auto threshold, declassification plus:		
	NN	88.01	463.31
	mode of 5 NN's	88.61	554.18
	Markov relaxation		
	image (order 8)	91.66	5658.16
	map	92.10	7019.48
	CONAN (9 by 9)	91.60	797.83
	(15 by 15)	91.63	998.49
	edge/noise		
	noise by NN	86.43	–
	edges by NN	84.48	–
	both by NN	84.40	–
minimum distance	none	80.07	287.21
	3 by 3 mode	83.14	532.18
	5 by 5 mode	85.29	553.57
	7 by 7 mode	86.33	585.34
	9 by 9 mode	87.02	618.68
	11 by 11 mode	87.64	667.17
	13 by 13 mode	87.53	738.30
	15 by 15 mode	87.59	789.84
	SAR 1 plus:		
	NAL neighbour	80.20	316.31
	mode of patch	80.18	507.52
	NN	80.16	350.19
	mode of 5 NN's	80.18	441.06
	SAR 4 plus:		
	NAL neighbour	80.38	316.31
	mode of patch	81.90	507.52
	NN	80.70	350.19
	mode of 5 NN's	81.14	441.06

classifier	enhancement	accuracy %	time seconds
Minimum Distance	SAR 8 plus:		
	NAL neighbour	86.18	316.31
	mode of patch	64.87	507.52
	NN	85.89	350.19
	mode of 5 NN's	86.32	441.06
	Prewitt on B5, auto threshold, declassification plus:		
	NN	79.60	412.55
	mode of 5 NN's	80.31	503.42
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	74.57	436.99
	mode of 5 NN's	74.60	527.86
	Roberts on B5, auto threshold, declassification plus:		
	NN	79.75	412.55
	mode of 5 NN's	80.67	503.42
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	74.54	436.99
	mode of 5 NN's	74.42	527.86
	Prewitt on classified image auto threshold, declassification plus:		
	NN	80.78	412.55
	mode of 5 NN's	82.04	503.42
	Roberts on classified image auto threshold, declassification plus:		
	NN	82.32	412.55
	mode of 5 NN's	83.60	503.42

classifier	enhancement	accuracy %	time seconds
Minimum Distance	Markov relaxation		
	image (order 9)	85.28	6396.34
	map	90.50	6968.72
	CONAN (9 by 9)	87.44	747.07
	(15 by 15)	89.16	947.73
	edge/noise		
	noise by NN	79.35	–
	edges by NN	77.83	–
	both by NN	77.75	–
box	none	66.12	164.98
	3 by 3 mode	69.00	409.95
	5 by 5 mode	70.99	431.34
	7 by 7 mode	71.26	463.11
	9 by 9 mode	71.03	496.45
	11 by 11 mode	69.57	544.94
	13 by 13 mode	68.21	616.07
	15 by 15 mode	66.74	667.61
	SAR 1 plus:		
	NAL neighbour	87.47	194.08
	mode of patch	75.20	385.29
	NN	87.44	227.96
	mode of 5 NN's	87.93	318.93
	SAR 4 plus:		
	NAL neighbour	87.65	194.08
	mode of patch	69.71	385.29
	NN	87.57	227.96
	mode of 5 NN's	88.40	318.93
	SAR 8 plus:		
	NAL neighbour	82.52	194.08
	mode of patch	40.93	385.29
	NN	91.43	227.96
	mode of 5 NN's	91.66	318.93
	Prewitt on B5, auto threshold, declassification plus:		
	NN	88.10	290.32
	mode of 5 NN's	89.00	381.19

classifier	enhancement	accuracy %	time seconds
Box	as above, but declassification of pixels within a radius of 2, plus:		
	NN	82.03	314.76
	mode of 5 NN's	81.95	405.63
	Roberts on B5, auto threshold, declassification plus:		
	NN	88.04	290.32
	mode of 5 NN's	88.77	381.19
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	82.52	314.76
	mode of 5 NN's	81.43	405.63
	Prewitt on classified image auto threshold, declassification plus:		
	NN	88.23	290.32
	mode of 5 NN's	88.67	381.19
	Roberts on classified image auto threshold, declassification plus:		
	NN	87.78	290.32
	mode of 5 NN's	88.83	381.19
	Markov relaxation		
	image (order 13)	87.46	6274.11
	itn.2 (order 10)	88.46	12955.62
	itn.3 (order 9)	88.62	19064.75
	itn.4 (order 10)	88.24	25746.26
	map	90.33	6846.49
	itn.2	90.70	13528.00

classifier	enhancement	accuracy %	time seconds
Box	CONAN (9 by 9)	84.87	624.87
	(15 by 15)	86.45	825.50
	edge/noise		
	noise by NN	88.09	-
	edges by NN	84.40	-
	both by NN	85.60	-
decision tree	none	58.23	69.45
	3 by 3 mode	61.17	314.42
	5 by 5 mode	63.72	335.81
	7 by 7 mode	64.84	367.58
	9 by 9 mode	65.10	400.92
	11 by 11 mode	65.13	449.41
	13 by 13 mode	64.47	520.54
	15 by 15 mode	64.28	572.08
	SAR 1 plus:		
	NAL neighbour	77.36	98.55
	mode of patch	68.01	289.76
	NN	77.87	132.43
	mode of 5 NN's	78.19	223.30
	SAR 4 plus:		
	NAL neighbour	77.73	98.55
	mode of patch	59.48	289.76
	NN	78.20	132.43
	mode of 5 NN's	79.34	223.30
	SAR 8 plus:		
	NAL neighbour	73.87	98.55
	mode of patch	27.05	289.76
	NN	77.56	132.43
	mode of 5 NN's	78.39	223.30
	Prewitt on B5, auto threshold, declassification plus:		
	NN	77.39	194.79
	mode of 5 NN's	78.29	285.66
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	72.85	219.23
	mode of 5 NN's	72.43	310.10

classifier	enhancement	accuracy %	time seconds
Decision Tree	Roberts on B5, auto threshold, declassification plus:		
	NN	77.78	194.79
	mode of 5 NN's	79.07	285.66
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	71.46	219.23
	mode of 5 NN's	70.95	310.10
	Prewitt on classified image auto threshold, declassification plus:		
	NN	78.38	194.79
	mode of 5 NN's	79.30	285.66
	Roberts on classified image auto threshold, declassification plus:		
	NN	77.24	194.79
	mode of 5 NN's	78.59	285.66
	Markov relaxation		
	image (order 9)	82.47	6178.58
	map	81.46	6750.96
	CONAN (9 by 9)	72.92	529.31
	(15 by 15)	80.59	729.37
	edge/noise		
	noise by NN	77.29	–
	edges by NN	75.80	–
	both by NN	75.72	–
look up table	none		
	(on spec. shape)	59.91	5.48
	(on o. bright.)	73.59	5.48
	3 by 3 mode	59.20	250.45
	5 by 5 mode	58.44	271.61
	7 by 7 mode	57.76	303.61

