THREE-DIMENSIONAL GEOLOGICAL STRUCTURAL FEATURES FROM REMOTELY-SENSED IMAGES AND DIGITAL ELEVATION MODELS

KEVIN PETER MORRIS

University College London

Thesis submitted for the degree of
Doctor of Philosophy

University of London

December 1994
For Andrea and Aiden.

So I dream of Columbus
Every time that the panic starts
I dream of Columbus
With my maps and my beautiful charts
I dream of Columbus
And there's peace in a travelling heart
I dream of Columbus

Written by Noel Brazil (1989)
Sung by Mary Black
Abstract

Accurate mapping of geological structures is important in numerous applications, ranging from mineral exploration through to hydrogeological modelling. Remotely sensed data can provide synoptic views of study areas enabling mapping of geological units within the area. Structural information may be derived from such data using standard manual photo-geologic interpretation techniques, although these are often inaccurate and incomplete. The aim of this thesis is, therefore, to compile a suite of automated and interactive computer-based analysis routines, designed to help the user map geological structure. These are examined and integrated in the context of an expert system.

The data used in this study include Digital Elevation Model (DEM) and Airborne Thematic Mapper images, both with a spatial resolution of 5m, for a 5 x 5 km area surrounding Llyn Cowlyd, Snowdonia, North Wales. The geology of this area comprises folded and faulted Ordovician sediments intruded throughout by dolerite sills, providing a stringent test for the automated and semi-automated procedures.

The DEM is used to highlight geomorphological features which may represent surface expressions of the sub-surface geology. The DEM is created from digitized contours, for which kriging is found to provide the best interpolation routine, based on a number of quantitative measures. Lambertian shading and the creation of slope and change of slope datasets are shown to provide the most successful enhancement of DEMs, in terms of highlighting a range of key geomorphological features. The digital image data are used to identify rock outcrops as well as lithologically controlled features in the land cover. To this end, a series of standard spectral enhancements of the images is examined. In this respect, the least correlated 3 band composite and a principal component composite are shown to give the best visual discrimination of geological and vegetation cover types.

Automatic edge detection (followed by line thinning and extraction) and manual interpretation techniques are used to identify a set of 'geological primitives' (linear or arc features representing lithological boundaries) within these data. Inclusion of the DEM data provides the three-dimensional co-ordinates of these primitives enabling a least-squares fit to be employed to calculate dip and strike values, based, initially, on the assumption of a simple, linearly dipping structural model.

A very large number of scene 'primitives' is identified using these procedures, only some of which have geological significance. Knowledge-based rules are therefore used to identify the relevant. For example, rules are developed to identify lake edges, forest boundaries, forest tracks, rock-vegetation boundaries, and areas of geomorphological interest. Confidence in the geological significance of some of the geological primitives is increased where they are found independently in both the DEM and remotely sensed data.

The dip and strike values derived in this way are compared to information taken from the published geological map for this area, as well as measurements taken in the field. Many results are shown to correspond closely to those taken from the map and in the field, with an error of < 1°. These data and rules are incorporated into an expert system which, initially, produces a simple model of the geological structure. The system also provides a graphical user interface for manual control and interpretation, where necessary. Although the system currently only allows a relatively simple structural model (linearly dipping with faulting), in the future it will be possible to extend the system to model more complex features, such as anticlines, synclines, thrusts, nappes, and igneous intrusions.
Chapter 1 - Introduction

1.1 Traditional Rôle of Remote Sensing in Geological Applications

1.2 Rôle of Digital Elevation Data in Geological Mapping

1.3 Requirements for Automation

1.4 Requirements for a Semi-Automated/Fully Automated Geological Mapping System

1.5 Structure of the Thesis

Chapter 2 - Advances in Geological Applications of Remote Sensing
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.3 Spatial Analyses and Lineament Detection</td>
<td>42</td>
</tr>
<tr>
<td>2.2.4 Spectral Mixture Models</td>
<td>44</td>
</tr>
<tr>
<td>2.3 Application of Digital Elevation Data</td>
<td>46</td>
</tr>
<tr>
<td>2.3.1 Drainage Networks</td>
<td>47</td>
</tr>
<tr>
<td>2.3.2 Inferring Structural Geology</td>
<td>48</td>
</tr>
<tr>
<td>2.4 Image Understanding</td>
<td>54</td>
</tr>
<tr>
<td>2.4.1 Understanding the Image Formation Process - Topographic Effects</td>
<td>55</td>
</tr>
<tr>
<td>2.4.2 Expert Systems</td>
<td>57</td>
</tr>
<tr>
<td>2.4.3 Examples of Knowledge-Based Systems Applied to Object Recognition in Images</td>
<td>58</td>
</tr>
<tr>
<td>2.5 Computer Vision</td>
<td>60</td>
</tr>
<tr>
<td>2.5.1 Image Segmentation</td>
<td>61</td>
</tr>
<tr>
<td>2.5.2 Edge Detection and Pattern Recognition</td>
<td>64</td>
</tr>
<tr>
<td>2.6 Aim and objectives</td>
<td>67</td>
</tr>
<tr>
<td>2.6.1 Summary of literature review</td>
<td>67</td>
</tr>
<tr>
<td>2.6.2 Aim</td>
<td>69</td>
</tr>
<tr>
<td>2.6.3 Objectives</td>
<td>69</td>
</tr>
<tr>
<td>Chapter 3 - Study Area</td>
<td>70</td>
</tr>
<tr>
<td>3.1 Study Area</td>
<td>70</td>
</tr>
<tr>
<td>3.2 Geology of the study area</td>
<td>70</td>
</tr>
<tr>
<td>Chapter 4 - Deriving Structural Geological Information from Remotely Sensed Images</td>
<td>75</td>
</tr>
<tr>
<td>4.1 Pre-processing of remotely sensed images</td>
<td>76</td>
</tr>
<tr>
<td>4.1.1 Special characteristics of the data</td>
<td>77</td>
</tr>
<tr>
<td>4.1.2 Radiometric calibration of the data</td>
<td>80</td>
</tr>
<tr>
<td>4.1.3 Geometric Correction</td>
<td>83</td>
</tr>
<tr>
<td>4.2 Extracting geological information from remotely sensed data</td>
<td>92</td>
</tr>
<tr>
<td>4.2.1 Colour enhancement of lithological features</td>
<td>92</td>
</tr>
<tr>
<td>4.2.2 Enhancement of monochrome images</td>
<td>100</td>
</tr>
<tr>
<td>4.2.3 Segmenting the scene into geologically/non-geologically</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5 - Digital Elevation Models - generation and extraction of 'primitive' geological features

5.1 Generation of digital elevation models (DEMs)
  5.1.1 Methods for generating DEMs
  5.1.2 Evaluation of DEM accuracy
    5.1.2.1 Comparison of spot heights
    5.1.2.2 Checking the integrity of the DEM by recontouring
    5.1.2.3 Comparison of DEMs by height differencing
    5.1.2.4 Identification of significant morphological features through drainage network simulation
    5.1.2.5 Identification of minor artefacts by Lambertian shading
    5.1.2.6 Summary
  5.2 Extracting geological information from DEMs
    5.2.1 Techniques to enhance geological information in digital elevation data
    5.2.2 Evaluation of Techniques for Segmenting Digital Elevation Data
    5.2.3 Extracting primitive information from DEMs
  5.3 Assessing success of DEMs in extracting useful geological information
    5.3.1 Manual interpretation using geomorphologically enhanced images
    5.3.2 Segmenting geological and non-geological features using geomorphological descriptors
    5.3.3 Evaluation of primitives automatically identified from elevation data
  5.4 Site specific conditions and consequences for other study areas
  5.5 Summary
Chapter 6 - Structural Mapping from Remotely Sensed Images and DEMs

6.1 Deriving structural measurements
   6.1.1 Estimating dip and strike
   6.1.2 Implementation of dip and strike estimation
   6.1.3 Evaluation of dip and strike results
   6.1.4 Error Analysis of Dip and Strike

6.2 Methods for Evaluating the Geological Significance of Detected Primitive Features
   6.2.1 Representing Knowledge
   6.2.2 Implementation of binary and confidence based knowledge-based rules
   6.2.3 Design and implementation of further knowledge-based rules

6.3 Production of a Geological Model
   6.3.1 The Initial Model
   6.3.2 Building a More Complex Model
   6.3.3 Map Products
   6.3.4 Improvements for the geological modelling procedure

6.4 Conclusions

Chapter 7 - Progress Towards a Structural Mapping Expert System

7.1 The Proposed Expert System
7.2 Benefits of Expert Systems
7.3 Knowledge-based Expert Systems
7.4 Components of an Expert System
   7.4.1 Knowledge Acquisition
   7.4.2 Control Strategies for the Application of Knowledge-Based Rules - the Inference Engine
   7.4.3 Machine Learning
   7.4.4 Intelligent User-interface

7.5 Current Implementation of the Proposed System
   7.5.1 Database of input parameters
Chapter 8 - Conclusions

8.1 Requirements for an Automated Geological Mapping System
8.2 Evaluation of Data and Techniques to Extract Geological Structural Measurements
8.3 Development of an Expert System
8.4 Problems Encountered During this Study
8.5 Areas Requiring Further Work
8.6 General Conclusions

Glossary and Abbreviations

Acknowledgements

References

Appendix A - Calculating Dip and Strike Values from Best Fit Coefficients
List of Figures

Chapter 1

1.1 SPOT satellite image of Djebel Armour, Algeria, showing a series of parallel anticlines and synclines 22
1.2 Landsat MSS data of an area of the Anti Atlas Mountains in Morocco, merged with Heat Capacity Mapping Mission data (from Colwell 1983) 22

Chapter 2

2.1 The electromagnetic spectrum (from Colwell 1983) 28
2.2 Example laboratory mineral spectra (after Grove et al. 1992) 30
2.3 Effects of terrain relief on SAR images (from Lillesand and Keifer 1979) 36
2.4 Example of layover effect in radar images - ERS-1 data showing Mount Vesuvius and surrounds. 37
2.5 The principal component transformation. I1 and I2 are the DNs of two of the original bands. I1' and I2' are the transformed bands such that there is the least correlation between them (from Rees 1990) 40
2.6 Conventional image contrast stretching (a-c) and decorrelation stretching (d-f). Conventional contrast stretching is performed parallel to the input band axes - stretching parallel to the principal component axes allows a much fuller use of the display colour space (from Rothery 1987a) 40
2.7 Satellite image data combined with geophysical data using an IHS transformation. The geophysical data is provided as hue information while the satellite image is shown as intensity. Both data sets are more interpretable when combined. 43
2.8 Binary encoding of mineral spectra according to Taranik and Kruse (1989). If the reflectance of the mineral decreases with increasing wavelength then it is encoded as -1, if it remains the same then the code is 0, or if it increases then a value of 1 is assigned. 45
2.9 Illustrating the derivation of relative absorption band depth images from reflectance values taken on the shoulders and within the trough of an absorption feature. After Crowley et al. (1989) 45
2.10 Principles of elementary profile analysis demonstrating how critical points along a profile are shown by sequences of point slope determinations. A - flat bottom, V - sharp valley, C - sharp change in slope, E - plateau, E1 - upgrading horizontal step, E2 - downgrading horizontal step, S - crest or summit, P - regular slope, U - concave slope, N - convex slope (from Chorowicz et al. 1989)

2.11 (a) model of veers in directions along a topographic edge at the intersection of a valley to form a "V", (b) more typical example of valley-Vs encountered in the image leading to errors of omission (from Wadge et al. 1990)

2.12 Schematic cross section to illustrate local surface illumination geometry, location of mapped topographic edges and the relationship to the planar bedding surfaces. I = angle of incidence, e = angle of emittance, and g = phase angle. (from Wadge et al. 1990)

2.13 Component modelling of sky radiance - direct sky component FD, diffuse sky component Fd, direct ground component FgD, diffuse ground component Fgd (after Woodham and Gray 1987)

2.14 Figure showing how three points may be used to define a plane, which may in turn be used to extract dip and strike values

2.15 Different type edges displayed one dimensionally

2.16 The Sobel filters. The results of both filters can be combined to produce edge strength and edge orientation images

2.17 Schematic illustration of the Hough transform method. (a) there are three nearly colinear points in the (x,y) plane and the corresponding curves in the (p,θ) plane are shown in (b)

Chapter 3

3.1 The Llyn Cowlyd study area
3.2 Views of the Llyn Cowlyd study area
3.3 Geology of the Llyn Cowlyd area, derived from the published 1:50,000 geology map (BGS 1985)

Chapter 4

4.1 The irradiant and radiant flux components which must be measured to determine reflectance accurately (after Woodham and Gray 1987)
4.2 ATM image of Llyn Cowlyd
4.3 Spectral curves for ATM bands using Spectron wavebands
4.4 Lambertian shaded DEM produced from stereo-matching of two aerial photographs of the Llyn Cowlyd study area. The areas of uniform grey tone indicate errors where the stereo-matcher has failed. The area shown is centred on GR 272700 361500 and covers an area approximately 800 x 500 m (north is towards the top of the page).
4.2 ATM image of Llyn Cowlyd

Source - David Allison, University College London.

4.5 Orthoimage of aerial photography of Llyn Cowlyd area, corrected using a stereo-matched DEM. The white areas indicate errors where the stereo-matcher has failed. The area shown is centred on GR 272700 361500 and covers an area approximately 300 x 400 m (north is towards the top of the page). Source - David Allison, University College London.

4.6 Orthoimage of ATM data (band 7) of Llyn Cowlyd area, corrected using a stereo-matched DEM. The black areas indicate errors where the stereo-matcher has failed. The area shown is centred on GR 272700 361500 and covers an area approximately 800 x 500 m (north is towards the bottom of the page). Source - David Allison, University College London.

4.7 Control points identified for the polynomial warp

4.8 Geometrically corrected image of the Llyn Cowlyd area (5 x 5 km) showing the 1.5 x 1.5km extract area as a black outline.

4.9 Perspective view of errors in X and Y, from the polynomial warp

4.10 Colour enhancements of the ATM images, Part 1

4.11 Colour enhancements of the ATM images, Part 2

4.12 Primitives identified manually from colour enhancements

4.13 Fourth principal component

4.14 Figure 4.14 - Comparison of contrast stretching techniques applied to band 7, showing automatically derived primitives as a solid line overlay.

4.15 Training areas and classified images

4.16 Probability images derived from the maximum likelihood classification. The lighter tones indicate a higher probability of that class occurring at that location.

4.17 Removing non-geological primitives using segmentation techniques

4.18 Procedure for extracting primitives

4.19 Comparison between whole-scene processing and window-based processing. More primitives are identified using the latter techniques due to the use of local thresholds.

4.20 Comparison of edge detection algorithms

4.21 Primitives automatically derived from remotely sensed images

4.22 Lithological boundary identified between vegetation and exposed rock

4.23 ATM band 11, showing strong contrast at the vegetation-rock boundaries

4.24 The band ratios 9/10 and 9/2

4.25 Band ratio 7/5 - the vegetation index

4.26 Principal components 1 and 2

4.27 Methods for extracting geological information from regions
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Typical relationships between topography and geology</td>
</tr>
<tr>
<td>5.2</td>
<td>Samples of different drainage patterns (after Argialas et al. 1988)</td>
</tr>
<tr>
<td>5.3</td>
<td>Overview of methods used in Chapter 5</td>
</tr>
<tr>
<td>5.4</td>
<td>Line drawing showing the derivation of the parallax equation (from Colwell 1983)</td>
</tr>
<tr>
<td>5.5</td>
<td>Finding the closest points in each quadrant and determining the Euclidean distance</td>
</tr>
<tr>
<td>5.6</td>
<td>Example of a triangulated irregular network created between digitized contour points</td>
</tr>
<tr>
<td>5.7</td>
<td>Creation of Theissen polygons around digitized contour points</td>
</tr>
<tr>
<td>5.8</td>
<td>Creation of new Theissen polygon around new grid cell showing area captured from each polygon</td>
</tr>
<tr>
<td>5.9</td>
<td>Semi-variogram - plot of semi-variance ($C(h)$) against vector distance ($h$). $K(0)$ represents the variance of the data, and $e$ is a small value below which $C(h)$ is considered to be equal to the variance</td>
</tr>
<tr>
<td>5.10</td>
<td>Semi-variogram of digitized contour data</td>
</tr>
<tr>
<td>5.11</td>
<td>Digital Elevation Model for a 5x5 km area surrounding Llyn Cowlyd</td>
</tr>
<tr>
<td>5.12</td>
<td>Recontoured DEM showing the smoothing of 5 metre contour lines</td>
</tr>
<tr>
<td>5.13</td>
<td>Standard interpolation routines use the closest data points found within a search radius, rather than matching features in successive contours</td>
</tr>
<tr>
<td>5.14</td>
<td>Height differences between DEMs 1 and 2. Brighter tones indicates larger differences between the two DEMs. The figure highlights a stepping effect caused by the interpolation, between neighbouring contours</td>
</tr>
<tr>
<td>5.15</td>
<td>Stream profile showing the stepping effect introduced by interpolation</td>
</tr>
<tr>
<td>5.16</td>
<td>Diagram showing the flow of raindrops over the DEM</td>
</tr>
<tr>
<td>5.17</td>
<td>Diagram indicating the increase in stream size as more raindrops are added</td>
</tr>
<tr>
<td>5.18</td>
<td>Diagram showing the anticlockwise search for the lowest neighbour. The random starting points means that different paths may be chosen if more than one neighbour has the lowest height</td>
</tr>
<tr>
<td>5.19</td>
<td>Digitized and simulated drainage networks for various DEMs</td>
</tr>
<tr>
<td>5.20</td>
<td>Simulated drainage networks for various DEMs</td>
</tr>
<tr>
<td>5.21</td>
<td>Diagram showing the phase angle between the illumination angle and the normal to the slope</td>
</tr>
<tr>
<td>5.22</td>
<td>Lambertian shaded DEMs produced from methods 2, 5, 6, and 7, showing the artefacts introduced by the interpolation methods</td>
</tr>
<tr>
<td>5.23</td>
<td>Vector summation of two orthogonal edges to produce slope magnitude and direction</td>
</tr>
</tbody>
</table>
Chapter 5

5.24 Images illustrating the effect of changing the solar zenith angle 172
5.25 Images illustrating the effect of changing the solar azimuth angle 173
5.26 Simulated drainage network showing that streams relate to lithological boundaries (A, B, and C) 174
5.27 Comparison of different perspective views 176
5.28 Comparison of perspective view with ground photograph 177
5.29 Perspective view illustrating that values of dip can be measured directly from the image 178
5.30 Products derived from the DEM 182
5.31 Primitives identified manually 184
5.32 Primitives identified manually from Lambertian shaded images 185
5.33 Primitives identified using a perspective view of the terrain 187
5.34 Texture measures applied to the DEM 188
5.35 Masked Lambertian shaded DEMs using the standard deviation and contrast texture measures to derive the masks. Each image shows the relative success of using DEM texture to segment geological from non-geologically important areas 189
5.36 Texture measure averaged over eight Lambertian shaded images 191
5.37 Primitives identified automatically 192
5.38 Diagram illustrating how the position of edges may shift due to the illumination angle 194
5.39 All automatically identified features 196

Chapter 6

6.1 Planar surface fit to 3D primitive data 199
6.2 Figure showing how three points may be used to define a plane, which may in turn be used to extract dip and strike values 201
6.3 Expected accuracy according to the relative location of three points 202
6.4 Diagram illustrating the arbitrary fit through a linear feature 204
6.5 Colour coding of individual primitives for the display of attribute information. Values increase from magenta, through blue, green, yellow, to red. Here the colours indicate increasing $R^2$ values from 0.34 (magenta) to 0.99 (red). 206
6.6 A typical interactive display showing two digitized primitives with dip and strike values overlaid. In the background is a second window illustrating some of the attributes associated with one of the primitives 208
6.7 Published 1:50,000 geological map 210
6.8 Dip and strike measurements taken in the field 211
6.9 Dip and strike results derived automatically from the DEM 213
6.10 Dip and strike results derived automatically from the slope data 214
6.11 Dip and strike results derived automatically from the aspect data 215
6.12 Dip and strike results derived automatically from the Lamb-45 DEM 216
6.13 Dip and strike results derived automatically from the Lamb-90 DEM 216
6.14 Dip and strike results derived automatically from band 7

6.15 Dip and strike results derived automatically from band 9

6.16 Dip and strike results derived automatically from the vegetation index

6.17 Dip and strike results derived automatically from the first PC

6.18 Dip and strike results derived manually from the slope data

6.19 Dip and strike results derived manually from the best ITI

6.20 Perspective view showing geological units on the eastern slopes of Llyn Cowlyd

6.21 A two step procedure used to evaluate the accuracy in dip measurements obtained from two different study areas using stereo SPOT (after Berger et al. 1992)

6.22 Hemispherical DEM - Lambertian shaded

6.23 Simulated geological unit overlaying the DEM

6.24 Modifications to the hemispherical DEM

6.25 Addition of random noise to the hemispherical DEM

6.26 Addition of random noise to positioning of simulated primitive

6.27 Calculation of curvature, by fitting a second-order polynomial to the projected primitive

6.28 Dip v. dip accuracy

6.29 Curvature v. DEM shape factor at different dip angles

6.30 Dip accuracy v. DEM shape factor at different dip angles

6.31 DEM shape v. dip accuracy (D/S 45/135)

6.32 Error in DEM v. dip accuracy (peaked hemisphere - D/S 60/135)

6.33 Error in DEM v. dip accuracy (hemisphere - D/S 60/135)

6.34 Length of arc v. dip accuracy (D/S 45/135) - lower end at 0.3

6.35 Perspective view of errors in X, Y (both derived from the geometric correction, and Z (from the spot height evaluation)

6.36 Fuzzy set description for the 'flatness' of a primitive

6.37 Gaussian fuzzy sets used in the lake-edge identification rule

6.38 Number of lake edges identified by varying two thresholds

6.39 Lake edges (in red) identified using a probabilistic approach

6.40 Lake edges removed using a threshold approach. N.B. primitives A, B, and C from Figure 6.39 have been removed erroneously

6.41 Results of the forest boundary identification rule. The orange and red primitives (mid-right) have identified forest boundaries successfully

6.42 Results of the vegetation identification rule. Most of the primitives shown here border vegetation

6.43 Results of the lake identification rule. The red and yellow primitives show that the lake edges have been identified successfully

6.44 Results of the rock identification rule. Most rock features are identified, however, there is some confusion at the edges of areas of shadow

6.45 Results of the roughness rule. Most of the rock exposures have a high value for the roughness rule as expected
6.46 Results of the combination rule. The yellow to orange primitives are identified most from different input data sets 263

6.47 An example stereonet (the Wulff net) (after Hobbs et al. 1976) 265

6.48 The principle of stereographic projection (after Hobbs et al. 1976) 266

6.49 The stereographic projection of a normal (pole) to a plane (after Hobbs et al. 1976) 267

6.50 Contouring the poles on the stereonet (above) and example structure and corresponding plots (below) (after Hobbs et al. 1976) 268

6.51 Sample stereonet of primitives derived from band 7 269

6.52 Final confidence map 271

6.53 Thresholded confidence map (threshold set at 0.3) 271

6.54 Simple structures in which hydrocarbons accumulate. (A) Anticlinal fold; (B) Unconformity; (C) Stratigraphic, by change of facies; (D) Fault. 272

6.55 Creation of the initial geological model 275

6.56 Bedding thicknesses derived by calculating the perpendicular distance between those primitives having the most common orientation (MCO). 276

6.57 Error planes emanating from the primitive centres. Overlapping planes are used to combine primitives which are considered to belong to identical lithological units 277

6.58 Revised bedding thicknesses after combination of similar planes 278

6.59 Geological model using above bedding sequence 278

6.60 Stereonet showing two concentrations, indicating a folded structure 280

6.61 Different local fold scenarios - illustrating the difficulty in matching bedding thicknesses 281

6.62 Comparison of bedding sequences on either side of a fold 282

6.63 Manual segmentation of the study area into structural units 284

6.64 Solving each planar equation for each pixel in the image to determine which geological layer is exposed at the surface 285

6.65 Models 1 and 2 286

6.66 Models 3 and 4 287

6.67 A simplified version of the published 1:50,000 geological map, with the manually interpreted structural segments. A dolerite intrusion occurs at X 288

6.68 A 3D voxel representation of an ore body (after Raper 1989) 291

6.69 Solving each planar equation for each pixel at each depth in the output volume to determine which geological layer is present 292

6.70 A cross-section derived from the 3D model 293

6.71 A solid 3D block diagram showing the geological structure 293

6.72 Block diagram with surface visible through transparent 'soil' layer 294

6.73 Geological unit projected into the above ground space for visualization purposes 294

6.74 Geological phenomena which can confuse structural interpretations 296
Chapter 7

7.1 Schematic flow chart of the proposed expert system 300
7.2 The standard architecture of a knowledge-based expert system (after Harmon and King 1985) 305
7.3 General pattern of professional development (after Harmon and King 1985) 307
7.4 Backward chaining through a search tree 309
7.5 Components of a blackboard system (after Hallam 1990) 311
7.6 An example of rule induction for classifying mammals 313
7.7 A node is connected to several other nodes by links. Its state depends on the values of other nodes and on the connective strengths and thresholds (after Hart 1990) 314
7.8 A multi-layer neural network (after Hart 1990) 314
7.9 System and user models (after Bass and Coutaz 1988) 316
7.10 Windows features, used to design user-friendly interfaces 318
7.11 Errors of fit for individual vertices on a primitive 319
7.12 Library index for object-specific rules 322
7.13 Four terrain classes used in the induced learning example 326
7.14 Number of primitives found for each terrain class 327
7.15 Interactive display of input data (DEM on left and best ITI on right) 329
7.16 Interactive editing of primitives. Colours are used to represent different lithologies (N.B. orange is used to represent faulting) 330
7.17 Attribute information displayed interactively 331
7.18 Display of dip and strike values 332
7.19 GUI allowing modification of parameters 333
7.20 Interactive display of perspective views 335
7.21 GUI allowing real-time manipulation of a wireframe perspective view 336
7.22 GUI allowing real-time manipulation of Lambertian shading 337
7.23 Display of single or multiple simulated streams 338
7.24 Current processing stages within the expert system 339
7.25 The 'completed', started, and untouched parts of the proposed expert system 342
7.26 Possible sub-systems for the proposed blackboard architecture 343
Chapter 2

2.1 Characteristics of Visible to SWIR satellite sensors used in geological applications 32
2.2 Characteristics of Visible to SWIR future satellite and current airborne sensors used in geological applications 33

Chapter 4

4.1 Wavelengths of the ATM AADS 1268 scanner 77
4.2 Specifications of the ATM AADS 1268 scanner 80
4.3 Calibration data for the Daedalus ATM scanner - calibrated 13th July, 1989 81
4.4 Geometric correction accuracy as a function of the order of polynomial used 89
4.5 ATM Wavebands and image processing techniques generally considered important in mapping selected surface characteristics 94
4.6 Matrix of correlation coefficients for all two-band combinations 96
4.7 Assessment of colour enhancement techniques 96
4.8 Statistical results for the three maximum likelihood classifications 109
4.9 Subjective analysis of edge detection algorithms 118
4.10 Eigenvectors and eigenvalues derived from the principal component analysis of all eleven ATM bands 125

Chapter 5

5.1 Interpolation routines used in the evaluation of DEM creation techniques 142
5.2 Processing times for the various interpolation routines. All interpolations were performed on a Sun Microsystems
SPARCStation 2, with the exception of method 5 which could only be executed on a VAX 11/750

5.3 Statistical results of random spot height sampling
5.4 Input data for DEM feature extraction and types of features likely to be extracted

Chapter 6

6.1 The database structure for the storage of primitive features
6.2 Parameters used to test the accuracy of the techniques to estimate dip and strike.
6.3 Results of the evaluation of dip and strike estimation techniques.
6.4 Example of error estimation for individual dip and strike
6.5 The combination of certainty factors to produce an overall likelihood measure for the lake-edge knowledge-based rule
6.6 Description of object specific knowledge based rules
6.7 Description of the validation knowledge based rules
6.8 Weighting factor assigned to each knowledge based rule when combining using the Shortliffe method

Chapter 7

7.1 Five models comprising PROSPECTOR (from Harmon and King 1985)
7.2 List of derived products
Chapter 1 - Introduction

This research investigates methods designed to improve mapping of geological structures. Remotely sensed images are used in conjunction with topographical data to describe the three dimensional properties of the earth's surface and, hence, to infer subsurface geological structure. A number of automated computer techniques are introduced to aid the interpretation procedures in terms of speed, accuracy, and repeatability.

Many different applications require accurate and detailed geological structural maps. As world resources become more scarce and demands on these resources increase, the search for minerals, oil, and ground water intensifies. Obvious surficial deposits have largely been exhausted (Evans 1980) necessitating a better understanding of geological structures in order to discover sub-surface resources. Such resources are not only required for commercial benefit (as with oil and minerals), but also in many African countries, where the population continues to rise and water resources continue to decline, to maintain the search for replenishable ground water supplies (Drury 1991). Furthermore, expanding use of nuclear power has created an imperative to find locations to safely dispose of nuclear waste; that is, locations which will not be disturbed by earthquakes or leak through ground water circulation. Accurate geological structural maps are important for each of these applications.

Without close physical examination of every layer of rock which comprises the earth's surface, it is not possible to fully understand or describe a geological structure. Studying cores from boreholes and interpretation of seismic surveys allow geologists to produce reasonably accurate maps of the geology and its structure. These methods are very expensive and often only cover local areas, or rather segments or points (Khan 1976). They may be supplemented by field mapping to give wider area coverage, but again this type of survey is expensive. A cheaper method of survey employs remotely sensed images acquired from either satellites or aircraft (Curran 1985). These images can provide synoptic coverage of an area, useful in reconnaissance mapping, or more detailed local mapping dependent on the properties of the scanning device and the
remote platform. For instance, Landsat Multispectral Scanning System (MSS) has been used to produce geological maps at a scale of 1:200,000 (Drury 1987), while Airborne Thematic Mapper (ATM) data have been used in this and other studies to produce maps at a scale of 1:10,000 (Greenbaum 1987). These images provide information about the way in which radiation is reflected/scattered from the earth's surface at different wavelengths in the electromagnetic spectrum (Curran 1985). In geological applications, this can facilitate the mapping of dominant minerals within individual rock units. Remotely sensed data also lends itself to a degree of automation, in that the data are acquired automatically and are stored in a digital form suitable as input to automated image processing and image understanding techniques (Muller 1988). As computers are refined and developed, giving increased power and decreased costs, these automated procedures are likely to become more accurate, timely, and cost effective.

As demand for scarce resources escalates further in the future, the planets in our solar system, and possibly beyond, may become attractive sites for exploration. In these hostile environments, remote sensing and automated interpretation of planetary images assumes greater importance, as ground based surveys would be difficult, dangerous, and very costly. The automated interpretation procedures would need to be modified for each planet, as the geological processes are different from that of the Earth. Much of the geology on the Earth is controlled by the presence of running water, while on planets such as Mars and Venus the dominant process is believed to be vulcanism (Brown and Musset 1981).

The techniques developed in this study are designed to help solve present day problems, and could potentially be extrapolated as useful stepping stones for techniques which will conceivably be used for many years into the future. This thesis will not investigate any of the applications of structural mapping described above, but examines the success with which geological structures can be mapped from remotely sensed images and digital topographic data. A brief introduction to the traditional use of these datasets is provided here; this is followed by a more detailed review of the literature in Chapter 2. The short-falls of these techniques highlight the need to develop new methods to map geological structures from remotely sensed images, both automatically and manually within the framework of an expert system. Further requirements for automation and the current developments of expert systems are also described here.

1.1 Traditional Rôle of Remote Sensing in Geological Applications

Remote sensing has been used successfully as a tool in many geological applications. Motivated primarily by the requirements of oil and mineral exploration, applications
include lithological mapping (Drury 1988, Rothery 1987), geobotanical studies (Labovitz et al. 1983, Ager et al. 1989) and structural analyses (both local and tectonic (Stefouli and Osmaston 1984, McFall and Singhroy 1989)). Specific applications range from gold exploration (Bedell 1990), hydrocarbon exploration (Dekker 1989), detection of metal stress in vegetation (Cox and Beckett 1989), searching for ground water in arid environments (Finch 1990, Drury 1991), prediction of volcanic eruptions (Oppenheimer 1993), environmental impact studies (Legg 1990), through to monitoring oil seepages in the oceans (Dean et al. 1989).

Traditionally, fairly standard image processing techniques have been adopted in geological applications of remote sensing. In general, these are designed to enhance the visual appearance of the remotely sensed images (Figures 1.1 and 1.2), often for manual extraction of the desired geological information using conventional photo-interpretation methods (Drury and Hunt 1988). These enhancements include:

- band ratios (Podwysocki et al. 1985),
- optimal band selection (Liu and McMahon-Moore 1989),
- Intensity-Hue-Saturation (IHS) transformation (Gillespie 1980),
- three-channel hue transformation (Liu and McMahon-Moore 1989),
- principal component analysis (Kaufmann 1988), and
- decorrelation stretching (Rothery 1987).

Many of these techniques have been used extensively in geological applications with much success. However, the major disadvantage of these methods is the empirical or qualitative nature of the results, such that many of these methods produce inconsistent results between images. Consequently, each image must be interpreted differently.

More recently, greater emphasis has been placed on the quantitative use of remotely-sensed data, particularly in the analysis of mineral and rock spectra (Taranik and Kruse 1989). With the advent of high spectral resolution airborne data, these spectra have been employed within spectral mixture models to ascertain the mineral composition included within each pixel of the image (Drake 1990).

1.2 Rôle of Digital Elevation Data in Geological Mapping

The relationship between topography and geology is very important as surface and sub-surface geology are often closely related to the morphology of the resultant landscape (Krishnamurthy et al. 1992). In particular, the drainage pattern and drainage density of an area are often related to lithological type and the predominant strike of
Figure 1.1 - SPOT satellite image of Djebel Armour, Algeria, showing a series of parallel anticlines and synclines

Figure 1.2 - Landsat MSS data of an area of the Anti Atlas Mountains in Morocco, merged with Heat Capacity Mapping Mission data (from Colwell 1983).
folding and faulting (Argialas et al. 1988), while a break of slope may indicate a lithological boundary or fault line (Butzer 1976). Topographic shading of natural scenes, in both aerial photographs and digital images, has often been used to give an indication of the structural geology of the area, including lineament mapping and an approximation to the local strike of the lithology.

Many other useful morphological features can be derived from an analysis of digital elevation models (DEM). However, the use of such data has only recently been investigated. In these studies, DEMs have been used to:

- aid standard classification procedures (Isaksson 1990),
- determine dip and strike interactively (McGuffie et al. 1989),
- aid manual interpretation (Sauter et al. 1989), and
- classify morphological patterns on the surface (Chorowicz et al. 1989).

McGuffie et al. (1989) show that accurate structural information may be derived interactively from remotely sensed data and DEMs, while Sauter et al. (1989) indicate how DEMs may be used on their own, to highlight geological information. Both of these studies relied on manual interpretation of data to derive relevant geological information. By comparison, Chorowicz et al. (1989) developed several automated techniques to extract morphological information from DEMs and used this to infer geological structures.

One aim of this thesis is to develop further the work described above by attempting to produce reliable techniques for geological structural mapping using both remotely sensed and digital elevation data. It will introduce a number of automated techniques, applied to both data sets, designed specifically for structural mapping. A number of data enhancement techniques are also described which enable better manual interpretation of geological information.

### 1.3 Requirements for Automation

Historically, vast quantities of remotely sensed data and other spatial data sets (such as DEMs, cartographical, geophysical and geochemical data) that are available to geologists, have not been used to their full potential (Barr 1990). Although several studies have combined and analysed these data within Geographical Information Systems (GIS) using statistical and knowledge-based approaches (Harris 1989, Rao et al. 1989), fully automated procedures have yet to be developed.
The conventional manual interpretation approach is, by definition, a subjective one and results will vary considerably between interpreters (Parsons and Yearly 1986). One way to overcome this is to incorporate various automated procedures into the analysis of images. The aim here is to create knowledge-based/expert systems which include the standard image processing methods, in addition to sophisticated algorithms more closely related to human perception and expert knowledge. An expert system such as this, designed to 'understand' the geology of a scene, could include not only the structural geological information, which is the subject of this research, but also other aspects of geology such as, rock composition, mineralization and geobotany.

An additional requirement for an automated geological image understanding system stems from the increases in volume of data that will be produced by the next generation of remote sensing devices, such as the Earth Observing System (EOS) instruments on board the NASA, ESA and NASDA Polar Platforms (Butler 1987, Truss 1988, Hara 1988). These sensors will record data over a wider range of wavelengths and in many more wavebands than current devices. Data will also be acquired at many different sensor look angles and spatial resolutions. It has been calculated that the NASA Polar Platform, EOS-A, will download approximately $10^{13}$ bits of data every day (EosDIS 1988). For many applications there is little hope of interpreting such large volumes of data manually. Although geological applications do not have the same time-critical needs as, for example, agricultural monitoring, they will nevertheless need to make best use of the copious information available in order to improve final outputs and decision making.

In addition to the increase in the volume of data available to users, the expectations of the usefulness of such data will also be raised dramatically (Muller 1988). New procedures must therefore be developed to automate and, hence, accelerate the image interpretation stages in all remote sensing applications, including those geological.