classifier	enhancement	accuracy %	time seconds
Look up Table	9 by 9 mode	57.27	336.95
	11 by 11 mode	57.05	385.44
	13 by 13 mode	56.69	456.67
	15 by 15 mode	56.02	508.11
	SAR 1 plus:		
	NAL neighbour	59.88	34.58
	mode of patch	59.83	225.79
	NN	59.82	68.46
	mode of 5 NN's	59.79	159.33
	SAR 4 plus:		
	NAL neighbour	60.22	34.58
	mode of patch	59.14	225.79
	NN	59.98	68.46
	mode of 5 NN's	60.09	159.33
	SAR 8 plus:		
	NAL neighbour	57.20	34.58
	mode of patch	48.12	225.79
	NN	44.41	68.46
	mode of 5 NN's	57.79	159.33
	Prewitt on B5, auto threshold, declassification plus:		
	NN	59.85	130.82
	mode of 5 NN's	60.25	221.69
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	53.56	155.26
	mode of 5 NN's	52.67	246.13
	Roberts on B5, auto threshold, declassification plus:		
	NN	59.90	130.82
	mode of 5 NN's	60.16	221.69
	as above, but declassification of pixels within a radius of 2, plus:		
	NN	54.29	155.26
	mode of 5 NN's	52.82	246.13

classifier	enhancement	accuracy %	time seconds
Look up Table	Prewitt on classified image auto threshold, declassification plus:		
	NN	60.08	130.82
	mode of 5 NN's	60.25	221.69
	Roberts on classified image auto threshold, declassification plus:		
	NN	60.00	130.82
	mode of 5 NN's	60.47	221.69
	Markov relaxation		
	image (order 6)	60.98	4074.72
	map	59.91	6687.49
	CONAN (9 by 9)	46.88	465.34
	(15 by 15)	55.95	666.00
	edge/noise		
	noise by NN	58.92	–
	edges by NN	57.09	–
	both by NN	56.95	–
non-probabilistic relaxation:			
by rank	–	81.43	711.78
by distance	–	81.32	711.78

Notes:

Spec. shape = spectral shape classifier
O. bright = overall brightness image
SAR = small area replacement
NAL = nearest along line
NN = nearest neighbour

Table A2a Iterative nearest neighbour clustering

iteration	normalised accuracy %	
0	80.07	(minimum distance)
1	67.58	
2	65.69	
3	63.76	

Table A2b: Histogram clustering

normalised accuracy (%): 77.25

Table A2c: Simulated per-field classifier

classifier	% accuracy alone	per-field
maximum likelihood	99.90	90.00
deviant distance	99.00	100.00
minimum distance	99.40	100.00
box	99.60	100.00
decision tree	99.90	90.00
look up table	90.00	40.00

Table A3: RESULTS FOR SALISBURY DATA SET

classifier and enhancement	accuracy	% decrease in error due to enhancement
box	38.69	-
9 by 9 mode	36.96	(-2.82)
9 by 9 CONAN	37.95	(-1.21)
SAR (8,5)	37.35	(-2.20)
minimum distance	93.99	-
9 by 9 mode	99.51	91.85
9 by 9 CONAN	99.68	94.68
SAR(8,5)	99.89	98.17
maximum likelihood	96.94	-
9 by 9 mode	99.40	80.39
9 by 9 CONAN	99.58	86.27
Markov relaxation	79.90	(-556.86)

Table A4: RESULTS FOR YENEN DATA SET

Classifier	Contextual Enhancement	Accuracy (%)
Minimum Distance	-	35.53
	9*9 CONAN	26.69
	Markov Relaxation (9)	41.98
	Small Area Replacement (8,1)	43.86
	9*9 Mode Filter	45.13
Maximum Likelihood	-	2.26
	9*9 CONAN	63.31
	Markov Relaxation (9)	64.27
	9*9 Mode Filter	67.44
	Small Area Replacement (8,1)	70.06

Table A5: RESULTS FOR NORTH WALES DATA SET

Classifier	Contextual Enhancement	Accuracy (%)
Deviant Distance	-	94.08
Minimum Distance	-	95.45
	Small Area Replacement (8,1)	95.98
	Markov Relaxation (10)	96.52
	9*9 Mode Filter	97.02
	9*9 CONAN	97.32
Maximum Likelihood	-	95.27
	Small Area Replacement (8,1)	97.02
	Markov Relaxation (10)	97.52
	9*9 Mode Filter	97.91
	9*9 CONAN	98.27

Table A6: Confusion matrices for selected classifications

Table A6a: Peak District Data Set

Confusion matrix for Maximum Likelihood Alone

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	0.	784.	0.	1.	0.	0.	0.	0.	0.	0.	0.	785.
2	0.	0.	1173.	2.	0.	0.	0.	0.	2.	0.	0.	1177.
3	0.	0.	21.	725.	0.	3.	0.	0.	19.	2.	1.	771.
4	0.	0.	0.	5.	982.	4.	0.	17.	0.	0.	0.	1008.
5	0.	0.	0.	3.	3.	234.	0.	10.	12.	0.	0.	262.
6	0.	0.	59.	0.	0.	0.	47.	0.	6.	0.	0.	112.
7	0.	0.	0.	0.	1.	86.	0.	1365.	15.	0.	0.	1467.
8	0.	0.	0.	17.	0.	223.	2.	24.	1330.	54.	0.	1650.
9	0.	0.	0.	6.	0.	0.	0.	0.	15.	478.	7.	506.
10	0.	0.	0.	1.	0.	0.	0.	0.	0.	8.	337.	346.
	0.	784.	1253.	760.	986.	550.	49.	1416.	1399.	542.	345.	

Normalised accuracy is   92.21920   %

Confusion matrix for Minimum Distance Alone

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	0.	785.	0.	0.	0.	0.	0.	0.	0.	0.	0.	785.
2	0.	0.	1059.	1.	0.	0.	111.	0.	6.	0.	0.	1177.
3	0.	0.	0.	570.	10.	84.	10.	8.	89.	0.	0.	771.
4	0.	3.	0.	0.	979.	1.	0.	25.	0.	0.	0.	1008.
5	0.	0.	0.	1.	2.	219.	0.	39.	1.	0.	0.	262.
6	0.	0.	0.	0.	0.	0.	111.	0.	1.	0.	0.	112.
7	0.	0.	0.	0.	24.	384.	0.	1053.	6.	0.	0.	1467.
8	0.	0.	0.	67.	21.	298.	157.	75.	901.	131.	0.	1650.
9	0.	0.	0.	0.	0.	0.	5.	0.	28.	465.	8.	506.
10	0.	0.	0.	0.	0.	0.	0.	0.	6.	9.	331.	346.
	0.	788.	1059.	639.	1036.	986.	394.	1200.	1038.	605.	339.	

Normalised accuracy is   80.07175   %

Confusion matrix for Deviant Distance Alone

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	0.	784.	0.	0.	1.	0.	0.	0.	0.	0.	0.	785.
2	0.	0.	1169.	1.	0.	0.	1.	0.	6.	0.	0.	1177.
3	0.	0.	16.	651.	0.	5.	0.	1.	98.	0.	0.	771.
4	0.	0.	0.	4.	936.	0.	0.	68.	0.	0.	0.	1008.
5	0.	0.	0.	41.	0.	187.	0.	30.	4.	0.	0.	262.
6	0.	0.	0.	0.	0.	0.	105.	0.	7.	0.	0.	112.
7	0.	0.	0.	22.	10.	160.	0.	1260.	15.	0.	0.	1467.
8	0.	0.	0.	163.	5.	193.	6.	55.	1227.	1.	0.	1650.
9	0.	0.	0.	0.	0.	0.	0.	0.	109.	396.	1.	506.
10	0.	0.	0.	0.	0.	0.	0.	0.	16.	7.	323.	346.
	0.	784.	1185.	882.	952.	545.	112.	1414.	1482.	404.	324.	

Normalised accuracy is   87.06086   %

Confusion matrix for Box classifier (+/- 2 sd) alone

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	171.	614.	0.	0.	0.	0.	0.	0.	0.	0.	0.	785.
2	120.	0.	1057.	0.	0.	0.	0.	0.	0.	0.	0.	1177.
3	517.	0.	1.	222.	0.	0.	0.	0.	31.	0.	0.	771.
4	224.	0.	0.	1.	756.	0.	0.	27.	0.	0.	0.	1008.
5	229.	0.	0.	1.	0.	26.	0.	2.	4.	0.	0.	262.
6	33.	0.	0.	0.	0.	0.	76.	0.	3.	0.	0.	112.
7	388.	0.	0.	2.	0.	74.	0.	982.	21.	0.	0.	1467.
8	595.	0.	0.	15.	0.	19.	0.	5.	1016.	0.	0.	1650.
9	140.	0.	0.	0.	0.	0.	0.	0.	41.	325.	0.	506.
10	64.	0.	0.	0.	0.	0.	0.	0.	6.	5.	271.	346.
	2481.	614.	1058.	241.	756.	119.	76.	1016.	1122.	330.	271.	

Normalised accuracy is 66.11826 %

Confusion matrix for LUT classifier (zero threshold), on spectral shape image.

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	0.	779.	0.	0.	0.	0.	0.	0.	6.	0.	0.	785.
2	2.	11.	1093.	13.	20.	0.	0.	0.	38.	0.	0.	1177.
3	0.	1.	1.	107.	13.	0.	0.	103.	542.	4.	0.	771.
4	0.	2.	0.	0.	903.	0.	0.	13.	70.	20.	0.	1008.
5	0.	0.	0.	0.	1.	0.	0.	70.	191.	0.	0.	262.
6	0.	0.	59.	0.	0.	0.	0.	0.	53.	0.	0.	112.
7	0.	0.	0.	0.	0.	0.	0.	376.	1091.	0.	0.	1467.
8	0.	0.	0.	0.	0.	0.	0.	83.	1567.	0.	0.	1650.
9	0.	3.	0.	0.	76.	0.	0.	104.	305.	18.	0.	506.
10	0.	1.	0.	0.	336.	0.	0.	0.	9.	0.	0.	346.
	2.	797.	1153.	120.	1349.	0.	0.	749.	3872.	42.	0.	

Normalised accuracy is 59.90846 %

Confusion matrix for manually defined Decision Tree alone

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	34.	750.	0.	0.	1.	0.	0.	0.	0.	0.	0.	785.
2	305.	0.	813.	0.	0.	0.	49.	0.	8.	2.	0.	1177.
3	198.	0.	0.	390.	5.	4.	1.	0.	173.	0.	0.	771.
4	115.	0.	0.	0.	252.	21.	0.	620.	0.	0.	0.	1008.
5	68.	0.	0.	1.	0.	171.	0.	19.	3.	0.	0.	262.
6	22.	0.	0.	0.	0.	0.	89.	0.	1.	0.	0.	112.
7	406.	0.	0.	0.	5.	169.	0.	865.	22.	0.	0.	1467.
8	418.	0.	0.	21.	14.	195.	0.	32.	962.	8.	0.	1650.
9	185.	0.	0.	0.	0.	0.	0.	0.	20.	301.	0.	506.
10	227.	0.	0.	0.	0.	0.	0.	0.	3.	2.	114.	346.
	1978.	750.	813.	412.	277.	560.	139.	1536.	1192.	313.	114.	

Normalised accuracy is 58.22613 %

Confusion matrix for look up table classifier on overall brightness image

```
                  Classified as ------>
          U    1     2     3     4    5    6     7     8    9    10
   1   0. 784.   0.    0.    1.   0.   0.    0.    0.   0.    0. 785.
   2   2.   8.1144.    0.    2.   0.  14.    0.    7.   0.    0.1177.
   3   0.   0.    2.  398.    2.   0.  10.  171.   61.   0.  127. 771.
   4   0.   1.    0.    0.  903.   0.   0.  104.    0.   0.    0.1008.
   5   0.   0.    0.   59.    0.   0.   0.  203.    0.   0.    0. 262.
   6   0.   0.    5.    0.    0.   0.  96.    0.   11.   0.    0. 112.
   7   0.   0.    0.   48.   65.   0.   0.1354.    0.   0.    0.1467.
   8   0.   0.    0.  265.   18.   0.  11.  247.1003.   0.  106.1650.
   9   0.   0.    0.    3.    0.   0.   1.    0.  466.   0.   36. 506.
  10   0.   0.    0.   20.    0.   0.   0.    0.   59.   0.  267. 346.
       2. 793.1151.  793.  991.   0. 132.2079.1607.   0.  536.
Normalised accuracy is    73.58981      %
```

Confusion matrix for Maximum Likelihood + Markov relaxation from map and ground data

```
                  Classified as ------>
          U    1     2     3     4    5    6     7     8    9    10
   1   0. 784.    0.    1.    0.   0.   0.    0.    0.   0.    0. 785.
   2   0.   0.1177.    0.    0.   0.   0.    0.    0.   0.    0.1177.
   3   0.   0.    0.  771.    0.   0.   0.    0.    0.   0.    0. 771.
   4   0.   0.    0.    0.1008.   0.   0.    0.    0.   0.    0.1008.
   5   0.   0.    0.    3.    3. 235.   0.    0.   17.   4.    0. 262.
   6   0.   5.   65.    0.    0.   0.  15.    0.   25.   2.    0. 112.
   7   0.   0.    0.    0.    0.   0.   0.1467.    0.   0.    0.1467.
   8   0.   0.    0.    0.    0. 118.   0.   17.1483.  32.    0.1650.
   9   0.   0.    0.    0.    0.   0.   0.    0.    0. 506.    0. 506.
  10   0.   0.    0.    0.    0.   0.   0.    0.    0.   0. 346. 346.
       0. 789.1242.  775.1011. 353.  15.1484.1525. 544.  346.
Normalised accuracy is    96.38793      %
```