1.4 Requirements for a Semi-Automated/Fully Automated Geological Mapping System

This research aims to develop procedures for the manual and automatic derivation of geological structural parameters from remotely sensed images and digital elevation data. Since useful morphological descriptors may be obtained from DEMs and lineaments can be identified in remotely sensed images, it is possible to determine structural parameters (e.g., dip and strike measurements of bedding and fault planes) from an integrated analysis of these two data sets.
To achieve this goal a number of requirements must be met. Firstly, a DEM must be produced of comparable accuracy and resolution to the remotely sensed images. A suite of algorithms are then required to enhance the geological content of both data sets. This three-dimensional (3D) information must then be extracted, either manually or automatically, to enable the derivation of dip and strike estimates. As these techniques may also identify non-geological features, geological and image interpretation knowledge must be used to distinguish the required geological information. The geological data must then be incorporated into a geological model, again using geological knowledge. Finally, an expert system can be designed to automate the whole procedure.

1.5 Structure of the Thesis

The need has been identified for automated approaches to the analysis and interpretation of remotely sensed images and digital elevation data for geological applications. This results from the inefficient use of the existing data sources and the probable under-use of data which will be available in the near future. More generally, it has been noted that little use is currently made of the relationship between topography and the underlying geological structure. One long-term aim of this research is to provide a complete expert system capable of performing all geological interpretation tasks better than any human interpreter. However, within the scope of this thesis, it is hoped that the research will provide many techniques and ideas which may be used within an expert system framework to analyse the structural geology of a scene, both manually and automatically.

In Chapter 2 a review is provided of current research into the geological applications of remotely-sensed data and digital elevation data. This covers the spectral response of geological materials at different wavelengths, conventional image processing techniques, the use of DEMs in geology, and image understanding. A number of areas requiring further research are identified, including the relationship between geology and topography, and the use of expert systems in geological applications. Chapter 3 describes the study area used in this research. Chapter 4 describes various tools which may be used to extract useful geological information from remotely sensed images, while Chapter 5 outlines how similar information may be gained from topographic data, either separately, or in conjunction with the remotely sensed images. Chapter 6 illustrates how these data may be combined to derive geological structural parameters and defines several knowledge-based rules designed to separate geological information from non-geological results. Finally, Chapter 7 outlines the need for an expert system
for geological mapping and proposes a number of possible solutions. It also specifies further research that is required within this field before these solutions may be achieved.
Chapter 2 - Geological Applications of Remote Sensing

This chapter reviews past and current work in geological remote sensing ranging from traditional manual interpretation of images and standard image processing techniques (such as band ratios and principal components) through to the more novel methods of spectral mixture modelling and integrated analysis with DEMs. This is followed by a discussion of several advances in other areas of image processing and image understanding which are less commonly associated with geological applications of remote sensing, but which could readily be incorporated to provide automated geological structural mapping procedures. These techniques include image segmentation, edge detection, knowledge-based rules and expert systems.

2.1 Basic Principles

Remotely sensed data have been used as a cost effective source of lithological and structural information. Due to their synoptic character they can be used to identify large scale phenomena which may not be readily perceived from the ground. Remote sensing in geology originated with interpretation of black and white aerial photography (Lillesand and Keifer 1979). Over the years, many geologists have preferred to interpret aerial photography visually because of its high spatial resolution and because it is frequently possible to distinguish local texture and lineaments in images (Drury 1986). With advances in sensor technology, similar properties are now afforded by several satellite sensors which often have the added bonus of providing multispectral data.

All geological materials reflect or emit energy differently at different wavelengths in the electromagnetic spectrum (Figure 2.1). Images recorded at these various wavelengths will therefore exhibit different geological features, for instance mineralogical assemblages, structural information, or the effect of the underlying geology on the surface vegetation. Each portion of the electromagnetic spectrum has different advantages and disadvantages in examining these phenomena. Although not all of these
Figure 2.1 - The electromagnetic spectrum (from Colwell 1983).
commonly used parts of the spectrum are employed within this thesis, it is important to understand the basic principles involved in geological remote sensing and the previous work undertaken in this field. The following sections will describe the uses of the visible and near infrared, thermal infrared, and microwave regions, within geological applications.

2.1.1 Visible Through to Short Wave Infrared Wavelengths

If lithological units are to be classified correctly then it is important to understand fully the relationship between electromagnetic radiation and geological materials. Much work has been undertaken concerning the spectral response of rocks and minerals (Burns 1970, Hunt 1977, Williams 1983, Davis et al. 1987), the majority of which was carried out in the laboratory. These studies show that the spectral signatures of minerals are affected by various absorption bands throughout the electromagnetic spectrum, including those caused by water, Fe$^{2+}$, Fe$^{3+}$ and OH$^-$ ions (Drury 1987, Pontual 1987). A useful and concise description of spectral responses and absorption bands is given by Goetz et al. (1983) and a comprehensive study of the spectral response of many rocks and minerals is given in a series of papers by Hunt et al. (1970-1974). Examples of such spectra are illustrated in Figure 2.2, which also highlights some of the more common absorption features. These laboratory spectral curves indicate how distinct many of the minerals are. However, the spectral response received at the sensor relies upon a series of complex factors, each of which may mask or confuse these curves, including:

- spatial resolution of sensor,
- spectral bandwidth of sensor,
- soil and vegetation cover (Lillesand and Keifer 1979),
- sun and view angle geometry with the target (Barnsley and Morris 1990a),
- topographic effects and shadow,
- moisture content,
- atmospheric effects (Colwell 1983), along with
- the rock type and weathering (Abrams 1980, Pontual 1987).

Many attempts have been made to correct many of the effects in images (e.g., topographic effects (Proy et al. 1989, Newton et al. 1991), sun-sensor geometry effects (Morris and Barnsley 1990b), and atmospheric effects (Chavez 1989)). Varying degrees of success have been achieved with these methods and none are perfect. In many geological applications, for example structural mapping, such corrections are not imperative. For other applications, such as mineral mapping, corrections for the
Figure 2.2 - Example laboratory mineral spectra (after Grove et al. 1992)

Reflectance

Wavelength (um)

Mineral/Element
- Kaolinite
- Illite
- Goethite
- Gypsum
- Quartz
- Jarosite
- Baryte
- Magnetite
- Haematite
- Sulphur
- Malachite
- Calcite
atmosphere, view angle and topography may be more important. Despite these effects and the difficulties in correcting for them, useful geological information may still be derived from remotely sensed data. In fact some effects such as topographic shading may actually enhance structural information (Sauter et al. 1989).

Pontual (1987) points out that radiation from the visible and infrared portions of the spectrum does not penetrate further than the upper few μm of a rock surface and that most rock exposures exhibit strong weathering. Therefore, when mapping lithological units, it is not only necessary to consider rock and rock mineralogy spectra but also spectra of minerals to which the rock might weather.

Several researchers have found that vegetation can be strongly influenced by underlying geological phenomena such as mineralization, lithology and fractures (Cannon 1960, Brooks 1972) resulting in the discipline of geobotany and its inclusion in a remote sensing framework (Harris 1987). Certain elements when taken up by plants, cause changes in reflectance at specific wavelengths. For instance, Cu and Pb cause increases in reflectance at near infrared and mid-infrared wavelengths, allowing possible mapping of zones of mineralization (Labovitz et al. 1983, Curtiss and Maecher 1991, Singhroy and Kruse 1991). In many climates, rocks are substantially covered by a layer of lichen. Cloutis (1989) has found that lichen has a similar spectral response to vascular vegetation, therefore concealing surface geology; however, due to a much suppressed response in the near infrared (red edge) and given sufficient spectral resolution the two can be distinguished.

Tables 2.1 and 2.2 list a number of satellite and airborne sensors which have been used (or are planned) to acquire images within the visible through to short-wave infrared (SWIR) portions of the spectrum, and the various geological applications for which the data have been employed (or planned). Also described are a number of important characteristics of each sensor including its spectral and spatial resolutions. The tables indicate the wide variety of geological applications for which these sensors have been used. Although the Landsat Thematic Mapper has undoubtedly been the most commonly used sensor, no one sensor provides the required characteristics for all geological applications. As a result, many studies use a combination of images acquired from different sensors (e.g., Kaufmann 1984, Cetin and Warner 1993, Yésou et al. 1993).
<table>
<thead>
<tr>
<th>Sensor</th>
<th>HIRIS</th>
<th>Daedalus AADS 1268 ATM</th>
<th>AIS</th>
<th>AVIRIS</th>
<th>Geoscan AMSS Mark II</th>
<th>GERIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Resolution (m)</td>
<td>30</td>
<td>5 at a height of 2km</td>
<td>9 at a height of 5km</td>
<td>17 at a height of 20km</td>
<td>4 at a height of 2km</td>
<td></td>
</tr>
<tr>
<td>Spectral Resolution (µm)</td>
<td>0.4-2.5</td>
<td>0.43-0.45</td>
<td>0.45-0.52</td>
<td>0.4-2.5 (224 bands)</td>
<td>0.522-0.955 10 bands</td>
<td>0.499-1.083 24 bands</td>
</tr>
<tr>
<td></td>
<td>192 bands @ 10nm spacing</td>
<td>0.52-0.60</td>
<td>0.605-0.625</td>
<td>0.52-2.52</td>
<td>2.044-2.352 8 bands</td>
<td>1.080-1.800 7 bands</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.63-0.69</td>
<td>0.695-0.75</td>
<td>0.64-1.05</td>
<td>8.64-11.28 6 bands</td>
<td>1.98-2.494 32 bands</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.76-0.90</td>
<td>0.91-1.05</td>
<td>0.75-1.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.55-1.75</td>
<td>2.08-2.35</td>
<td>1.75-2.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.50-13.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Mineral mapping</td>
<td>b. Observation of volcanic thermal features (SWIR)</td>
<td>b. Observation of volcanic thermal features (SWIR)</td>
<td>b. Hydrothermal alteration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. Lineament mapping</td>
<td>d. Hydrothermal alteration</td>
<td>d. Lineament mapping</td>
<td>d. Hydrothermal alteration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>e. Sedimentary facies mapping</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. Smithurst et al. (1987)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HIRIS = High Resolution Infrared Spectrometer, ATM = Airborne Thematic Mapper, AIS = Airborne Imaging Spectrometer, AVIRIS = Airborne Visible Infrared Imaging Spectrometer, GERIS = Geophysical and Environmental Research Imaging Spectrometer.

Table 2.2 - Characteristics of Visible to SWIR future satellite and current airborne sensors used in geological applications
<table>
<thead>
<tr>
<th>Sensor</th>
<th>Landsat MSS</th>
<th>Landsat TM</th>
<th>SPOT</th>
<th>MOMS-02</th>
<th>IRS</th>
<th>Fuyo-1 (JERS-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Res</td>
<td>80</td>
<td>30 (0.45-2.36µm)</td>
<td>10 (for panchromatic)</td>
<td>4.5</td>
<td>72.5</td>
<td>18.3 x 24.2</td>
</tr>
<tr>
<td>(m)</td>
<td></td>
<td>120 (10.4-12.5µm)</td>
<td>20 (for multispectral)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spectral Res</td>
<td></td>
<td>0.45-0.52</td>
<td>0.50-0.59</td>
<td>0.472</td>
<td>0.45-0.52</td>
<td>0.52-0.69</td>
</tr>
<tr>
<td>(µm)</td>
<td>0.50-0.60</td>
<td>0.52-0.60</td>
<td>0.61-0.68</td>
<td>0.552</td>
<td>0.62-0.68</td>
<td>0.52-0.59</td>
</tr>
<tr>
<td></td>
<td>0.60-0.71</td>
<td>0.63-0.69</td>
<td>0.79-0.89</td>
<td>0.662</td>
<td>0.77-0.86</td>
<td>0.76-0.86</td>
</tr>
<tr>
<td></td>
<td>0.69-0.80</td>
<td>0.76-0.90</td>
<td>1.55-1.75</td>
<td>0.790</td>
<td></td>
<td>1.60-1.71</td>
</tr>
<tr>
<td></td>
<td>0.80-1.10</td>
<td>2.00-2.36</td>
<td>Panchromatic</td>
<td>0.640 (3 stereo sensors)</td>
<td></td>
<td>2.01-2.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.4-12.5</td>
<td>2.57-2.82</td>
<td></td>
<td></td>
<td>2.13-2.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.57-3.92</td>
<td></td>
<td></td>
<td>2.27-2.40</td>
</tr>
<tr>
<td>Geological Uses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Microwave 1275 MHz</td>
</tr>
<tr>
<td>a. HYDRO</td>
<td>a. Hydrothermal alteration mapping</td>
<td></td>
<td>a. Stereoscopic studies.</td>
<td>a. Spectral mapping of minerals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>thermal</td>
<td>b. Observation of volcanic thermal features (SWIR)</td>
<td></td>
<td>b. Large scale mapping</td>
<td>b. Topographic mapping</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alteration</td>
<td>c. Hazard mapping</td>
<td></td>
<td>c. Lithological mapping</td>
<td>c. Lithological mapping</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mapping</td>
<td>d. Lithotological mapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e. Groundwater mapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>f. Vegetation stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>g. Weathering minerals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>h. Mapping environmental effects of mining</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>i. Weathering minerals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams and Carter (1976)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Goetz et al. (1975)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Segal (1983)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salomonson et al. (1980)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Podwysoki et al. (1984), Davis et al. (1987), White et al. (1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Rothery et al. (1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Murphy and Vita-Finzi (1991)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Finch (1990)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Isakkson and Anderson (1990)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. Podwysoki et al. (1985), Kaufmann (1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h. Legg (1990)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. Pontual (1987)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Day and Muller (1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. (Yésou et al. (1993), Jutz and Chorowicz (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Chorowicz et al. (1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Huadong and Pinliang (1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Soofi et al. (1991)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Gouyet et al. (1991)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bodechtal et al. (1989)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. &amp; b. Bodechtal et al. (1985)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Ramasamy et al. (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Denniss et al. (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MSS = Multi Spectral Scanner, TM = Thematic Mapper, SPOT = Systeme Pour l'Observation de la Terre, MOMS = Modular Optoelectronic Multispectral Scanner, IRS = Indian Remote Sensing system, JERS = Japanese Earth Resources Satellite.

Table 2.1 - Characteristics of Visible to SWIR satellite sensors used in geological applications.
2.1.2 Thermal Infrared Wavelengths

Thermal infrared radiation may be used to map lithological units where there is a contrast in the thermal emissivity of various rock types (Short and Stuart 1982). The relative difference between rock units is easy to identify in thermal images; however, measurement in terms of quantifiable units (i.e., Kelvin) is difficult due to a number of important factors (Szekiel'da 1988):

- atmospheric conditions; cloud cover during the daily cycle may either prevent or trap heat, while increasing wind speed may cool the surface by turbulent convection,
- surface orientation; changes in slope and aspect allow differential illumination by the sun,
- surface moisture; evaporation results in a lower temperature of a moist surface,
- atmospheric absorption and re-emission by water vapour and aerosols, and
- the amount of vegetative cover.

A useful measure of the thermal properties of rocks is their thermal inertia (TI), which is equal to $\sqrt{kC}$, where $k$ is the thermal conductivity and $C$ is the heat capacity per unit volume (Szekiel'da 1988, Wood et al. 1990). For instance, sand, which has a low thermal inertia, has large amplitude radiative flux variations, while mafic rocks have relatively small variations and, therefore, a high thermal inertia (Short and Stuart 1982).

Relatively few geological applications have used satellite thermal data (Watson 1975, Schott 1989), partly due to the difficulties mentioned above and partly due to the relatively coarse resolution of thermal sensors (Thematic Mapper (TM) = 120m, Heat Capacity Mapping Radiometer (HCMR) = 600m and Advanced Very High Resolution Radiometer (AVHRR) = 1.1km). Increased resolution is afforded by airborne sensors, such as ATM (Table 2.2) and the thermal infrared multispectral scanner (TIMS), which has six spectral channels in the thermal region (Palluconi and Meeks 1985) and a variable ground resolution, dependent on the height of the aircraft (e.g., 25 meters at a flying height of 10km, Lang et al. 1987). TIMS data have been used successfully to map both lithologies and minerals (Hook et al. 1992, Gillespie 1992a).

2.1.3 Microwave Wavelengths

Microwave sensors are commonly active sensors, rather than the passive sensors mentioned above, which actively send out a signal to the Earth's surface and record the
response. These sensors measure the amplitude, phase shift, time of arrival, and Doppler shift of returning signals (Szekielda 1988, Rees 1990). These signals may be combined and processed as part of a synthesized aperture radar (SAR) system to increase the spatial resolution of the sensor (Rees 1990). SAR has a number of advantages over other image data. Firstly, microwave wavelengths are not absorbed by water vapour and can therefore penetrate cloud\(^1\). This means that SAR may be acquired at any time, day or night. In some areas, such as Indonesia, SAR may be the only data available of an area, due to almost permanent cloud cover (Trevett 1986). Secondly, geological structural information is well displayed in SAR images due to its side-looking characteristics and the strong shadows present in the data. In particular, lineaments are easily identified in SAR imagery (Borengasser and Taranik 1988, Roy et al. 1993, Yésou et al. 1993) and, in airborne SAR, the direction of the illumination may be easily modified to enhance features of different orientations. The strength of the returning signal is strongly dependent on the surface roughness and therefore lithological units may be mapped by their surface textural properties (Lynne and Taylor 1986).

Satellite SAR sensors include Seasat (Jordan 1980; which unfortunately only remained operational for 100 days in 1978), ERS-1 (Cox and Joyce 1984), launched in 1991, and JERS-1 launched in 1992 (Nishidai 1993); associated with these sensors have been the shuttle imaging radars (SIR-a and SIR-B, Ford et al. 1983) which were placed on board two shuttle missions. Each have been used in geological applications (Curlis et al. 1986). In the period between Seasat and ERS-1, exploration companies turned instead to airborne SAR (e.g., Rao et al. 1989, Soofi and Payot 1991, Mercer et al. 1991), which has the additional advantages of providing greater spatial resolution and the ability to change the illumination angle to an orientation most able to highlight the geological structure of interest (Trevett 1986).

One problem associated with radar imagery is due to the geometric properties of the data. Relief in the landscape suffers from a 'layover' effect caused by the fact that the tops of hills and mountains are effectively projected onto the surface the behind topographic feature (Figures 2.3 and 2.4). As a result it is not a simple process to relate features in the image to ancillary data, such as maps and DEMs. These effects may be rectified using a radar imaging model, a DEM, and several ground control points (Toutin et al. 1992).

This section has described the uses and applications of satellite and airborne sensors within the visible to SWIR, thermal infrared, and microwave regions. There are a great many sensors available and care must be taken when deciding which sensor or

\(^1\) N.B. Heavy rainfall can affect the radar signal.
Radar depression angle
Pulse direction
Terrain slopes steeper than these lines will be imaged with layover

Layover
Layover
Layover
Layover

Deformation

Resulting image (ground range format)
Weak return
Shadow
Shadow
Shadow

Figure 2.3 - Effects of terrain relief on SAR images (From Lillesand and Keifer 1979)
Figure 2.4 - Example of the layover effect in radar images - ERS-1 data showing Mount Vesuvius and surrounds.
combination of sensors to use for particular applications. For instance, planning decisions may be dependent on the scale of geological mapping, the type of application, the geological setting, the climate, the topography, percentage vegetation cover, accessibility, and cost.

The application of this thesis involves the small-scale mapping of geological structures in a mountainous area of Snowdonia. Ideally then, with unlimited funding and availability the project might employ high spatial (5m or less) and high spectral (approximately 2nm) resolution data to map lithological units and identify mineral components, stereo aerial photography to provide elevation data (see section 5.1.1), multiband thermal data to help discriminate lithologies, and multilook SAR data to map geomorphological features and differences in surface roughness.

2.2 Conventional Image Processing Techniques

This section considers image processing techniques which are commonly used in geological remote sensing projects. It discusses the pros and cons of a number of manual and semi-automated image enhancement techniques. More recent and advanced techniques are discussed in sections 3.4 and 3.5. In the past, geological lineaments have been the only structural measurement obtainable from remotely sensed images, due to the 2-dimensional (2D) nature of the data. Several techniques are described which attempt to identify these features automatically. Finally, recent work on incorporating DEMs into structural mapping are discussed.

2.2.1 Visual Enhancements

In most geological studies, the main motivation is to enhance remotely sensed images for subsequent manual interpretation by geology experts. Consequently, the techniques developed for this purpose attempt to concentrate as much information as possible into three bands, for display on a conventional Red/Green/Blue (RGB) colour monitor. As an example, the Thematic Mapper has seven spectral bands and there are therefore 210 possible permutations of displaying three of these bands at any one time (Drury and Hunt 1988).

Although many of the bands are highly correlated and exhibit little new geological information, such that there is a considerable degree of redundancy in the data, there remains a great deal of information to be evaluated. Sheffield (1985) attempted to measure the significance of each band by ranking three-band combinations according to the determinants of their 3 by 3 variance-covariance matrices within a seven-dimensiona
data space. Chavez et al. (1982) designed a similar statistic known as the Optimal Index Factor (OIF) which is the ratio of the sum of standard deviations to the sum of the absolute values of the correlation coefficients for three band combinations. Both methods can be affected by a linear stretch of the grey-level range in each bands. The method described by Liu and McMahon-Moore (1989), the index of optimal band-triplet selection (IOBS), overcomes this by examining the correlation coefficients of each three band combination. However, none of these methods makes any use of the amount of geological knowledge contained in each band, and many authors have preferred to choose band combinations by trial and error techniques, making subjective decisions on how well images represent the known geology (Wadge and Quarmby 1988, Rothery 1987b, Crosta and McMahon-Moore 1989).

2.2.2 Mathematical Transformations

Other methods of reducing multispectral data sets into readily displayable units include band ratioing and principal component analysis. Band ratioing is a standard image processing technique which suppresses spatial radiance variations that are proportionally constant between bands (Crippen et al. 1988), and according to Cracknell and Saraf (1989) are attributable to terrain illumination, ground albedo and look-angle effects. However, it has been shown that the radiometric effects of view-angle are not constant between bands (Barnsley 1983), and similarly that terrain illumination is different in each band (Proy et al. 1989). Cracknell and Saraf (1989) found that a ratio of band 3/band 7 (0.57\mu m/0.88\mu m) of the Daedalus AADS-1286 scanner gives a particularly good differentiation between rock units in an area of Scotland almost totally covered in vegetation. This ratio is the inverse of a typical vegetation index (Schowengerdt 1983) and is claimed to retain some terrain information in less rugged areas. There are many possible combinations for band ratios each of which may enhance some geological feature (Rowan et al. 1974, Goetz et al. 1975, Vincent 1977, Podwysocki et al. 1983, Segal 1983, Gillespie et al. 1988, Ramasamy et al. 1993). However, some of the more commonly used ratios are:-

- (0.55\mu m/0.65\mu m) for distinguishing hydrothermally altered rocks and gossans (Rothery and Milton 1981), and
- 1.65\mu m/2.22\mu m to enhance clay minerals (Podwysocki et al. 1985, Fraser and Green 1987).

Principal component analysis (PCA) determines perpendicular (or statistically independent) axes of maximum inter-band covariance in a multispectral feature space (Schowengerdt 1982, Figure 2.5) and is an obvious method for reducing data redundancy. This method has been used successfully in many geological applications.
Figure 2.5 - The Principal Components Transformation. $I_1$ and $I_2$ are the DNs of two of the original bands. $I_1'$ and $I_2'$ are the transformed bands such that there is the least correlation between them (from Rees 1990).

Figure 2.6 - Conventional image contrast stretching (a-c) and decorrelation stretching (d-f). Conventional contrast stretching is performed parallel to the input band axes - decorrelation stretching parallel to the principal component axes allows a much fuller use of the display colour space (from Rothery 1987b).
However, it suffers from two major problems, a loss of potentially useful information in unused components, and the difficulty in interpreting colour composites of components (Chavez and Kwarteng 1989). Chavez and Kwarteng (1989) describe a modified PCA technique known as directed principal component analysis (DPCA). The DPCA method uses two ratios as input, one of which should contain information regarding the component of interest (i.e., a geological discriminant) and the other should contain enhanced information regarding the spectrally interfering component e.g., vegetation. Output should then have the first principal component (PC) along the axis of maximum variance due to vegetation and the second PC should align itself along the maximum variance due to geology (Fraser and Green 1987). A similar method by Chavez and Kwarteng (1989) selects two bands as input and surmises that the second PC represents spectral contrast between the two bands, hence enhancing geological spectral differences. This method is complementary to that mentioned above in that it has been most successful in areas where vegetation is a minor cover type, otherwise the spectral contrast will be dominated by vegetation contrast. The success of these techniques can be extremely unpredictable as the statistical calculations are dependent upon the range and distribution of digital numbers (DN) within an image. Hence, the results will vary from image to image.

A further step to PCA is the ‘decorrelation stretch’ (Rothery 1987, Gillespie 1992b), where each principal component is contrast stretched before the inverse PCA transform takes place. This enables a more complete use of the RGB feature space than other techniques (Figure 2.6). Ferrari (1992) describes an ‘improved decorrelation stretch’ which additionally carries out a high pass filter (sharpening, Schowengerdt 1983) on the first principal component to enhance fine detail, and low pass filters (smoothing, Schowengerdt 1983) on the second and third components to remove noise apparent in these data. The data are then inversely transformed into standard feature space. Ferrari (1992) reports good visual enhancement of images, but again these will vary between different scenes.

The results of most band ratio and principal components are visually confusing and bear little relation to the original appearance of the image. A group of methods which visually enhance information and attempt not to alter appearance are the colour space transformation techniques. The methods transform the original red-green-blue (RGB) co-ordinate system into one of the following:

- principal component colour space (Soha and Schwartz 1978, Rothery 1987 a,b, Drury and Hunt 1988, Kaufmann 1988)

• NTSC Y-I-Q space (Rothery 1987b)

The images are then contrast stretched in the new colour space and inversely transformed to the original colour space. Matrices for the colour space transformation are discussed in Pratt (1978). Liu and McMahon-Moore (1989) describe a method of combining three hue bands, derived from three different colour composites, into a hue colour composite image (Hue-Red-Green-Blue - HRGB). The three original composites are chosen using optimum band selection. It is claimed that the shadow is totally suppressed using this technique and that spectral differences are well enhanced. As with most of the enhancement techniques mentioned in this section, each of the images produced have a unique colouring scheme and may therefore be difficult to interpret on an image by image basis.

So far in this section only remotely sensed images have been mentioned; however, ancillary spatial data sets (e.g., geological maps and geophysical and geochemical data) may be enhanced and combined in a similar manner, e.g., using an IHS transformation (Figure 2.7) (Harris 1989, Harris et al. 1990, and Rao et al. 1989) and principal components (Gibson 1993). These techniques may similarly be used to combine remotely sensed data sets acquired from different platforms, e.g., SeaSat and SPOT (Yésou et al. 1993), airborne SAR and Landsat MSS (Koopmans and Forero 1993), and aerial photographs and Landsat MSS (Grasso 1993).

2.2.3 Spatial Analyses and Lineament Detection

The surface expression of folds, faults, joints, strikes, lithological contacts and other geological features can often be in the form of lineaments (Harris 1987). These can either exist at a local scale or over much larger areas indicating tectonic structures (usually overlooked in ground surveys, Trevett 1986; see also Biju-Duval et al. 1976, Boccaletti et al. 1980, 1982). Lineaments are therefore one of the most important features used in the mapping of geological structures from remotely sensed data. They can be enhanced in an image using a standard high-pass filter (Schowengerdt 1983) and although often interpreted manually (Yésou et al. 1993) can easily be detected using edge detection and pattern recognition techniques (Koopmans 1986, Wadge and Cross 1989, Abramson and Schowengerdt 1993).

To avoid the common misregistration of linear features between different image interpreters, Stefouli and Osmaston (1984) have classified geological lineaments in
Figure 2.7 - Satellite image data combined with geophysical data using an IHS transformation.

The geophysical data is provided as hue information while the satellite image is shown as intensity. Both data sets are more interpretable when combined.
terms of their appearance and context within an image. Lineaments can be defined as "any linear image formed by points or groups of points which possess certain similarities of relief, tone, texture or pattern, different from those of the surrounding area; or a boundary which divides two areas that differ in one or more of these properties" (Stefouli and Osmaston 1984). It can be seen from this definition that image processing techniques such as edge detection, texture analysis and digital elevation models have a major role to play in the determination and classification of lineaments. Traditionally, the first two of these have been used in lineament analysis and elevation data has largely been ignored. A variety of edge detection filters has been used, the most common of which is the gradient filter (Podwysocki et al. 1975, Smithurst and Vaughan 1987, Harris 1987). Section 2.5.2 reviews developments of edge detection algorithms some of which have been used in geological applications e.g., the Haar transform (Majumdar and Bhattacharya 1988), the Hough transform (Skingley and Rye 1986, Wadge and Cross 1989), the Fourier transform (Eppes and Rouse 1974, Hilrose and Harris 1985) and others (Frost et al. 1983, Moore and Waltz 1983). Lineaments have also been detected using an analysis of texture within the image (Oldfield and Elgy 1987, Brunner and Veck 1985). Section 2.5.1 describes various texture measures, some of which detect a directional component of the texture, which could be extremely useful in geological applications.

2.2.4 Spectral Mixture Models

In the past very little quantitative work has been carried out in geological remote sensing, however, with the advent of higher spectral resolution data recent work has concentrated on modelling the proportions of various minerals contributing to the radiance of each pixel in an image. If the spectra of minerals present in the ground resolution element of a pixel are known, then the proportions of each mineral present in the resulting spectra can be derived by using a procedure known as mixture modelling (Marsh et al. 1980, Wadge and Quarmby 1988, Carrere 1989, Mackin et al. 1990, Rubin 1991, Gillespie 1992a, Settle and Drake 1993, Murphy and Wadge 1994), which attempts to 'mix' the laboratory spectra until a match to the image spectra is attained. Taranik and Kruse (1989) describe a binary encoding technique for matching different spectra (Figure 2.8), such that the shape is described by its slope (i.e., whether reflectance increases or decreases at any given point on the curve) rather than the absolute magnitude. This has the advantage over other techniques, in that it rapidly identifies absorption features. The size of absorption features is often measured using absorption band depths (Crowley et al. 1989, Clarke et al. 1990) calculated either using least-squares fits with library spectra (Clarke et al. 1990) or as images (Crowley et al. 1989). These images are created by identifying a number of spectral channels on the
Figure 2.8 - Binary encoding of mineral spectra according to Taranik and Kruse (1989). If the reflectance of the mineral decreases with increasing wavelength then it is encoded as -1, if it remains the same then the code is 0, or if it increases then a value of 1 is assigned.

Figure 2.9 - Illustrating the derivation of relative absorption band depth images from reflectance values taken on the shoulders and within the trough of an absorption feature. After Crowley et al. (1989).
shoulder of the feature and in the trough (Figure 2.9); the relative band depth (RBD) is then defined as:

\[ RBD = \frac{(a + b + c + d)}{(e + f)} \]  \hspace{1cm} (2.1)

There are several methods for unmixing image spectra, including linear models (Brown 1982, Shimabukuro and Smith 1991, Settle and Drake 1993) and non-linear models (Pech et al. 1986), which differ in the way that individual spectra combine to give the final signal. The individual spectra can be taken from known laboratory spectra of minerals or from pure 'end-members' identified from the image (a priori knowledge of the scene is required for this method). The latter technique is best used when absolute reflectances are not available, due to a lack of calibration or atmospheric correction (Smith et al. 1985, Settle and Drake 1993).

Clark and Cañas (1993) have recently introduced a new method of spectral matching using artificial neural networks. Neural networks will be discussed briefly in Chapter 7 but basically provide a method for computer learning by simulating the trial and error processes of the brain. Preliminary analysis of these techniques have shown them to be extremely effective in most circumstances (Clark and Cañas 1993).

Apart from the need to know the precise cover types of an area and their spectral responses, a problem with these unmixing techniques is that several combinations of cover types can produce similar results and may also disguise an important unknown cover type. However, the quantitative results that these techniques provide are definitely the way forward for lithological and mineralogical mapping and with the future launch of high resolution imaging spectrometers, such data will become generally available (Goetz 1989, Rivard and Arvidson 1992).

2.3 Application of Digital Elevation Data

Digital elevation models (DEMs) have been widely used in remote sensing applications, but rarely in geological applications. In particular, they have been used in the fields of topographic correction (see Section 2.4.1), terrain visualization (Gelberg and Stephenson 1987, Wolff 1987, Day 1988a, Muller 1988b), the identification and assessment of hydrological features (Haralick 1983, Haralick et al. 1985b, Chorowicz et al. 1989, Riazanoff et al. 1988), and to improve multispectral classification accuracy (e.g., Isaksson and Andersson 1990). The different ways in which DEMs may be created are discussed in Chapter 5, which also provides an evaluation of a number of
these methods for use in geological applications and their place within an automated system. This section will describe how DEMs have been used within geological and geomorphological studies, firstly by examining how DEMs may be used to derive simulated drainage networks of an area and further to extract structural information from such data.

2.3.1 Drainage Networks

The prospective use of digital terrain models in the analysis of three-dimensional geology (from remotely sensed images) has already been introduced. The use of ridges, valleys, slopes and aspects in the role of geological image segmentation would be extremely advantageous, as terrain is often closely related to underlying geology (Hobbs 1903, Chorowicz et al. 1989). For instance, an escarpment topography is indicative of gently dipping sedimentary sequences. Furthermore, valley lines represent drainage patterns which also rely heavily on the structural geology, and the porosity and permeability of rocks and soils. Parvis (1950) and Howard (1967) give comprehensive empirical descriptions of some thirty different drainage patterns relating to various forms of geology. Argialas et al. (1988) trimmed this number to eight major types and describe their significance in geological interpretation. Argialas et al. (1988) also describe an expert system designed to recognize different drainage patterns on the basis of such characteristics as angles between 'branches' in the system, distance between branches, and concentration of branches. This system has been used on real drainage patterns extracted manually from aerial photographs with great success and could presumably be extended to include input from the automatic extraction of drainage patterns.

The extraction of ridge and valley lines from DEMs has been the subject of some recent research (Haralick 1983, Bevacqua and Floris 1987, Riazoneff et al. 1988, Jenson 1985, Skidmore 1990). Haralick (1983) uses a derivative approach, identifying 'turning points' in a series of directions in various sized windows. The disadvantage of this approach is that isolated valleys of a small number of pixels can occur, and therefore a technique is required which relates each valley point to the overall drainage pattern of the image. Riazoneff et al. (1988) describe three algorithms which follow valley lines within a DEM. Several 'seed' points are selected, such as saddle points or minimum points in the DEM, and a valley line is followed if it obeys certain rules, e.g., the steepest slope from a saddle point. Jenson (1985) enhanced this method to use every element of the DEM as a seed point. As each pixel in the DEM is used a counter for that pixel is incremented, resulting in an image of cumulative flow. Individual streams can then be selected using thresholding techniques.
2.3.2 Inferring Structural Geology

Few studies have made use of the additional information contained within DEMs to identify structural geological features in remotely sensed images. Instead most have concentrated solely on the spectral contrast of rocks and minerals and on lineament detection in two dimensions (2D) only. Recently, greater attention has been focused on the inclusion of the 3rd dimension (elevation data) in geological applications in remote sensing (McGuffie et al. 1987, Theissen et al. 1989, Chorowicz et al. 1989, 1991, Wadge et al. 1990). Several areas of research in this area will now be described to provide an understanding of different methods of mapping geology in three-dimensions.

Theissen et al. (1989) have introduced the notion of three-dimensional lineament analysis by including a DEM in their geologic spatial analysis (GSA) system. The system defines lineaments (topographic, image and mapped) as vectors in a three-dimensional space and determines whether each lineament occupies the same plane as any of the remaining lineaments. Thiessen et al. (1989) use these results in conjunction with other geological information including geophysical data and underground acoustic imaging, to determine sub-surface geology. Similarly, Sauter et al. (1989) have applied edge detection routines to DEMs, and used rose diagrams to identify the most predominant orientations of slope azimuth. Furthermore, they employed ridge and valley delineation algorithms and Lambertian shading techniques to identify of other geological features and suggest that these may be used as part of a semi-automatic geomorphological analysis.

Chorowicz et al. (1989, 1991) introduced a DEM encoding technique to recognize strike ridge geomorphology and fluvial deposits. They attempted to classify the DEM in terms of simple geomorphological elements such as:-

- crest lines and thalwegs,
- smooth or steep, concave, convex, or regular slopes, and
- marked changes in the topographic slope, forming a shoulder or hollow shape.

The methods used include analysis along profiles, analysis by derivatives computed along four directions at each point in the image, and multidirectional topographic surface analysis. Several rules have been used to classify the geomorphological elements as shown in Figure 2.10. These simple classes have been combined to define various patterns, such as a strike-ridge pattern:-

\[ V, n1 \ast P, S, n2 \ast P, C, n3 \ast U, V \]  

(2.2)
Figure 2.10 - Principles of elementary profile analysis demonstrating how critical points along a profile are shown by sequences of point-slope determinations. A - flat bottom, V - sharp valley, C - sharp change in slope, E - plateau, E1 - upgrading horizontal step, E2 - downgrading horizontal step, S - crest or summit, P - regular slope, U - concave slope, N - convex slope (from Chorowicz et al. 1989).

Figure 2.11 - (a) model of veers in direction along a topographic edge at the intersection of a valley to form a "V", (b) more typical example of valley-Vs encountered in the image leading to errors of omission (from Wadge et al. 1990).
where \( n_1, n_2, \) and \( n_3 \) signify repetition of these elements.

An algorithm is then used to search for this pattern in the DEM. Results for the strike-ridge and fluvial deposit patterns are reported to be good in the area chosen, although the procedure does fall down if the ridge is heavily dissected by gullies. Such a system would appear to have inherent difficulties in coping with more complex folding and faulting.

Wadge et al. (1990) describe a semi-automatic structural mapping system which does not involve the use of a DEM, but instead uses the technique of 'shape-from-shading' to determine the local slope. The procedure firstly segments the remotely sensed images using a set of texture measures. Topographic edges (i.e., edges caused by shadow) are then identified using a Sobel filter and lines extracted using seed points and a line following procedure. This is achieved by following the line to the maximum edge strength of the surrounding pixel within 45° of the edge orientation of the original pixel, providing the new edge strength is above a user-specified threshold. In the strike ridge topography of the test area chosen, such lines should follow the strike of the more resistant rock units. These lines are then searched for 'V' shapes (Figure 2.11), caused by gullies cutting through the ridge, which indicate the direction of dip of the rock unit. Local slope is determined using the natural shading of the scene. Shading is a function of incidence (solar illumination direction relative to the local surface normal), exitance (sensor direction relative to the surface normal) and phase angle (solar illumination angle relative to the sensor direction) (Figure 2.12). If the solar illumination is known, values for the local slope can be determined; however, such values are not unique and need to be constrained in some way. The dip directions determined from the 'V' shapes can be used for this purpose. Initial results give very accurate results for an area of arid terrain in the Atlas mountains of Tunisia (Wadge et al. 1990). Problems are likely to occur in areas of more complex geology and more complex surface cover. The process also makes some assumptions which are not necessarily true in all or many areas of the world, namely:

- uniform surface cover over the terrain and within each pixel; occurrences of vegetation in an otherwise arid landscape will produce erroneous results, similarly if a rock type has a different spectral response in some areas due to, for example, iron staining, then the 'shading' or radiance value of that pixel will be different.
- the surface behaves as a Lambertian reflector; little work has been carried out on the bi-directional reflectance properties of rock surfaces. However, it has been shown that similar surfaces such as soils and
Figure 2.12 - Schematic cross section to illustrate local surface illumination geometry, location of mapped topographic edges and the relationship to the planar bedding surfaces. I = angle of incidence, e = angle of emittance, and g = phase angle. (from Wadge et al. 1990).

Figure 2.13 - Component modelling of sky radiance - direct sky component FD, diffuse sky component Fd, direct ground component FgD diffuse ground component Fgd (after Woodham and Gray 1987).
tarmac exhibit distinctly non-Lambertian reflectance properties (Barnsley and Morris 1990a).

- the light impinging on the surface is solely due to the direct solar illumination; this will never be the case because, as can be seen in Figure 2.13, light is scattered by the atmosphere and also by other land surfaces before it impinges on the target (Woodham and Gray 1987). Similarly, light is reflected onto other surfaces and scattered in the atmosphere before it reaches the sensor.

Efford (1993) has suggested a number of improvements to shape from shading techniques by adapting the radiative transfer models of Hapke (1984) and including surface roughness parameters. This may solve the latter two points listed above but cannot overcome the fact that different cover types have different reflectance and scattering properties. It may be possible to include a set of techniques which firstly attempt to identify areas of similar spectral response and then use shape from shading techniques to derive a DEM. However, these areas are likely to be too small and irregular, particularly in vegetated and populated areas.

A method which uses both remotely sensed images and DEMs was introduced by McGuffie et al. (1987). This work developed interactive software which enables a geologist to identify three points along a lithological boundary which are used to calculate the dip and strike of the geological unit. The software makes use of the fact that just as any line can be defined by two points, any plane is defined by three points (Figure 2.14). The X, Y, & Z co-ordinates are determined from the 2D image and the 3rd dimension of elevation.

Of all the structural mapping methods described in this section, the latter technique would appear to provide the most reliable and accurate way of determining dip and strike. This is partly due to the fact that it is a manual technique, rather than an automated process, and the human brain is better at interpreting a scene than the automated and semi-automated techniques described above.

The following sections will describe ways in which automated techniques may be improved by attempting to better understand two important processes. Firstly, the way in which input data is formed and the processes involved in that formation. This enables an improved understanding of what the data represent and how it can be used. The second important process relates to the way in which the mind interprets images and scenes and how information is gathered about objects in that scene. If these processes can be understood then automated procedures may be developed to perform similar tasks.
Figure 2.14 - Figure showing how three points may be used to define a plane, which may in turn be used to extract dip and strike values.
2.4 Image Understanding

Geological remote sensing has traditionally employed image enhancement techniques which simply improve image interpretability rather than attempt to classify or model an image automatically. This is in part due to lack of sufficient rock exposures, particularly in temperate climates, and partly due to the complex nature of geology both structurally and lithologically (Rothery 1987a). Lithological contacts often have gradational boundaries and facies boundaries which cross geochronological boundaries (Anderton et al. 1979). Furthermore, when an image is classified and used to produce a thematic map, all textural and contextual information is lost (Rothery 1987a, Greenbaum 1987). However, with advances in expert systems, geographic information systems, image processing, computer processing and data storage, such problems could be overcome and an automated geological interpretation system designed.

An image understanding system in remote sensing is defined by Muller (1988c) as "the development of techniques and computational systems for the automated extraction of scene properties from satellite and aerial imagery for specialist domain". Essentially, it is the amalgamation of a series of fully automated procedures and knowledge-based rules designed to perform the interpretative tasks of a human, such that the process is objective, accurate, fast and repeatable.

As mentioned in Chapter 1, automation is partly necessitated by the vast increase in data available to users and the raised expectations of the usefulness of such data. Without appropriate automation of scene analysis the potentially enormous accumulation of data provided by the Earth Observation System on the Polar Platform, to be launched 1998-2008 (producing in the order of $10^{13}$ bits of data per day) will probably never be fully used, since traditional methods (i.e., visual interpretation of conventional classification techniques) are too time consuming. These conventional techniques, such as supervised and unsupervised maximum likelihood classification (Schowengerdt 1983), which may be used to classify lithological units, suffer from several major limitations:-

- local spectral signatures are difficult to extrapolate to larger areas (Muller 1988)
- each pixel usually contains more than one cover type/spectral property within its field-of-view and many objects have similar spectral properties
- measured radiance at the sensor is contributed to by many extraneous factors, for example, differences in irradiance, transmittance and path radiance (Mather 1987), differences in topographic slope and aspect (Holben and Justice 1981, Hugli and Frei 1983, Jones et al. 1988), the
radiometric effects of sun-view angle geometry (Barnsley 1984 a, b), and geometric distortion (see Lillesand and Kiefer 1979 for satellite images and see Dowman 1984 for ATM images)

- no regard is taken of textural or shape information, or the inter-relationships between pixels and/or regions (i.e., context, Tailor et al. 1986)
- no knowledge about the scene is usually included, such as previous classification, height and slope information or ancillary data from maps, records, and local expertise (Tailor et al. 1986).

An image understanding or knowledge-based system could therefore include:

"models of the image formation process, knowledge and/or models of the types of structure to be expected within the scene and their relationships, other data sets such as previous interpretations or map data (perhaps resident within a geographic information system), and the analyst's own expertise" (Tailor et al. 1986).

Muller (1988c) has offered additional criteria by including three-dimensional information and reflectance models of surface cover. Matsuyama (1987) points out that image noise and errors of analysis are common, and suggests that any recognition process in an image understanding system should be able to cope with them in a flexible manner. Also an image understanding system should have versatile capabilities of geometric reasoning (Matsuyama 1987). Reid et al. (1985) divides specialist expertise into two parts, that of engineering or physical models and experiential 'rules of thumb'. McKeown et al. (1985) state that it is imperative that any system should be based on cartographic co-ordinates (e.g., latitude, longitude and elevation) rather than on an image based co-ordinate system. It is much easier to relate models to the real world and to integrate data from different segmentation or analysis methods and multitemporal images of the same area, using cartographic co-ordinates, so that the location of each pixel in an image can be known to some degree of accuracy.

2.4.1 Understanding the Image Formation Process - Topographic Effects

Several of the image formation procedures mentioned above are addressed in Chapters 4 and 5, such as calculation of reflectance, geometric correction and the inclusion of ancillary height and slope information. One of the procedures most relevant to the mapping of geological structures is the effect of topography within a remotely sensed image. Such effects can be used to map geological features but shading may also obscure or alter the spectral response of such features, resulting in poor automatic classification accuracies (Hutchinson 1982, Stohr and West 1985, Hall-Köynves 1987).
The topographic effect can be defined as the variation in response from inclined surfaces compared to the response from a horizontal surface as a function of the orientation of the surface relative to the light source and sensor position (Holben and Justice 1981, Justice et al. 1981). Although such topographic effects can be extremely useful for the identification of geological lineaments, particularly at low sun-angles, when attempting to classify or segment the image into lithological units in terms of rock or vegetation spectral signatures these effects will seriously reduce the chances of correct identification of unit boundaries.

Several techniques have been proposed to correct for this effect, including band ratioing (Kriegler et al. 1969, Crane 1971, Vincent 1973, Justice et al. 1981), which has not proved entirely successful, due to shading affecting images differently at various wavelengths. There are also models which correct solely for changes in illumination due to topographic differences, based on assumptions of wholly Lambertian reflectance and models which correct also for the variation due to differences in slope orientation with respect to Sun and sensor geometry (Justice and Holben 1979, Cavayas 1987, Jones et al. 1988, Proy et al. 1989). Newton et al. (1991) incorporate spectral properties of the irradiance hemisphere and atmospheric effects into their correction model to better estimate the effects of shading in a scene. Conese et al. (1993) also use an atmospheric model in their correction procedure. No single method has been shown to remove all topographic effects completely and it has been concluded that this is due to the fact that each land cover type exhibits different terrain effects (Cavayas 1987, Jones et al. 1988). This indicates differences in the bi-directional reflectance distribution function (BRDF)\(^2\) of each cover type. Furthermore, within a given cover type topographic effects vary with respect to changes in wavelength (Justice and Holben 1979, Rochon et al. 1979, Cavayas 1984).

Cavayas (1987) recognizes the importance of an expert system in the successful correction of topographic effects and modelling of the entire image formation process. Cavayas (1987) proposed a system to include a library of reflectance information for intervals of solar elevation and azimuth angles and a database of multitemporal imagers of the BRDFs of each cover type, and for vegetation would need similar information for each stage of growth and soil moisture conditions. Again this is a substantial undertaking involving much expert knowledge and automated procedures, but will no doubt be achievable in the near future.

\(^2\) The BRDF describes the transfer of incident radiation from any angle direction \((\Theta_i, \Phi_i)\) into any direction \((\Theta_r, \Phi_r)\) by reflection at a given point on a surface; where \(\Theta_i, \Phi_i, \Theta_r, \) and \(\Phi_r\) are the zenith and azimuth angles of the incident and reflected radiation respectively (Nicodemus et al. 1977).
The image formation processes which result in a digital number for each pixel in an image are varied and complex. It is therefore virtually impossible to correct totally for all of these effects and some assumptions regarding these processes will always need to be made. It is often sufficient, and perhaps more beneficial, to understand the formation processes which result in the data and the limits that imposes on the data, rather than model each process in turn.

2.4.2 Expert Systems

Geological analysis of remotely sensed data is normally effected manually by a photo interpreter or by a specialist in particular aspects of geology (e.g., mineral exploration, hydrogeology, or civil engineering). A specialist uses knowledge of the image, ancillary data, and the particular application to aid in the extraction of relevant information (Tailor et al. 1988). The incorporation of these skills into an automatic process has been termed a knowledge-based system (KBS) (Tailor et al. 1986, Nicolin and Gabler 1987) or expert system (Wilkinson and Fisher 1984, Nazif and Levine 1984). The eye uses many clues to detect and recognize objects, many of which are described in a set of psychological Gestalt principles (Wertheimer 1938, Sandford 1985, Ahyja and Tuceryon 1989). These include the region parameters of similarity, proximity, uniform density, continuity, closure, texture, context, shape (Tailor et al. 1986) and spectral properties (Zobrist and Thompson 1975). The amount of prior knowledge of the scene or past experience of objects within it, that a human brain uses in its interpretation process, is a matter of some debate, but there is no doubt that once you have learned to recognize an object, it is then easier to recognize that same object again in a different setting. Much research in the fields of computer vision (e.g., Nazif and Levine 1984) and remote sensing (e.g., Tailor et al. 1986, McKeown 1987, Muller et al. 1987a, Nicolin and Gabler 1987, Goodenough et al. 1987) has recently been undertaken to find techniques and methods to design an intelligent KBS which can successfully mimic the complex procedures of the human brain, in the recognition and interpretation of images. Wilkinson and Fisher (1984) suggest that such a system could eventually out-perform a human at making diagnoses from given inputs, due to the rules being soundly based on accumulated statistics. Unfortunately, none of the above research examples are related to geological applications although many of the same principles would apply. Wadge et al. (1990) in their semi-automatic system for deriving geological structural parameters from images, using shape-from-shading techniques, describe some methods ideal for a fully automated knowledge-based system such as line following and texture-based segmentation. Computer vision techniques such as these, designed to simulate the process in which the brain interprets a scene, are discussed in section 2.5. Prior to this discussion, expert systems will be briefly introduced and described using a number of examples from the literature.
Typically, an expert system includes four major components: a set of facts and rules (the knowledge of the system), a database upon which the rules operate (can include images, geographic information system, and data tables), a rule interpreter or inference engine acting as a scheduler or control mechanism (Goodenough et al. 1987, Tailor et al. 1986) and a user friendly interface for operators who are specialists in their own field, but not necessarily in remote sensing or image understanding (Jackson and Mason 1986, Goodenough et al. 1987). A general procedure incorporated in most image-based expert systems is the segmentation of a scene into its constituent components, to describe their structures and mutual relationships and to produce an interpretation of the scene within a specified application. This process can be performed within the constraint of a defined set of models, with the images being used to confirm the expectation of the model, termed 'top-down' processing or performed using 'bottom-up' processing where the data 'drive' the system to produce an appropriate interpretation (Tailor et al. 1986). Nicolin and Gabler (1987) and Matsuyama (1987) recognize a bi-directional control mechanism which is a mixture of both top-down and bottom-up processing. Knowledge-based rules will be introduced more fully in Chapter 6, while Chapter 7 includes a detailed description of the proposed expert system resulting from the work in this thesis.

To describe expert systems it is useful to give a brief description of several research projects related to remote sensing. A system designed by McKeown et al. (1985) to recognize airports and their constituent parts (e.g., runways and hangars) within aerial photography, will be described, followed by a review of the paper by Brooks (1981) which explains an expert system used to identify types of aircraft. Finally, the SIGMA system of Matsuyama (1987) will be described, which attempts to model the urban environment from aerial photography.

2.4.3 Examples of Knowledge-Based Systems Applied to Object Recognition in Images

The MAPS system designed by McKeown and Denlinger (1984, see also McKeown 1985, McKeown 1991) uses map knowledge to predict the appearance and the position of the expected object (e.g., runways) within the image. This system has been integrated into SPAM (System for Photo interpretation of Aerial Imagery, McKeown et al. 1985) which incorporates models of expected airport configurations and attempts to fit image segments representing such objects as runways, terminals, hangars, and grassy areas, into these models. SPAM uses a 'blackboard' approach to its control strategy. The blackboard is a database containing information on region parameters (e.g., shape and spectral properties) and those objects already recognized. The controlling process
contains three mechanisms, namely the focus of attention, conflict resolution and image segmentation error correction. The focus of attention system directs an object-detection subsystem into an appropriate region of the image. Each type of object has its own specific detection subsystem. For example, elongated regions are searched for roads. The conflict resolution system arbitrates between opposing hypotheses for the same region, assigns a reliability measure to each and removes the less favoured postulation. If, for instance, the shape of a region is not as expected for a particular object but that all other parameters are correct, the segmentation error correction module will attempt to split or merge the segment with neighbouring regions in order to satisfy the shape parameter expectation.

The SPAM system has proved successful in identifying some airports. However, Matsuyama (1987) points to some limitations of the system, namely:-

- all of the knowledge regarding each suspected object is buried within the program structure and it is not therefore apparent what 'knowledge' is used, and
- the system uses only two-dimensional criteria to recognize objects, therefore failing to fully utilize the fact that an image is a two-dimensional representation of a three-dimensional scene.

Another approach to image understanding is that of 'symbolic representation' which is the principle behind ACRONYM developed by Brooks and Binford (Brooks 1981). The ACRONYM system uses generalized cylinders to describe the structure of complex three-dimensional objects (e.g., aeroplanes) in aerial photography. Knowledge of the position of the camera, in relation to the scene, is used to calculate the expected two dimensional appearance of, for instance, a Boeing 747 in the image. This knowledge is stored in a 'frame' data structure which consists of a set of 'slots' containing attribute information relating to the object, as well as relationships between objects and computational procedures. Object description is hierarchical; e.g., an aeroplane consists of fuselage and wings, and a fuselage consists of a nose, body, stabilizers and rudder etc. Criticisms of the ACRONYM system are that, firstly, the recognition strategy is purely bottom-up and that no attempt is made to return to the segmentation stage and look for missing objects (Matsuyama 1987), and secondly, it would seem difficult to integrate spatial knowledge into the system (McKeown et al. 1985).

Matsuyama (1987) combines both bottom-up and top-down processing in the image understanding system SIGMA. It consists of three 'experts', the geometric reasoning expert which examines spatial relationships between objects, the model selection expert which identifies appropriate search models for each object, and the low-level vision
expert which performs knowledge-based segmentation at the behest of the model selection expert. The system uses a 'frame' data structure (as in ACRONYM) and accumulates evidence and hypotheses into an iconic database where such data are also interpreted. The iconic database contains iconic representations of objects within the scene which can be manipulated as an entity rather than a group of pixels.

Although the SIGMA system and the methods proposed by Tailor et al. (1988a) overcome the criticisms of SPAM and ACRONYM it is not clear how each system would cope with a geological application. ACRONYM and SIGMA contend with the three-dimensional problem by calculating a corresponding two-dimensional appearance at an appropriate view direction. However, once the object is recognized it is registered as belonging to a two-dimensional world/surface. In a geological application each object needs to be registered into a three-dimensional world so that models can be created for structures below the surface and for structures above the surface which have been eroded away. It would therefore seem appropriate to represent segment attributes and model hypotheses within a three-dimensional data structure; for instance, a three-dimensional geographic information system (3D-GIS).

In a geological image understanding system, attempting to 'understand' the structural geology of the scene, a bottom-up/top-down combination is essential. Structural primitives (e.g., dip and strike measurements) may be combined to deduce a simple structural model, and the model may then be used to guide further processing. The process may be repeated until a satisfactory model or models have been achieved. These basic ideas will be elaborated upon in Chapter 7 which describes a proposed expert system for structural mapping and those parts of the system which have already been developed.

2.5 Computer Vision

To simulate the complex interpretation processes of the human brain a series of automated techniques are required to identify the component parts of a scene. Such methods come under the broad title of computer vision techniques. Many of these techniques relate to a set of rules known as the Gestalt principles (Wertheimer 1958, Sandford 1985), mentioned in section 2.4.2. It is believed that the human eye recognises lines and areas of uniform texture before the brain pieces these bits of information together and 'understands' the scene (Gregory 1990). The following sections examine image processing techniques for the automated extraction of these phenomena, namely image segmentation and edge detection.
2.5.1 Image Segmentation

Segmentation may be defined as "a partitioning of an image into segments, each of which is a group of adjacent pixels defined by properties which lie within a certain range" (Tailor et al. 1988b). The aim of segmentation is the extraction of pertinent and stable areas; where pertinence is defined as the agreement of the segment with physical or semantic properties of the object and stability is defined as the robustness of segmentation to changes in the processes of image formation such as those found in multitemporal data (Cheevasuvit et al. 1986). Segmentation can also be thought of as "a pre-processing operation which is applied prior to image classification in order to improve classification accuracy from that achievable by classifying pixels individually on the basis of their spectral signatures" (Cross et al. 1988).

Segmentation is both application dependent (for example geological mapping on a regional scale would require a much broader segmentation than that of small-scale mapping) and sensor dependent (in so far as changes in spatial and spectral resolution will alter the precise location and definition of segments to some degree). Most of the literature on image segmentation relates to high resolution images such as aerial photography (Matsuyama 1987, McKeown et al. 1987) and airborne thematic mapper images (Tailor et al. 1988 a, b), usually applied to an environmental or urban mapping problem.

Traditionally, there are two broad areas of image segmentation, those of region based techniques, for example clustering (Seddon and Hunt 1985, Townshend and Justice 1980, Korsnes 1993) and the split-and-merge process (Cross et al. 1988, Cheevasuvit et al. 1986, Laprarde 1988), and edge based techniques (Nevatia and Babu 1980, Peacegood et al. 1986). Clustering merges adjacent pixels iteratively until a similarity criterion is satisfied for each region. Split-and-merge techniques are similar in that a region is recursively split or merged with other regions depending on the state of a heterogeneity measure (Horowitz and Pavlidis 1974, 1976). Finally, edge detection concentrates on the location of points with high-intensity gradients across them and linking these points to describe complete segment boundaries (Nevatia and Babu 1980). A more detailed review of edge detection algorithms is given in section 2.5.2. Both segmentation techniques may be variably more successful than the other, in certain types of scenes depending on the content of the images. In most cases, however, a combination of the two can provide superior results.

Segmentation techniques are prone to many errors, which are manifested as over segmentation and under segmentation, and are due primarily to the data-driven nature of the methods (Tailor et al. 1986, 1988b) and the per-pixel treatment on which they are
based. Adaptive thresholding techniques partially overcome this problem. For instance, a technique modified from Chow and Kaneko (1972), reported by Yanowitz and Bruckstein (1989), identifies points of high-intensity gradient within each column of an image. The algorithm then fits a threshold surface through these points so that a threshold value can be defined for each image pixel. Furthermore, Cheevasuvit et al. (1986) have modified the basic split-and-merge process by using an array of threshold values and then selecting those segments which remain stable through a specified range of thresholds. Gooding et al. (1991) combined a number of techniques within their segmentation algorithm, including histogram splitting and edge detection, to improve results. Segmentation can also be markedly improved by the inclusion of external knowledge, such as maps or specialist domain information, and internal knowledge including such segment parameters as shape, texture, and adjacency within a heuristic segmentation procedure (Nazif and Levine 1984, Korsnes 1993).

Nagao and Matsuyama (1980) designed a 'segmentation by recognition' system in which the fitting of regions to an expected shape parameter and rules of adjacency control the split-and-merge process. Another bottom-up system is that of Nazif and Levine (1984) which utilizes general knowledge about image formation and the laws of perceptual grouping, both of which are independent of the specialist domain and the semantic content of the image. The image segmentation process applied by Laprade (1988) divides the image into areas in which the image intensity surface can be closely approximated by the least-squares fit of a series of planar surfaces. Regions are combined or separated on the basis of an F-test on the residuals of the least-squares calculation.

Cross et al. (1988) describe a two-dimensional segmentation technique, based on segmentation by texture analysis (using average grey value differences (Weszka et al. 1976)) and tone (average grey level over the specified region). Texture/tone vectors are assigned to each pixel within a quadtree data structure (quadtrees are discussed by Burrough 1986). Beginning at a specified level in the quadtree, the sons (or sub-regions) of a region are examined for similarity (based on certain constraints, e.g., a texture measure) to determine whether they can be merged. Conversely regions can be split using similar constraints. A major disadvantage of the texture measure proposed by Cross et al. (1988) within a geological domain is that it is insensitive to the directional aspect of texture which can be present in heavily jointed and strongly stratified rock units. A similar texture-based segmentation system designed by Raafat and Wong (1986) uses the grey level and a gradient vector to describe region texture. The gradient vector includes both gradient magnitude and gradient directionality and may prove useful in geological studies. Another measure of texture directionality is that of the local directed standard deviation (LDSD) measured in four directions (vertical,
horizontal, diagonal and anti-diagonal) and the mean LDSD measured over a broader region (Wang et al. 1986).

A number of texture measures may also be derived from a co-occurrence matrix of a window passed across the image (Haralick et al. 1973, Gotlieb and Kreyzig 1990). These measures may also be directional if required. Hsiao and Sawchuk (1989) describe an unsupervised texture-based segmentation technique based on Laws' texture energy measures (Laws 1980). They further describe a probabilistic relaxation method which is used to reduce local ambiguities in the segmentation process.

Several authors have used a fractal operator as a segmentation constraint (Hyde et al. 1985, Korsnes 1993). The theme of fractal geometry has been described as that of self similarity and the "degree to which objects appear similar, but nevertheless differ at varying levels of magnification" (McLaren and Kennie 1989). Fractal geometry has been used to simulate landscape and terrain (Mandlebrot 1975, Yokoya and Yamamoto 1989). However, such an application implicitly suggests that the spatial structures within the landscape repeat themselves over a range of scales. Although this technique has been used to produce aesthetically pleasing and realistic looking images (McLaren and Kennie 1989), their attractiveness does not necessarily reflect an accurate representation of the image subject. As Burrough (1985) points out "most landscapes are not the result of a single dominant process but are the result of the complex interaction and superimposition of many processes". Having said this there is no doubt that fractal measures may be extremely useful as a texture measure (Dodd 1987). Yokoya and Yamamoto (1989) suggest that fractal dimensions can be useful for interpolating terrain models to higher resolutions, although they do point out that such interpolation must have an upper limit, where fractals no longer describe the surface accurately. They further point out that fractals may be used anisotropically to pick out directional features in the terrain (see also Olsen et al. 1993).

Sali and Wolfson (1992) point out that no one texture measure sufficiently describes all features in a remotely sensed scene. They therefore propose the application of a series of texture measures to a scene followed by an unsupervised clustering of the resulting texture feature space. The combination of first and second-order statistical measures and fractal based measures resulted in an improved segmentation of a natural vegetated scene.

This section has described a number of different segmentation methods, each having their own advantages and disadvantages. As with the image enhancement techniques there is no one method which stands out as the most successful and no group of methods which appear particularly suited to geological applications. It has been worth
mentioning each technique to give an indication of the wide range of techniques and what is possible in computer vision.

2.5.2 Edge Detection and Pattern Recognition

A plethora of edge detection and pattern recognition algorithms can be used to aid the identification of objects. Matsuyama (1987) suggests the use of an image processing expert, which has available to it various image processing algorithms, but which also contains knowledge about the usefulness and relevance of each algorithm to each application and knowledge concerning the appropriate parameters to pass on to the algorithm chosen in relation to that algorithm. This section attempts to take a brief look at those algorithms which might prove useful in a geological application and does not intend to include a comprehensive guide to edge detection and pattern recognition (such a review can be found in Blicher 1985).

Once the brain has recognized an object in a scene it is easy to recognize how that object is made up of a series of edges. It would therefore seem trivial to define boundaries within a digital image by finding sharp changes in grey level around homogeneous regions. However, an edge rarely presents itself in this ideal way, especially in remotely sensed images. The 'ideal' homogeneous regions are affected by topography, illumination, view angle, noise and variations in its spectral properties. Edges can disappear in shadows, occlusions and between objects having similar radiometric responses. Various methods have been introduced to enhance edge or boundary detection such as feature following, line joining and the inclusion of low- and high-level knowledge of the scene. These will be discussed later. First some standard edge detection algorithms will be reviewed.

Edges can most easily be imagined in one-dimension where they appear as step edges, roof edges or pulse functions (Figure 2.15). In two dimensions similar representations can be considered with the inclusion of corner edges. As can be seen from Figure 2.15 edges occur where the gradient of the intensity function is at its highest with the exception of the roof edge. However, this is less common in remote sensing (Blicher 1985). Therefore, by finding the gradient or first order derivative of the intensity function, the strength of the edges can be found. The Sobel operator (Figure 2.16) is a good example of a first-order derivative filter (Schowengerdt 1983). The filter works by calculating the gradient between neighbouring pixels in both the horizontal and vertical directions. This basic model can be adapted to a variety of operators including linear, non-linear, directional or rotationally symmetric detectors (Torre and Poggio 1986). To date, no attempt has been made in the literature to include the third dimension of height into the operator. This could be effected by adapting the weighting functions of the
Figure 2.15 - Different type of edges displayed one dimensionally.

Sobel filters

\[
\begin{array}{ccc}
-1 & -2 & -1 \\
0 & 0 & 0 \\
1 & 2 & 1 \\
\end{array}
\]

Identifies N-S edges

\[
\begin{array}{ccc}
-1 & 0 & 1 \\
-2 & 0 & 2 \\
-1 & 0 & 1 \\
\end{array}
\]

Identifies E-W edges

Figure 2.16 - The Sobel filters. The results of both filters can be combined to produce edge strength and edge orientation images.
operator to the three-dimensional distances between neighbouring pixels and the pixel under examination.

Calculation of the second derivative will map edges to zero-crossings (i.e., where a sign change occurs) in the resulting image. An example of this is the Laplacian operator which also gives negative values at the foot of a slope and positive values at the crest (Mather 1987). The weighting of a second derivative can be calculated using a Hessian function matrix (Blicher 1985) and it is the use of various properties of this matrix which has stimulated the design of several more sophisticated edge detectors (Marr and Hildreth 1980, Beaudet 1978, Canny 1986, and Petrou and Kittler 1988, Abramson and Schowengerdt 1993).

Despite their sophistication, these operators are sensitive to noise (creating false edges) and can produce many responses to a single edge. To obtain an optimal performance Canny (1986) created three performance criteria by which to measure the performance of edge detection operators. These criteria include the signal-to-noise ratio, localization (i.e., the accuracy with respect to the actual edge) and proximity to a single response per edge, of which the signal-to-noise ratio is the most significant (Petrou and Kittler 1988). Canny derived quantitative measures for these criteria by which operators could be evaluated. He proposed the derivative of the Gaussian function as the best approximation to the optimal solution. Later developments by Spacek (1986) and Petrou and Kittler (1988) derived cubic splines and the definitive optimal operator respectively as the optimal solution.

A completely different type of edge detector, which could be extremely useful in a geological application, is that of the entropy operator (Shiozaki 1986). The operator can be applied to work in 'colour space' and returns low values in areas where hue and brightness change rapidly and high values in homogeneous areas. It could therefore be used following a suitable colour enhancement algorithm; e.g., the decorrelation stretching technique of Rothery (1987b).

The edge detectors discussed thus far operate at a local scale; i.e., within a local window surrounding the pixel in question. An example of a global edge detector is provided by the Hough transform (Hough 1962, Blicher 1985), which is used to detect curves, straight lines or arbitrary, but known, shapes. Cross (1988) and Parrot and Taud (1992) used the Hough transform in an attempt to detect circular geological features, for example igneous intrusions. In their work on the automatic interpretation of synthetic aperture radar (SAR) images, Quegan et al. (1988) recommend the use of the Hough transform as an initial line detection technique prior to image segmentation. The Hough technique remaps all possible edges in the image in terms of the parameters $p$ and $\theta$. 
and assign a value of the mean intensity along the line to each \((p, \theta)\) point. Features are then represented by light or dark points in the new parameter space.

It is worth noting here that most authors use thresholding techniques to separate 'real' edges from spurious noise or false edges. However, within the context of an expert system, where segments are split or merged with respect to confidence levels assigned to them, then it is useful to retain the 'knowledge' of edge strength and orientation produced by the edge detectors rather than assign some arbitrary threshold value (Lewis 1987). This edge knowledge can then be used in conjunction with feature-following algorithms which predict where extensions to lines should occur and try to match several distinct edges into a meaningful line or feature (Peacegood and Wilkinson 1985). McKeown and Pane (1985) and Weiss and Boldt (1986) use a rule-based approach to the linking of edges; i.e., only link two edges if they have similar intensity gradients and orientations. Other work in this area includes the river network detector described by Haralick et al. (1985a) and the road detectors described by Fischler et al. (1981) and Zhu and Leh (1986).

2.6 Aim and objectives

2.6.1 Summary of literature review

It is clear from this literature review that there are many sources of data for geological mapping and that there are also a large number of techniques available to enhance and manipulate these data. However, there is no clear definition of which data or which techniques are suitable for different geological environments. Many methods mentioned in this chapter appear to have certain advantages and disadvantages depending on the application for which it is used and the data upon which it is executed. Chapter 4 therefore includes an evaluation of a number of these techniques when applied to remotely sensed data and their suitability for incorporation into an expert system. This evaluation is by no means exhaustive due to the lack of available software, but does include the most commonly used techniques.

Little use has been made of elevation data, particularly in geological applications, although the important work of Chorowicz et al. (1987) and McGuffie et al. (1987) will be built on in Chapter 5. Chapter 6 describes how data derived from both remotely sensed and elevation data may be used to produce geological structural models. A number of automated procedures and knowledge-based rules will also be introduced, and Chapter 7 will outline how some of the techniques reviewed in this chapter may be
Figure 2.17 - Schematic illustration of the Hough Transform method. (a) There are three nearly colinear points in the \((x,y)\) plane and the corresponding curves in the \((\rho,\theta)\) plane are shown in (b).
combined with those developed in this thesis within the framework of a proposed expert system.

2.6.2 Aim

The aim of this thesis is to provide tools for the quantitative mapping of three-dimensional geological structures, both remotely and automatically.

2.6.3 Objectives

Following on from the work outlined in the literature review several objectives are defined to reach the specified aim. These are:-

- to assess a range of traditional image enhancement methods in providing data suitable for incorporation into automated procedures
- to assess various methods of creating DEMs,
- to produce tools for extracting the geological structural information from images and elevation data, both manually and automatically,
- to assess the accuracy of these measurements,
- to develop automated techniques and knowledge-based rules to produce structural models, and
- to propose an expert system which will automatically perform an analysis of the three-dimensional geological structure of a scene.
3.1 Study Area

The site used in this research features an area of mountainous terrain located near Capel Curig, Snowdonia National Park, U.K. (Figure 3.1). The location was selected partly because of the close relationship between the topography of the area and its geology (Figure 3.2A), but was more directly dictated by the availability of suitable airborne remotely sensed data (acquired on 19th July 1989, more details of which are given in Chapter 4). The centre of the area is dominated by a substantial reservoir, Llyn Cowlyd, situated at the base of a large glaciated valley (Figure 3.2B). Although the topographic expression of the geology is masked in some areas by the effects of glaciation, in general the surface morphology can be used extensively to map the geology. This is particularly so in areas where there is an inter-layering of hard and soft rocks, at the summits of mountains, and in areas of steeper slope (Figures 3.2C and 3.2D).

The vegetation cover includes coniferous forest (Forestry Commission), bilberry, rowan tree, heather, bracken, and sheep-grazed grassland. The grassland becomes boggy in some low-lying areas and also in the higher areas where the softer geology has caused hollows between the harder rocks. The weather prior to the flight had been very dry, resulting in the low-water level in the reservoir at the time of data collection. As a result the grassland was quite dry, but the bogs were not fully devoid of water. Differences in underlying geology were not reflected in vegetation changes across the area and only marginally apparent in the wetness of the grassland due to the lack of rainfall.

3.2 Geology of the study area

The district around Capel Curig forms part of the Welsh Basin, in which thousands of metres of sediment accumulated during Lower Palaeozoic times (Howells 1979). The basin is thought to have been separated from a proto-Atlantic ocean to the north by an Irish Sea land mass, possibly related to an earlier subduction zone (Dewey 1969). The land mass is presumed to be the source of much of the sedimentation in the basin,
Figure 2.1 – The Llyn Cowlyd Study Area
Figure 3.2 - Views of the Llyn Cowlyd study area
particularly the coarser clastics of the Carneddau Group. Figure 3.3 shows a
diagrammatic representation of the sedimentary sequence indicating the relative
thicknesses of each unit. Sediments in the area comprise mudstones, siltstones, and
sandstones, indicating a continuing fairly shallow marine environment in which
sedimentation kept pace with subsidence of the basin. The sandstones are greywackes
which may signify instability at the margins of a subsiding basin. Towards the top of
the Carneddau Group the siltstones grade through to mudstones, indicating a deepening
of the sea bed (Dewey 1969). This deeper water environment is interspersed with tuffs
(predominantly acidic in composition) during the Capel Curig Volcanic Formation and
the Crafnant Volcanic Formation, which caused local shallowing in places. The entire
sedimentary sequence is intruded in places by dolerite sills which are locally
transgressive, together with a few dykes trending in various directions.

From personal field surveys, it is clear that the topographic expression of this geology is
manifested as quite distinct geomorphological features caused by the inter layering of
hard and soft rocks. The harder and more resistant rocks include dolerite, ash-flow tuff
and sandstone, and are generally located along ridges or prominent features such as
cliffs; while the soft rocks comprise slate, mudstone, siltstone and tuff, and are found in
more low-lying areas, which are more likely to be covered by vegetation.