Confusion matrix for Minimum Distance + Markov relaxation from map and ground data

```
                  Classified as ------>
          U    1     2     3     4    5    6     7     8    9    10
   1   0. 785.    0.    0.    0.   0.   0.    0.    0.   0.    0. 785.
   2   0.   0.1170.    0.    0.   0.   7.    0.    0.   0.    0.1177.
   3   0.   0.    0.  740.    0.   0.   0.    0.   31.   0.    0. 771.
   4   0.   0.    0.    0.1008.   0.   0.    0.    0.   0.    0.1008.
   5   0.   0.    0.    1.    1. 172.   0.   79.    9.   0.    0. 262.
   6   0.   5.    0.    0.    0.   0. 106.    0.    1.   0.    0. 112.
   7   0.   0.    0.    0.    0.  92.   0.1374.    1.   0.    0.1467.
   8   0.   0.    0.   16.   14. 173.  90.  148.1109. 100.    0.1650.
   9   0.   0.    0.    0.    0.   0.   0.    0.    0. 506.    0. 506.
  10   0.   0.    0.    0.    0.   0.   0.    0.    0.   0. 346. 346.
       0. 790.1170.  757.1023. 437. 203.1601.1151. 606.  346.
Normalised accuracy is    90.49976      %
```

Confusion matrix for Maximum Likelihood + 9*9 mode filter

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	0.	784.	0.	1.	0.	0.	0.	0.	0.	0.	0.	785.
2	0.	0.	1177.	0.	0.	0.	0.	0.	0.	0.	0.	1177.
3	0.	0.	0.	771.	0.	0.	0.	0.	0.	0.	0.	771.
4	0.	0.	0.	0.	1008.	0.	0.	0.	0.	0.	0.	1008.
5	0.	0.	0.	1.	0.	261.	0.	0.	0.	0.	0.	262.
6	0.	4.	50.	0.	0.	0.	36.	0.	21.	1.	0.	112.
7	0.	0.	0.	0.	0.	8.	0.	1459.	0.	0.	0.	1467.
8	0.	0.	0.	0.	0.	226.	0.	0.	1393.	31.	0.	1650.
9	0.	0.	0.	0.	0.	0.	0.	0.	0.	506.	0.	506.
10	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	346.	346.
	0.	788.	1227.	773.	1008.	495.	36.	1459.	1414.	538.	346.	

Normalised accuracy is   95.75705   %

Confusion matrix for box classifier + Small area replacement (8,5)

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	0.	696.	0.	0.	0.	0.	10.	0.	19.	60.	0.	785.
2	0.	0.	1177.	0.	0.	0.	0.	0.	0.	0.	0.	1177.
3	0.	0.	0.	568.	0.	0.	0.	0.	9.	194.	0.	771.
4	0.	0.	0.	0.	975.	0.	0.	1.	32.	0.	0.	1008.
5	0.	0.	0.	0.	38.	0.	0.	0.	221.	3.	0.	262.
6	0.	1.	0.	0.	0.	0.	88.	0.	23.	0.	0.	112.
7	0.	0.	0.	0.	0.	0.	0.	1455.	12.	0.	0.	1467.
8	0.	0.	0.	0.	0.	0.	0.	0.	50.	1600.	0.	1650.
9	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.	505.	506.
10	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	346.	346.
	0.	697.	1177.	568.	1013.	0.	98.	1506.	1917.	762.	346.	

Normalised accuracy is   91.66254   %

Confusion matrix for deviant distance + Small area replacement (8,5)

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10	
1	0.	785.	0.	0.	0.	0.	0.	0.	0.	0.	0.	785.
2	0.	0.	1177.	0.	0.	0.	0.	0.	0.	0.	0.	1177.
3	0.	0.	0.	719.	0.	0.	0.	0.	52.	0.	0.	771.
4	0.	0.	0.	0.	996.	0.	0.	12.	0.	0.	0.	1008.
5	0.	0.	0.	1.	0.	231.	0.	0.	30.	0.	0.	262.
6	0.	0.	0.	0.	0.	0.	107.	0.	5.	0.	0.	112.
7	0.	0.	0.	11.	0.	59.	0.	1382.	15.	0.	0.	1467.
8	0.	0.	0.	72.	0.	217.	0.	47.	1314.	0.	0.	1650.
9	0.	0.	0.	0.	0.	0.	0.	0.	69.	437.	0.	506.
10	0.	0.	0.	0.	0.	0.	0.	0.	4.	0.	342.	346.
	0.	785.	1177.	803.	996.	507.	107.	1441.	1489.	437.	342.	

Normalised accuracy is   92.65215   %

Confusion matrix for box classifier + 15*15 CONAN

```
                 Classified as ------>
        U    1     2     3     4     5    6     7     8    9    10
 1   0. 708.   0.    0.    0.   77.   0.    0.    0.   0.    0.  785.
 2   0.   0.1174.    0.    0.    3.   0.    0.    0.   0.    0.1177.
 3   0.   0.    0.  597.    0.  174.   0.    0.    0.   0.    0.  771.
 4   0.   0.    0.    0.  912.   96.   0.    0.    0.   0.    0.1008.
 5   0.   0.    0.    0.    0.  262.   0.    0.    0.   0.    0.  262.
 6   0.   0.    0.    0.    0.   11.  69.    0.   32.   0.    0.  112.
 7   0.   0.    0.    0.    0.  217.   0.1250.    0.   0.    0.1467.
 8   0.   0.    0.    0.    0.  544.   0.    0.1106.   0.    0.1650.
 9   0.   0.    0.    0.    0.   10.   0.    0.    0. 496.    0.  506.
10   0.   0.    0.    0.    0.    8.   0.    0.    0.   0.  338.  346.
     0. 708.1174. 597. 912.1402.  69.1250.1138. 496. 338.
Normalised accuracy is    85.50223     %
```