The main deformation phase occurred during the early stages of the Caledonian orogeny
and caused major folding and faulting along a north-easterly orientation. The folds are
gentle to isoclinal and have a gentle plunge to the north-east and axial planes which dip
steeply to the north-west. A second, less pronounced deformation occurred along a
south-easterly direction. A simplified geological map of the area is shown in Figure 3.3.

The Llyn Cowlyd area provides a good training ground for the objectives set out in
Chapter 2. There is a strong relationship between geology and topography although this
is masked in places by the effects of glaciation. There is only an overall exposure of
less than 20% and the structure is a little more complex than ideal for a first study area.
Many of the techniques developed in this research to map geological structure are in
some ways unavoidably specific to the Llyn Cowlyd area. Future data collection would
hopefully include a full range of geological and climatological environments allowing
techniques to be developed which are more generally applicable.
Figure 3.3 – Geology of Llyn Cowlyd area, derived from the published 1:50,000 geology map (BGS 1985)

Legend

- Dolerite
- Mudstones (thin sdst. bands) – CyM
- Trefriw tuff (basic tuff)
- Mudstone (black) – CaS
- Acid ash flow tuff – M/UCV
- Siltstones/mudstones – CV
- Hyaloclastite, basalt – DV
- Acid ash flow tuff – LCV
- Sandstone – CEi
- Siltstone – CEi
- Acid tuff, tuffite – CEi
- Lakes/reservoirs

CyM – Conwy Mudstone
CaS – Cadnant Shales
M/UCV – Middle/Upper Crafnant Volcanic Formation
LCV – Lower Crafnant Volcanic Formation
CEi – Cwm Eigiau Formation

Scale 1:35,000

0 1km
Chapter 4 - Deriving Structural Geological Information from Remotely Sensed Images

The primary focus of Chapter 4 is to examine ways in which structural geological information can be derived from remotely sensed images. Unfortunately the information inherent in such images is complex, related as it is to different land-cover types present in a particular scene as well as to geological phenomena. Here the aim is to extract the geological information, separating it from information relating to other items, such as anthropogenic features and changes in vegetation type. To perform these tasks, image processing techniques are required to enhance features of interest in an image so that they may be identified manually or extracted automatically. For each identified feature it is important to obtain a measure, or confidence statistic, as to whether the feature represents a true geological phenomenon, before it can be included in a geological structural model.

Several of the techniques described in Chapter 2 can be used to enhance and extract geological features from remotely sensed images. Further techniques are also available specifically to segment images into areas which may contain geological information and those which do not. Examination and evaluation of a wide range of these techniques is carried out in this Chapter. Although emphasis is placed on their suitability for use with the images used within this study, their possible uses in other study areas and application to different remotely sensed data is also examined. The techniques encompass three broad types of data manipulation, namely, enhancement, segmentation, and line extraction. In particular, the following techniques are examined in detail:-

Enhancement
- Red-Green-Blue (RGB) Colour Enhancements
- Hue-Saturation-Intensity (HSI) Transformations
- Band Ratios
- Principal Component Analysis (PCA)
- Decorrelation Stretches
Segmentation

- Multispectral Classification
- Region Growing

Line extraction
- Sobel Filter
- Compass Filter
- Fourier Analysis
- Canny Filter

These techniques have been selected for analysis as they are often mentioned in the literature, in either geological or image vision contexts. Each technique is evaluated in terms of its success in deriving structural information measured relative to ground survey data of the study area. The enhancement techniques are initially evaluated subjectively, assessing the clarity with which each resultant image highlights certain geological features within the image. This is important to a user when manually interpreting a scene. Subsequent evaluation assesses the number of features identified both manually and automatically from each of the enhanced images. The accuracy of these features is derived through comparison with the appropriate geological map.

To identify geological features automatically from remotely sensed data it is fundamental that the extracted lines should be positioned accurately and not fragmented. Line extraction techniques are therefore evaluated in terms of the number of features they identify, the accuracy to which the features are identified, and the comparative length of each identified feature.

Prior to examination and evaluation of these techniques it is necessary to apply several pre-processing steps to the data in order to convert the image to radiance and to geometrically correct the data to a suitable map projection.

4.1 Pre-processing of remotely sensed images

Pre-processing of remotely sensed data is required to convert the images, originally acquired as arbitrary digital numbers (DN), into physical units of radiance (W⁻⁷cm⁻²μm⁻¹sr⁻¹) or reflectance (%) and to project the image space into a real world co-ordinate system. Conversion of the image into reflectance, the ratio of outgoing radiance to the incoming irradiance, allows a comparison of pixels in the image with known library spectra of cover types and a more direct inter-comparison between wavelengths, e.g., for band ratios. To determine reflectance (or more accurately bidirectional reflectance), information is required regarding the levels of irradiant flux impinging on the surface.
from all angles within the illumination hemisphere, the amount of radiant flux which is then absorbed by and refracted through the atmosphere, and a full description of the directional components of the radiant flux (Nicodemus et al. 1977) (Figure 4.1). It is impossible to record radiant and irradiant flux at the infinitesimal angle increments which are required to describe the bidirectional reflectance distribution function (BRDF) of individual cover types. It is recognized that these points combine to result in an error in the calculation of reflectance but that the methods used to calculate reflectance are the best that can be achieved given the data available.

4.1.1 Special characteristics of the data

The image data used (Figure 4.2) in this investigation were acquired by the Natural Environment Research Council (NERC) funded Daedalus Airborne Thematic Mapper (ATM) scanner (AADS-1268) which records data in 11 separate wavebands from the visible to the thermal infra-red wavelengths (Table 4.1). The image was obtained at 9:15 GMT on 19th July 1989. The altitude of the aircraft was approximately 2000m giving a nominal spatial resolution of 5m (Table 4.2) in the nadir viewing position, although the rugged terrain and the wide scan angle of the ATM (86°) results in extremely variable spatial resolution throughout an image (Barnsley and Kay 1990).

<table>
<thead>
<tr>
<th>Spectral Channel</th>
<th>Wavelength μm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.42-0.45</td>
</tr>
<tr>
<td>2</td>
<td>0.45-0.52</td>
</tr>
<tr>
<td>3</td>
<td>0.52-0.60</td>
</tr>
<tr>
<td>4</td>
<td>0.605-0.625</td>
</tr>
<tr>
<td>5</td>
<td>0.63-0.69</td>
</tr>
<tr>
<td>6</td>
<td>0.695-0.75</td>
</tr>
<tr>
<td>7</td>
<td>0.76-0.90</td>
</tr>
<tr>
<td>8</td>
<td>0.91-1.05</td>
</tr>
<tr>
<td>9</td>
<td>1.55-1.75</td>
</tr>
<tr>
<td>10</td>
<td>2.08-2.35</td>
</tr>
<tr>
<td>11</td>
<td>8.50-13.00</td>
</tr>
</tbody>
</table>

Table 4.1 - Wavelengths of the ATM AADS 1268 scanner

In its conventional mode of operation the ATM scanner has an S-bend correction which corrects for the angular distortion introduced by the rotating scanner, resulting in each ground resolution element (GRE) being separated by an equal distance (Barnsley and Kay 1990). Due to the requirements of another project carried out using the same data
$\text{ID} = \text{Direct Irradiant Flux}$
$\text{Id} = \text{Diffuse Irradiant Flux}$
$\text{RD} = \text{Direct Radiant Flux}$
$\text{Rd} = \text{Diffuse Radiant Flux}$
$\text{Sd} = \text{Diffuse Sky Radiant Flux}$

**Figure 4.1.** The irradiant and radiant flux components which must be measured to determine reflectance accurately (After Woodham and Gray, 1987).
Figure 4.2 - ATM image of Llyn Cowlyd
the S-bend correction was not applied to the data. Therefore, each pixel across the track was separated by an equal angle of view, i.e. scale was not constant across the scan line. This combines with the geometric distortions inherent in the image due to the roll, pitch, and yaw of the aeroplane and the distortions introduced by the mountainous nature of the terrain.

<table>
<thead>
<tr>
<th>Instantaneous Field-of-view</th>
<th>2.5 mrad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digitized Field-of-view</td>
<td>85.92°, 72° after S-bend correction</td>
</tr>
<tr>
<td>Roll Correction</td>
<td>±15°</td>
</tr>
<tr>
<td>Height of Aircraft</td>
<td>2000m</td>
</tr>
<tr>
<td>Nominal Ground Resolution</td>
<td>5m</td>
</tr>
</tbody>
</table>

Table 4.2 - Specifications of the ATM AADS 1268 scanner

Measurements of total irradiance were acquired on the ground during image acquisition. The measurements were taken at a distance of approximately two kilometres from the study area due to the requirements of a coincident study, but the clear sky and constant illumination conditions suggest that the measurements should be reasonably applicable to the present study area. A Spectron SE-590 spectro-radiometer, provided by the NERC Equipment Pool for Field Spectroscopy, was used to measure the irradiance (Rollin and Milton 1988). The Spectron was fitted with a 'cosine head' (a fish eye lens designed to integrate flux over the hemisphere), such that incoming radiation has a correction of \( \cos(\text{zenith angle}) \) applied to it. A measure is therefore obtained of the total incoming irradiance. The Spectron records data in 252 wavebands at intervals of 2.8nm between 400nm and 1100nm and can therefore only be used in conjunction with ATM bands 1 to 8.

4.1.2 Radiometric calibration of the data

To calculate reflectance the digital numbers (DN) must first be converted to radiance. This involves modification of the DN by separate gain and offset values for each band (Table 4.3) (Wilson 1985), using the following formula:-

\[
\text{RADIANCE} = \text{GAIN} \times (\text{DNVALUE} - \text{BASE})
\]  

(4.1)

where,

- \( \text{GAIN} = n / (\text{DNON} - \text{DNOFF}) \) at gain setting 1,
- \( \text{BASE} = \text{DNOFF} \) at gain setting 1,
- \( \text{DNON} = (\text{VCAL} \times 256) / 4000 \) DN value for calibration source,
DNOFF = ( VO * 256 ) / 4000  DN value for zero input source,
N = Average panel radiance in W\(^{-7}\) cm\(^{-2}\) \(\mu\)m\(^{-1}\) sr\(^{-1}\),
VCAL = Sensor voltage from calibration source mV DC (Max 4000mV),
VO = Sensor voltage from zero input source mV DC (Min 0mV).

<table>
<thead>
<tr>
<th>Band</th>
<th>N</th>
<th>VCAL</th>
<th>VO</th>
<th>DON</th>
<th>DOFF</th>
<th>GAIN</th>
<th>BASE</th>
<th>Gain Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.96</td>
<td>518</td>
<td>266</td>
<td>33.15</td>
<td>17.02</td>
<td>0.370</td>
<td>17.02</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>11.36</td>
<td>603</td>
<td>246</td>
<td>38.59</td>
<td>15.74</td>
<td>0.497</td>
<td>15.74</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>21.55</td>
<td>858</td>
<td>307</td>
<td>54.91</td>
<td>19.65</td>
<td>0.611</td>
<td>19.65</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>28.79</td>
<td>1298</td>
<td>523</td>
<td>83.07</td>
<td>33.47</td>
<td>0.580</td>
<td>33.47</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>35.11</td>
<td>1256</td>
<td>297</td>
<td>80.38</td>
<td>19.01</td>
<td>0.572</td>
<td>19.01</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>41.64</td>
<td>1369</td>
<td>300</td>
<td>87.62</td>
<td>19.20</td>
<td>0.609</td>
<td>19.20</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>48.12</td>
<td>1297</td>
<td>160</td>
<td>83.01</td>
<td>10.24</td>
<td>0.661</td>
<td>10.24</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>49.33</td>
<td>1928</td>
<td>295</td>
<td>123.39</td>
<td>18.88</td>
<td>0.478</td>
<td>18.88</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>22.00</td>
<td>3351</td>
<td>234</td>
<td>214.46</td>
<td>14.98</td>
<td>0.110</td>
<td>14.98</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>9.34</td>
<td>6508</td>
<td>320</td>
<td>416.51</td>
<td>20.48</td>
<td>0.023</td>
<td>20.48</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 4.3 - Calibration data for the Daedalus ATM scanner - calibrated 13th July, 1989

Only the first ten bands have been calibrated as the NERC does not provide calibration figures for band 11, the thermal infra-red band.

Following the conversion to radiance an approximate measure of reflectance may be obtained by ratioing the radiance data with the coincident measurements of irradiance. As the irradiance was measured in different wavelengths and bandwidths to the ATM data, a number of bands in the irradiance data were combined to simulate the ATM bands. Figure 4.3 shows the spectral response of the ATM sensor as a function of wavelength and these graphs were used to provide weighting factors for each of the Spectron bands within the spectral response of each ATM band.

As a result of the limitations of the calibration data and the irradiance data, reflectance data can only be achieved for ATM bands 1 to 8, radiance can be calculated for bands 9 and 10, and band 11 can only be used as DN.
Figure 4.3 - Spectral curves for ATM bands using Spectron wavebands.
4.1.3 Geometric Correction

Airborne remotely sensed images tend to suffer geometric distortion from a number of sources, such as platform vibration, roll, pitch, and yaw of the aircraft, terrain relief, and the instrument scanning methods. There is a number of algorithms available for the geometric correction of such images, and these will be described briefly here. These include:

- Manual selection of control points followed by polynomial warping.
- Manual selection of control points followed by Delauney triangulation.
- Automatic matching of line features in the image to digitized map features.
- Automatic correction of each scan line using accurate inertial navigation system (INS) measurements recorded on the aircraft in conjunction with digital elevation data of the area.
- Automatic control point generation followed by stereo-matching the images to coincident aerial ortho-photography.

Polynomial warping is probably the most commonly used method for the geometric correction of remotely sensed data, both satellite and airborne. Unambiguous ground control points (GCPs) are identified within the image, such as road intersections and river features, which can be also recognized on a map of the area. The warping method then involves fitting polynomials of x and y to the control point data using a least squares approach (Schowengerdt 1983). These equations are then used to resample the data to the corrected image space. The polynomials may be of any order, providing there are more (or at least as many) control points than unknown coefficients in the polynomial equations. For instance the following 2nd order polynomials require six or preferably more control points to perform the least squares fit.

\begin{align*}
    x' &= a.x^2 + b.x.y + c.y^2 + d.x + e.y + f \\
    y' &= g.x^2 + h.x.y + i.y^2 + j.x + k.y + l
\end{align*}

Higher-order polynomials may be used to correct more severe distortions in an image and conversely, lower-order polynomials for correcting lesser distortions. However, the higher-order polynomials tend to become more unreliable away from the specified control points. ATM images are susceptible to high frequency distortions due to the instability of the platform and this, coupled with the added distortions introduced by the
mountainous terrain, means that high-order polynomials are required to correct such data.

Deveraux et al. (1990) have developed a geometric correction system designed for airborne images. The system segments the image using Delauney triangulation between GCPs. A linear interpolation is then used to find ground co-ordinates of the centre of each triangle. Resulting new co-ordinates are then added to the list of original GCPs enabling a second triangulation. This process is repeated until all triangles are below a specified size. A linear interpolation is then performed within each triangle to resample the image data. This piecewise linear approach will not model the high frequency distortions in ATM imagery unless GCPs are identified at intervals of several scan lines and several pixels across track (Allison and Muller 1992). This is clearly impractical, especially in a scene such as this due to lack of possible control points. Furthermore the linear interpolation is not suitable for images acquired without S-bend correction, as was the case with the Llyn Cowlyd image. A more appropriate interpolation would include a \( \tan \theta \) function (where \( \theta \) is the across track view angle) in the across track direction.

Dowman et al. (1983) recognized that high frequency distortions in airborne images are most apparent in linear features within an image. Furthermore, if these linear features can be matched to digitized map features then the remainder of the image is likely to be corrected fairly accurately. Successful results obtained with this technique could therefore be expected with images containing numerous linear features, such as roads and field boundaries. However, the study area in question incorporates few linear features. The method also suffers from the inability to detect the linear features with any reliability and continuity (Dowman et al. 1983).

A more accurate and reliable method of geometric correction could be achieved by using high precision data from an inertial navigation system (INS) which can obtain positional, roll, pitch, and yaw data for each scan line that is recorded in the image. Schwarz et al. (1993) have shown that INS data combined with data from Global Positioning Systems (GPS) can produce positional accuracies of 10-15cm, more than sufficient for 5m resolution ATM data of flat terrain. If these data are combined with distance to ground measurements (for which a DEM is required) a ground position for each pixel in a mountainous terrain could be calculated accurately. Unfortunately, these data are not generally available and were not acquired for the image used in this study.

Finally, a promising technique, still to be fully researched, involves stereo-matching the airborne images to coincident aerial ortho-photography (Allison and Muller 1992). Allison and Muller have used the Llyn Cowlyd data to test their methods and some initial results will be shown here. Aerial photography suffers less from geometric...
distortion than airborne images as it is acquired instantaneously and therefore not
affected in the same way as scanner data by movements of the aircraft, i.e., on a line-by-
line basis. Residual distortions are due to the lens and the terrain. Lens distortions can
be removed using a camera model (a set of parameters describing the geometry of the
camera and the lens) (Zemerly et al. 1992). Terrain distortion can be eliminated using a
DEM of the area, also derived from stereo pairs of aerial photography (Figure 4.4).
Stereo-matching techniques (described more fully in section 5.1.1) can also be used to
derive the DEM. Once these distortions are known an orthoimage of the aerial
photography can be produced (Figure 4.5). The ATM image may subsequently be
matched to the orthoimage using the same stereo-matching technique, to produce an
orthoimage of the ATM data. Due to the high frequency distortions in the ATM data the
matching process often fails. Allison et al. (1991) therefore proposed an intermediate
stage in which the airborne data are warped using the polynomial method and a set of
automatically identified control points. This removes the gross distortions in the ATM
data leaving the stereo-matcher to account for the higher frequency distortions. Figure
4.6 shows an orthoimage of the ATM imagery (band 5). It can be seen from this Figure
and Figures 4.4 and 4.5 that these techniques do not as yet produce a complete corrected
image, with holes appearing in the data sets. This is due to misregistration in the stereo-
matching process resulting from large local distortion caused by the terrain and by
radiometric differences between the image and the photography. Unfortunately, it is
these large terrain distortions that are of interest for geological mapping. It is hoped that
future developments of the matching procedure will solve some of these problems and
that such techniques may be used on an operational basis.

Of all the geometric correction techniques mentioned above, the polynomial warping
and Delauney triangulation techniques are the only two suitable for use in this study.
Due to problems associated with Delauney triangulation as mentioned above, the
polynomial warping method was selected to correct the ATM image.

A set of control points was identified in both the image and Ordnance Survey 1:10,000
scale base maps. For the most part, the control points were identified at locations
obvious in both image and map. However, these were mainly along tracks, streams, and
the reservoir shore and do not therefore describe the higher ground. Rock exposures
were also used as control points on the higher ground where these could be identified

1 These control points are identified using the Foerstner interest
operator (Foerstner and Gulch 1987), which uses a patch based
matching technique to identify similar points between the ATM
data and the aerial photography.

2 A program to perform the polynomial warping was written as part
of this thesis.
Figure 4.4 - Lambertian shaded DEM produced from stereo-matching of two aerial photographs of the Llyn Cowlyd study area. The areas of uniform grey tone indicate errors where the stereo-matcher has failed. The area shown is centred on GR 272700 361500 and covers an area approximately 800 x 500 m (north is towards the top of the page). Source - David Allison, University College London.
Figure 4.5 - Orthoimage of aerial photography of Llyn Cowlyd area, corrected using a stereo-matched DEM. The white areas indicate errors where the stereo-matcher has failed. The area shown is centred on GR 272700 361500 and covers an area approximately 300 x 400 m (north is towards the top of the page). Source - David Allison, University College London.
Figure 4.6 - Orthoimage of ATM data (band 7) of Llyn Cowlyd area, corrected using a stereo-matched DEM. The black areas indicate errors where the stereo-matcher has failed. The area shown is centred on GR 272700 361500 and covers an area approximately 800 x 500 m (north is towards the bottom of the page). Source - David Allison, University College London.
accurately on the map\textsuperscript{3}. An additional attempt was made to match points in the remotely sensed image to corresponding points in a Lambertian-shaded DEM image (see section 5.1.2 for a description of this type of data). Although many surface features could be identified, no single points could be located with sufficient accuracy, probably due to lack of detail in the DEM (a higher resolution DEM, =1 metre is required for this purpose) and unrealistic shading properties of the Lambertian model.

<table>
<thead>
<tr>
<th>Order of polynomial</th>
<th>RMS Error in pixels</th>
<th>Maximum Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.61</td>
<td>54.63</td>
</tr>
<tr>
<td>2</td>
<td>9.99</td>
<td>24.53</td>
</tr>
<tr>
<td>3</td>
<td>5.44</td>
<td>11.59</td>
</tr>
<tr>
<td>4</td>
<td>4.04</td>
<td>7.17</td>
</tr>
<tr>
<td>5</td>
<td>1.89</td>
<td>3.74</td>
</tr>
</tbody>
</table>

Table 4.4 - Geometric correction accuracy as a function of the order of polynomial used

The forty seven control points used in the geometric correction are shown in Figure 4.7. Several different order polynomials were tested to examine which best modelled the distortions within the data. The results of each warp are shown in Table 4.4. During the warping procedure estimates are obtained of the error for each control point in relation to the polynomial equations used. This gives an indication of the success of the geometric correction. Individual errors are combined to give an overall root mean square (RMS) error for the warping procedure. Results show that lower-order polynomial functions are insufficient to describe high frequency distortions within the image. Polynomials of order six or higher tend to over compensate in areas between the control points\textsuperscript{4}. The most accurate geometric correction is achieved by using a 5th order polynomial in X and Y. It should be recognized that this result, at =9.5 metres, is still very poor, but that it is at present the best that can be achieved. Figure 4.8 shows a standard false colour composite of the corrected image for the 5x5km study area. Subsequent processing and investigation will concentrate in a 1.5x1.5km sub-scene for three reasons: firstly the corrected image covers this area totally; secondly, to increase

\textsuperscript{3} The positional accuracy of these control points on the map is less certain as these map features are simply hand-drawn interpretations of the ground features rather than rigorously surveyed points.

\textsuperscript{4} An attempt was made to use sixth order polynomials in the geometric correction program but unfortunately the program crashed presumably due to problems introduced by these problems and the large numbers produced in the matrix inversion routines which are possibly be beyond the precision of the computer.
Figure 4.7 - Control points identified for the polynomial warp
Figure 4.8 - Geometrically corrected image of the Llyn Cowlyd area (5 x 5 km) showing the 1.5 x 1.5km extract area as a black outline.
processing speeds; and thirdly, this area exhibits some exposed geology and prominent structures.

Poor results from the geometric correction will have cumulative effects on any comparison between the remotely sensed images and the DEM and on any products derived from these data. One method to assess these effects is to use the error statistics derived from the geometric correction. Such error values may be interpolated over the entire image area to give an estimate of geometric accuracy for each pixel. Figure 4.9 shows the interpolated errors in the X and Y directions respectively. As will be described later in Chapter 6, these may be used to assess subsequent errors and as a confidence statistic for any products derived from the data.

4.2 Extracting geological information from remotely sensed data

This section investigates a number of methods to enhance geological features within remotely sensed data, techniques to extract this information from images, and ways in which knowledge can be gained from the data as to which areas are more likely to contain geologically relevant information.

4.2.1 Colour enhancement of lithological features

Each band of ATM data provides different information regarding the reflectance of the earth's surface. However, the images produced in certain wavebands are often highly correlated. This indicates a degree of redundancy in the data set, such that a selected sub-set may contain a large proportion of the total variance in the original eleven wavebands. To extract useful geological information from the data, knowledge is required regarding which band (or combination of bands) provides the most effective information. The bands chosen should be able to discriminate between different lithologies or to enhance other features which may relate to the underlying geology (e.g., variations in vegetation type, soil, surface temperature or surface moisture). Differences in lithologies can often be masked by weathering of the rock surface or a covering of lichen or other surficial material. Consequently, band combinations are required to enhance differences in these surfaces. Table 4.5 shows the major surface characteristics of the study area, indicating which band combinations and techniques are generally considered most important in discriminating each. The Table highlights eight of the eleven bands considered useful in this application, six of which are used for distinguishing lithologies. The only surface characteristic that cannot be measured with the existing data is surface moisture (as both day and night thermal images are required to measure the thermal inertia of the surface, related to surface moisture).
Figure 4.9 - Perspective View of Errors in X and Y, from the Polynomial Warp.
Specified band combinations are used in band ratio techniques which can subsequently become indices for the surface characteristics. For example, 7/5 is often used as a vegetation index while 9/10 is used as a weathering (or clay minerals) index. These monochrome indices are ideal for automated techniques such as edge detection whereas a combination of various band sets in a colour image is preferable for visual interpretation of the data.

<table>
<thead>
<tr>
<th>Surface characteristic</th>
<th>Band/band combination</th>
<th>Image processing technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithologies</td>
<td>3 &amp; 5(b), 9 &amp; 10, 9 &amp; 2, 5 &amp; 8 (d)</td>
<td>Ratio, RGB, HSI, PCA</td>
</tr>
<tr>
<td>Vegetation changes</td>
<td>7 &amp; 5, 7 &amp; 3(a), 9 &amp; 8(d)</td>
<td>Ratio, RGB</td>
</tr>
<tr>
<td>Surface temperature</td>
<td>11</td>
<td>Linear/Pseudo stretch</td>
</tr>
<tr>
<td>Surface moisture</td>
<td>11 (day and night(d))</td>
<td>Stretch, difference image</td>
</tr>
<tr>
<td>Weathering</td>
<td>9 &amp; 10(c)</td>
<td>Ratio, RGB, HSI, PCA</td>
</tr>
</tbody>
</table>

(a) Cracknell and Saraf (1989)  
(b) Rothery and Milton (1981)  
(c) Podwysocki et al. (1985)  
(d) Drury (1987)

**Table 4.5 - ATM Wavebands and image processing techniques generally considered important in mapping selected surface characteristics**

A small selection of the most common techniques for enhancing geological features, suggested in the literature, have been included in Table 4.5. This highlights a major problem in geological mapping: there are no 'standard' formulae employed to enhance remotely sensed data. This stems from two considerations. First, each study area has different specific characteristics which influence the type of technique used (for example, an arid environment would preclude the use of a vegetation index). Second, the techniques that have commonly been used are heavily scene-dependent (for instance, results derived from a PCA rely on statistics derived from the image and will therefore be different for each individual scene).

Due to the subjective nature of mapping geological features manually and the scene dependency of some of the commonly used image enhancement techniques, it is necessary to assess the applicability of a number of the techniques for this particular study area. Each of the following techniques have been assessed for the purposes of manual interpretation:-
- A true colour RGB composite (TCC), using bands 5, 3, and 2
- A standard false colour RGB composite (FCC), using bands 7, 5, and 3.
- An HSI transformation of the above composite (Gillespie 1980), using a histogram equalization contrast stretch of the saturation band.
- An RGB composite of three hue images, from RGB images 11-10-9, 8-7-5, and 4-3-2 (Liu and McMahon-Moore 1990).
- An RGB composite of the first three principal components (Rothery 1987a,b).
- An RGB composite of decorrelation stretched principal components, taking the three least correlated bands from above (Rothery 1987a).

These techniques reflect the most common methods employed in previous geological remote sensing studies.

Of these RGB composites, the true colour and false colour composites should be the easiest to interpret as most users would be familiar with the resulting colours in the images and what they represent. A linear stretch of the RGB composite of three least correlated bands, the HSI transformed composite, and the two PCA techniques should all provide reasonably well enhanced images. By examining the image statistics these techniques attempt to display the most information content, the first two examine three band combinations, while the latter finds the maximum directions of image variance in the full band set. However, the resulting colours within these images may be difficult to interpret on their own. When used in conjunction with standard colour composites they may become more manageable. The three hue RGB technique has been shown to discriminate lithological units in semi-arid terrain (Liu and McMahon-Moore 1990) and should perform well in areas in which lithologies and other surficial materials are spectrally distinguishable. However, in areas with vegetation cover and/or similar rock units this technique may fail. Similarly, the band ratios technique should produce good results in arid environments, but where the geology is masked by vegetation the only useful ratio is likely to be the vegetation index. Unfortunately, then, it is anticipated that the latter two techniques may not work well in the present study area.

The least correlated three band combination was selected using the techniques described by Liu and McMahon-Moore (1989). Table 4.6 displays a matrix of correlation coefficients for all two-band combinations and an indicator of three-band intercorrelation (ITI) is given by averaging the three pairwise correlations for the three bands. The highest ITI value (0.996) results from bands 6, 7, and 8, which is not
surprising as all three are near infra-red bands and would be expected to be highly correlated. The least correlated three bands are 2, 7, and 11 with an ITI value of 0.695. Again, the selection of these three bands is not unexpected, as they are each showing different properties of the earth's surface, i.e., band 2 is looking at visible radiance, band 7 is displaying radiance at near infrared wavelengths at which vegetation reflects strongly, while band 11 is exhibiting energy emitted from the surface at thermal wavelengths.

<table>
<thead>
<tr>
<th>Band</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.681</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.686</td>
<td>0.915</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.684</td>
<td>0.914</td>
<td>0.982</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.655</td>
<td>0.824</td>
<td>0.960</td>
<td>0.964</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.542</td>
<td>0.600</td>
<td>0.831</td>
<td>0.828</td>
<td>0.933</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.526</td>
<td>0.754</td>
<td>0.812</td>
<td>0.810</td>
<td>0.919</td>
<td>0.977</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.528</td>
<td>0.582</td>
<td>0.815</td>
<td>0.816</td>
<td>0.920</td>
<td>0.993</td>
<td>0.997</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.615</td>
<td>0.761</td>
<td>0.920</td>
<td>0.925</td>
<td>0.969</td>
<td>0.937</td>
<td>0.930</td>
<td>0.939</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.657</td>
<td>0.896</td>
<td>0.957</td>
<td>0.967</td>
<td>0.941</td>
<td>0.802</td>
<td>0.786</td>
<td>0.796</td>
<td>0.931</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.537</td>
<td>0.627</td>
<td>0.789</td>
<td>0.806</td>
<td>0.876</td>
<td>0.884</td>
<td>0.884</td>
<td>0.895</td>
<td>0.906</td>
<td>0.833</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Most correlated three band combination - 6, 7, 8 - average correlation = 0.996
Least correlated three band combination - 2, 7, 11 - average correlation = 0.695

Table 4.6 - Matrix of correlation coefficients for all two-band combinations

<table>
<thead>
<tr>
<th>Assessment</th>
<th>TCC</th>
<th>FCC</th>
<th>ITI</th>
<th>ITI/HSI</th>
<th>HRGB</th>
<th>PCA</th>
<th>DPCA</th>
<th>Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Lithological discrimination</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Lithological boundaries</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Vegetation boundaries</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ease of interpretation</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

1 = good, 2 = average, 3 = poor.

Table 4.7 - Assessment of colour enhancement techniques

Each colour composite listed above is included in Figures 4.10 and 4.11. These colour enhancements are produced to help the geologist manually interpret images. By its very nature manual interpretation is a subjective process. Therefore, each enhancement is
A - True colour composite

B - Standard false colour composite

C - Three least correlated bands

D - HSI transformation of C

Figure 4.10 - Colour enhancements of the ATM imagery, Part 1
Figure 4.11 - Colour enhancements of the ATM imagery, Part 2
evaluated subjectively by assessing how well each technique enhances the geological information in the scene and by the number and accuracy of the manually identified features. The results of this assessment are summarized in Table 4.7.

Of the colour images, PCA enhancement (Figure 4.11.G) gives the best colour separation for all cover types, but most importantly achieves the best separation of geological features. Particular features which are reasonably enhanced are lithological boundaries in the bracken area at A and scree at B. The reason that PCA gives such good enhancement of features is that the method takes information inherent in all eleven bands and calculates a new series of axes in multispectral feature space, such that each new image successively contains the majority of the remaining variance from all eleven bands, i.e., PCA1 contains the most variance, PCA2 contains majority of the remaining variance and so on. The PC axes are orthogonal to one another and hence the transformed images are uncorrelated (Schowengerdt 1983). A significant problem with this technique, however, is that it is strongly scene-dependent. Therefore, a given colour in one scene will not necessarily represent the same features or cover types in another scene because the variance of the second scene will be different. An interpreter cannot therefore use PCA images as a standard product; each new PCA image must be interpreted afresh with little experience gained from any previous interpretations (Crippen 1991).

Of the remaining colour images, the best ITI colour image (Figure 4.10.C) allows clearest interpretation. The rock exposures are well distinguished, as shades of magenta and dark blue, while the different vegetation types are also well delineated. Furthermore, the colours apparent in this image are generally repeatable between different images, although the colours may change slightly due to varying contrast stretches of individual bands.

The standard false colour composite (FCC) was originally designed to enhance vegetation within a scene and this is demonstrated in the FCC of this study area (Figure 4.10.B); the areas of bracken (A), grass (B), and forest (C) are well distinguished. The areas of exposed rock, however, are not as apparent as in the previous two colour images. Again this colour image is a standard product, with colours which can be easily understood, and repeated from scene to scene.

From Figures 4.10 and 4.11 it is evident that the remaining colour enhancements are less useful for mapping the geology within this study area, due to their poor colour separation and levels of noise. However, these techniques have been shown to be important in other areas with perhaps less vegetation cover and more distinguishable lithological types (Gillespie 1980, Drury 1987, Rothery 1987a, Liu and McMahon-
Moore 1990). It is therefore important that the interpreter should have these techniques available in case the nature of the scene warrants their use.

The three colour images (PCA, ITI, and FCC) which show, subjectively, the most geological information have each been interpreted manually. The interpreted geological features have been digitized and overlaid onto an image showing the geological units of the area (Figure 4.12). The lithological boundaries are shown as solid lines and the interpreted faults are displayed as bold lines. The first point to notice is that the identified features do not match the geological map precisely but do follow the general pattern of the map. A second important feature to note is that many lithological boundaries are identified within units considered to be single lithological units on the map. This indicates a significant advantage associated with the use of the remotely sensed data - in some cases, more detailed geological information can be derived than is currently available from existing paper maps. Of the three maps presented, the ITI and PCA images produce the most geological information. However, each enhances features in areas where the others do not. In conclusion, each method makes a contribution to the manual interpretation of the scene and is therefore important for geological mapping. Chapter 6 describes how primitive features such as these may be used to derive estimates of dip and strike.

### 4.2.2 Enhancement of monochrome images

For the semi-automatic extraction of features from remotely sensed images, using edge detection techniques, single band monochrome images are required. Here, the following single band images have been evaluated:

---

5 This mismatch is probably due to one or more of the following reasons:

i) the geological map was originally produced at a scale of 1:50,000, which is not readily comparable to the 5 metre resolution of the image data,

ii) possible inaccuracies in the original mapping of the geology in this area,

iii) errors in the geometric correction of the ATM image, and

iv) positional errors in the interpretation stage.

It is unclear at this stage which of these reasons are responsible for the mismatch between the two data sets. However, the maximum error of approximately 50 metres is clearly beyond the scope of any errors remaining after the geometric correction procedure (i.e. an average error of 9.5 metres and a maximum error of 18.7 metres - see Table 4.4). Any positional errors would be expected to be equal to or less than a pixel (i.e. 5 metres) as the user can produce sub-pixel accuracy when digitizing by zooming in to the image. Therefore, it can be assumed that the majority of the error is due to inaccuracies in the geological map itself. As a result, this fact should be borne in mind whenever comparing any features identified from the image with the geological map.
Figure 4.12 - Primitives identified manually from colour enhancements.
• ATM bands 2, 3, 5, 7, 8, 9, 10, and 11.
• The first four principal component images.
• The band ratios 9/10, 9/2, and 5/8.

The selected sub-set of eight bands from the original eleven ATM bands have been chosen to coincide with those specified in Table 4.5. Results derived from these individual bands may be similar due to the high correlation between images and the comparable shading in each image. However, as each surface cover type has different spectral properties and selected bands cover a wide range of spectral zones, each image should produce slightly different but complementary results. The individual principal component images should enhance more subtle spectral differences inherent in the original bands and should therefore add significant information. The first four principal component images have specifically been chosen here, as beyond the fourth component the images become very noisy (Figure 4.13) and contain little useful information. Finally, the band ratio images as mentioned above, will probably not provide any additional information due to the high vegetation cover, but are included for completeness.

Any of these monochrome images may be further enhanced using standard histogram stretching techniques (Schowengerdt 1983), such as:-

• Histogram equalization.
• Logarithmic stretch.
• Histogram normalization.

Linear stretch and piecewise linear stretches are not used here. This is because the linear stretch has no effect on the thresholded edge image, which uses a histogram percentage threshold, and the piecewise linear stretch is an interactive process, not conducive to the desired automated approach.

The monochrome enhancement techniques are evaluated both subjectively, for the purposes of manual interpretation, and objectively, using the line extraction techniques discussed later in section 4.2.4. Again, the methods are assessed in terms of the number and accuracy of the identified features.

The results of the monochrome contrast stretching techniques, applied to band 7, are shown in Figure 4.14. From the point of view of manual interpretation, the linear and histogram equalized stretches appear to enable more geological information to be
Figure 4.13 - Fourth principal component.
Figure 4.14 - Comparison of contrast stretching techniques applied to band 7, showing automatically derived primitives as a solid line overlay.
extracted. The linear stretch gives the maximum contrast between rock and vegetation and also between shaded and non-shaded parts of the image. The histogram equalized image preferentially enhances features which were in the mid-grey tone range in the original image. For instance, the lithological boundaries in the area of bracken are enhanced, enabling the better mapping of geology in this area. The hyperbola stretch is in fact very similar to the histogram equalized stretch and does not add any additional information. The Wallis and histogram normalized stretches appear poor in quality with much of the shadow in the scene being washed out. The logarithm stretch attempts to enhance the darker area of the scene while reducing the contrast of the brighter areas. As there is very little detail in these shaded and lake areas no additional geological information can be interpreted and indeed many of the major features are less obvious. Although of little use in this area, the logarithm stretch has been shown to be extremely valuable in other areas, such as the Antarctic (Morris et al. 1992), where the darkest features are the geological outcrops exposed through the ice and snow fields.

Each of these techniques could be used within the automated line extraction techniques and are therefore evaluated here by extracting line features from the enhanced images to ascertain which techniques best identify the geological features. The results of the line extraction techniques are shown as vector overlays in Figure 4.14. Each stretch identifies the major geological features in areas A and B. However, the histogram equalized and hyperbola stretches identify substantially more of the minor geological features in area C than do the others. Although not as visually effective, the Wallis and normalized stretches identify some lithological boundaries in the bracken area, which are not extracted from the other enhanced images. So once again, although two of the stretches provide about 90% of the possible extracted information, the others can produce additional data which could prove vital in creating an accurate model of the geology. It is therefore unwise to single out one particular technique for all subsequent processing but better to include all techniques that can provide useful information. In this case the linear, histogram equalization, Wallis, and histogram normalization stretches have been shown to be effective in extracting useful and unique geological information.

From the assessment of both the colour and monochrome enhancement techniques, it is apparent that no single method can be chosen to help extract all relevant geological information from a scene. It would therefore be reasonable to include many such techniques within a geological mapping system. Geological applications of remote sensing are different to agricultural applications in terms of time dependence. Many agricultural applications require a short turn around period from initial data acquisition through to the production of final results, and therefore only the 'best' enhancement method might be used. By contrast, interpretations for a geological site may only be
performed once. Therefore, the best interpretation might be achieved by including a number of possible techniques.

4.2.3 Segmenting the scene into geologically/non-geologically relevant areas

The multitude of techniques and possible input data sets described in the previous section will result in the generation of a profusion of primitive features, many of which will bear no relation to any geological phenomenon, for example, forest boundaries and tracks. It is therefore desirable to gain as much knowledge as possible about the properties of the surface cover in order that these superfluous edges can either be removed or assigned a low significance. One way to achieve this is to segment the image into discrete, homogeneous regions on the basis of the apparent land cover. This will help to identify, for example, field boundaries and other linear features such as road networks that might be identified by the line extraction techniques. Essentially, the aim here is to isolate those primitive features that have real geological significance from the mass of features identified.

There are a number of possible image segmentation techniques that might be used, including multispectral classification and region growing. The most common method of mapping surface cover into thematic units is to use supervised multispectral classification to classify an image using training statistics from a number of manually identified regions (Schowengerdt 1983). Since this is not a fully automated procedure, its inclusion in the design of an expert system could in some respects be disadvantageous. However, completely automated techniques, such as unsupervised classification and region growing, produce segmented images with no specific information regarding the content, meaning, or significance of each segment. To segment geologically relevant primitives, it is important to gain knowledge about the surface. Therefore, the supervised classification technique has been selected for further study. A number of different classification algorithms may be used, including the parallelepiped, Euclidean distance, Mahalanobis distance, maximum likelihood, and Bayesian maximum likelihood algorithms (Lillesand and Keifer 1979, Schowengerdt 1983). It is widely acknowledged (Lillesand and Keifer 1979) that the maximum likelihood algorithm usually provides the most accurate results and has the added advantage that images indicating the 'confidence', or probability of correct classification, may be produced for each pixel in the classified image. Confidence statistics can then be derived for each identified feature relating to the likelihood that the feature is geological. For example, this could be achieved by averaging the confidence values of a 'rock' class for each pixel comprising the primitive feature under examination. Such techniques are discussed in greater detail in Chapter 6.
A maximum likelihood algorithm has been used to classify the image in this study. Seven classes have been identified within the scene (Figure 4.15); these are water, shadow, forest, bracken, grass, rock_1, and rock_2. The rock exposure class has been split into two due to the significant spectral difference between illuminated and shaded exposures (see Figure 4.10.C, where the illuminated exposures are magenta and those that are shaded are dark blue). Once again, several combinations of bands have been used as inputs into the classification in order to assess which combination produces the most useful results. The three selected band sets are:

- ATM bands 2, 3, 5, 7, 8, 9, 10, and 11.
- The principal component images.

Results of the maximum likelihood classification on the three different image data sets are assessed both visually and using standard confusion matrices. Figure 4.15 shows training areas used for each class in solid colours, and validation areas used to test the results in identical colours but with a hashed pattern so that each area appears to be transparent with the background image partly visible. Confusion matrices are produced by comparing results of the classified image within each test area. The confusion matrices comprise tables indicating the number of pixels classified both correctly and incorrectly in each class. Each row signifies classification accuracy for each class. The number of pixels falling in the appropriate class column indicates the accuracy that this class is classified and pixels falling in other columns are pixels of commission. The remaining three columns (e.g., Table 4.8) of the matrices indicate the total number of test pixels in each class, the classification accuracy for that class, and the total number of commission pixels.

Figure 4.15 depicts the results of the three classifications and Table 4.8 the confusion matrices produced for each classification. From a simple visual comparison of the classified images it is apparent that most accurate classes are produced from the original band data, and this is borne out by the high classification accuracy shown in the confusion matrix. The classified image derived from the ratio data is extremely noisy, although broadly speaking, the majority of the classified pixels are in correct general areas. The noise results in a low classification accuracy. The high classification accuracy produced by the PCA images is slightly misleading as the classified image shows large areas of bracken on the south-west facing slopes and only small areas of exposed rock. The original bands therefore produce the most accurate classification results both visually and statistically.
Figure 4.15 - Training areas and classified images.
### a) Confusion Matrix for Original Band Data Set

<table>
<thead>
<tr>
<th>Class</th>
<th>Water</th>
<th>Shadow</th>
<th>Forest</th>
<th>Bracken</th>
<th>Grass</th>
<th>Rock_1</th>
<th>Rock_2</th>
<th>t_pix</th>
<th>cl_ac</th>
<th>com_pix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>1185</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1189</td>
<td>99.7%</td>
<td>4</td>
</tr>
<tr>
<td>Shadow</td>
<td>0</td>
<td>346</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>346</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>Forest</td>
<td>0</td>
<td>0</td>
<td>316</td>
<td>0</td>
<td>92</td>
<td>1</td>
<td>0</td>
<td>409</td>
<td>77.3%</td>
<td>93</td>
</tr>
<tr>
<td>Bracken</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>140</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>151</td>
<td>92.7%</td>
<td>11</td>
</tr>
<tr>
<td>Grass</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>355</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>355</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>Rock_1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>123</td>
<td>0</td>
<td>124</td>
<td>99.2%</td>
<td>1</td>
</tr>
<tr>
<td>Rock_2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>82</td>
<td>95</td>
<td>86.3%</td>
<td>13</td>
</tr>
</tbody>
</table>

Average Classification Accuracy = 93.59%

### b) Confusion Matrix for Principal Component Images

<table>
<thead>
<tr>
<th>Class</th>
<th>Water</th>
<th>Shadow</th>
<th>Forest</th>
<th>Bracken</th>
<th>Grass</th>
<th>Rock_1</th>
<th>Rock_2</th>
<th>t_pix</th>
<th>cl_ac</th>
<th>com_pix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>1120</td>
<td>69</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1189</td>
<td>94.2%</td>
<td>69</td>
</tr>
<tr>
<td>Shadow</td>
<td>24</td>
<td>321</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>346</td>
<td>92.8%</td>
<td>25</td>
</tr>
<tr>
<td>Forest</td>
<td>1</td>
<td>255</td>
<td>7</td>
<td>126</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>409</td>
<td>62.4%</td>
<td>154</td>
</tr>
<tr>
<td>Bracken</td>
<td>0</td>
<td>0</td>
<td>119</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>151</td>
<td>355</td>
<td>78.8%</td>
<td>32</td>
</tr>
<tr>
<td>Grass</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>331</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>124</td>
<td>58.9%</td>
<td>51</td>
</tr>
<tr>
<td>Rock_1</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>7</td>
<td>73</td>
<td>0</td>
<td>0</td>
<td>95</td>
<td>19.0%</td>
<td>77</td>
</tr>
<tr>
<td>Rock_2</td>
<td>0</td>
<td>0</td>
<td>56</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>18</td>
<td>95</td>
<td>19.0%</td>
<td>77</td>
</tr>
</tbody>
</table>

Average Classification Accuracy = 71.3%

### c) Confusion Matrix for Ratio Images

Key:
- t_pix = Total number of pixels in each class
- cl_ac = Classification accuracy for each class
- com_pix = Number of pixels of commission in each class

Table 4.8 - Statistical results for the three maximum likelihood classifications
The usefulness of the classified images and the derived confidence images is also assessed by attempting to remove non-geological features from a set of automatically identified primitives. This is achieved by simply overlaying the primitive features (from the linear stretch image used in the previous section) over the classified image and thresholded confidence images.

The confidence images produced for several classes and derived from the original band data are shown in Figure 4.16. Lighter areas in these images signify places where the confidence value for the class is high. By overlaying the automatically derived primitives onto a thresholded confidence image of the forest class it can be seen how such images might be used to remove non-geological features (Figure 4.17.A). It should be noted, however, that primitives should not necessarily be discarded just because they fall within a certain class; many geologically relevant primitives may occur along boundaries between two vegetation types due to the effect that the underlying geology may have on the surface type. The classified image can be used in a similar way (Figure 4.17.B); however, the confidence images allow more flexibility, as the threshold value may be altered to broaden or tighten areas for each class. Confidence images may also be used to assign confidence statistics to each feature rather than to remove features on the basis of only one fact. Again this will be discussed in greater detail in Chapter 6.

4.2.4 Automated line extraction techniques

Geological features in remotely sensed images may be identified either manually or automatically. The principal disadvantages of manual interpretation are that it is both time-consuming and subjective. By contrast, automated procedures can provide data in a timely and reproducible manner. Automatic identification of line features in images relies on edge detection techniques, which highlight sharp differences in grey-levels within an image. This section will evaluate a number of edge detection techniques and methods for extracting useful line information from the resulting edge images.

An edge-detection filter applied to an image results in an image which can be thought of as indicating 'edge strength'. This can then be thresholded, to select the strongest edges, and 'thinned', to produce definite edges, one pixel in width, representing the identified feature. Many edge detection algorithms exist, varying in size, complexity, and performance. Techniques which have been used in previous geological applications include standard gradient filters (Smithurst and Vaughan 1987, Harris 1987, Blondel et al. 1992), the Hough transform (Skingley 1986, Wadge and Cross 1989), and the Fourier transform (Eppes and Rouse 1974, Hilrose and Harris 1985). Additional edge detection algorithms exist which could prove useful in geological applications, such as the filters designed by Canny (1986) and Petrou and Kittler (1987).
Figure 4.16 - Probability images derived from the maximum likelihood classification. The lighter tones indicate a higher probability of that class occurring at that location.
Figure 4.17 - Removing non-geological primitives using segmentation techniques.
The previous geological studies, mentioned above, used edge detection filters mainly to identify geological lineaments within a scene which are often assumed to occur as straight lines. Here, it is intended to extract edges relating to both lithological boundaries and fault lineaments, which are more likely to be curvilinear in nature. Each technique will therefore be assessed with this in mind. They will also be examined with a view to discovering the most reliable line extraction method for geological features in this and other study areas. The Hough transform is not evaluated here, as this has been used in other studies to identify features with regular shapes such as straight lines and circles, whereas in this study the geological features are most likely to be irregular in shape, following the uneven terrain.

Four edge detection routines, representing the major types available, have been used here:-

- a 3*3 Sobel filter
- a 3*3 Compass filter
- a Fourier transform
- a Canny filter

The first two methods are both local area gradient filters (Schowengerdt 1983) which identify edges within a 3*3 neighbourhood. The resulting edge images can therefore be precise in the positioning of each edge, as the edge strength value is assigned to the centre pixel. However, edges can be discontinuous, with some edges (as the human would see them) breaking up into small parts, due to the local nature of the calculations. The Sobel filter standardly uses two passes of the filter applied in orthogonal directions while the Compass uses four directional gradients separated by 45° angles. Each pass is summed to give an overall edge strength image.

The Fourier transform is used to identify features in images which are repeated throughout the image at a given frequency. For instance, waves on the ocean surface may have a certain frequency which would be apparent in the power spectrum of the frequency domain (Wyatt 1989). Images of exposed geology can also exhibit a certain frequency caused by the inter-layering of lithological units in a sedimentary sequence. The power spectrum could therefore be filtered to remove all other frequencies from the image.

The Canny filter is an 'optimal' edge detector (Canny 1986). It involves three stages, firstly a smoothing of the image using a Gaussian filter, secondly application of a gradient filter, and thirdly a suppression of non-maximal pixels in the edge strength image. The disadvantage of this technique is that the smoothing process may remove
fine detail from the edges. However, its major advantage is the suppression stage, which results in edges that tend to be rather more continuous than those produced using simple gradient filters. This is useful in geological applications where it is necessary to identify as much of each lithological boundary as possible.

To produce primitives that can be used to estimate dip and strike, individual line entities must be extracted from the edge images. Figure 4.18 outlines diagrammatically the procedure used here to extract such primitives.

Application of the Canny edge detection algorithm produces an image which indicates strengths of the edges for all pixels in the image which are determined to be a local maxima. The product of other edge detectors, such as the Sobel operator, is edge strengths for every pixel in the image. If edges are to be used as entities, decisions need to be made as to what constitutes a reasonable edge and what is 'background noise'. The simplest method of achieving this, and that which has been used here, is to threshold the image. This results in a binary image where all edge strengths greater than or equal to the selected threshold are assigned a value of 1 and all those below are assigned a value of 0.

After thresholding, the binary images are 'thinned'. This process recursively thins a binary image until each edge has a width of one pixel, while retaining the integrity of any branching structure in the edge (O'Gorman 1990). The thinning algorithm used here also identifies other properties such as end points, bifurcations, and other branching points, which can then be used to describe each line as an individual identity. Although the Canny detector succeeds in producing edges one pixel in width, the results are processed using the thinning algorithm so that the branching structure of each edge can be determined.

An algorithm (line_extract) has been developed as part of this study to create entities from thinned edges. Although other algorithms exist to vectorize such image data (ESRI 1992) it is desirable here to extract entities which follow one path through the branching structure and therefore have no branching structure of their own. The line_extract program is similar to an algorithm described by Blondel et al. (1992), which searches an eight grey-level image for 'straightish' lines by evaluating whether neighbouring pixels are within certain limits of the overall direction of the line. Line_extract searches the thinned image for endpoints and then follows each edge looking for the longest path through the branching structure of the edge. This is achieved by searching a 3*3 window (starting at the detected end-point) for the next point in the edge and then moving the window to this new point. Pixel values in the thinned image indicate whether each point is a branching point (for example a pixel
Figure 4.18 - Procedure for extracting primitives.
value of 8 indicates a bifurcation point while a pixel value of 16 indicates a four way split in the line). When a branching point is found, line_extract first follows one branch, and any subsequent sub-branches, and then the other. For each branch a record is kept of the length and co-ordinates of each path so that the longest path may be extracted once the branching structure has been investigated fully. Here, it is assumed that the longest path will be the most useful in determining dip and strike (future studies will hope to examine how justifiable this is and ways in which some geological knowledge may be incorporated into the line_extract program). Each edge entity is then stored as a series of 2-D co-ordinates representing each pixel along the path, to be used later in the study for the calculation of dip and strike (see Chapter 6).

One problem associated with thresholding the image as a whole is that only the strongest n% of the edges are selected. Edges which are not necessarily strong, but which are pronounced in relation to the surrounding area may therefore be missed. The threshold value may be relaxed or tightened to include or exclude these other features but one consequence of this is that with a decrease in the threshold the stronger edges may become too broad and may lose detail. A way of overcoming this is to carry out the thresholding over smaller sub-scenes extracted from the entire edge image. The top n% of edge strengths within this extract will then be selected. A C shell script has been written which successively extracts the appropriate sub-images until the whole scene has been covered. An overlap of 20% for each sub-image is allowed in order that edges do not have an abrupt termination at the boundary. Figure 4.19 shows a set of primitives derived directly from the entire image and a similar set of data derived using the windowing technique.

Each edge-detection technique has been tested using band 7 of the ATM data as this bands highlights vegetation/rock boundaries, where many of the lithological boundaries occur. When identifying lithological boundaries an edge detection algorithm should be capable of delineating edges which are:-

- **accurate** - inaccuracies at the edge detection stage will propagate through subsequent processing and cause errors in the calculation of dip and strike,
- **continuous** - edges which are broken are less likely to give a full picture of the geology and less likely to allow accurate dip and strike estimates to be made,
- **thin** - if after the thresholding stage the edge is more than a few pixels in width, then the thinning process may introduce additional branches which may cause the line_extract program to extract an erroneous path through the branching structure, and
Figure 4.19 - Comparison between whole-scene processing and window-based processing. More primitives are identified using the latter techniques due to the use of local thresholds.
- numerous - it is important to identify as many of the geological features as possible.

To assess the merits of each edge detection algorithm these factors must be considered. By visually examining a number of edge features in each edge image (Figure 4.20) a relative assessment was made as to how well the above factors are satisfied by each algorithm. Table 4.9 gives the results of this evaluation with the figures representing a qualitative number between 1 and 3, i.e., ranging from good, to average, to poor.

Each method, apart from the Fourier analysis gives an accurate representation of edges present within the image. There is, however, a much wider variation in the continuity and edge width properties. In terms of continuity, the Compass filter gives poor results as the calculated gradients are averaged over four directions and can therefore appear to smooth out some of the detail in the edges. The optimal edge detector (Canny) gives the best results for this application in that it produces precise, one pixel width edges which are also the most continuous of all techniques tested. This is due to suppression of non-maxima in the edge-strength image and smoothing of the data prior to the application of a gradient filter (Canny 1986). This smoothing process is the only slightly worrying feature of the Canny filter as this could remove fine detail from the edges. However, from Figure 4.21.D this does not appear to be the case. The Sobel filter gives good to average results throughout and in some areas identifies edges not apparent in the Canny edge image.

<table>
<thead>
<tr>
<th>Edge detector</th>
<th>Accuracy</th>
<th>Continuity</th>
<th>Width</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sobel</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Compass</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Fourier</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Canny</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.9 - Subjective analysis of edge detection algorithms

The Canny filter appears to provide the best quality edge information, while the Sobel filter provides complementary data. Both are used in subsequent processing. The compass filter and Fourier transform provide no additional information. Although the results presented here represent just one study area, the conclusions made about each of the detectors are probably sufficiently general as to be applicable to most study areas.
Figure 4.20 - Comparison of edge detection algorithms.
Figure 4.21 - Primitives automatically derived from remotely-sensed imagery.
In order to evaluate the importance of different data input to the line extraction process the Canny filter has been used to derive primitive information from the following images:

- individual ATM bands 2, 3, 5, 7, 9, 10, and 11,
- band ratios 9/10, 9/2, 7/5, and
- principal components 1, 2, 3, and 4.

Each of these data sets have been chosen as previous studies have shown them to be useful in geological applications (see Table 4.5).

The most notable results of the line extraction procedure applied to these images are shown in Figure 4.21. The first noteworthy point is that the spectral data exhibit many different types of edge in addition to those produced by geological features. These include edges of lakes, forests, tracks, roads, changes in vegetation and various other anthropogenic features. Besides obvious lithological boundaries apparent in some of the larger exposed areas, many geological features are derived from the natural shading of the scene producing edges along geomorphological features. Most other geological features are a result of a boundary between exposed rock and vegetation rather than lithological boundaries per se. Figure 4.22 illustrates how lithological boundaries can be deduced from the contact between vegetation and rock exposures. The geology in this area includes an inter-layering of hard and soft rocks where the softer rocks are often overlain by vegetation, thus the border between the harder rocks and the vegetation will often reflect the lithological boundary between the two rock units.

With the exception of band 11 the edges derived from the individual bands tend to highlight the natural shading in the scene along with the most obvious land cover changes, i.e., forest, lake, and heavily shaded areas. Band 11 exhibits many stronger edges at the vegetation-rock boundary (Figure 4.23).

The band ratios 9/10 and 9/2 (Figure 4.24) are very disappointing due to the amount of noise present in these images. None of the individual bands exhibits noticeable noise but when ratioed together the noise becomes more apparent. Additionally, these ratios do not enhance the exposed geology in the area; 9/10 manages to suppress the differences in vegetation and the shading in the scene, while 9/2 actually masks most of the rock exposures. As a result the edges derived from these two ratios are noisy, inconsistent, and do not follow geological features. The poor performance of these band ratios could be partly due to the high proportion of vegetation cover in the study area. The ratios have been used successfully before, but these applications have generally been in areas with arid or semi-arid climates (Drury 1987).
Exposed rock

Vegetation cover

Key:

- Harder dolerites, sandstones, tuffs etc.
- Softer shales, mudstones etc.

Figure 4.22 - Lithological boundary identified between vegetation and exposed rock.
Figure 4.23 - ATM band 11, showing strong contrast at the vegetation-rock boundaries.
Figure 4.24 - The band ratios 9/10 and 9/2.
Conversely, band ratio 7/5 is far less noisy and gives a reasonable differentiation between geological and non-geological cover. This ratio is often referred to as the vegetation index and it is apparent from Figure 4.25 that the grey-levels do relate to the concentration/vigour of the vegetation, with outcrop areas being delineated as dark areas in the image. The strongest edges from this ratio tend to follow the boundaries of the lake and shaded areas, and tracks in the forest. Many of the moderate edges do, however, follow the vegetation-rock boundaries. Although the vegetation index might not be the clearest image from which to derive geological primitives, the index can be useful in other ways. The index can define a confidence statistic relating to likelihood that edges derived from other sources have some geological significance, in a similar manner to the confidence images derived from the image classification.

<table>
<thead>
<tr>
<th>PC</th>
<th>eigenvector</th>
<th>eigenvectors (for each band)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>eigenvalue</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>89.07</td>
<td>0.616</td>
</tr>
<tr>
<td>2</td>
<td>8.340</td>
<td>-0.024</td>
</tr>
<tr>
<td>3</td>
<td>1.840</td>
<td>-0.007</td>
</tr>
<tr>
<td>4</td>
<td>0.306</td>
<td>-0.604</td>
</tr>
<tr>
<td>5</td>
<td>0.186</td>
<td>0.046</td>
</tr>
<tr>
<td>6</td>
<td>0.143</td>
<td>-0.407</td>
</tr>
<tr>
<td>7</td>
<td>0.051</td>
<td>0.005</td>
</tr>
<tr>
<td>8</td>
<td>0.023</td>
<td>-0.002</td>
</tr>
<tr>
<td>9</td>
<td>0.015</td>
<td>-0.118</td>
</tr>
<tr>
<td>10</td>
<td>0.010</td>
<td>0.269</td>
</tr>
<tr>
<td>11</td>
<td>0.009</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Table 4.10 - Eigenvectors and eigenvalues derived from the principal component analysis of all eleven ATM bands

The principal component images produce some of the most useful edges, highlighting many geological features present in the scene which are not identified with other techniques. The eigen-vectors for each PC are shown in Table 4.10 and these can be used to indicate how each band relates to each PC. The first principal component results from a combination of a number of bands (i.e., most significantly bands 1, 2, 6, 8, 9, and 10) and it is demonstrated in Figure 4.26.A that this produces an image which displays most of the natural shading of the scene but suppresses some of the rock exposures and differences in the lower-lying vegetation. The edges derived from this image therefore
Figure 4.25 - Band ratio 7/5 - the vegetation index.
Figure 4.26 - Principal components 1 and 2.
follow the geomorphological features of the scene. The second PC (which results from 2, 3, and 5 - Figure 4.26.B) does not exhibit any shading and also suppresses the forest. Consequently, much of the geological information is enhanced well; each rock exposure is shown as a dark area and a number of geological units are identified in the north and east of the scene which are not identifiable in other enhancements. This is an excellent example of how slight changes in the spectral reflectance of vegetation can represent changes in sub-surface geology and also of how lithological contacts may not necessarily be reflected in the surface geomorphological expression. The third PC also suppresses much of the vegetational differences in the scene and complements PC 2 very well by highlighting many of the geological exposures missed by PC 2. The remaining PCs hold very little additional information, for example Figure 4.13 shows PC 4 which is extremely noisy and particularly ineffective for automatic edge detection techniques.

The result of the line extraction procedures of these four data sets are compared with the geological map in Figure 4.21. Again the primitive features do not match the geological map precisely due to the reasons outlined in section 4.2.1. The automatically identified primitives include many features which are not geologically relevant and therefore on first appearance do not seem successful when compared to the manually identified primitives (e.g., Figure 4.12). However, on closer examination many of the major geological features have been delineated (e.g., areas A and B) in conjunction with a number of minor geological features, not identified in the manual interpretation (e.g., at C and D).

From the images tested here, the most promising results were obtained from PCs 2 and 3, band 11, and the ratio 7/5. Although the two PC images exhibited the best results for this scene it should be remembered that for another scene the first PC, or any combination of PCs, may be preferred. This has consequences within the framework of an expert system, which will be described in Chapter 7, as there will always be uncertainty as to what the edges relate from each of the PCs. Alternatively, band 11 and the vegetation index provide edges relating to a known physical relationship, e.g., an edge from the vegetation index indicates a strong difference in vegetation vigour across the edge. This fact does not exclude the use of PCs, however, because other external information such as the vegetation index and the texture of the DEM (see section 5.2.2) can be used to describe the terrain in which a PC edge occurs and therefore further knowledge can be obtained regarding the nature of the edge.
4.3 Site specific versus general tools

Many of the results discussed in this Chapter are likely to be specific to the Llyn Cowlyd study area. This section will therefore discuss a number of the environmental conditions which make the Llyn Cowlyd area quite different from other areas, and also which results will be applicable to other study areas around the world.

The major environmental factors which affect the Llyn Cowlyd area and Snowdonia as a whole are the high percentage vegetation cover and the effects of glaciation in this mountainous terrain. The vegetation cover masks much of the geological information. If the vegetation and soil cover were completely removed then techniques such as band ratios and mixture modelling might conceivably be more successful. Other geographical areas of interest may not be so mountainous, therefore showing less shading in the scene from which geomorphological and geological information may be derived. In these areas spectral enhancements may be more appropriate than deriving monochrome edges. Conversely, in non-glaciated mountainous areas less geomorphological expression will have been removed and the reverse could be true.

In the design of a generally applicable mapping system it is therefore unwise to rule out some of the evaluated techniques simply because they do not appear successful in the Llyn Cowlyd area, particularly since many of the techniques have been reported to be useful in other study areas. Rather it would be better either to have each technique available for a user (if the study area was suitable for that method), to provide an automated way of determining whether a technique is suitable, or to apply each technique blindly and then automatically assess whether it has been successful. These issues will be discussed more thoroughly in Chapter 7.

4.4 Possible future improvements in identifying geological features

Segmented images have been employed to signify the land-cover type most likely at any point in a scene. It is hoped that future work could examine the possibility of using such images for a quite different purpose by studying the shape of each resulting segment. Here the segments could be used to extract further geological information. Previous work by Barr and Barnsley (1995) and McKeown and Harvey (1984, 1985) used the shape and size of such objects to map urban environments. It is reasonably easy to see how the often regular and repetitive shapes in an urban scene could be used successfully to map urban components, however; when trying to map natural phenomenon individual objects are likely to be more complex. For instance, the shape of the largest forest object in Figure 4.17b is not necessarily indicative of forest terrain, due to irregular and
unrecognizable shape. Some of the rock objects, however, exhibit a linear shape which could be used to identify lithological boundaries. Figure 4.27 shows an idealized rock object which represents one rock unit arcing across the terrain. There are several ways in which such objects could be used in future work to retrieve or to approximate a lithological boundary. Figure 4.27 shows a number of paths around or through the object (between the two points on the object which are furthest apart, A and B) which could resemble the actual boundary. Of course, any real objects identified with a classification or segmentation procedure, are most unlikely to have a similar shape to that of the object in Figure 4.27 and most will in fact have fairly erratic shapes which bear little resemblance to any geological feature. However, it is possible that a series of rules could be defined to follow different paths through the objects depending on their shapes and relationship to other neighbouring objects.

Despite the success of the Canny and Sobel edge detectors, it is possible that these could be improved upon by using line following techniques. For example, a line can be followed through areas of low edge strength provided that certain conditions are met regarding the strength and orientation of neighbouring pixels (McKeown 1985). Geological knowledge could also be added to this procedure. For instance, if it was known that the local geological structure had a particular dip and strike, then the line following conditions could be biased to look for such orientations.

4.5 Conclusions

This chapter has evaluated a number of techniques for the identification and extraction of geological features from remotely sensed data. Even though each analysis has been reasonably thorough and has provided a plethora of image products and results, each one could easily be the subject of a separate detailed research project. Only the most obvious techniques for this particular application have been chosen, yet still there is a mass of information provided by these techniques most of which is complementary. At the end of these evaluations the main conclusion that can be drawn is that there is no standard method for deriving geological information from remotely sensed data.

The evaluations have also shown that useful geological features can be derived both manually and automatically from remotely sensed data. Some enhancement, segmentation, and line extraction techniques have been shown to be more successful than others, but this success could be peculiar to the Llyn Cowlyd study area. Again there is no way of categorically stating which methods should be employed.

The only response to these points is to use all available data and techniques. A geologist is unlikely, however, to know necessarily what each technique can provide or how to
1. Shortest edge between two furthest points
2. Shortest direct route within region
3. Shortest path to turning points
4. Mid-line of region

Figure 4.27 - Methods for extracting geological information from regions.
use it. The task is therefore to create a computer system which can guide the user through each of the methods, suggesting which should be used for a particular kind of terrain or geological environment. Similarly, with the mass of possible automatically identified primitive data the system could be designed to combine all of the data while removing identical features, leaving only the unique information. These issues are discussed in Chapters 6 and 7.

Nevertheless, several problems still remain with the remotely sensed data which could not be solved within this study, namely, inaccurate conversion of radiance values into reflectance, and geometric correction of the image. A more accurate measure of reflectance could be made by including atmospheric correction methods and by using several angular radiance measurements of each pixel in the scene. Similarly, it is hoped that the geometric correction could be improved using future developments of the stereo-matching correction method, outlined in section 4.1.3.

Other future work could evaluate the techniques described in this chapter for other study areas varying in environmental conditions and altering the remotely sensed data used. For instance, data acquired at different resolutions will identify geological information at different scales; what effect will this have on the subsequent creation of a structural model? Also, the use of hyperspectral images would allow the use of mixture modelling techniques which could be used to identify individual rock types.
Chapter 5 - Digital Elevation Models - generation and extraction of 'primitive' geological features

There is often a close link between the morphology of an area and its structural geology (Lillesand and Kiefer 1979, Strahler 1975, Twidale 1976). Therefore, geological structures may be revealed by a detailed analysis of a digital elevation model (DEM) representation of the surface topography. Just as Chapter 4 investigated methods of extracting geological information from remotely sensed data, this chapter investigates similar ways of extracting such information from DEMs.

It is important here to briefly illustrate how geomorphology and structural geology can be related. Escarpments, plateaux, and ridges (Figure 5.1) can all be simple indicators of sub-surface geology and its structure. Smaller scale changes in slope can also reflect the sub-surface geology. Similarly, drainage patterns and drainage density of an area are often related to lithological type and the predominant strike of folding and faulting (Parvis 1950, Howard 1967). Figure 5.2 illustrates the use of various drainage patterns that might be used to classify a selection of structural/petrological environments. Each of these geomorphological features and patterns may be extracted from a digital model of the topography.

Chapter 5 will discuss ways in which DEMs may be used to identify such geomorphological features and how these relate to geological features, for example lithological boundaries and faults. It will also discuss methods by which a DEM and remotely sensed images may be used together to derive such information. An overview of the layout of Chapter 5 is illustrated in Figure 5.3. A number of methods for the creation and evaluation of DEMs are assessed. The path shown in bold in Figure 5.3 indicates the most accurate DEM creation method, given the available data. The following sections investigate several ways of enhancing and extracting the geological information held within them. The information derived from these techniques combined with those from the remotely sensed data forms the basis for subsequent calculation of dip and strike measurements in Chapter 6.
Figure 5.1 - Typical relationships between topography and geology
Figure 5.2 - Samples of different drainage patterns (after Argialas et al. 1988)
Figure 5.3 - Overview of methods used in Chapter 5
Each of the DEM enhancement methods will be evaluated with respect to its success in identifying geological features, both manually and automatically. As with remotely sensed images, this involves a subjective assessment of results, followed by a more objective evaluation of the number of features identified and their accuracy when compared to the geological map. Perspective views of the remotely sensed images are evaluated subjectively, to ascertain whether an interpretation of the scene is enhanced using this technique.

Segmentation techniques are evaluated both in terms of how well they separate geological features from non-geological features derived from the DEM, and also the success of separating such features derived from remotely sensed images. Line extraction techniques are not further assessed here as this was covered using remotely sensed data. Results derived from the DEM and the enhancements of the DEM are, however, assessed by comparison of the data to the geological map.

5.1 Generation of digital elevation models (DEM)

If any of the geomorphological features or patterns mentioned above are to be derived from an analysis of a DEM, then creation of a detailed and accurate DEM is essential. This section therefore evaluates several possible methods for creating a DEM and assesses their accuracy. The methods considered include:-

- Stereo-matching of stereo images or aerial photographs,
- Radar or laser altimetry data, and
- Interpolation from manually or automatically digitized contours.

If accurate measurements are to be extracted from a DEM, the DEM must be a precise representation of the surface not only in terms of individual spot heights, but also in terms of the geomorphology it describes. It is essential to preserve the context of the DEM in order that simulated drainage networks are accurate (Lee et al. 1992) and other features, such as ridges and steep slopes that are detected in the data, retain their positional accuracy and shape. The accuracy of DEMs is usually tested using a series of randomly identified spot heights within the DEM and comparing these with heights derived from a contour map (Day and Muller 1988, Sasowsky et al. 1992); however, this does not allow a full appreciation of the differences between each method. Four additional techniques are introduced and investigated here, each of which attempts to examine more closely the geomorphological or contextual accuracy of the DEM generation methods and to highlight any artefacts which may be introduced. The following techniques are used here and are discussed in section 5.1.2.
• Comparing random spot heights taken from each DEM with the original map data.
• Comparing recontoured DEMs with the original map data.
• Comparing drainage networks derived from each DEM with the map data.
• Subtracting DEMs to identify areas of significant differences between individual DEMs.
• Using Lambertian shading techniques to highlight any artefacts introduced by the creation techniques.

These techniques allow a subjective assessment of DEM accuracy and a visual comparison of the quality of the DEMs. As a result the final assessment and conclusions are also somewhat subjective, but they highlight the major differences between the DEM creation methods.

5.1.1 Methods for generating DEMs

Stereo-matching has been used successfully in previous studies to derive accurate DEMs from stereo satellite images such as SPOT data (Day and Muller 1988, Hanaizumi 1990, Brockelbank and Tam 1991, Heipke 1992). The technique relies on the disparities present in stereo images caused by topographic expression of the surface (Figure 5.4). Image matching techniques are used to identify coincident points in each image and the relative disparity between each set of matched points used to determine height values. The relative measurements can be made absolute by manually identifying several control points in both images.

Matching can be performed on a patch basis (Gruen and Baltsavias 1987, Otto and Chau 1989) resulting in one height value per patch, therefore producing a DEM with a lower resolution than the original images. Typically, resolutions of 50 to 100 metres can be achieved from stereo SPOT panchromatic images which have a resolution of 10 metres (Day and Muller 1988). To produce a DEM resolution of 5 metres for direct comparison with the ATM images, stereo images of at least 1 metre resolution would be required. Stereo ATM images of the study area were acquired due to requirements of another research project using the same data. However, these images could not be used to produce a DEM, partly due to the poor resolution it would produce (of order 25 metres), but mainly because of severe distortions present in the images. A camera model, describing geometric properties of the sensor, is required to remove all distortions apart from terrain distortions. For stable platforms such as satellite sensors and aerial photography, the definition of a camera model is a straightforward task.
Figure 5.4 - Line drawing showing the derivation of the parallax equation (from Colwell 1983).
However, complex attitudinal and altitudinal distortions present in ATM data prohibit its description by such a model.

Aerial photography acquired in conjunction with the ATM images, can and has been used to derive elevation data for the study area (Allison and Muller 1992). Photography is produced as hardcopy paper prints (or film negatives) which firstly need to be digitized into a raster grid, using either a flat-bed scanner or a digitizing camera. Due to the fine grain of the photography, the print may be digitized at a resolution of approximately 10μm, which, with the flying height of the aircraft, gives a ground resolution of approximately 10cm. The photography can therefore be used to produce a DEM of 0.5 - 1 metre in resolution, ideal for the purposes of this study.