Confusion matrix for box classifier + Markov relaxation (map),
second iteration

```
                 Classified as ------>
        U    1     2     3     4     5    6     7     8    9    10
 1   0. 722.   0.    0.    0.    0.   0.    0.   63.   0.    0.  785.
 2   0.   0.1177.    0.    0.    0.   0.    0.    0.   0.    0.1177.
 3   0.   0.    0.  640.    0.    0.   0.    0.  131.   0.    0.  771.
 4   0.   0.    0.    0.  997.    0.   0.   11.    0.   0.    0.1008.
 5   0.   0.    0.    0.   26.    0.   0.   12.  224.   0.    0.  262.
 6   0.   9.    0.    0.    0.    0.  68.    0.   35.   0.    0.  112.
 7   0.   0.    0.    0.    0.    0.   0.1455.   12.   0.    0.1467.
 8   0.   0.    0.  103.    0.   33.   0.   93.1421.   0.    0.1650.
 9   0.   0.    0.    0.    0.    0.   0.    0.    0. 506.    0.  506.
10   0.   0.    0.    0.    0.    0.   0.    0.    0.   0.  346.  346.
     0. 731.1177. 743.1023.  33.  68.1571.1886. 506. 346.
Normalised accuracy is    90.69768     %
```

Confusion matrix for box classifier + 15*15 CONAN

```
                 Classified as ------>
        U    1     2     3     4     5    6     7     8    9    10
 1   0. 708.   0.    0.    0.   77.   0.    0.    0.   0.    0.  785.
 2   0.   0.1174.    0.    0.    3.   0.    0.    0.   0.    0.1177.
 3   0.   0.    0.  597.    0.  174.   0.    0.    0.   0.    0.  771.
 4   0.   0.    0.    0.  912.   96.   0.    0.    0.   0.    0.1008.
 5   0.   0.    0.    0.    0.  262.   0.    0.    0.   0.    0.  262.
 6   0.   0.    0.    0.    0.   11.  69.    0.   32.   0.    0.  112.
 7   0.   0.    0.    0.    0.  217.   0.1250.    0.   0.    0.1467.
 8   0.   0.    0.    0.    0.  544.   0.    0.1106.   0.    0.1650.
 9   0.   0.    0.    0.    0.   10.   0.    0.    0. 496.    0.  506.
10   0.   0.    0.    0.    0.    8.   0.    0.    0.   0.  338.  346.
     0. 708.1174. 597. 912.1402.  69.1250.1138. 496. 338.
Normalised accuracy is    85.50223     %
```

Confusion matrix for box classifier + edge declassification and
reclassification

```
                Classified as ------>
        U    1    2    3    4    5    6    7    8    9   10
 1   0. 723.   0.   0.   0.   0.   0.   0.  49.  13.   0. 785.
 2   0.   0.1174.   3.   0.   0.   0.   0.   0.   0.   0.1177.
 3   0.   0.   0. 637.   0.   0.   0.   0. 133.   1.   0. 771.
 4   0.   0.   0.   3. 969.   1.   0.  34.   0.   1.   0.1008.
 5   0.   0.   0.   7.   4. 169.   0.  18.  63.   1.   0. 262.
 6   0.   0.   0.   0.   0.   0. 108.   0.   4.   0.   0. 112.
 7   0.   0.   0.   6.   0. 151.   0.1266.  44.   0.   0.1467.
 8   0.   0.   0.  71.   0. 166.   0.  37.1376.   0.   0.1650.
 9   0.   0.   0.   0.   0.   0.   0.   0.  57. 449.   0. 506.
10   0.   0.   0.   0.   0.   0.   0.   0.  12.  10. 324. 346.
     0. 723.1174. 727. 973. 487. 108.1355.1738. 475. 324.
Normalised accuracy is   89.00297     %
```

Confusion matrix for manually derived Decision Tree + Markov
relaxation, order 9, transition probabilities estimated from
classified image

```
                Classified as ------>
        U    1    2    3    4    5    6    7    8    9   10
 1   0. 785.   0.   0.   0.   0.   0.   0.   0.   0.   0. 785.
 2   0.   0.1167.   0.   0.   0.   3.   0.   7.   0.   0.1177.
 3   0.   0.   0. 636.   2.   0.   0.   0. 133.   0.   0. 771.
 4   0.   0.   0.   0. 325.   0.   0. 673.  10.   0.   0.1008.
 5   0.   0.   0.   2.   0. 109.   0. 121.  30.   0.   0. 262.
 6   0.   6.   0.   1.   0.   0.  91.   0.  14.   0.   0. 112.
 7   0.   0.   0.   1.  16.  10.   0.1409.  31.   0.   0.1467.
 8   0.   0.   1.  40.  26. 116.   1. 137.1305.  24.   0.1650.
 9   0.   0.   0.   0.   0.   0.   0.   0.   3. 503.   0. 506.
10   0.   0.   0.   0.   0.   0.   0.   0.   9.   0. 337. 346.
     0. 791.1168. 680. 369. 235.  95.2340.1542. 527. 337.
Normalised accuracy is   82.47155     %
```

Confusion matrix for look up table classifier + Markov
relaxation, order 6, transition probabilities estimated from
classified image.

```
                Classified as ------>
        U    1    2    3    4    5    6    7    8    9   10
 1   0. 784.   0.   0.   0.   0.   0.   0.   1.   0.   0. 785.
 2   0.   0.1175.   1.   0.   0.   0.   0.   1.   0.   0.1177.
 3   0.   0.   0.  94.   0.   0.   0.  61. 616.   0.   0. 771.
 4   0.   0.   0.   0. 983.   0.   0.   0.  25.   0.   0.1008.
 5   0.   0.   0.   0.   0.   0.   0.  48. 214.   0.   0. 262.
 6   0.   0.  73.   0.   0.   0.   0.   0.  39.   0.   0. 112.
 7   0.   0.   0.   0.   0.   0.   0. 306.1161.   0.   0.1467.
 8   0.   0.   0.   0.   0.   0.   0.  62.1588.   0.   0.1650.
 9   0.   0.   0.   0.  60.   0.   0.  83. 363.   0.   0. 506.
10   0.   0.   0.   0. 346.   0.   0.   0.   0.   0.   0. 346.
     0. 784.1248.  95.1389.   0.   0. 560.4008.   0.   0.
Normalised accuracy is   60.98466     %
```

Confusion matrix for decision tree derived by SuperExpert: 40 examples per class.

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10
1	0.	508.	0.	0.	277.	0.	0.	0.	0.	0.	0.
2	0.	0.	1020.	157.	0.	0.	0.	0.	0.	0.	0.
3	0.	0.	2.	725.	10.	4.	0.	4.	20.	0.	6.
4	0.	57.	0.	0.	948.	1.	0.	2.	0.	0.	0.
5	0.	0.	0.	9.	10.	65.	0.	173.	5.	0.	0.
6	0.	0.	0.	112.	0.	0.	0.	0.	0.	0.	0.
7	0.	26.	0.	0.	501.	147.	0.	789.	4.	0.	0.
8	0.	0.	0.	620.	62.	83.	0.	187.	604.	89.	5.
9	0.	0.	0.	71.	0.	1.	0.	0.	5.	386.	43.
10	0.	0.	0.	5.	0.	0.	0.	0.	0.	0.	341.
	0.	591.	1022.	1699.	1808.	301.	0.	1155.	638.	475.	395.