Work by Allison and Muller (1992) demonstrates the production of DEMs from aerial photography, but several problems experienced with the techniques means that such DEMs may not yet be produced routinely. These problems are related to the high resolution of the aerial photography and harsh shading present in the data. Both factors increase the difficulty in identifying coincident points in the stereo images. Increasingly severe terrain distortion present in the images gives rise to greater difficulty in the matching process required to identify these points. The high resolution actually makes many small topographic features appear relatively large (within the confines of each patch) often causing the matcher to fail. A second problem, of shading, also results in failure, as there is simply no detail within individual patches which can be used for matching purposes.

Figure 4.4 depicts a Lambertian shaded DEM produced by Allison and Muller (1992) using stereo-matching techniques. The DEM is not complete due the problems mentioned and cannot therefore be used within this study. However, it is worth mentioning these early results, as similar techniques will eventually be used routinely to produce high resolution DEMs. Furthermore, the techniques are almost totally automated and would therefore fit well into an expert system in which the user need not be concerned with the manual creation of a DEM.

A second method of deriving a DEM is to acquire altimetry data. Altimetry data can be acquired in one of two ways, namely using laser profiling or radar techniques (Cohen et al. 1987, Rees 1990, Seshamani 1993). Satellite radar altimetry data is presently acquired globally at a spatial resolution of 7.3km. Better spatial resolution can be achieved using airborne instruments (i.e., 5 metres for an airborne laser profiler (Rees 1990); however, these data are recorded as a profile and not an image. Unfortunately, it would be extremely difficult to combine the number of profiles needed to create a DEM of the appropriate resolution and coverage in this way due to the distortions introduced
by the aircraft motion. Therefore the resolution and intrinsic one dimensionality of altimetry data preclude its use within this study. Future sensors will be capable of producing global DEMs at resolution of as little as several metres, e.g., use of a tethered interferometric synthetic aperture radar (Moccia 1992). Such a system is planned to include two SAR receivers tethered up to 100km apart which enables the high accuracy and resolution required.

A more conventional method of producing DEMs is to interpolate digitized map contours onto a regular grid. The remainder of this section will therefore concentrate on an assessment of this technique.

Topographic base maps are a widely available source of elevation data. In the U.K., they are produced by the Ordnance Survey using manual interpretation of stereo aerial photography. This procedure may introduce some errors into the fine detail of the maps, but the presence of a number of control points surveyed at ground level within each map affords very accurate height information at, and near, these points.

The digitization of contours from a base map can be done in one of two ways, firstly, by manually digitizing points along each contour line, and secondly, by digitizing the map into a raster grid and using line-following algorithms to digitize contour features (Howman and Woodsford 1978). The latter technique is faster, but the requisite software was unavailable for this study. The manual technique is extremely time consuming and does not fit well into an expert system environment, but it should only need to be performed once for each study area. It is used here merely as a means of deriving the elevation data required for further analysis.

Several interpolation techniques are commonly available which may be used to convert digitized contour data into a raster DEM. The techniques evaluated in this study are listed in Table 5.1. Other interpolation techniques exist, but are often only variations on those tested here. Each of the techniques used here is described below.

The first three methods are all grid-based techniques. Methods 1 and 2 grid the irregular input data first. If a data point lies within a grid cell, the data value is assigned to that grid location; if the cell contains more than one point, an average value is calculated and assigned to the grid location. This method has the advantage of increased speed of calculation over many of the other routines (Table 5.2). This is, however, offset by a loss of accuracy, particularly in areas of steep slope, where a slight shift in the digitized points can result in the height value being assigned to a neighbouring grid cell. Any subsequent edge detection routine applied to the DEM would also be affected, as would the calculation of dip and strike measurements.
<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Method No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted average interpolation</td>
<td>UNIRAS</td>
<td>1</td>
</tr>
<tr>
<td>Double-linear interpolation</td>
<td>UNIRAS</td>
<td>2</td>
</tr>
<tr>
<td>Polynomial surface fitting</td>
<td>UNIRAS</td>
<td>3</td>
</tr>
<tr>
<td>Triangulated irregular networks</td>
<td>UNIRAS (UNIRAS 1989)</td>
<td>4</td>
</tr>
<tr>
<td>Theissen polygons</td>
<td>MAPICS (MAPICS 1986)</td>
<td>5</td>
</tr>
<tr>
<td>Kriging</td>
<td>University College London, (Day 1991)</td>
<td>6</td>
</tr>
<tr>
<td>Continuous curvature</td>
<td>General Mapping Tools (Wessel and Smith 1992)</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 5.1 - Interpolation routines used in the evaluation of DEM creation techniques

<table>
<thead>
<tr>
<th>Method</th>
<th>Processing Time (CPU Hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.22</td>
</tr>
<tr>
<td>2</td>
<td>4.80</td>
</tr>
<tr>
<td>3</td>
<td>6.01</td>
</tr>
<tr>
<td>4</td>
<td>7.15</td>
</tr>
<tr>
<td>5</td>
<td>22.67</td>
</tr>
<tr>
<td>6</td>
<td>8.74</td>
</tr>
<tr>
<td>7</td>
<td>4.32</td>
</tr>
</tbody>
</table>

Table 5.2 - Processing times for the various interpolation routines. All interpolations were performed on a Sun Microsystems SPARCStation 2, with the exception of method 5 which could only be executed on a VAX 11/750

Method 1 (weighted-average interpolation) searches for the closest point in each of four quadrants surrounding the grid point in question, providing that each point lies within a specified radius. A weighting factor is applied to each point, which is inversely proportional to the 2D Euclidean distance between the control points and the grid location (Figure 5.5). The grid value, \( Z_0 \), is then calculated using:-
Figure 5.5 - Finding the closest points in each quadrant and determining the Euclidean distance
\[ Z_o = \frac{\sum_{i=0}^{n} W_i \times Z_i}{\sum_{i=0}^{n} W_i} \]  \hspace{1cm} (5.1)

where \( W_i \) is the weighting factor for each data point, \( Z_i \) is the data value, and \( i=1, \ldots, n \) is the number of data points.

Method 2 interpolates the control points using a double-linear interpolation technique. This involves a linear interpolation along a line joining two points in each neighbouring quadrant to their intersections with the X and Y axes passing through the grid location. This is followed by a second linear interpolation along each axis to the grid location itself. This method has the disadvantage (at least in this application) that the interpolated DEM is not fitted through the control points and is therefore likely to result in the generation of a smooth surface albeit closely related to these points.

The final grid-based method (method 3) employs two windows which are passed simultaneously across the original data. A polynomial of \( n^{th} \) degree is used to fit (using a least-squares approach) all data points in the larger window (\( n \) is determined from the number of data points found within the window). This is, in turn, used to interpolate all grid locations in the smaller window. A method of this nature is extremely dependent on the size of window chosen. For instance, if the larger window is too small, no data points will be found and the interpolation will fail. On the other hand, if too many points are found, a polynomial of very high-order will be required which may occasionally introduce spurious features into the resultant DEM.

Method 4 creates a triangulated irregular network (TIN) from the data points using Delauney triangulation (Figure 5.6). A fifth-order polynomial in X and Y is then used to interpolate within each triangle. Each polynomial is calculated using five partial derivatives determined at the vertices of each triangle by examining the value of each point and its closest neighbours. This attempts to provide a smooth continuation of the surface between neighbouring triangles.

Method 5 is closely related to Method 4 in that it produces a set of Theissen (or Voronoi) polygons around the data points (Figure 5.7). For each grid location, weighting factors are determined for the points adjacent to the polygon. These are calculated using the proportion of the area of a new Theissen polygon, introduced by the grid location, which is included in each original polygon sited in that area (Figure 5.8) (Sibson 1980).
Figure 5.6 - Example of a triangulated irregular network created between digitized contour points.
Figure 5.7, Creation of Theissen polygons around digitized contour points

Figure 5.8 - Creation of new Theissen polygon around new grid cell showing area captured from each original polygon.
Kriging (Method 6) is described in detail in Davis (1973) and is based on the regionalized variable theory (see also Oliver 1990). This theory assumes that a variable is statistically homogeneous throughout the area of study (i.e., the same pattern of variation can be observed at all locations). This variation is modelled using a semi-variogram which plots the semi-variance, the average squared difference between pairs of values, against the distance or lag between them (Figure 5.9). Semi-variance $\gamma$, between two points $Z(x)$ and $Z(x+h)$, distance $h$ apart, is calculated using the following equation:-

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}\{Z(x_i) - Z(x_i + h)\}^2 \tag{5.2}$$

where $n$ is the number of pairs of sample points separated by distance $h$.

A smooth function is fitted to the semi-variogram which may be modelled using a linear, polynomial or spherical relationship (here a simple linear model has been used, as this model most readily fits the semi-variogram, Figure 5.10). All points lying within a search radius are used to calculate the interpolated value. Weights for these points are calculated using a series of equations of the form:-

$$\sum_{j=1}^{n} a_j \gamma(u - x_j) = \gamma(u - z) - \lambda \tag{5.3}$$

where, $a_{1...n}$ are the weighting factors,

- $n$ is the number of points within the search radius,
- $\gamma$ is the semi-variance from the semi-variogram for the calculated distance,
- $\lambda$ is an estimate of the error of interpolation, and
- $Z$ is the location of the unknown point.

The unknowns, $a_{1...n}$ and $\lambda$, are calculated by solving the series of linear equations which are produced from Equation 5.3.

The search radius, defining points to be included in each semi-variogram, and the number of directions investigated through the grid location, may be changed to suit the input data. If either of these parameters is too small then no data points will be found, and an interpolation cannot be performed. A major advantage of this technique is that error statistics ($\lambda$) may be produced for each grid-cell in the resulting DEM. Conversely, a disadvantage is that the same semi-variogram model is used throughout the interpolation procedure while alternative models might be appropriate for different parts of the terrain. This, in conjunction with large variances in some areas may introduce noise into the final interpolated grid.

147
Figure 5.9 - Semivariogram - plot of semi-variance \( C(h) \) against vector distance \( h \). \( K(0) \) represents the variance of the data, and \( e \) is a small value below which \( C(h) \) is considered to be equal to the variance.
Figure 5.10 - Semi-variogram of digitized contour data.
The final method for interpolating digitized contours onto a regular grid (Method 7) is an adjustable-tension continuous-surface gridding algorithm (Smith and Wessel 1990). Standard minimum-curvature interpolation techniques grid the data whereby the resulting surface has continuous second derivatives and minimal total squared curvature (Briggs 1974, Swain 1976). This method can introduce large oscillations and extraneous inflexion points. Smith and Wessel (1990) have modified the minimum curvature technique by adding a tension component, where gridded values $z(x,y)$ are determined by solving:

$$(1-T) \times L_{(||z||)} + T \times L_z = 0$$

where $T$ is a tension factor between 0 and 1 and $L$ indicates the Laplacian operator. When $T = 0$ a "minimum curvature" solution is found; Smith and Wessel suggest a value of 0.35 for $T$ when interpolating over steep topography; this suppresses the large oscillations. Values for $T$ have been tested (i.e., 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, and 1) and results indicate that the most accurate DEM is produced with a value for $T$ of 0.2.

Each of the interpolation routines described in this section have certain advantages and disadvantages. Those that have been reported in the literature for the generation of DEMs include methods 5, 6, and 7 (Sibson 1980, Day 1991, Smith and Wessel 1992, respectively). No comparison has previously been made between these methods; it is therefore difficult to judge which method provides the most accurate results simply from a study of the available literature.

### 5.1.2 Evaluation of DEM accuracy

DEM's have been created in this study for a $25\text{km}^2$ area centred on Llyn Cowlyd (G.R. 272500 362500) (© Ordnance Survey 1976), using each of the interpolation routines outlined in the preceding section. Figure 5.11 shows an example DEM of the area; light tones indicate areas of high elevation while darker pixels signify lower land. It is difficult to ascertain which derived DEM is most accurate solely on the basis of a simple visual inspection. Therefore, a series of tests must be performed to determine which exhibits the highest geometric fidelity when compared with the original data. Although it is true that errors are caused during digitization, and also that the original contoured map is not without error, at this stage the aim is merely to evaluate how interpolation routines retain information inherent in the original digital contour data.

---

1 A resolution of 5 metres has been chosen at the interpolation stage. This corresponds with the resolution of the ATM data so that both data sets may be compared directly.
Figure 5.11, Digital Elevation Model for a 5 x 5 km area surrounding Llyn Cowlyd.
Certain procedures (listed in the introduction to section 5.1) have been used here to measure the accuracy of each DEM. These attempt to measure not only the absolute height accuracy of the DEM but also how well it describes, in broader terms, the morphology of the area. These are discussed below.

5.1.2.1 Comparison of spot heights

The first evaluation method compares spot heights randomly identified in DEMs with those of corresponding points identified manually from a map. The root mean square (RMS) error and the standard deviation (SD) of the height differences are calculated for these points. This is intended to provide an indication of the absolute accuracy of each interpolation routine and the variation in resultant errors. Height differences have also been regressed against slopes estimated from the map to investigate any systematic relationship between spot height errors and concentration of digitized points and the rate of change of the height value. These might be caused by the gridding procedures of methods 1 and 2 or the high polynomial interpolations of method 3.

Table 5.3 shows the results of this analysis, using 60 randomly-located points. It illustrates that method 3 gives a high RMS error and SD, while the others are very similar with RMS errors ranging between 1.5 and 2.5 metres. Visual inspection of method 3 reveals that it results in a blocky texture due to very high-order polynomials required in some areas of the DEM. It can be concluded from the regression analysis that none of the spot height errors are particularly related to the slope of the terrain and hence the concentration of points.

<table>
<thead>
<tr>
<th>Method</th>
<th>Root Mean Square Error</th>
<th>Standard Deviation</th>
<th>R² v. Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.992</td>
<td>2.007</td>
<td>7.0</td>
</tr>
<tr>
<td>2</td>
<td>1.531</td>
<td>1.542</td>
<td>7.4</td>
</tr>
<tr>
<td>3</td>
<td>9.930</td>
<td>9.940</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>2.528</td>
<td>2.609</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>1.772</td>
<td>1.769</td>
<td>3.8</td>
</tr>
<tr>
<td>6</td>
<td>1.555</td>
<td>1.560</td>
<td>7.2</td>
</tr>
<tr>
<td>7</td>
<td>2.112</td>
<td>1.954</td>
<td>3.4</td>
</tr>
</tbody>
</table>

R² = Coefficient of Determination of height differences versus DEM parameter.

Table 5.3 - Statistical results of random spot height sampling
5.1.2.2 Checking the integrity of the DEM by recontouring

For most applications, and for this study in particular, it is imperative that a DEM retains all of the information held within the shape of the map contours, as this contains information that is of considerable significance in geomorphological, and hence geological terms. One way to examine whether the DEM has retained this information is to recontour the gridded DEM data and to compare these contours with those digitized from the map (assuming that no further artefacts are introduced by the recontouring technique). DEMs may be recontoured to the heights of the original map contours to assess the accuracy with which their shape is retained. In addition, recontouring intermediate heights allows an examination of how slope morphology is preserved in areas where there is little or no digitized data.

Figure 5.12 shows a series of contours derived using method 1. The 10 metre contours, in the recontoured data, coincide with comparable contours in the original data set, as, with the exception of method 2, each procedure is designed to fit the original data. However, if the 5 metre contours are examined, it is apparent that the interpolation procedures can smooth out information held within digitized contours. This results in a stepped effect, particularly along ridges and valleys. The reason for this effect is that the interpolation routines search for the closest points to the grid location (Figure 5.13) regardless of their position along a contour line or how neighbouring contours may be interrelated. The interpolation procedures require some knowledge of drainage patterns and ridge lines to overcome this problem. More recent algorithms which attempt to overcome this problem, which have not been evaluated here, are discussed later.

5.1.2.3 Comparison of DEMs by height differencing

Results of spot height analysis and visual comparison of the recontoured data provide limited information relating to differences between the derived DEMs. Another potentially useful mechanism with which to highlight these differences is to subtract one DEM from another. Although this evaluation makes no direct comparison with the map data, it indicates areas of disagreement between DEMs.

The height difference images also indicate that many of the routines tested here produce a stepped effect between areas surrounding the digitized contours. As an example, Figure 5.14 (Method 1 subtracted from Method 2) shows this effect quite markedly especially in areas with sparse data, and provides a qualitative assessment of the spatial distribution of relative differences between these two interpolation routines. In the case of Figure 5.14, method 2 exhibits a gross stepping effect, while method 1 is smoother. However, each interpolation routine appears to exhibit this stepped effect to varying
Figure 5.12 – Recontoured DEM showing the smoothing of 5 metre contour lines
Figure 5.13 - Standard interpolation routines use the closest data points found within a search radius, rather than matching features in successive contours.
Figure 5.14 - Height differences between DEMs 1 and 2. Brighter tones indicates larger differences between the two DEMs. The figure highlights a stepping effect caused by the interpolation, between neighbouring contours.
degrees. Consequently, each will produce subsequent errors in products derived from the DEM. Figure 5.15 illustrates a profile of height values along a valley (taken from the Kriged DEM). It demonstrates the stepping effect between the 10 metre contours and indicates the magnitude of the error as being of order 1-2 metres.

5.1.2.4 Identification of significant morphological features through drainage network simulation

The methods for evaluating DEMs discussed thus far have examined absolute and relative height values and retention of contour information at local scales. A further, separate technique is required to assess the accuracy of the geomorphological content of each DEM on a larger scale. Such a method should be able to assess the contextual accuracy of each element in the DEM by its relationship with neighbouring pixels and the DEM as a whole. One option is to simulate the drainage network for each DEM and compare this with the true (or actual) drainage network. This method is useful because the drainage network also implies geological information. Parvis (1950) and Howard (1967) give comprehensive empirical descriptions of some thirty different drainage patterns relating to various forms of structural geology. For example, trellis patterns are often indicative of sedimentary strata, while radial patterns may indicate an igneous environment (Figure 5.2).

A variety of algorithms have been developed to extract ridge and valley lines from a DEM. Most employ a 3*3 kernel which is passed across the image to identify different geometric properties e.g., \( \cap \) or \( \cup \) shaped, sink holes or saddle points. The kernel-based approach tends to produce poor disconnected drainage patterns (Riazonoff et al. 1988). Riazanoff et al. (1988) and Skidmore (1990) provide good reviews of the algorithms available. Riazanoff et al. (1988) also introduce what they define to be a 'structuralist' approach, in which lines are followed through the DEM via the steepest downward path, starting from isolated points such as saddle points, and then, for ridge-line detection, climbing along the steepest slope. This method simply derives the major valley and ridge lines as only a few seed points are chosen. Jenson et al. (1985) developed this further by using every element within the DEM as a seed point and following the steepest path through the terrain. A similar algorithm, raindrop, has been developed as part of this project. Raindrop differs from the Jenson et al. (1985) algorithm in that it does not assume that all sink holes or depressions in the DEM are errors and does not attempt to fill in these areas by modifying the DEM. It should not necessarily be assumed that sink holes or depression are errors in the DEM. This is particularly true in the Llyn Cowlyd study area, where depression have been formed by the inter-layering or hard and soft rocks. The following paragraphs will describe the raindrop algorithm and the results derived from each DEM.
Figure 5.15 - Stream profile showing the stepping effect introduced by interpolation.
Raindrop 'drops' one raindrop on each pixel of the DEM, assuming that the surface of the DEM is completely impervious. The progress of each raindrop down the hill slope is simulated by searching for the steepest slope in a 3*3 window surrounding the pixel upon which the drop has fallen (Figure 5.16). Each drop is followed to its new position and another window is examined. This operation is repeated until the raindrop reaches the edge of the image or a specified height designated as sea-level. These steps are then duplicated for the next raindrop to fall on the DEM, and so on. As each raindrop passes through a point in the DEM a raindrop count for that point is incremented, and the final valley-line image is a count of the magnitude of the streams passing through each point (given the assumption that each pixel in the DEM has equal run-off properties) (Figure 5.17). If the raindrop reaches a point from which there are no negative slopes to neighbouring pixels (e.g., a sink hole) then a lake-filling algorithm is invoked. This algorithm recursively fills an area surrounding the sink hole until a negative slope is once again found around its boundary (i.e., the lake is full and water can flow out of the lake); the stream tracking can then be resumed. It is possible that lakes may be superseded by larger lakes, especially in wide valleys with little contour information, where interpolation routines can insert shallow hollows. When the 3*3 window is searched for the steepest path, it is possible that there will be two or maybe more equally steep paths emanating from the pixel in question. The window is searched in an anti-clockwise fashion and the first of the steep paths found is chosen. To avoid any bias in the process the starting pixel in the 3*3 neighbourhood is randomly selected (Figure 5.18). This has an added advantage in that deltas may be formed in flat areas, when one raindrop follows one path and the next follows another which is equally valid.

The raindrop algorithm has an advantage over other drainage network detectors in so far as the order or magnitude of each stream is determined. This stream ordering is directly analogous to the Strahler method developed in 1952 (Strahler 1952). This means that by employing a thresholding technique streams of a particular range of sizes may be extracted, which may be useful in many geological applications. In this study area, low-order streams towards the tops of hills and mountains often tend to follow lithological boundaries.

The results of the drainage network detection algorithm can be compared with the network shown on the topographic base map (Figure 5.19 and 5.20). Network analysis highlights some very interesting differences between the DEMs. Firstly, the digitized drainage network appears to be incomplete compared with the simulated networks. This is due to the generalization of features that occurs during the map making process, which means that the two data sets are not directly comparable. However, from a visual comparison and by interpreting the topographic map it appears that DEMs derived using methods 1 and 6 produce the most realistic drainage networks and tend to follow minor
Figure 5.16 - Diagram showing the flow of raindrops over the DEM

Figure 5.17 - Diagram indicating the increase in stream size as more raindrops are added.
Search direction

Heights in brackets.

Figure 5.18 - Diagram showing the anticlockwise search for the lowest neighbour. The random starting point means that different paths may be chosen if more than one neighbour has the lowest height.
Drainage network digitized from 1:10,000 base map

Method 1

Method 2

Method 3

Figure 5.19 - Digitized and simulated drainage networks for various DEMs.
Figure 5.20 - Simulated drainage networks for various DEMs.
variations in topography (which are closely related to the local geology). The drainage network derived using method 5 enhances horizontal and vertical artefacts in the DEM. It is not known how these artefacts are caused, although it could result from some of the digitized points being too close together (Campbell pers. comm. 1990). The DEMs created using methods 2 and 7 tend to follow the major streams, noted in the Ordnance Survey data, but they also produce a sheeting effect in areas of approximately uniform slope, where many of the streams run in parallel. Method 4 produces an extremely large lake due to the fact that erroneous large height values have been created along the eastern edge of the DEM, while the blocky effect of method 3 has been repeated in its drainage network.

5.1.2.5 Identification of minor artefacts by Lambertian shading

Figure 5.20 illustrates that some gross artefacts can be present in the DEMs which only become apparent within the simulated drainage networks. These large errors are introduced by interpolation routines. It is also reasonable to assume that there may be a large number of minor artefacts present. These may be highlighted using Lambertian shading techniques (Holben and Justice 1981). Again this evaluation technique has the added advantage that it can be employed to enhance geomorphological and hence geological features within the DEM (Sauter et al. 1989).

Lambert's law states that the intensity of reflected light is proportional to the cosine of the incidence angle \( \theta \) (Zhou 1992):

\[
l = l_0 k \cos(\theta)
\]

where,

\[
0 \leq \theta \leq \frac{\pi}{2}
\]

\( l_0 \) is the intensity of the light source, and 
\( k \) is a constant that denotes an approximation to the diffuse reflectivity of the surface.

Here the surface is assumed to have a reflectivity of 1.0 and the intensity of the light source is also set to one. The only variable is therefore the phase angle between the illumination source and the normal of the surface at any point in the DEM (Figure 5.21). The phase angle can be calculated using the following formula:

---

2 Other shading algorithms are described by Watt (1989) and Foley et al. (1990), but the method chosen here is computationally simple and succeeds in highlighting the features of interest.
Figure 5.21 - Diagram showing the phase angle between the illumination angle and the normal to the slope.
Figure 5.22 - Lambertian shaded DEMs produced from methods 2, 5, 6, and 7, showing the artefacts introduced by the interpolation methods.
\[ \gamma = \cos(\alpha) \times \cos(\theta) + \sin(\alpha) \times \sin(\theta) \times \sin(\beta - \varphi) \] (5.6)

Figure 5.22, shows Lambertian shaded images created using four of the interpolation methods (methods 2, 5, 6, and 7). Each image shows a number of artefacts that are a function of the relevant interpolation process. The gross stepping of method 2 is very apparent, as are the more subtle artefacts of the other three methods. Method 5 introduces a number of horizontal and vertical artefacts. Methods 6 and 7 exhibit less obvious artefacts. The Kriging method produces patches of noise within the DEM which appear as rough areas in the Lambertian shaded image, while with the continuous surface method certain digitized points are manifested as local peaks within a fairly uniform surrounding area. The artefacts resulting from methods 2 and 6 can be ascribed to previously mentioned inadequacies of these methods; however, the reasons for those produced by methods 5 and 7 are unclear.

5.1.2.6 Summary

When creating DEMs for geological applications, one requirement is to include as much fine detail as possible, as a subtle change of slope could indicate an important geological feature. Such features could range in scale from major tectonic displacements (such as the San Andreas fault) through to millimetre scale bedding planes on a rock exposure. Each extreme provides geological structural information which may help to build a more accurate model of the geology. In the case of this study area the level of detail achieved in the DEM is limited by the resolution of the contour data and the detail within each contour. Fine detail, such as small kinks in a contour line, may indicate geological features, particularly if they are repeated in neighbouring contour lines.

Digitizing each contour at a high density and interpolating data onto a fine grid (in this case 5 metres) can produce a record of the fine detail. One difficulty associated with this is that in areas with little geomorphological expression data points can be very widely spaced, and may be outside the search radius of some interpolation routines. This effect is most noticeable in DEMs derived from methods 3 and 6, especially within the three large lakes in the study area where no contour information is available. The Kriging software outputs zeros for grid locations where no data points have been found, while method 3 outputs 999.999. These areas of unknown height have been artificially filled in by recursively passing a mean filter across the DEM which calculates means only for the unknown pixels. This method smoothes out small problem areas quite satisfactorily, but over large areas can produce flat terrain. Routines which create a TIN between the data and the continuous surface method do not experience this problem.
A DEM should also be accurate both in terms of individual height values and geomorphological shape, not only so that useful primitives may be obtained directly from it, but also so that it may be integrated effectively with remotely sensed images. With the exception of the spot height method, it is difficult to quantify results from each evaluation method used, as they are each evaluated visually. A final decision as to which interpolation method produces the most accurate DEM is therefore rather subjective. However, it is possible to exclude the DEM derived using method 3 due to its large RMS error. Method 2 unfortunately masks much fine detail due to the double interpolation process. Methods 4 and 5 produce marked artefacts apparent in simulated drainage networks, and the initial gridding procedure of method 1 causes a shift in position of the digitized points. The remaining DEMs derived from methods 6 and 7 provide the most accurate results. There is little to choose between each method. The Kriging method introduces some noise into the DEM, apparent in the Lambertian shaded image, while the continuous-curvature method results in a relatively poor drainage network due to a slight smoothing of detail in the DEM. It must be concluded that none of the methods tested produce entirely satisfactory results, but in this evaluation, the method chosen as the most accurate is the Kriging method, due to its lower RMS error and greater detail apparent in the drainage network and Lambertian shaded images.

A number of techniques have been developed which may overcome the problems of stepping and introduction of artefacts, but which were not available for this evaluation. One such technique, produced by ESRI (1992b), uses manually digitized ridge and valley lines along which a smooth interpolation is fitted. This method removes any stepping effect along these features but requires a large amount of user interaction even for major ridges and valleys. It would take an inordinate amount of time to digitize all the minor features.

A more successful suggestion might be to match individual features in neighbouring contours. Figure 5.13 indicates how standard routines would interpolate a ridge which cuts diagonally through the contours. Most routines would interpolate using points A, B, C, and D and produce a smooth contour where there should be a ridge. However, an interpolation routine which employed some form of contour-matching, more closely related to the manner in which a human might interpret this area, would use points E, F, G, and H, to produce a more realistic interpolation. Such a technique has been developed by Tang (1992). This method automatically identifies specific geomorphological elements, such as peaks, pits, saddle points, ridge and drainage lines from a raster representation of the contour lines. An interpolation is then performed

---

3 This shift may be as much as half of the diagonal distance across the DEM element.
along these features. The reported results (Tang 1992) would appear to overcome the problems inherent in the methods tested here and might therefore be considered for future studies.

5.2 Extracting geological information from DEMs

Chapter 4 describes how geological features may be extracted from remotely sensed data. Geological information may also be derived from elevation data in a broadly similar manner. Although DEMs do not contain the spectral information of remotely sensed data, they do contain much valuable geological information (Chorowicz et al. 1989, Sauter et al. 1989, Wadge et al. 1990). This section will describe a number of ways in which such information may be extracted both manually and automatically. Each method will be evaluated using a combination of qualitative and quantitative techniques similar to those employed in Chapter 4.

5.2.1 Techniques to enhance geological information in digital elevation data

As mentioned in the introduction to this chapter, geological information may often be derived from the geomorphological expression of the terrain and from drainage patterns described by the terrain. Techniques are therefore required to enhance and extract this information. A discussion follows of several standard ways to enhance information within DEMs and further describes how DEMs and remotely sensed data may be used in combination to aid interpretation.

A good indicator of a surface or sub-surface geological boundary or structure is a change in slope (Figure 5.1). Slope images may be created by passing a gradient filter across a DEM (Sauter et al. 1989). In this study, a Sobel filter has been used in two orthogonal directions. The magnitude and direction of the slope may be derived from the results of these filters using a simple vector summation approach (Figure 5.23). Slope images may be described as an estimate of the first derivative of a DEM, as the resultant values indicate rate of change of height over the local area. To determine the rate of change of slope, a gradient filter may be applied in turn to the slope image to produce an estimate of the second derivative of the DEM. The resulting image will then highlight breaks of slope within the DEM. Both the slope and rate of change of slope images should highlight many geological features in this study area due to the strong relationship between the geomorphology and the underlying geology.

A useful by-product of the gradient filter and vector summation is an image showing the direction of slope for each pixel in the DEM. This is termed an aspect image (Jones et
**Sobel filters**

<table>
<thead>
<tr>
<th></th>
<th>-1</th>
<th>-2</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Identifies N-S edges

<table>
<thead>
<tr>
<th></th>
<th>-1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Identifies E-W edges

---

**Figure 5.23 - Vector summation of two orthogonal edges to produce slope magnitude and direction.**
The slope directions, in themselves, are not particularly useful in identifying geological features; however, a measure of change of direction would help to identify interesting features, as many changes in direction of slope occur at rock outcrops and breaks of slope. Such a measure can be obtained by passing a gradient filter over the aspect image to determine rate of change of direction of the slope for each pixel in the DEM.

Lambertian shading techniques (described in section 5.1.2) may also be used to enhance subtle changes of slope in the DEM, producing an effect similar to that of natural shading in remotely sensed images. However, the advantage of DEM shading is that rather than the standard and fixed shading of the scene at time of image acquisition, any illumination position may be simulated to enhance different directional features in a DEM. Use of a small zenith angle highlights the major geomorphological features in all directions, while use of a large zenith angle enhances more of the minor features which are perpendicular to the azimuth angle of the illumination (Figure 5.24). Figure 5.25a shows simulated shading of the DEM at the illumination angles equivalent to that during the acquisition of the remotely sensed images (c.f. Figure 4.8 - the false colour composite of the same area). By comparison, Figure 5.25b depicts a shaded scene with the illumination positioned to the north-east (albeit an artificial phenomenon for this study area). Additional features are highlighted by this second illumination, particularly on the south-east facing slopes. This demonstrates that many different features are enhanced by altering the angle of illumination. Success of these images in highlighting geological features is discussed in section 5.2.3.

Drainage networks derived from DEMs may also provide a useful source of geological information. The overall pattern of drainage is generally indicative of the nature of geological environment present (Figure 5.2); for example, any strongly linear features in the network often suggest the presence of a fault (Parvis 1950, Howard 1967). It has also been noticed, from a comparison with the remotely sensed data, that many low-order streams in this study area, particularly at, or near, summits of mountains, tend to follow lithological boundaries (Figure 5.26). These low-order streams may be extracted simply by applying a threshold to the drainage network to extract all pixels in the network image within specified bounds. Again these minor features are only likely to be useful in areas similar to Llyn Cowlyd, where an inter-layering of hard and soft rocks controls the erosion of the terrain locally. More uniform geological environments and more mature drainage patterns are less likely to facilitate the identification of such features (Lillesand and Kiefer 1979).

---

4 Here, streams do not necessarily refer to actual bodies of surface water, but simply to drainage channels identified by the simulation process.
Solar zenith angle of $1^\circ$.

Solar zenith angle of $80^\circ$.

Figure 5.24 - Images illustrating the effect of changing the solar zenith angle.
A. Simulated natural shading of the DEM (zen=45.1°, az=119.6°).

B. Simulated shading with azimuth = 45°.

Figure 5.25 - Images illustrating the effect of changing the solar azimuth angle.
Figure 5.26 - Simulated drainage network showing streams that relate to lithological boundaries (A, B, and C).
When in the field a geologist may often infer structural information by examining the geomorphology of the surrounding area. A geologist may also measure dip and strike values at some distance from an exposure or at a lithological boundary. This can be achieved by viewing the exposure along strike and measuring the dip using a compass clinometer. These field methods may also be emulated by computer using a combination of DEM and remotely sensed data. Such facilities may save on extensive field surveys which can be both time consuming and expensive.

DEM and remotely sensed data can be combined to produce a 3D visualization of the terrain by overlaying images on top of the DEM (McLaren and Kennie 1989). Figures 5.27 and 5.28 show the results of such a process and suggests ways in which different views may help to understand different aspects of the scene. Figure 5.27b gives a general indication of the topography and clearly shows the effects of glaciation in the area, while Figure 5.27a shows how the geology dips into the side of the mountain. Each feature in the scene may be viewed from any number of directions, which aids interpretation. Figure 5.28 illustrates how a perspective view of the terrain can be created which approximates the view of an actual ground photograph, while Figure 5.29 shows how the view may be manipulated to measure dip and strike directly from the screen. Such views can to some extent replace the need for expensive ground surveys or at least reduce the time and cost of such surveys. Manual interpretations have been performed on both perspective views, such as these, and on Lambertian-shaded DEMs. These are compared with a British Geological Survey (BGS) geological map (1:50,000 Series, Sheet 106, © Crown copyright 1985) of the area in section 5.2.3.

5.2.2 Evaluation of Techniques for Segmenting Digital Elevation Data

Just as remotely sensed data can be segmented into areas likely to contain geologically relevant information, using multispectral classification techniques, DEMs may also be used to perform a similar function due to the geomorphological information held within such data. Variation in elevation or slope can be used as an indicator to those areas providing most geological information, i.e., rough areas are likely to correspond to areas where geology is exposed or at least only partially covered. This may not necessarily always be the case, where there is little relationship between geological structure and topography, but is indeed true of the Llyn Cowlyd study area. There are two reasons why it is important to know where these areas exist:-

- to segment the DEM and the remotely sensed images in order that computer processing time can be cut, or
- to obtain a measure of likelihood that an edge identified from an image or DEM actually represents a feature of geological significance.
Figure 5.27 - Comparison of different perspective views
Simulated perspective view

Ground photograph

Figure 5.28 - Comparison of perspective view with ground photograph.
Figure 5.29 - Perspective view illustrating that values of dip can be measured directly from the image
Suitable algorithms exist to measure texture over a local area (Haralick et al. 1973) which, in the case of DEMs, would be equivalent to a measure of local roughness. Olsen et al. (1993) also suggest the use of fractal dimensions to classify different types of landscape. However, four standard texture measures are chosen here, including standard deviation, inverse distance moment, entropy, and contrast (Haralick et al. 1973). When applied to a DEM, these techniques should highlight areas where there is a large difference in height within the neighbouring pixels. However, this will also highlight areas of steep slope, but which do not necessarily exhibit any texture or geomorphological information. In such circumstances, confusion may arise between the highlighted areas produced in this way, and the pertinent areas of interest. To avoid this problem, texture measures may be applied to Lambertian shaded images which enhance changes in slope rather than changes in height. However, as Lambertian images highlight features with a directional component, several directions of illumination need to be combined to give a complete measure of the geomorphological features within each DEM. In this case eight azimuth angles have been defined between 0° and 315° separated by 45° and with a zenith angle of 45°. Texture is measured for each Lambertian image and the resulting texture images are combined by extracting the maximum value for each pixel from the eight directional measurements. Each texture measure is evaluated in section 5.3.2 with reference to the success with which geological and non-geological primitives may be separated.

The degree of roughness of terrain is dependent on both the scale of the data and the type of environment present in a study area. At extremely high resolutions roughness could relate to microscopic surface features (e.g., soil or individual blades of grass) rather than geology, while at extremely low resolutions roughness could relate to plate movements. A roughness measure is required which relates to the inter-layering of individual lithologies, which for this study area ranges from several decimetres to approximately 30 metres. The texture measures are calculated for each pixel in the DEM by examining pixels within the surrounding neighbourhood or window. It is therefore necessary to choose an appropriate window size for the task. A window size of 15×15 (i.e., 75×75 metres) has been employed, as this covers an area including a number of inter-layered units and provide a useful texture measure.