Normalised accuracy is   66.63   %

Confusion matrix for decision tree derived by SuperExpert: 25 examples per class.

Classified as ------>

	U	1	2	3	4	5	6	7	8	9	10
1	0.	508.	0.	0.	277.	0.	0.	0.	0.	0.	0.
2	0.	0.	1105.	57.	0.	0.	15.	0.	0.	0.	0.
3	0.	0.	5.	536.	27.	8.	8.	11.	172.	4.	0.
4	0.	57.	0.	0.	944.	6.	0.	1.	0.	0.	0.
5	0.	0.	0.	6.	5.	71.	0.	179.	1.	0.	0.
6	0.	0.	0.	0.	0.	0.	111.	0.	1.	0.	0.
7	0.	15.	0.	1.	111.	682.	0.	657.	1.	0.	0.
8	0.	0.	0.	93.	48.	114.	476.	198.	681.	40.	0.
9	0.	0.	0.	0.	0.	1.	129.	0.	10.	366.	0.
10	0.	0.	0.	0.	0.	0.	6.	0.	1.	53.	286.
	0.	580.	1110.	693.	1412.	882.	745.	1046.	867.	463.	286.

Normalised accuracy is   65.13   %

Table A6b: YEMEN DATA SET

9*9 Conan on Maximum Likelihood Classifier

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	55.	0.	0.	0.	0.	1.	0.	1.	57.
2	0.	0.	23.	34.	0.	80.	0.	0.	22.	159.
3	0.	0.	0.	222.	0.	0.	0.	0.	0.	222.
4	0.	0.	0.	0.	108.	0.	632.	22.	0.	762.
5	0.	0.	0.	4.	0.	878.	163.	0.	1.	1046.
6	0.	0.	0.	0.	0.	0.	258.	62.	0.	320.
7	0.	0.	0.	0.	0.	0.	11.	220.	0.	231.
8	0.	0.	0.	0.	1.	0.	2.	3.	29.	35.
	0.	55.	23.	260.	109.	958.	1067.	307.	53.	

Normalised accuracy is   63.3121 %

9*9 CONAN on Minimum Distance Classifier

```
                    Classified as ------>
          1     2     3     4     5     6     7     8     9 Total
    1    0.   52.    0.    0.    0.    0.    0.    4.    1.   57.
    2    0.    0.    0.   54.   15.   49.    0.    0.   41.  159.
    3    0.    0.    0.  221.    1.    0.    0.    0.    0.  222.
    4    0.    0.    0.    0.   84.    0.  426.  252.    0.  762.
    5    0.    0.    0.    0.  268.    4.  774.    0.    0.1046.
    6    0.    0.    0.    0.    0.    0.  314.    6.    0.  320.
    7    0.    0.    0.    0.    0.    0.  151.   80.    0.  231.
    8    0.    0.    0.    9.   22.    0.    2.    1.    1.   35.
         0.   52.    0.  284.  390.   53.1667.  343.   43.
```
Normalised accuracy is   26.6949 %

Deviant Distance Classifier

```
                    Classified as ------>
          1     2     3     4     5     6     7     8     9 Total
    1    0.   46.    2.    0.    3.    2.    2.    2.    0.   57.
    2    0.   11.   73.   20.    3.    5.    0.    0.   47.  159.
    3    0.    0.    9.  168.   27.   10.    0.    0.    8.  222.
    4    0.    1.    0.   15.   96.  149.  335.  166.    0.  762.
    5    0.    0.    1.   44.   51.  509.  388.   52.    1.1046.
    6    0.    0.    0.    0.    2.   11.  232.   75.    0.  320.
    7    0.    0.    0.    0.    3.   37.  123.   68.    0.  231.
    8    0.    2.   15.    5.    2.    2.    0.    0.    9.   35.
         0.   60.  100.  252.  187.  725.1080.  363.   65.
```
Normalised accuracy is   42.4082 %

Markov Relaxation (order 9) on Maximum Likelihood
Classifier

```
                    Classified as ------>
          1     2     3     4     5     6     7     8     9 Total
    1    0.   52.    0.    0.    0.    1.    4.    0.    0.   57.
    2    0.    0.   20.    7.    1.  131.    0.    0.    0.  159.
    3    0.    0.    0.  222.    0.    0.    0.    0.    0.  222.
    4    0.    0.    0.    0.   54.   74.  634.    0.    0.  762.
    5    0.    0.    0.    0.    0.1006.   40.    0.    0.1046.
    6    0.    0.    0.    0.    0.    0.  320.    0.    0.  320.
    7    0.    0.    0.    0.    0.    0.   85.  146.    0.  231.
    8    0.    0.    0.    0.    0.   34.    1.    0.    0.   35.
         0.   52.   20.  229.   55.1246.1084.  146.    0.
```
Normalised accuracy is   64.2655 %

Markov Relaxation (order 9) on Minimum Distance Classifier

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	52.	0.	1.	0.	2.	2.	0.	0.	57.
2	0.	0.	0.	31.	0.	128.	0.	0.	0.	159.
3	0.	0.	0.	222.	0.	0.	0.	0.	0.	222.
4	0.	0.	0.	0.	40.	110.	605.	7.	0.	762.
5	0.	0.	0.	0.	38.	545.	463.	0.	0.	1046.
6	0.	0.	0.	0.	0.	0.	320.	0.	0.	320.
7	0.	0.	0.	0.	0.	0.	221.	10.	0.	231.
8	0.	0.	0.	13.	0.	21.	1.	0.	0.	35.
	0.	52.	0.	267.	78.	806.	1612.	17.	0.	

Normalised accuracy is   41.9845 %

Maximum Likelihood Classifier Alone

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	11.	0.	531.	712.	1594.	818.	374.	0.	4040.
2	0.	0.	0.	31.	63.	36.	163.	77.	0.	370.
3	0.	12.	2.	3.	116.	22.	26.	28.	0.	209.
4	0.	24.	4.	67.	18.	57.	0.	0.	0.	170.
5	0.	0.	0.	2.	4.	47.	0.	1.	0.	54.
6	0.	0.	0.	0.	11.	3.	22.	61.	0.	97.
7	0.	0.	0.	0.	1.	1.	10.	23.	0.	35.
8	0.	9.	0.	6.	155.	25.	192.	133.	0.	520.
	0.	56.	6.	640.	1080.	1785.	1231.	697.	0.	

Normalised accuracy is   2.25660 %

9 by 9 Mode Filter on Maximum Likelihood Classifier

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	56.	0.	0.	0.	0.	1.	0.	0.	57.
2	0.	0.	21.	6.	1.	131.	0.	0.	0.	159.
3	0.	0.	0.	222.	0.	0.	0.	0.	0.	222.
4	0.	0.	0.	0.	119.	37.	584.	22.	0.	762.
5	0.	0.	0.	0.	0.	1018.	28.	0.	0.	1046.
6	0.	0.	0.	0.	0.	0.	253.	67.	0.	320.
7	0.	0.	0.	0.	0.	0.	10.	221.	0.	231.
8	0.	0.	0.	0.	0.	31.	1.	3.	0.	35.
	0.	56.	21.	228.	120.	1217.	877.	313.	0.	

Normalised accuracy is   67.4435 %

Minimum Distance Classifier Alone

```
                  Classified as ------>
          1     2     3     4     5     6     7     8    9 Total
   1    0.   44.    3.    0.    5.    0.    2.    3.   0.   57.
   2    0.    0.   19.   31.    7.  102.    0.    0.   0.  159.
   3    0.    0.    0.  169.   35.   18.    0.    0.   0.  222.
   4    0.    0.    0.   16.  143.   89.  323.  191.   0.  762.
   5    0.    0.    0.   38.  143.  392.  387.   86.   0.1046.
   6    0.    0.    0.    0.    4.    6.  222.   88.   0.  320.
   7    0.    0.    0.    0.   18.   20.  119.   74.   0.  231.
   8    0.    0.    2.    9.    4.   20.    0.    0.   0.   35.
        0.   44.   24.  263.  359.  647.1053.  442.   0.
```
Normalised accuracy is   37.5353 %

9 by 9 Mode Filter on Minimum Distance Classifier

```
                  Classified as ------>
          1     2     3     4     5     6     7     8    9 Total
   1    0.   53.    0.    3.    0.    0.    1.    0.   0.   57.
   2    0.    0.    0.   32.    0.  127.    0.    0.   0.  159.
   3    0.    0.    0.  222.    0.    0.    0.    0.   0.  222.
   4    0.    0.    0.    0.   79.   80.  533.   70.   0.  762.
   5    0.    0.    0.    0.   59.  553.  434.    0.   0.1046.
   6    0.    0.    0.    0.    0.    0.  320.    0.   0.  320.
   7    0.    0.    0.    0.    0.    0.  180.   51.   0.  231.
   8    0.    0.    0.   11.    1.   22.    1.    0.   0.   35.
        0.   53.    0.  268.  139.  782.1469.  121.   0.
```
Normalised accuracy is   45.1271 %

SAR (8,1) on Maximum Likelihood Classifier

```
                  Classified as ------>
          1     2     3     4     5     6     7     8    9 Total
   1    0.   57.    0.    0.    0.    0.    0.    0.   0.   57.
   2    0.    0.   23.   13.    0.  123.    0.    0.   0.  159.
   3    0.    0.    0.  220.    0.    2.    0.    0.   0.  222.
   4    0.    0.    0.    0.  222.   43.  497.    0.   0.  762.
   5    0.    0.    0.    2.    0.1000.   44.    0.   0.1046.
   6    0.    0.    0.    0.    0.    0.  270.   50.   0.  320.
   7    0.    0.    0.    1.    0.    0.   38.  192.   0.  231.
   8    0.    0.    0.    0.    0.   27.    4.    4.   0.   35.
        0.   57.   23.  236.  222.1195.  853.  246.   0.
```
Normalised accuracy is   70.0565 %

SAR (8,1) on Minimum Distance Classifier

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	55.	0.	1.	0.	0.	1.	0.	0.	57.
2	0.	0.	14.	19.	0.	126.	0.	0.	0.	159.
3	0.	0.	0.	221.	0.	1.	0.	0.	0.	222.
4	0.	0.	0.	0.	96.	0.	666.	0.	0.	762.
5	0.	0.	0.	0.	89.	533.	424.	0.	0.	1046.
6	0.	0.	0.	0.	0.	0.	320.	0.	0.	320.
7	0.	0.	0.	1.	0.	0.	227.	3.	0.	231.
8	0.	0.	0.	4.	1.	26.	4.	0.	0.	35.
	0.	55.	14.	246.	186.	686.	1642.	3.	0.	

Normalised accuracy is   43.8559 %

Table A6c: North Wales Data Set

9 by 9 CONAN on Minimum Distance Classifier

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	4040.	0.	0.	0.	0.	0.	0.	0.	4040.
2	0.	0.	370.	0.	0.	0.	0.	0.	0.	370.
3	0.	0.	0.	209.	0.	0.	0.	0.	0.	209.
4	0.	0.	130.	2.	37.	0.	1.	0.	0.	170.
5	0.	0.	0.	0.	0.	54.	0.	0.	0.	54.
6	0.	0.	0.	0.	0.	0.	97.	0.	0.	97.
7	0.	0.	0.	0.	0.	0.	0.	21.	14.	35.
8	0.	0.	0.	0.	0.	0.	0.	0.	520.	520.
	0.	4040.	500.	211.	37.	54.	98.	21.	534.	

Normalised accuracy is   97.3248 %

9 by 9 CONAN on Maximum Likelihood Classifier

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	4040.	0.	0.	0.	0.	0.	0.	0.	4040.
2	0.	0.	370.	0.	0.	0.	0.	0.	0.	370.
3	0.	0.	0.	209.	0.	0.	0.	0.	0.	209.
4	0.	0.	10.	4.	155.	0.	1.	0.	0.	170.
5	0.	0.	34.	0.	0.	0.	0.	0.	20.	54.
6	0.	0.	0.	0.	1.	0.	96.	0.	0.	97.
7	0.	0.	11.	1.	0.	0.	0.	10.	13.	35.
8	0.	0.	0.	0.	0.	0.	0.	0.	520.	520.
	0.	4040.	425.	214.	156.	0.	97.	10.	553.	

Normalised accuracy is   98.2712 %

Deviant Distance Classifier Alone

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.4039.	1.	0.	0.	0.	0.	0.			0.4040.
2	0.	0.	362.	0.	4.	0.	0.	3.	1.	370.
3	0.	0.	2.	166.	1.	0.	34.	0.	6.	209.
4	0.	0.	118.	2.	46.	0.	0.	0.	4.	170.
5	0.	0.	0.	0.	0.	53.	0.	0.	1.	54.
6	0.	0.	0.	4.	0.	0.	93.	0.	0.	97.
7	0.	0.	24.	1.	0.	0.	1.	3.	6.	35.
8	0.	0.	111.	0.	1.	0.	0.	0.	408.	520.
	0.4039.	618.	173.	52.	53.	128.	6.	426.		