These were chosen due to availability of software and as they are commonly used techniques.

From experimentation it was found that this zenith angle tends to highlight most of the interesting features within the Llyn Cowlyd study area. The inclusion of additional zenith angles would not significantly add to the texture measure but would increase the time taken to process the DEM.
5.2.3 Extracting primitive information from DEMs

The surface expression of folds, faults, joints, strikes, lithological contacts and other geological features is often manifested in the form of lineaments (Harris 1987). Such features have been extracted successfully from the remotely sensed data of the study area using Sobel and Canny edge detection filters. The same techniques can also be applied to a DEM. Edges derived from spectral data are commonly used and are therefore relatively easily understood. By contrast, edges derived from DEMs have not been so widely used. Table 5.4 describes the enhanced products used in the extraction of primitive features and the feature characteristics likely to be identified.

<table>
<thead>
<tr>
<th>Input data</th>
<th>Identifiable features derived from edge processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM</td>
<td>Edges will identify steep slopes within the DEM. At individual rock outcrops these are likely to follow lithological boundaries, however, in areas such as cliffs on the SE side of the reservoir, these edges will become broad and may encompass more than one geological feature.</td>
</tr>
<tr>
<td>Slope image</td>
<td>Edges will highlight marked changes in slope such as upper and lower reaches of cliffs. These should frequently relate to lithological boundaries between soft and hard lithologies. Edges should also pick out more subtle breaks of slope in areas which on a broader scale have a uniform slope. Such features may represent boundaries between more similar lithological types.</td>
</tr>
<tr>
<td>Aspect image</td>
<td>Edges will delineate sharp changes in direction of slope such as ridges and valleys. These do not necessarily relate to lithological boundaries but on a local scale may follow exposures and hence local bedding planes.</td>
</tr>
<tr>
<td>Lambertian shaded images</td>
<td>These images enhance breaks of slope within the DEM which in turn relate to lithological boundaries. By altering the illumination angle, resulting edges should cover a wide range of geological features.</td>
</tr>
</tbody>
</table>

Table 5.4 - Input data for DEM feature extraction and types of features likely to be extracted

One problem peculiar to the aspect image is that it represents azimuth angles between 0-360°. Therefore, if there is a small change in angle between 359° and 1° the edge
detector will produce a large edge value. This can be overcome by assigning all angle differences \( A \) which are \( \geq 180^\circ \) to \((360^\circ - A)\), during the edge detection process.

5.3 Assessing success of DEMs in extracting useful geological information

Each DEM enhancement technique is evaluated against its success in aiding the manual interpretation of the geology within a scene and also against its success in automatically identifying geological features. In both cases the results are compared with the geological map which is assumed to be an accurate representation of the geology of the area. Each segmentation method is assessed visually by comparing primitives against geologically/non-geologically relevant areas.

5.3.1 Manual interpretation using geomorphologically enhanced images

Figure 5.30 depicts slope, change of slope and aspect images derived from the DEM. Each image reflects real physical properties of the surface (providing the DEM is assumed to be accurate), which is not always the case with many of the enhancements that have been applied to the remotely sensed data (This should be remembered when interpreting grey levels in each image). Brighter pixels in the slope image represent steep slope (of order 80% slope) while black pixels represent level ground. This image highlights steeper slopes in the terrain very well, and consequently allows a good interpretation of geomorphological/geological features. The change of slope image represents sharp changes as bright pixels and uniform slope as black areas. Many features within this image occur as double parallel lines, highlighting the upper and lower parts of steep slopes (or cliffs). These features more readily relate to lithological boundaries than those in the slope image as more prominent lithological boundaries occur in these locations rather than half way down a slope. In the aspect image, white pixels relate to slopes dipping to the North and decreasing grey levels signify a clockwise change in slope direction through to black which has a bearing of 359.9°. This image is more difficult to interpret and few geological features are apparent. However, it does highlight, in part, the drainage pattern in the area, which can sometimes be useful in identifying faults. The change of direction of slope image depicts more definable features, many of which reflect the geological structure of the area. Care is required when interpreting this image, since many features occur at ridges and valleys which do not necessarily follow geological features, while others highlight exposed rock features very well. Inspection of this image alone is unreliable as aforementioned geological features cannot readily be distinguished from the less important features.
Figure 5.30 - Products derived from the DEM.
A manual interpretation of each product has been performed and results given in Figures 5.31 - 5.34. The slope and change of slope products are excellent images for identifying each major geological feature in the area. Comparison with manual interpretation of the remotely sensed data reveals that the slope and change of slope interpretations provide a closer match to individual exposures rather than following a more general lithological outline. This is explained by the tendency of the eye to group objects together and treats them as a whole, when interpreting spectral images (Gregory 1986). Features identified manually from the two slope images are more comparable to the primitives derived automatically from the remotely sensed images. Both automatically derived datasets give a more accurate and realistic representation of many geological features in the study area, while manual interpretation of the remotely sensed data has provided a more general overview.

In addition to potential lithological boundaries identified in the slope images, faults have also been identified (e.g., A-A' - Figure 5.31). The linearity of a series of slopes dipping in similar directions helps to highlight such features. Faults identified here are not as complete and as numerous as those identified from the spectral data due to lack of surface cover information.

Neither the aspect image nor the change of direction images can be interpreted in isolation, but can be used in conjunction with other data sets. Results shown in Figure 5.31 were obtained using a combined analysis of the aspect images and the slope images. Very little additional information is derived from these products and their use is therefore limited.

The Lambertian shaded images have also been interpreted manually, with results as shown in Figure 5.32. Rather than showing separate interpretations for each illumination direction, the results of the entire sequence of images have been combined. The results again identify all major features but also delineate many minor features not apparent in the slope images. This is due to the success of this technique in enhancing subtle changes of slope in particular directions. One problem encountered in interpreting these images, and indeed the other DEM products, is the precision with which individual features can be identified. This reflects a lack of fine detail within the DEM. Although the DEM has been interpolated to a resolution of 5 metres, the interpolation process resulted in a smoothing of detail originally present in the contour data. Ideally a higher resolution DEM is required so that these features may be identified more precisely. A DEM resolution of approximately 1 or 2 metres would probably be more appropriate for comparison with the remotely sensed data and mapping of the geology in the Llyn Cowlyd study area, although this would require
Figure 5.31 - Primitives identified manually.
Figure 5.32 - Primitives identified manually from Lambertian shaded DEMs.
further evaluation. Such resolutions could not be generated from digitized 1:10,000A contours but may be feasible from stereo-matching of aerial photography.

In general, the slope products and the Lambertian images identify the same major features recognized in the remotely sensed data. They also highlight many additional features, covered by uniform surface material and not apparent in the remotely sensed data, and therefore provide useful supplementary information.

Figure 5.33 illustrates a manual interpretation obtained using perspective views of the remotely sensed data and Lambertian shaded images. Software developed as part of this study allows the user to digitize features from a perspective view such as this, created at any view angle. The operator may therefore 'roam' around the terrain to obtain any number of required views7. Results from this process identify many more geological features than from any manual results shown previously. In terms of the number of geological features recognized, benefits are definitely gained by using both the remotely sensed data and DEM products, rather than say simply the remotely sensed images. Interpretability of both data sets is also increased using the perspective view technique. This allows simulation of views only previously possible during ground surveys, also views of areas that may have been inaccessible on the ground, and any possible oblique aerial view. Geologists can gain a better understanding of the geology and topography than from a planimetric view and can even measure dip and strike values directly from the screen (Figure 5.29).

5.3.2 Segmenting geological and non-geological features using geomorphological descriptors

Segmentation techniques have been used with two aims in mind. Firstly to segment the study area into those sectors likely to contain geological information and those which are not, and secondly to provide a confidence measure relating to the likelihood that a primitive represents a geological feature, which may be assigned to automatically identified features.

An example of each texture measure, when applied to a Lambertian-shaded image, is shown in Figure 5.34. A visual assessment of these images reveals that the standard deviation and contrast texture measures identify those areas of the terrain which best

---

7 At present the software allows the user to alter the view of a wireframe model of the terrain in real-time but does not as yet allow an alteration of the full perspective view in real-time. The creation of the perspective view takes approximately 60 CPU seconds on a SPARCStation 10/30. The increasing performance of computers will eventually make this real-time facility commonplace.
Figure 5.33 - Primitives identified using a perspective view of the terrain.
Figure 5.34 - Texture measures applied to the DEM.
Figure 5.35 - Masked Lambertian shaded DEMs using the standard deviation and contrast texture measures to derive the masks. Each image shows the relative success of using DEM texture to segment geological from non-geologically important areas.
describe the geology. The standard deviation and contrast measure are therefore the only methods that are of use in this segmentation process. To evaluate which of these measures provides the most successful segmentation a texture composite image was created for each measure. This involved combining texture measures from each Lambertian image such that the pixels in the final output texture map were the maximum texture derived from each of the eight illumination angles. The resulting texture image was then thresholded at a 50% level\(^8\), and used to mask a Lambertian shaded image. This masking process allows a visual comparison between 'geological' and 'non-geological' areas for each texture measure. Figure 5.35 illustrates the results of this masking process for two of the measures. The standard deviation measure provides slightly better segmentation, with less of the steep slope chosen at A, where there is little geological detail, and more at B where there is greater geological exposure. The standard deviation measure is also quicker to apply to an image (i.e., on a SparcStation 10/30 the standard deviation measure took less than a minute to execute, while the contrast measure took several hours).

The standard deviation texture measure is also used to define confidence statistics for primitives identified automatically. Figure 5.36 depicts the texture image derived using this technique; the brightness of each pixel is therefore assumed to signify the likelihood of geological features being present at that pixel and its surrounds. The method of assigning a confidence statistic to a primitive using the grey levels is discussed in Chapter 6.

5.3.3 Evaluation of primitives automatically identified from elevation data

Primitive features have been derived automatically from the DEM and its products, using the techniques outlined in section 4.2.4, i.e., edge detection, thresholding, thinning, and line extraction. The results of these procedures are evaluated by comparing primitives with manual interpretation results and with those primitives derived from the remotely sensed data.

Results for the automatic procedures are displayed in Figure 5.37. Primitives derived from the DEM and the slope image are very similar to those identified manually. The automatic primitives are however more numerous as image processing techniques identify everything which incorporates certain characteristics while the human will assign less significance to smaller and less contiguous features (Gregory 1986). These

---

\(^8\) This threshold was chosen arbitrarily but kept constant for each texture measure to allow direct comparison. A 50% threshold indicates that 50% of the pixels within the image have been selected. This was found to give a reasonable separation between rough and uniform areas of the terrain in this study area.
Figure 5.36 - Texture measure averaged over eight Lambertian shaded images.
Figure 5.37 - Primitives identified automatically.
lesser features may or may not have geological significance. Verification of their significance (or lack of it) may be established by ascertaining whether structural measurements derived from these features fit the geological model of the surrounding area (again this is discussed in Chapter 6).

Lambertian images also produce results similar to, but more numerous than, the equivalent manual interpretation. Lack of detail in the DEM and the precision with which features could be identified, noticed during manual interpretation, is not a problem here as the edge filters will identify precisely where the edge is located. This precision can however be affected by the illumination zenith angle. If a feature is not sharp, and has a rounded appearance (as is the case here due to the smoothing of the interpolation procedure), then an edge identified between the shaded and illuminated parts of the feature will vary as a function of the illumination angle (Figure 5.38).

5.4 Site specific conditions and consequences for other study areas

A number of environmental conditions specific to the Llyn Cowlyd study area and other similar sites affect the use of many techniques mentioned in this Chapter. One main factor is the glaciation of the area, which has resulted in U-shaped valleys and a smoothing of the topographic expression. Fortunately, the inter-layering of hard and soft rocks has allowed some geomorphological description of the geology to remain. Other areas may either have a stronger or weaker relationship between the topography and the underlying geology. The relative importance of DEMs and their products against remotely sensed data will therefore vary depending on local conditions.

Further conditions specific to Llyn Cowlyd are the type of geological structure and the frequency of inter-layering of different rock units. Although the geological structure in Llyn Cowlyd is intruded by dolerite sills, these sills occur as layers similar to sedimentary rocks, and the resulting structure is reasonably simple. Some areas could however contain more complex structures, such as nappes, chevron folding, heavy faulting, and the intrusions of dykes or batholiths. Metamorphic structures can be of greater complexity (Hobbs et al. 1976) due to recrystallization and extreme distortions possible in such rocks. Study areas with more complex structures may prove more difficult to interpret without a higher resolution DEM, less vegetation cover and/or greater spectral separability between rock units. The frequency of inter-layering of rock units also has implications for the scale of data required, as do specific objectives of the mapping project. The 5 metre resolution chosen in this study allows major geological units to be mapped. Further investigations are required to evaluate dependency of scale
Figure 5.38 - Diagram illustrating how the position of edges may shift due to the illumination angle.
upon the types of geology and geological structures that can be mapped using techniques described in this chapter.

5.5 Summary

This chapter has discussed a number of ways in which geological information may be derived from digital elevation data. A number of products can be derived from DEMs to enhance geological features in the DEM. The most successful of these products are the slope, change of slope, and Lambertian shaded images, both in terms of automated and manual mapping. As in Chapter 4, there is a tremendous amount of information which can be produced using the techniques. Figure 5.39 shows all of the primitives identified during the evaluation procedures in Chapters 4 and 5. Although many primitives are identified by all techniques each technique adds extra data which could provide valuable information when creating a model of the geology. Therefore, rather than simply using the most successful technique, it is concluded that all potentially useful techniques should be used and that further processing techniques be developed to incorporate this data into a sensible model of the geology.

As with remotely sensed data, problems have been encountered with the accuracy of the elevation data. Inaccuracies and artefacts introduced by each interpolation procedure used have been noticed in the DEM. These could be overcome by using alternative DEM creation techniques, such as contour matching or stereo matching. Furthermore, a lack of fine detail in the DEM is apparent, particularly in the Lambertian shaded images. This may be overcome using the above improved creation technique or simply by obtaining a DEM of increased resolution.

Despite these problems, elevation data can provide a considerable amount of useful geological information. It also adds a further dimension to remotely sensed images and aids interpretation of such data. The third dimension also means that each point in a primitive can be described in terms of three dimensions rather than two. This enables an estimation of dip and strike values for each primitive, as will be described in the next chapter.
Figure 5.39 - All automatically identified primitives.
Work described in chapters 4 and 5 has demonstrated that geological primitives can be derived from a combination of remotely sensed images and digital elevation data. This chapter discusses how these primitives can be used to create a geological structural model. Structural values are calculated for each identified primitive and are subsequently assessed to determine whether the primitive represents a geological feature. All of the geological features are then combined within a structural model.

It has already been noted that the methods used to identify geological primitives also identify non-geological features, such as lake and forest boundaries. Several methods have been proposed to eliminate these unwanted features; for example, the results of multispectral classification and textural analysis of the elevation data can be used to distinguish likely geological features. Additional methods can be incorporated which attempt to identify specific objects within the scene (such as lake boundaries) and to assess the likelihood that a primitive demarcates a geological feature. Such methods are termed knowledge-based rules as they incorporate interpretative 'knowledge' into the decision making process (Rich 1983, Harmon and King 1985, Mirzai 1990). Each of these rules may be designed using a binary approach, where a threshold is used to provide a 'yes' or 'no' answer. Alternatively, a fuzzy logic approach can be used (Zadeh 1965, Forsyth 1989) where each measure is combined to produce a confidence statistic relating to the likelihood that a feature is geological. Both approaches are discussed in detail (using a lake boundary identification rule as an example) and each rule is evaluated against a manual interpretation of the features and by comparison with the topographic base map.

It was also noted in the previous chapter that, for certain primitives, it is insufficient to assess geological likelihood, and that greater reliability of results can be achieved by determining whether the data 'fit' the geological structure of the surrounding area. Thus, it is important to determine the structure (dip and strike) of each primitive. These values may be calculated using height values for each point comprising the primitive. A
novel least-squares approach is introduced to estimate the dip and strike for both planar and curved surfaces. The accuracy of these measurements is compared to those obtained from the corresponding geological map and also to measurements taken in the field. Two techniques are used to assess the likely errors inherent in these estimates:-

- through simulation against a known surface and structure (i.e., a plane cutting through a hemisphere)
- by using error estimates derived from the geometric correction of the remotely sensed images and spot height analysis used to assess the accuracy of the DEM.

To produce a structural model of the study area only those primitives considered most likely to be geological are included. This involves the application of a threshold to the confidence values assigned to each primitive. As the value is a relative figure, the threshold must be set subjectively and may therefore not be optimal. The threshold may require alteration to include more or less structural measurements in the model. Primitives are stored in a database in conjunction with associated attributes referring to the results of the knowledge-based rules and structural measurements. In this way individual measurements may be incorporated into the model by a simple inquiry to the database.

A geological model of the study area can be produced from individual dip and strike values. The model may be created either manually, by interpreting each measurement laid out on a map, or by using semi-automated techniques to build a model recursively. Automated techniques begin by finding the most common dip and strike value and assume a linearly dipping structure over the entire study area. This initial model is then improved recursively by adding further dip and strike values and manually interpreted fault planes into the model.

6.1 Deriving structural measurements

Primitives derived from edge detection techniques or from manual digitization of geological phenomena represent features which lie on a 3D surface. Although each feature is identified in the 2D plane of an image or DEM, the third dimension is derived from the associated elevation of each pixel in the feature (Figure 6.1). If the primitive represents a lithological boundary or a fault, points along it may be used to calculate dip and strike of the bed or fault. At this stage dip and strike values are calculated for each primitive regardless of whether the primitive represents a geological feature. Selected
Height of each vertex in the primitive

**Figure 6.1 - Planar surface fit to 3D primitive data.**
accuracy estimates are used as indicators to geological likelihood. These are therefore calculated for each primitive and stored in a database.

Several methods exist to derive structural measurements both manually and automatically from remotely sensed images and DEMs. This section seeks to explore the advantages and disadvantages of previous methods, notably the three-point method (McGuffie et al. 1989, Berger et al. 1992) and to introduce a new technique, based on a least-squares approach, developed as part of this study. Several techniques have also been designed to test the accuracy of this new method, and to evaluate the likely errors which might occur due to inaccuracies inherent in the input data.

6.1.1 Estimating dip and strike

Previous studies (McGuffie et al. 1989, Gamsjäger 1991, Berger et al. 1992) employ a three-point approach to determine dip and strike values from satellite imagery or aerial photography. This method draws upon the fact that any three points define a plane, so that any three points on a lithological boundary can be used to calculate dip and strike for that plane (Figure 6.2). In all three of these studies the points are identified manually; although Chorowicz et al. (1991) have identified points automatically for the purpose of deriving dip and strike information from a digitized geological map. Berger et al. (1992) and Chorowicz et al. (1991) conclude that the most accurate structural measurements are obtained by identifying one of the three points either significantly up or down dip of the other two (Figure 6.3). Unfortunately, it is not always possible to identify three points on a lithological boundary in this idealized way, as boundaries often follow reasonably straight lines along a slope rather than across a ridge. In these cases, greater success may be achieved by identifying as many points along the boundary as possible to define the plane more accurately. A second disadvantage of the three-point method is that it does not provide any indication of accuracy. Therefore, if one point is positioned slightly off target, the user will have no indication of the possible error introduced. Finally, all three points must lie on a linearly dipping segment of the structure; if a fold exists, three points are not sufficient to describe its structure.

The method used here facilitates identification of geological features both manually and automatically using a number of points to define the lithological boundary. A least-squares approach is used to fit the data to a surface. This also allows a measure of the relative accuracy of each point in the fit and an overall measure of the 'goodness of fit' of the surface using the coefficient of determination ($R^2$). A significant advantage of this method over the three-point approach is that the fitted surface may be either a planar or curved surface and can therefore be used to identify folded geological features.
Figure 6.2 - Figure showing how three points may be used to define a plane, which may in turn be used to extract dip and strike values.
Examples of three point sets which give accurate dip and strike values.

Examples of three point sets which give poor dip and strike values.

Figure 6.3 - Expected accuracy according to the relative location of three points.
A planar surface has an equation of the form:

\[ z = ax + \beta y + \chi \]  

where \( x, y \) and \( z \) are the 3D co-ordinates of each point along the feature and \( a, \beta, \) and \( \chi \) are the coefficients estimated from the fit.

A curved surface may be defined by a number of different equations including sine, log, and exponential terms. Here a simple second-order polynomial is used to describe surfaces over a local area. Any higher-order polynomials produce complex surfaces which are difficult to justify when fitting one 3D line\(^1\). Such surfaces can also become unstable away, and between, the known points, if there is a significant distance between them. The second-order polynomial surface has an equation of the form:

\[ z = ax^2 + \beta y^2 + \chi xy + \delta x + \epsilon y + \phi \]  

where \( a, \beta, \chi, \delta, \epsilon, \) and \( \phi \) are the estimated coefficients.

A minimum of three points is required to solve the simultaneous equations describing the planar surface (with a minimum of six points for the curved surface) although more points help to provide a more complete definition of the feature and allow calculation of residuals and \( R^2 \) values. Once the equations have been determined, it is an elementary process to calculate dip and strike measurements for the feature in question (see Appendix A). For the planar fit, the dip and strike measurement is valid for each point along the feature. However, a separate dip and strike estimate must be derived for each point on a curved surface.

A major difficulty with these surface-fitting techniques (and similarly for the three-point method) is that if all data points lie along a straight line in 3D space then any number of planes fit the data, where the line belongs to the plane (Figure 6.4). The resulting dip and strike measurement would therefore depend on minor perturbations in the internal precision of the CPU used to calculate the fit. One test to determine whether this is happening is to investigate and compare the coefficient of determination (\( R^2 \)) derived from the planar fit with that derived from a linear fit of the data in a new co-ordinate system defined by the planar surface. If the linear \( R^2 \) is \( \geq \) the planar \( R^2 \), then the edge should not be used for the calculation of dip and strike, as it is more likely to define a line than a plane. A number of thresholds may also be defined to reject those lines that have a low planar \( R^2 \) or low curved surface \( R^2 \), as well as those that have a high linear

---

\(^1\) Such surfaces may be of greater use when fitting a number of geological features to a surface, when the complex surface may describe a number of folds in a sequence.
Figure 6.4 - Diagram illustrating the arbitrary fit through a linear feature
A high linear fit may indicate that the geological feature in question does not describe the structure particularly well at that point, or may indicate that the feature is not geological at all. Rather, it may represent some anthropogenic feature, such as a road, a field boundary or a forest boundary.

6.1.2 Implementation of dip and strike estimation

Dip and strike values are estimated for each primitive whether derived manually or automatically. Three least-squares fits are performed on each feature, a linear, a planar, and a curved surface fit, and three $R^2$ values are calculated for each fit. The relative value of these $R^2$ values determines whether the primitive is considered to represent a line, plane, or curved surface. If the linear $R^2$ value is greater than both the planar $R^2$ and the curved surface $R^2$ then the primitive is discarded, as the feature is unlikely to provide a useful geological measurement. Due to the nature of least-squares fits the curved surface equations will always give a value for $R^2$ which is higher than, or equal to, the value for the planar fit. Curved surfaces are also more difficult to include within a geological model as separate dip and strike values must be determined for each point on the primitive. Therefore, the $R^2$ value derived from the curved surface fit must be significantly higher than the planar $R^2$ to justify use of a curved surface for a particular feature. Initially, a primitive is only considered to represent a curved surface if the curved surface $R^2$ is at least 0.1 greater than the planar $R^2$. This figure is set somewhat arbitrarily, but in practice provides a reasonable separation between curved and planar surfaces. The figure may be modified to include more or less curved surfaces, if necessary, during creation of a structural model, as all results are stored in a database.

As no definitive decisions have yet been made as to which primitives represent geological features, all results are stored in a database for later manipulation during creation of a structural model. The database is a simple flat file containing co-ordinates of each primitive and a number of associated attributes. At this stage the basic structure of the database will be described, but further attributes are added later, relating to knowledge-based rules applied to each primitive (section 6.2.1). Table 6.1 indicates the database structure designed for each feature identified either manually or automatically from images or elevation data. Each structure contains compulsory fields associated with the feature, i.e., a unique identification code, a colour code (used for displaying results according to the required attribute (e.g., Figure 6.5) or combination of attributes), and map co-ordinates for each point on the feature. Each entry also holds certain optional attributes derived for each feature e.g., dip and strike values. DIP and STK attributes are determined from the planar fit, irrespective of whether CR2 is significantly greater than PR2, as dip and strike for curved surfaces are only determined during creation of the model. The ORG field is used to describe the originating image.
Figure 6.5 - Colour coding of individual primitives for the display of attribute information. Values increase from magenta, through blue, green, yellow, to red. Here the colours indicate increasing $R^2$ values from 0.34 (magenta) to 0.99 (red).
within which the feature was identified. This becomes important when comparing features identified from different sources.

<table>
<thead>
<tr>
<th>Code / variable type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td>Unique identification code.</td>
</tr>
<tr>
<td>n2</td>
<td>Colour code.</td>
</tr>
<tr>
<td>n3</td>
<td>Number of points describing the feature.</td>
</tr>
<tr>
<td>x_i y_i ... x_n3 y_n3</td>
<td>X and Y map co-ordinates for each point in the feature.</td>
</tr>
<tr>
<td>@ DIP x</td>
<td>Dip value.</td>
</tr>
<tr>
<td>@ STK x</td>
<td>Strike value.</td>
</tr>
<tr>
<td>@ CSF x_a x_b x_c x_d x_e x_f</td>
<td>Coefficients describing the curved surface fitted to the data.</td>
</tr>
<tr>
<td>@ LR2 x</td>
<td>R^2 for the linear fit.</td>
</tr>
<tr>
<td>@ PR2 x</td>
<td>R^2 for the planar fit.</td>
</tr>
<tr>
<td>@ CR2 x</td>
<td>R^2 for the curved surface fit.</td>
</tr>
<tr>
<td>@ ORG string</td>
<td>The originating image, e.g., IMAGE, DEM, SLOPE, ASPECT.</td>
</tr>
</tbody>
</table>

Table 6.1 - The database structure for the storage of primitive features

The table also shows the ASCII notation used to format the primitive data. Each primitive is stored on a separate line and always contains data up to "x_n3 y_n3". Each optional attribute data is preceded by a @ symbol followed by a three letter code indicating the type of attribute information to be described (e.g., LR2 indicates a value for the linear coefficient of determination is to follow). The number of values or text items which follow the code is dependent on the type of attribute indicated. In this way any number of attributes may be added to a particular feature as and when required, and in any order.

For manual interpretation of images and DEMs it is important for the geologist to be able to view results graphically, not only to assist interpretation of the results themselves, but also to aid current and future interpretations of the scene. An interactive display and interpretation software package has been written, as part of this study\(^2\), which can be used to display the primitive features and any associated attributes. Figure 6.6 depicts the type of display available to the user. Each primitive is shown as a vector overlay on top of the image. Dip and strike values have been calculated for each

---

\(^2\) This package, called tracer, includes a number of edit and display features, a summary of which is given in Chapter 7.
Figure 6.6 - A typical interactive display showing two digitized primitives with dip and strike values overlaid. In the background is a second window illustrating some of the attributes associated with one of the primitives.
primitive and are shown as standard dip and strike symbols attached to the centre location of the primitive. Associated attributes for individual primitives can be displayed in a separate window. This allows the user to assess attributes such as planar $R^2$ and therefore determine the relevance of individual features to the overall structural model.

6.1.3 Evaluation of dip and strike results

The results of dip and strike estimation have been compared to similar measurements taken from the geological map and in the field. It should be noted that many more dip and strike values are identified from remotely sensed and elevation data than are available from the map and from the field survey. Therefore, each estimate cannot be evaluated individually. Instead, as the structure is constant in some local areas, a visual evaluation of results must suffice.

Figure 6.7 shows the published geological map for the study area. No dip and strike measurements are shown for this particular part of the map, but an idea of the structure is gained by the fold axes, several faults and one dip and strike measurement just outside the area (about 300m east of the north-east corner of the study area). The mapped structure clearly indicates some folding in the area but the degree and type of folding is not apparent. The results of the field survey provide a more complete picture of the geological structure than the published map (Figure 6.8). The field survey again shows the evidence of some strong folding (at A), but this is relatively local compared to the overall linearly dipping sequence/gentle folding of the remainder of the area.

Figure 6.9 through to 6.19 provide results derived from manual and automatic interpretations of several of the products derived from the remotely sensed and digital elevation data. From an initial study of these results it is unclear whether the measurements indicate the dip and strike of bedding or of a fault, or indeed whether they represent a geological feature at all. When related to the geological map, some of the results show a very good comparison. Despite this, there are many questionable results, due to:-

- non-geological features,
- noise in the data,
- the problem of fitting planar surfaces to straight lines or curved surfaces, and
- the fact that, in some cases, edges may cross geological boundaries.
Figure 6.7 - Published 1:50,000 geological map (BGS 1985)
Figure 6.8 - Dip and strike measurements taken in the field.
For results derived directly from the DEM and slope image, the most accurate results (Figures 6.9 and 6.10) occur where the steepest slopes cross a ridge and therefore describe an arc on the Earth's surface to which a plane may be readily fitted (e.g., in the south and central parts of the area). Surprisingly, all of the steepest slopes detected on the eastern slopes of Llyn Cowlyd (where geological features are most visually apparent in Figure 6.20) produce dips in the opposite direction to those taken from the map. This is partly due to the fact that these lines have a fairly high linear $R^2$ value and partly due to problems of simple thresholding. In this part of the DEM the slopes approach vertical cliffs in places, and as a result the steepest slopes may cross many geological boundaries. Here, it would be beneficial to know the spectral properties of the lithology that the edge represents. This knowledge could then aid a line-following process designed specifically to avoid crossing geological boundaries.

Edges derived from the aspect image (Figure 6.11) tend to enclose surface features (e.g., rock outcrops) and as such do not necessarily follow lithological boundaries. For instance, in this area outcrops of dolerite sills frequently occur at or near the top of ridges; here only one edge of the outcrop, usually the lowest, describes a geological boundary, while the others peter out into a bordering softer rock unit or into vegetation. As a result, very few of the dip and strike measurements derived from the aspect image correspond to the geological map. Nevertheless, as with the slope image, the aspect image contains a considerable amount of potentially useful information which has not been fully extracted using the edge detection techniques chosen here. The aspect edges may require segmentation into constituent parts before more accurate analyses can be undertaken.

From Figures 6.12 and 6.13 it is apparent that different features may be identified by altering the illumination angles. In addition to the accurate results in the south of the area derived from the slope image, the Lambertian shaded images give accurate results to the north and east of the centre where the changes in slope are more subtle, but are enhanced by the shading technique.

Many of the results obtained from the remotely sensed images and derived products correspond to those derived from the DEM image, due to natural scene shading of the topography. Additional features, including man-made objects, are caused by spectral changes in the image. The results of two individual bands, 7 and 9, are shown in Figures 6.14 and 6.15. Both bands are accurate in the south and east but also to some degree in the region of Creigiau Gleison (273000E 361500N to 273200E 361900N) where the DEM products failed. This is due to the spectral differences between rock units and vegetation particularly in band 9.
Figure 6.9 - Dip and strike results derived automatically from the DEM.
Figure 6.10 - Dip and strike results derived automatically from the slope data.
Figure 6.11 - Dip and strike results derived automatically from the aspect data.
Figure 6.12 - Dip and strike results derived from the Lamb-45 shaded DEM.
Figure 6.13 - Dip and strike results derived from the Lamb-90 shaded DEM.
Figure 6.14 - Dip and strike results derived automatically from band 7.
Figure 6.15 - Dip and strike results derived automatically from band 9.
The remaining automated results shown here are derived from the vegetation index and the first principal component (Figures 6.16 and 6.17 respectively). They have produced similar, though not identical, results to bands 7 and 9. Each of the illustrated results has contributed new and accurate dip and strike values to the overall picture of the geological structure. Both therefore have a valuable contribution to make in the production of structural data.

The success of the results in the south and central parts of the study area are probably due to the nature of the rock outcrops in that area. The harder rocks tend to cause quite distinct changes in slope and the lithological boundaries form arcs where they are exposed at the surface. This compares to the slopes to the east of Llyn Cowlyd where the extremely steep slopes and heavy shadow make it difficult to identify individual boundaries (Figure 6.20). The obvious effects of glaciation in this valley have resulted in the soft and hard rocks being indistinguishable in terms of topographic expression.

Figures 6.18 and 6.19 illustrate examples of results from the manual identification of features. The first point of note is that the primitives identified manually have a greater extent than the automatically identified features. This is due to the fact that the human eye finds it easy to extrapolate features and join different features which appear to belong to the same unit. This is most noticeable along the main ridge where the automated techniques fail to follow features across the ridge. These are easily identified manually and most of the resulting dip and strike values are accurate to within 2-3°. Even with manually identified primitives errors occur on the steepest slopes, probably due to the amount of error which can be introduced both by a vertex being one pixel out and the lack of detail in the DEM.

A further problem with results derived from the spectral data could be due to the misregistration of ATM images with respect to the DEM. The average spatial error between the two data sets is approximately 9.5 meters, which can precipitate a large error in the 3D co-ordinates produced for each line and, hence, in the final calculation of dip and strike. The following section will evaluate ways of quantifying how the geometric errors affect the dip and strike results.

6.1.4 Error Analysis of Dip and Strike

A key step in producing a structural model is to derive estimates of the error in dip and strike for each primitive. This section will discuss ways in which dip and strike estimation techniques may be evaluated, in addition to a method which can be used to estimate the error introduced through uncertainty in the input data. It may not always be
Figure 6.16 - Dip and strike results derived from the vegetation index.
Figure 6.17 - Dip and strike results derived from the first PC.
Figure 6.18 - Dip and strike results derived manually from the slope data.
Figure 6.19 - Dip and strike results derived manually from the best ITI.
Figure 6.20 - Perspective view of the eastern slopes of Llyn Cowlyd
possible to gain accuracy estimates of the input data, so that a measure is required which can indicate the expected error.

Berger et al. (1992) have performed error analysis on their manually derived stereoscopic measurements from SPOT images and aerial photography. They estimated likely errors in their measurements using the length (along dip) between their three chosen points and elevation accuracy of the stereo data. This analysis shows (Figure 6.21) the relationship between slope length, elevation accuracy, and the required accuracy of dip. Unfortunately, Berger et al. (1992) have used idealized circumstances to derive their error estimates, in so far as they assume that the three points are accurate and that the slope length is equal on both sides of the central point. However, a fundamental observation that can be drawn from their work is the relationship between the angle made by the three points on the plane and the accuracy of the dip and strike measurement. With the addition of more points describing the feature, the conclusion of the work by Berger et al. (1992) suggests that the accuracy is related to the curvature of the line. The following paragraphs test the validity of the curvature measure and investigate inaccuracies introduced by simulated errors in the input data and line detection techniques.

In this study, a novel set of techniques has been developed to ascertain the accuracy of the methods used to derive dip and strike. These operate by drawing features onto a hemispherical DEM (Figure 6.22) and determining their orientations. The orientation values are determined using the same techniques employed with the remotely sensed and elevation data of Llyn Cowlyd (section 4.2.4). An advantage of using a hemispherical DEM is that a plane of known orientation can be plotted precisely onto the hemisphere (Figure 6.23). Thus it is possible to simulate an exposed geological unit at the surface of a DEM. A disadvantage is that the hemisphere has a constant curvature and, therefore, the results produced by Berger et al. (1992) can only be tested by varying the length of the lines crossing the hemisphere; e.g., a quarter circumference as opposed to a half circumference. Alternatively, the hemisphere can be modified mathematically to resemble a cone or pancake shape (Figure 6.24). A series of experiments has been designed to test the accuracy of dip and strike measurements using these shapes. The parameters that are varied, the range of values chosen and the objective of each test is given in Table 6.2.
Figure 6.21 - A two step procedure used to evaluate the accuracy in dip measurements obtained from two different study areas using stereo SPOT (after Berger et al. 1992).
Figure 6.22 - Hemispherical DEM - Lambertian shaded.

Figure 6.23 - Simulated geological unit overlaying the DEM.
Figure 6.24, Modifications to the hemispherical DEM.
Table 6.2 - Parameters used to test the accuracy of the techniques to estimate dip and strike

<table>
<thead>
<tr>
<th>Test parameter</th>
<th>Range of values</th>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dip angle.</td>
<td>dip 0° - 90°</td>
<td>To evaluate increase/decrease in error as a function of dip angle.</td>
</tr>
<tr>
<td>Random noise added to heights in DEM.</td>
<td>0-50% of the radius of the hemisphere</td>
<td>To investigate the effect of errors in the elevation data (Figure 6.25).</td>
</tr>
<tr>
<td>Random noise added to positional accuracy of arc.</td>
<td>0-30° from known orientation</td>
<td>To test the effect of errors in the line detection procedure (Figure 6.26).</td>
</tr>
<tr>
<td>Length of arc.</td>
<td>5-50% of the circumference</td>
<td>To simulate the Berger et al.'s (1992) test by measuring the maximum length down slope between points on the primitive.</td>
</tr>
<tr>
<td>Random sampling of points on arc.</td>
<td>3 pixels to all pixels</td>
<td>To judge the number of points required to give accurate dip and strike values</td>
</tr>
<tr>
<td>Shape of the hemisphere.</td>
<td>cone like to pancake like</td>
<td>To analyse the effects of changing curvature on the accuracies produced⁴.</td>
</tr>
</tbody>
</table>

The results of each of these tests are evaluated by comparing the planar $R^2$ value with the accuracy obtained. The aim is to use the $R^2$ value to predict likely errors in the real primitives. Following the ideas behind Berger et al.'s (1992) work, accuracy is also compared with a measure of curvature of the primitive. To calculate the curvature of a primitive, the co-ordinates are re-projected into a new space defined by the fitted plane. A second-order polynomial in $x$ is then used to fit this new data (Figure 6.27). The $x^2$ term of this polynomial is subsequently used as a measure of curvature.