Normalised accuracy is   94.0855 %

Markov Relaxation (Order 10) on Maximum Likelihood Classifier

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.4040.	0.	0.	0.	0.	0.	0.			0.4040.
2	0.	0.	370.	0.	0.	0.	0.	0.	0.	370.
3	0.	0.	0.	209.	0.	0.	0.	0.	0.	209.
4	0.	0.	24.	8.	127.	0.	0.	0.	11.	170.
5	0.	0.	24.	0.	0.	0.	0.	0.	30.	54.
6	0.	0.	0.	4.	0.	0.	93.	0.	0.	97.
7	0.	0.	13.	0.	0.	0.	0.	0.	22.	35.
8	0.	0.	0.	0.	0.	0.	0.	0.	520.	520.
	0.4040.	431.	221.	127.	0.	93.	0.	583.		

Normalised accuracy is   97.5250 %

Markov Relaxation (Order 10) on Minimum Distance Classifier

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.4040.	0.	0.	0.	0.	0.	0.			0.4040.
2	0.	0.	370.	0.	0.	0.	0.	0.	0.	370.
3	0.	0.	0.	209.	0.	0.	0.	0.	0.	209.
4	0.	0.	145.	5.	16.	0.	0.	0.	4.	170.
5	0.	0.	0.	0.	0.	54.	0.	0.	0.	54.
6	0.	0.	0.	2.	0.	0.	95.	0.	0.	97.
7	0.	0.	0.	0.	0.	0.	0.	0.	35.	35.
8	0.	0.	0.	0.	0.	0.	0.	0.	520.	520.
	0.4040.	515.	216.	16.	54.	95.	0.	559.		

Normalised accuracy is   96.5241 %

Maximum Likelihood Classifier Alone

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	4038.	1.	0.	1.	0.	0.	0.	0.	4040.
2	0.	0.	359.	0.	5.	0.	0.	6.	0.	370.
3	0.	0.	1.	185.	1.	0.	20.	0.	2.	209.
4	0.	0.	57.	2.	108.	0.	0.	0.	3.	170.
5	0.	0.	34.	0.	0.	5.	0.	0.	15.	54.
6	0.	0.	0.	8.	0.	0.	89.	0.	0.	97.
7	0.	0.	25.	4.	0.	0.	0.	4.	2.	35.
8	0.	0.	64.	2.	7.	0.	0.	0.	447.	520.
	0.	4038.	541.	201.	122.	5.	109.	10.	469.	

Normalised accuracy is 95.2684 %

Maximum Likelihood Classifier plus 9 by 9 Mode Filter

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	4040.	0.	0.	0.	0.	0.	0.	0.	4040.
2	0.	0.	370.	0.	0.	0.	0.	0.	0.	370.
3	0.	0.	0.	209.	0.	0.	0.	0.	0.	209.
4	0.	0.	17.	6.	145.	0.	1.	0.	1.	170.
5	0.	0.	32.	0.	0.	0.	0.	0.	22.	54.
6	0.	0.	0.	1.	0.	0.	96.	0.	0.	97.
7	0.	0.	15.	3.	0.	0.	0.	0.	17.	35.
8	0.	0.	0.	0.	0.	0.	0.	0.	520.	520.
	0.	4040.	434.	219.	145.	0.	97.	0.	560.	

Normalised accuracy is 97.9072 %

Minimum Distance Classifier Alone

Classified as ------>

	1	2	3	4	5	6	7	8	9	Total
1	0.	4040.	0.	0.	0.	0.	0.	0.	0.	4040.
2	0.	0.	343.	0.	6.	0.	0.	1.	20.	370.
3	0.	0.	3.	176.	1.	0.	25.	0.	4.	209.
4	0.	0.	102.	3.	63.	0.	0.	0.	2.	170.
5	0.	0.	0.	0.	0.	51.	0.	2.	1.	54.
6	0.	0.	0.	6.	0.	0.	91.	0.	0.	97.
7	0.	0.	7.	2.	0.	0.	1.	21.	4.	35.
8	0.	0.	54.	2.	4.	0.	0.	0.	460.	520.
	0.	4040.	509.	189.	74.	51.	117.	24.	491.	

Normalised accuracy is 95.4504 %

Minimum Distance Classifier plus 9 by 9 Mode Filter

```
                Classified as ------>
         1     2     3     4     5     6     7     8     9 Total
    1  0.4040.   0.    0.    0.    0.    0.    0.    0.4040.
    2  0.    0. 370.    0.    0.    0.    0.    0.    0. 370.
    3  0.    0.    0. 209.    0.    0.    0.    0.    0. 209.
    4  0.    0. 130.    7.   31.    0.    1.    0.    1. 170.
    5  0.    0.    0.    0.    0.   54.    0.    0.    0.  54.
    6  0.    0.    0.    0.    0.    0.   97.    0.    0.  97.
    7  0.    0.    0.    1.    0.    0.    0.   10.   24.  35.
    8  0.    0.    0.    0.    0.    0.    0.    0. 520. 520.
       0.4040. 500. 217.   31.   54.   98.   10. 545.
Normalised accuracy is      97.0155 %
```

Small Area Replacement (8,1) on Maximum Likelihood Classifier

```
                Classified as ------>
         1     2     3     4     5     6     7     8     9 Total
    1  0.4040.   0.    0.    0.    0.    0.    0.    0.4040.
    2  0.    0. 370.    0.    0.    0.    0.    0.    0. 370.
    3  0.    0.    0. 209.    0.    0.    0.    0.    0. 209.
    4  0.    0.   32.   26.   95.    0.    3.    0.   14. 170.
    5  0.    0.   35.    0.    0.    0.    0.    0.   19.  54.
    6  0.    0.    0.    0.    0.    0.   97.    0.    0.  97.
    7  0.    0.    0.    0.    0.    0.    0.    0.   35.  35.
    8  0.    0.    0.    0.    0.    0.    0.    0. 520. 520.
       0.4040. 437. 235.   95.    0. 100.    0. 588.
Normalised accuracy is      97.0155 %
```

SAR (8,1) on Minimum Distance Classifier

```
                Classified as ------>
         1     2     3     4     5     6     7     8     9 Total
    1  0.4040.   0.    0.    0.    0.    0.    0.    0.4040.
    2  0.    0. 353.    0.    0.    0.    0.    0.   17. 370.
    3  0.    0.    0. 209.    0.    0.    0.    0.    0. 209.
    4  0.    0. 131.   15.   15.    0.    3.    0.    6. 170.
    5  0.    0.    0.    0.    0.   54.    0.    0.    0.  54.
    6  0.    0.    0.    2.    0.    0.   95.    0.    0.  97.
    7  0.    0.    0.    0.    0.    0.    0.    0.   35.  35.
    8  0.    0.   12.    0.    0.    0.    0.    0. 508. 520.
       0.4040. 496. 226.   15.   54.   98.    0. 566.
Normalised accuracy is      95.9782 %
```