Since copious data are produced by such experiments, due to the large number of variables, it is impractical to show every result here. Therefore only the most important points are summarized here (Table 6.3).

---

⁴ This is achieved using the following formula for height as a function of radius

$$h = r \times \sin(dip)^n$$

where, $h$ is the modified height at a distance $\sin(dip)$ from the centre of the hemisphere, $r$ is the radius of the hemisphere, and $n$ varies between 0.1 (pancake shaped), through 1.0 (hemisphere), to 10.0 (cone shaped). See Figure 6.24.
Random noise - 10% of radius.

Random noise - 30% of radius.

Figure 6.25, Addition of random noise to the hemispherical DEM.
Random noise - 2% of radius.

Random noise - 10% of radius.

Figure 6.26, Addition of random noise to positioning of simulated primitive.
Reprojection into new plane defined by the fitted surface

Figure 6.27 - Calculation of curvature, by fitting a second-order polynomial to the reprojected primitive.
<table>
<thead>
<tr>
<th>Test</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dip angle.</td>
<td>There is no systematic change in the error introduced at different dip angles (Figure 6.28), although there is a possible increase in error for features having a dip of less than 15°.</td>
</tr>
<tr>
<td>Shape of the DEM.</td>
<td>Varying the shape of the DEM allows an assessment of features having quite different curvatures (Figure 6.29). Figure 6.30 shows that the accuracy of derived dip and strike measurements is strongly related to the curvature of the line. Most measurements are more inaccurate for the pancake shaped DEMs, particularly for the low dipping features, where the error reaches more than 4.5°. Figure 6.31 depicts the errors produced for a dip angle of 45° and their relationship with curvature and the $R^2$ of the planar fit. Although those features exhibiting a high curvature do have a low error and some of the lower curvature features have high error, it can be seen that in this case it is not necessarily a direct proportional relationship. The $R^2$ values, on the other hand, show an inverse relationship to the error in dip.</td>
</tr>
<tr>
<td>Random noise in DEM.</td>
<td>Errors in dip increase in proportion to the error in the DEM (Figures 6.32 and 6.33). The error in dip is higher for the lower curvature DEM (Figure 6.33) and the error is increased at a lower level, i.e., 10 pixels rather than 20 - 30 pixels for the cone-shaped DEM.</td>
</tr>
<tr>
<td>Length of arc.</td>
<td>Errors are minimal in most cases but increase considerably at shorter lengths (Figure 6.34).</td>
</tr>
<tr>
<td>Positional accuracy of points on the arc.</td>
<td>As expected, errors in dip are proportional to the positional accuracy.</td>
</tr>
<tr>
<td>Random sampling of points on the arc.</td>
<td>For a smooth hemisphere, no significant errors are introduced by reducing the number of points on the arc. However, as additional errors are introduced into the DEM or the shape of the DEM is varied, greater errors are produced with fewer points. Therefore, in a real terrain, with errors in the height and positional accuracy of the DEM, it is important to find as many points along the geological feature as possible.</td>
</tr>
</tbody>
</table>

Table 6.3 - Results of the evaluation of dip and strike estimation techniques

Although the measure of curvature of a feature, suggested by the work of Berger et al. (1992), can be a useful indicator of dip accuracy, the relationship is not direct in all cases. Other factors are important in assessing the accuracy of dip and strike.
Figure 6.28 - Dip v. Dip Accuracy

Legend
- Curvature (*1000)
- R2 of Planar Fit (*4)
- Error in Dip Estimation
Figure 6.29 - Curvature v. DEM Shape Factor at Different Dip Angles
Figure 6.30 - Dip Accuracy v. DEM Shape Factor at Different Dip Angles

Error in Dip Estimation

log DEM Shape Factor (Broad to Peak)

Legend
- Dip = 80
- Dip = 70
- Dip = 60
- Dip = 50
- Dip = 40
- Dip = 30
- Dip = 20
- Dip = 10
Figure 6.32 - Error in DEM v. Dip Accuracy (peaked hemisphere - D/S 60/135)

Legend
- Curvature (*1000)
- R2 of Planar Fit (*4)
- Error in Dip Estimation

Error in DEM (pixels), where radius of hemisphere = 250 pixels
Figure 6.33 - Error in DEM v. Dip Accuracy (hemisphere - D/S 60/135)

Error in DEM (pixels), where radius of hemisphere = 250 pixels

Legend
- Curvature (*1000)
- R2 of Planar Fit (*4)
- Error in Dip Estimation
Figure 6.34 - Length of arc v. Dip Accuracy (D/S 45/135) - lower end at 0.3
measurements, such as the $R^2$ value, the length of the line, the number of points on the line, and the error in x-y-z co-ordinates of the feature.

Having discussed methods of predicting likely errors in estimation of dip and strike, it is also necessary to examine any errors introduced by inaccuracies in the input data, i.e., remotely sensed images and DEMs. For this study, estimates of error in each input data set have been obtained during geometric correction of the image and spot height evaluation of the DEM. These are point estimates, but may be interpolated over the study area to give an estimate for each pixel. Figure 6.35 shows perspective views of these interpolated data for the remotely sensed images and DEM. Interpolations of this type offer a measure of the inaccuracy of the location for each pixel in the X, Y, and Z directions. This can be used to give an estimate of the inaccuracy in calculation of dip and strike for each feature. This is done by taking, in turn, the maximum error for each direction (i.e., X, Y, and/or Z) or combination of directions (i.e., X & Y, X & Z, Y & Z, and X & Y & Z) and calculating the new dip and strike. The range of results then indicates the accuracy of the original dip and strike measurement. The maximum error from each of these tests is taken to indicate the possible error inherent in the dip and strike estimate. Table 6.4 shows an example of this procedure performed on an edge derived from the slope image. The errors introduced by the input data are typically small and are well within standard requirements. For example, Berger et al. (1992) quote a required accuracy of 2-3° for a typical oil exploration application.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Dip</th>
<th>Strike</th>
<th>Pole(^4) Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original measurement</td>
<td>41.4153°</td>
<td>46.0097°</td>
<td>0.0°</td>
</tr>
<tr>
<td>Use max. X error</td>
<td>0.0033°</td>
<td>0.1188°</td>
<td>0.0892°</td>
</tr>
<tr>
<td>Use max. Y error</td>
<td>0.0132°</td>
<td>0.0979°</td>
<td>0.0746°</td>
</tr>
<tr>
<td>Use max. Z error</td>
<td>0.0724°</td>
<td>0.2611°</td>
<td>0.2089°</td>
</tr>
<tr>
<td>Use max. X &amp; Y error</td>
<td>0.0098°</td>
<td>0.0209°</td>
<td>0.0185°</td>
</tr>
<tr>
<td>Use max. X &amp; Z error</td>
<td>0.0755°</td>
<td>0.3794°</td>
<td>0.2945°</td>
</tr>
<tr>
<td>Use max. Y &amp; Z error</td>
<td>0.0594°</td>
<td>0.1626°</td>
<td>0.1357°</td>
</tr>
<tr>
<td>Use max. X, Y, &amp; Z error</td>
<td>0.0625°</td>
<td>0.2809°</td>
<td>0.2198°</td>
</tr>
<tr>
<td>Maximum error</td>
<td>±0.0755°</td>
<td>±0.3794°</td>
<td>±0.2945°</td>
</tr>
</tbody>
</table>

Table 6.4 - Example of error estimation for individual dip and strike measurements

\(^4\) The pole is the direction of the normal to a plane.
Figure 6.35 - Perspective view of errors in X, Y (both derived from the geometric correction), and Z (derived from the spot height evaluation).
The main factors which affect accuracy of the derived dip and strike measurements are the measure of curvature, the planar $R^2$ value, the length of the feature, and the error in the co-ordinates of each feature. Each of these measures is assessed for every feature and stored in the database, for later analysis and accuracy assessment. This co-ordinate error estimate is calculated for each feature in the database and added to the attribute listing for each feature using the code `@ERR x` (see Table 6.1).

6.2 Methods for Evaluating the Geological Significance of Detected Primitive Features

The methods used to extract primitive features from remotely sensed images and digital elevation data, described in Chapters 4 and 5, result in a vast amount of data, of which only a percentage will relate to geological structural features. In order that these data can be used successfully in subsequent modelling of the geology, the non-geological primitives must be identified. Most edges derived from a DEM can be thought of as being geologically or geomorphologically relevant, unless any gross anthropogenic features are present in the scene, such as road and railway cuttings, dams, quarries or habitation. There are likely to be more non-geological features found in remotely sensed images due to the spectral content of these images which distinguish many different cover types; these may include many anthropogenic features, water bodies, and vegetation/agricultural boundaries in addition to the geological features of interest. Techniques are therefore required to identify features of this nature. Methods which perform the following automatic identifications are introduced and evaluated in this section:

- identification of lake primitives
- identification of forest boundary primitives
- identification of forest track primitives
- identification of rock exposure primitives

Primitives identified from a remotely sensed image which are coincident with primitives identified in a DEM are more likely to represent geological features than those which are not coincident. Such features might include lithological boundaries at a break of slope or faults along a valley. The likelihood that the feature is geological, is therefore increased. In addition to the object-specific algorithms mentioned above, each primitive is compared with primitives derived from other sources to identify coincident features.

---

5 Such features could be identified by comparison with digitized map features. However, areas of geological interest often occur in poorly mapped regions of the world. It is therefore desirable to develop generally applicable techniques.
Significant combinations occur between image primitives and those derived from DEM products, such as slope, change of slope, and Lambertian shaded images. Less important combinations occur between image primitives and aspect primitives, and primitives identified in a number of different image products.

Further analysis of each primitive can be performed by comparing the location of primitives against the segmentation and confidence images mentioned in Chapters 4 and 5. These include products of multispectral classification and geomorphological texture measure.

The above requirements and suggestions could all be performed manually, but the process would be time consuming, laborious, and prone to inconsistencies. These tasks are well suited to automation. Their automation requires the coding of interpretative skills into computer language. Such algorithms are termed knowledge-based rules as they incorporate expert knowledge of interpretation, geology, and remote sensing (Harmon and King 1985, McKeown et al. 1985, Tailor et al. 1986, Matsuyama 1987).

6.2.1 Representing Knowledge

There are five basic ways of representing knowledge (Harmon and King 1985, Edwards 1991), namely:

- Semantic networks
- Object-attribute-value triplets
- Rules
- Frames
- Logical expressions

A semantic network is a collection of nodes connected by links. Nodes may represent physical objects or conceptual entities, such as acts or abstract categories, or descriptors. Links relate objects and descriptors. For example, in the statement "Lake - has an - edge - which is - flat", 'has an' and 'which is' are links, while 'lake', 'edge', and 'flat' are all nodes. A series of hierarchical and interconnecting networks can be designed in this way to represent knowledge. The hierarchical nature of this method reduces redundancy but it is difficult to incorporate exceptions within this structure.

Object-attribute-value (O-A-V) triplets are specialized semantic networks. The network is replaced by a series of O-A-V relationships. Objects are the same as nodes in a semantic network and attributes are general characteristics or properties associated with objects. An O-A-V triplet describing the above statement could be - edge - height
difference - flat (or less than 5 metres). Objects may have any number of associated attributes and may also be stored in a hierarchical structure. One advantage of this method is the incorporation of certainty factors (CFs) which represents the confidence that exists in a piece of evidence. Typical values range between -1 and 1 (Buchanan and Shortliffe 1984). For example, if the edge is not perfectly flat and has a range of heights of less than 5 metres then the confidence of the edge being flat might be 0.8.

A rule can often take the form of:-

\[
\text{if ( condition )} \\
\quad \text{then ( a )} \\
\text{else} \\
\quad \text{then ( b )}
\]

where the condition can be an O-A-V triplet and \( a \) and \( b \) are conclusions drawn from the facts determined. Again CFs may be used to quantify a conclusion. CFs of related rules may be combined in a linear manner (Buchanan and Shortliffe 1984) or using weighting factors for each rule.

A frame, first proposed by Minsky (1975), is a description of an object containing slots for all information associated with the object. Slots may store values, pointers to other frames, sets of rules, or procedures by which values may be obtained. These facilities allow for a more sophisticated and functional representation of knowledge but are consequently more complex and difficult to develop.

There are two common forms of representing knowledge using logical expression; propositional logic and predicate calculus. Propositional logic uses statements which are either true or false, and which may be linked together with connectives, such as, AND, OR, NOT and IMPLIES. For instance, if \( X \) is true and \( Y \) is false, then \( X \) AND \( Y \) is false and \( X \) OR \( Y \) is true. Predicate calculus is simply an extension of propositional logic which provides statements about objects. For example, the statement \( \text{is-blue(sky)} \) asserts that the sky is blue, which again is either true or false (Harmon and King 1985). Although simple to implement, these techniques cannot allow uncertainty in their representation of knowledge and are therefore not totally suitable for the task in hand.

Any one of the five methods could be used to represent the interpretative knowledge that is required here. However, they differ in complexity and functionality. Knowledge-based rules allow a suitable degree of flexibility and are easy to implement. They are also one of the most well recognized forms of representing knowledge, and have been chosen for use in this study.
6.2.2 Implementation of binary and confidence based knowledge-based rules

Knowledge-based rules may be implemented using binary or fuzzy logic. Fuzzy logic uses the uncertainty measures indicated in the above definitions (Zadeh 1965, Graham and Jones 1988). Each implementation is evaluated using a rule, developed within this research, and designed to differentiate geological from non-geological primitives, e.g., a rule to identify lake primitives.

The lake rule employs several significant properties to discriminate this type of edge from any other:-

i) the edge has a constant elevation along its entire length,
ii) if the edge is complete it will form a closed shape, and
iii) one or both sides of the edge will have the distinctive spectral signature of water.

A single knowledge-based rule can be designed to incorporate these factors using a combination of different if-then-else statements. The following code written in C (Kernigian and Ritchie 1978) is an example of how the rules i) and iii) above may be specified, using a binary logic implementation.

(For simplicity of presentation and reading, a number of lines and features, necessary to make the code run, have been omitted.)

```c
for(n = 0; n < num_points; n++)
/* for each point in the primitive */
    {height = dem[primitive.x[n] * num_cols + primitive.y[n]];
     /* obtain the height value */
     if (height > height_max) height_max = height;
     if (height < height_min) height_min = height;
     /* find the max and min height values */
     y_diff = primitive.y[n + 1] - primitive.y[n];
     x_diff = primitive.x[n + 1] - primitive.x[n];
     /* calculate the local differences between the current point and the next */
     azimuth = acos(y_diff / (sqrt(x_diff*x_diff + y_diff*y_diff)));
     /* calculate the local direction between the two points */
     left = azimuth - 90.0;
     right = azimuth + 90.0;
     /* calculate directions to the left and right of the line */
    for(radius = 1; radius <=2; radius++)
```

247
/* repeat look for water at two different distances away from primitive */

    (new_x = primitive.x[n] + radius * sin(left);
    new_y = primitive.y[n] + radius * cos(left);
/* calculate co-ordinates for look points */

    reflectance3 = atm_band3[new_y * num_cols + new_x];
/* find the green reflectance value at look point */
    reflectance5 = atm_band5[new_y * num_cols + new_x];
/* find the red reflectance value at look point */
    reflectance7 = atm_band7[new_y * num_cols + new_x];
/* find the near infra-red reflectance value at look point */

    if (reflectance3 < water_threshold3
        && reflectance5 < water_threshold5
        && reflectance7 < water_threshold7)
left_water++;
/* if reflectance is below the threshold for water increment occurrence counter */

    new_x = primitive.x[n] + radius * sin(right);
    new_y = primitive.y[n] + radius * cos(right);
    reflectance3 = atm_band3[new_y * num_cols + new_x];
    reflectance5 = atm_band5[new_y * num_cols + new_x];
    reflectance7 = atm_band7[new_y * num_cols + new_x];
    if (reflectance3 < water_threshold3
        && reflectance5 < water_threshold5
        && reflectance7 < water_threshold7)
right_water++;
/* repeat this process for the right hand look */

    } /* end of loop */

if (height_max - height_min > level_threshold) lake = FALSE;
else lake = TRUE;
/* check if the range in height values is above a specified threshold */

if (left_water<num_water && right_water<num_water) lake=FALSE;
/* if the number of occurrences of water on both sides of the
primitive is less than the threshold then the primitive is not a lake edge */

/* if lake = TRUE then the primitive is a lake edge, else it is not */

A number of different thresholds, set prior to execution of the rule, have been
incorporated into the above code which result in a definitive answer to the question - *is
the primitive a lake edge?* The water_threshold thresholds are used to specify
reflectance values for a series of wavebands (in this case green, red, and near infra-red)
below which a pixel is deemed to represent water. Such thresholds can be derived from
a spectral library of common cover types or can be derived from any previous

---

6 The accuracy and effectiveness of these thresholds will be
dependent on a number of contributing factors, such as the
classifications of the image which include a water class. Alternatively, a classified image could be used directly to ascertain which pixels either side of the primitive are water. The *num_water* threshold is used to specify how many of the pixels to one side of the primitive should be water before the primitive is considered a lake edge. Ideally, it would be expected that every pixel would be water, but noise in the images and the nature of mixels (i.e., where a mixture of land cover types are represented in one pixel) may result in a number of the pixels being classified as non-water pixels. The *num_water* threshold therefore specifies the percentage of the neighbouring pixels that must be classified as water. Similarly, the *level_threshold* is used to allow for noise and errors in the DEM; e.g., height differences along a lake edge might be as much as 10 metres in this area due to steep inclinations either side of the reservoir. It can be seen from the description of these thresholds that their specification requires expert knowledge of the image formation processes (Muller 1988), in terms of surface reflectance properties and image noise, and knowledge of the creation of the DEM in terms of height errors. The results from a rule specified with thresholds in this way can be very sensitive to the threshold values set and different thresholds may prove better for certain areas of the image or for different images. For instance, a geological primitive positioned next to some waterlogged or boggy vegetation may or may not be identified as a lake edge with a slight change in one of the thresholds. It may therefore prove better to incorporate CFs into the rule so that confidence values may be assigned to each primitive. In this way such water logged features may be positioned somewhere between definite lake edges and non-lake edges and treated accordingly.

The alternative probabilistic\(^7\) approach to answering the question set above, uses certainty factors rather than thresholds, which results in a confidence value that each primitive is a lake edge. This method is less sensitive to small changes in the rule-base structure of any pre-defined thresholds. In the example of the lake rule, all thresholds are replaced by a certainty calculation. For example, the *level_threshold* value can be replaced by a measure of how level the primitive is (for example, standard deviation of height differences) and the reflectance thresholds can be replaced by the curve fitting routines used in spectral mixture modelling (Taranik and Kruse 1989, Drake 1990) which measure the similarity between image and library spectral signatures. Each measure may loosely be described as a probability measure since they specify the likelihood of each primitive representing a particular feature of a lake edge. These accuracy of the reflectances calculated for the remotely sensed image, the depth of the water, and the amount of suspended sediment in the water. The thresholds can be set to allow a certain degree of variability in the reflectance of water bodies. In this study area the water bodies are deep and clear and therefore cause few problems.

7 Here probability is not being used in its strict statistical sense but merely as an informal measure of confidence.
individual measures can then be combined to give an overall measure describing how well the primitive represents a lake edge.

The means by which these measures are combined depends greatly upon the individual significance of each of the measures, which again involves a certain degree of knowledge about a particular feature, in this case a lake edge. There are three different methods in general use which are used to combine such certainty or uncertainty measures (Graham and Jones 1988), including:-

- Bayesian representations
- certainty factors
- fuzzy sets

The Bayesian representation of rules uses a set of a priori probabilities to define the relevance of each measure. Often these probabilities are initialized as equal values and are then recursively modified to posterior probabilities given new evidence. The implementation of this approach therefore involves interaction with a user, which is not practical within this application, or a large amount of a priori data of knowledge about the subject. Barr (1990) describes how a number of remotely sensed and geophysical datasets are combined using this Bayesian approach to give the likelihood that a particular site would contain lead mineralization. Each of the individual datasets was assigned an a priori probability based on the previous discovery and mine working related to other lead mineralization in the area.

Certainty factors may be combined using a formula designed by Shortliffe (1976):

\[
MB(H:E1,E2) = MB(H:E1) + [MB(H:E2) \times (1-MB(H:E1))]
\]  

(6.3)

where MB is the measure of belief and H is the hypothesis given evidence E. This has certain advantages over the Bayesian approach in that it is symmetric with respect to the order with which evidence is gathered and it is cumulative asymptotically which, according to Graham and Jones (1988), conforms with the processes of intuition. The measure of belief in a piece of evidence is composed of two factors, the certainty with which the evidence has been measured (e.g., the flatness of the terrain), and the credibility of that piece of evidence in proving the hypothesis (e.g., if a primitive is flat then it is 0.7 certain that the primitive represents a lake). The product of these two factors, termed here as the certainty factor and the weighting factor, respectively, describes the measure of belief.
A fuzzy set is a mathematical approach which defines the likelihood that a piece of evidence is true (Zadeh, 1965, Graham and Jones 1988, Wu et al. 1988). For instance, Figure 6.36 shows a fuzzy description of the flatness of a primitive; if the standard deviation of height differences is zero then the certainty factor that the primitive is flat is 1.0, and similarly, if the standard deviation is above 20 metres then the primitive is not flat. The gradation between these two extremes describes the fuzzy set. Fuzzy sets may be combined using fuzzy set theory (Zimmerman 1986) which extends standard set theory into the domain of fuzzy logic and includes such components as products, unions, and intersections. Fuzzy sets and fuzzy logic can capture the richness of natural language and natural reasoning and can therefore be used to design both descriptive and adaptable rules.

Rules have been combined within this study using the methods defined by Shortliffe (1976) and fuzzy logic to define the certainty factor of each rule. Figure 6.37 shows the fuzzy sets for each of the four components (or pieces of evidence) defining the lake identification rule. A Gaussian curve is used to describe each set. Table 6.5 demonstrates how the resulting certainty factors may be combined to produce an overall confidence indicating the presence of a lake primitive.

<table>
<thead>
<tr>
<th>Lake parameter/attribute</th>
<th>Certainty factor</th>
<th>Weighting factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure of height differences</td>
<td>L = 0.6</td>
<td>( \omega_1 = 0.60 )</td>
</tr>
<tr>
<td>Likelihood of water spectra (Green)</td>
<td>R₁ = 0.4</td>
<td>( \omega_2 = 0.3 )</td>
</tr>
<tr>
<td>Likelihood of water spectra (Red)</td>
<td>R₂ = 0.5</td>
<td>( \omega_3 = 0.3 )</td>
</tr>
<tr>
<td>Likelihood of water spectra (Near Infrared)</td>
<td>R₃ = 0.9</td>
<td>( \omega_4 = 0.8 )</td>
</tr>
</tbody>
</table>

\[
MB\text{\ after 1st two pieces of evidence (MB1)} = L*\omega_1 + R_1*\omega_2 * (1-L*\omega_1) = 0.4368
\]

\[
MB\text{\ after 3rd piece of evidence (MB2)} = MB1 + R_2*\omega_3 * (1-MB1) = 0.5213
\]

\[
\text{Overall confidence after all evidence} = MB2 + R_3*\omega_4 * (1-MB2) = 0.866
\]

| Table 6.5 - The combination of certainty factors to produce an overall likelihood measure for the lake-edge knowledge-based rule |

Values of the weighting factors are designed to reflect the importance of each attribute. Reflectance of water in the near infrared is an excellent distinguishing feature, as water absorbs nearly all radiation at this wavelength. The flatness of lake features is also an important characteristic, while the reflectances in the green and red wavelengths are less so. Values are set by trial and error at present. In the future they could be incorporated...
Figure 6.36 - Fuzzy set description for the 'flatness' of a primitive.
Figure 6.37 - Gaussian fuzzy sets used in the lake-edge identification rule
in a learning process where an expert system recursively refines these factors until an optimum combination is found (Forsyth 1989).

Even with the use of the confidence approach, a threshold must be used at some stage during processing to finally decide whether a primitive should be rejected as non-geological or included within a model of the geology. A key issue is to decide at what stage this threshold should be applied. Results from the probabilistic version of the lake edge rule may be thresholded to remove all lake edges, or may be stored in the results database for future reference and manipulation. If the results remain in the database, the probabilities may be combined with results from other knowledge-based rules (such as road and forest identification rules) to give an overall measure of likelihood that the primitive represents a geological feature. In turn, this probability measure may be combined with accuracy measurements relating to the fit of a planar or curved surface to the primitive, and to measurements of how well the structural measurement fits into the current understanding or model of the geology. It is therefore theoretically possible for each primitive to be investigated fully (in terms of its relevance to the geological model) by continuously updating a probability measure associated with it. In practice, however, it is more realistic to make definite decisions and execute certain thresholds throughout the processing procedure. This ensures that the least likely geological primitives are rejected, saving computer processing time. Even with this thresholding the probability of, for instance, a primitive being a lake edge is still stored in the database so that if at any time the overall probability of the primitive being non-geological falls below a certain threshold then the primitive can be rejected.

Results from the threshold and confidence methods are summarized in Figures 6.38 - 6.40. Figure 6.38 depicts the relationship between the number of primitives identified as lakes and two attributes associated with each primitive, namely the levelness and the near-infrared DN. The graph shows how a slight change in a threshold can result in a large difference in the amount of lake edges identified. Each threshold also has a different sensitivity to the number of lake edges, making it difficult for the user to set appropriate threshold values. Conversely, the probabilistic approach allows the user to see graphically which primitives are most likely to be lake edges. The threshold can then be set simply and interactively by the user. Figures 6.39 and 6.40 demonstrate primitives identified as lake edges using the threshold and probabilistic approaches respectively (N.B. the colours in Figure 6.39 signify increasing attribute values from magenta through blue, green, yellow, to red). The Figures indicate that the probabilistic approach identifies lake edges with more success than the threshold approach, with the latter finding more 'lakes' away from the reservoir (e.g., primitives A, B, and C in Figure 6.39). A probabilistic approach can also be used in conjunction with other rules to aid
Figure 6.38 - No. of lake edges identified by varying two thresholds.
Figure 6.39 - Lake edges (in red) identified using a probabilistic approach.

Figure 6.40 - Lake edges removed using a threshold approach. NB primitives A, B, and C from Figure 6.39 have been removed erroneously.
identification of geological features. Therefore, this technique is used throughout the design of additional rules.

6.2.3 Design and implementation of further knowledge-based rules

Extreme care must be taken when designing and implementing rules like the lake-edge rule. They should be general rules which aim at universal application, and should not be made specific to a particular scene. This point can be illustrated with respect to identification of roads. In this area of Snowdonia, roads are generally low lying (following broad valleys and passes), are straight or slightly curved, have a gentle incline, are interconnected in a road network and are usually bordered by vegetation. However, in other more mountainous areas, roads may follow a more tortuous path obeying few of the rules set above. Similarly, in less hilly terrain roads might follow higher ground, and so forth. A road identification rule would therefore need to be adaptable to the type of terrain and the nature of the vegetation cover present. Both factors require a more extensive knowledge of a given area which may be gained from a DEM and from remotely sensed images. Further rules can be used to classify a DEM into terrain types (such as flat, rolling hills, hilly, or mountainous) and to classify a vegetation or climate type of an area (such as arid, semi-arid, temperate, savannah, forest, or tropical).

The preceding section described how knowledge-based rules may be designed and implemented, in particular for a lake-edge rule. Similar identification rules may be created for specific objects such as field boundaries, forest plantations, tracks, habitation or any other anthropogenic features. Further knowledge-based rules can be designed simply to judge the relevance of a primitive to the geological model rather than to identify a specific object. The simplest of these rules is applied to primitives derived from remotely sensed images and attempts to compare them with similar primitives derived from the DEM. If a spectral primitive coincides spatially with a primitive derived from the DEM then the likelihood that both primitives are geologically relevant is increased. The spatial location of each spectral primitive, which is stored in the results database, may be compared with each primitive derived from the DEM. The proportion of pixels or vectors along the primitive which overlap can be used to estimate the spatial similarity of the features.

The following tables provide an overview of all rules incorporated into this study and the methods used to implement them.
### Object Specific Rules

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>Features examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Identification of lake edges</td>
<td>a) height difference along a primitive - should lie on horizontal plane</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) spectral signature to one side of the primitive should be comparable to library</td>
</tr>
<tr>
<td></td>
<td></td>
<td>spectra of water in three bands - green (ATM band 3), red (ATM band 5), and near</td>
</tr>
<tr>
<td></td>
<td></td>
<td>infrared (ATM band 7)</td>
</tr>
<tr>
<td>1.2</td>
<td>Identification of forest boundaries</td>
<td>a) strength of vegetation edge - should be strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) edge should be linear or angular in planimetric view</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c) should be high textural inhomogeneity in vegetation index to one side of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>primitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d) boundaries can be steep</td>
</tr>
<tr>
<td>1.3</td>
<td>Identification of forest tracks</td>
<td>a) strength of vegetation edge - should be strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) edge should be linear, or sharp curves approaching 180° (i.e., hairpin bend)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c) should be high textural inhomogeneity in vegetation index to one side of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>primitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) boundaries can be steep</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d) should be spectral signature of rock or soil to one side of the primitive, in</td>
</tr>
<tr>
<td></td>
<td></td>
<td>three bands - blue (ATM band 2), near-infrared (ATM band 7), and thermal infrared</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ATM band 11)</td>
</tr>
<tr>
<td>1.4</td>
<td>Identification of rock primitives</td>
<td>a) classified rock pixels to one or both sides of the primitive</td>
</tr>
</tbody>
</table>

**Table 6.6 - Description of object specific knowledge based rules**

---

8 The thermal band is used simply as digital number rather than reflectance, due to the difficulties in calibrating this band (Wilson 1985).
II Validation rules

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>Features examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>II.1</td>
<td>Coincidence with slope</td>
<td>Primitives identified in spectral images are compared with those derived from the DEM as steepest slope primitives.</td>
</tr>
<tr>
<td>II.2</td>
<td>Coincidence with change of slope</td>
<td>Primitives identified in spectral images are compared with those derived from the DEM as greatest change of slope primitives.</td>
</tr>
<tr>
<td>II.3</td>
<td>Coincidence with change of slope direction</td>
<td>Primitives identified in spectral images are compared with those derived from the DEM as the greatest change in direction of slope.</td>
</tr>
<tr>
<td>II.4</td>
<td>Validation by vegetation index</td>
<td>The primitives are compared with a vegetation index image of the scene, and the geological relevance probability is modified according to the strength of vegetation surrounding the primitive.</td>
</tr>
<tr>
<td>II.5</td>
<td>Validation by texture of local terrain</td>
<td>The primitives are compared with a texture image derived from a series of Lambertian shaded images, which describes the roughness of the local terrain and therefore indicates the likelihood of geological features being present.</td>
</tr>
<tr>
<td>II.6</td>
<td>Accuracy assessment</td>
<td>The various indicators of dip and strike accuracy, including curvature, R², length, and x-y-z error, are combined to give an overall assessment of accuracy.</td>
</tr>
<tr>
<td>II.7</td>
<td>Comparison of dip and strike values within local area</td>
<td>Each dip and strike value should fit into the local structure. The pole is therefore compared with the mean pole of the local area and a confidence is assigned according to the number of standard deviations away from the mean.</td>
</tr>
</tbody>
</table>

Table 6.7 - Description of the validation knowledge based rules

Figures 6.41 to 6.46 indicate the results, derived from the PCT 2 image, for certain applications of the knowledge-based rules listed in Tables 6.6 and 6.7.
Figure 6.41 illustrates the results of the forest boundary identification rule; the colour\(^9\) of the primitive representing the confidence value that the primitive actually depicts a forest boundary. Virtually all of the true forest boundaries have a high confidence value (i.e., the red and orange coloured primitives to the right of the image); only two forest features have a low value and no non-forest features have high confidence values.

The vegetation rule (Figure 6.42) assigns high values to most of the features, with a few exceptions along the lake edge and at the edges of shadows. As most of the rock exposures in this area are bordered by vegetation only those features with the highest confidence value are important, therefore, the lower valued features can be ignored.

Results of the lake rule are shown in Figure 6.43. Although there are only two lake edges in this particular set of features, both have been identified correctly.

The rock identification rule (Figure 6.44) provides slightly more ambiguous results in that not all rock exposures are identified and other features such as shadow and lake edges have high confidence values. The strongly illuminated rock exposures are identified more successfully than those in shade and casting shadows, whose spectral properties either side of an edge will therefore be different and cause confusion in the rule. However, by combining the rock rule with other rules the ambiguity can be removed.

The roughness rule (Figure 6.45) highlights those features which occur in areas of rough terrain. Many of the rock exposures are identified in this way. Unfortunately, so are several of the lake and forest edges. Again, these latter features are identified/excluded using other rules. The most important result of this rule is that it assigns low confidence values to the areas of scree at A and therefore separates these features from the \textit{in situ} rock exposures. The roughness rule is the only rule that successfully separates these two types of features.

Figure 6.46 shows those primitives identified in both the remotely sensed data and the slope image. These are significant as the likelihood that they are geological features is increased dramatically by the dual

---

\(^9\) Colours range from magenta (the lowest confidence level), through blue, cyan, green, yellow, orange, and red (the highest confidence level).
Figure 6.41 - Results of the forest boundary identification rule. The orange and red primitives (mid-right) have identified forest boundaries successfully.

Figure 6.42 - Results of the vegetation identification rule. Most of the primitives shown here border vegetation.
Figure 6.43 - Results of the lake identification rule. The red and yellow primitives show that the lake edges have been identified successfully.

Figure 6.44 - Results of the rock identification rule. Most rock features are identified, however, there is some confusion at the edges of areas of shadow.
Figure 6.45 - Results of the roughness rule. Most the the rock exposures have a high value for the roughness rule, as expected.

Figure 6.46 - Results of the combination rule. The yellow to orange primitives are identified most from different input data sets.
identification. This helps to separate the main non-geological features derived from the remotely sensed images; for example, the forest and lake boundaries.

The final rule identifies those dip and strike orientations which do not fit well into the local structure and are therefore likely to be erroneous. In many geological applications dip and strike orientations are plotted onto what is loosely termed a stereonet\(^{10}\) (Hobbs et al. 1976) (Figure 6.47). These plots can be created by projecting the intersection of each plane with a hemisphere, onto a horizontal circular plane (Figure 6.48). Alternatively, to avoid clutter on the plot, a single point, representing the pole (or normal) of the surface can be projected onto the horizontal plane (Figure 6.49). The concentration of these points may then be contoured to highlight areas of common orientation. Figure 6.50 illustrates several such plots along side the types of features they represent. The stereonet plots can therefore be used to describe the geological structure present.

Figure 6.51 shows a stereonet of dip and strike values, derived from the band 7 image, with a concentration of values at 11°/198° (dip/strike) and several extraneous values away from this concentration. These latter values do not fit well into the local structure and may represent non-geological features. A confidence value assigned to these primitives is therefore decreased.

All of the rules described in this section are combined using the Shortliffe (1976) method outlined in section 6.2.1. The weighting factors for each rule are given in Table 6.8. Several rules require a cut-off point or threshold defining a confidence limit of zero or one, e.g., if the planar \(R^2\) is below 0.6 then a confidence of zero is given for the "Planar \(R^2\)" rule. The co-occurrence rules are determined by counting the number of coincident features found in other sources. If, for instance, a primitive is coincident with less than three features derived from different DEM sources, then the rule specifies a confidence value of zero. Otherwise, the confidence value is calculated as:-

\[
\text{confidence} = \frac{\text{links\_found} - \text{cut\_off}}{\text{dtm\_sources} - \text{cut\_off}}
\] (6.4)

where \(\text{dtm\_sources}\) is the number of possible links to different DTM sources (i.e., 11).

\(^{10}\) Stereonet is derived from 'stereographic projection net' which describes the projection method by which orientations are plotted onto a 2D surface.
Figure 6.47 - An example stereonet (the Wulff net) (after Hobbs et al. 1976)
Figure 6.48 - The principle of stereographic projection (After Hobbs et al. 1976)
Figure 6.49 - The stereographic projection of a normal (pole) to a plane (after Hobbs et al. 1976)
Figure 6.50 - Contouring the poles on the stereonet (above) and example structures and corresponding plots (below) (after Hobbs et al. 1976).
Figure 6.51 - Sample stereonet of primitives derived from band 7
### Table 6.8 - Weighting factor assigned to each knowledge based rule when combining using the Shortliffe method

<table>
<thead>
<tr>
<th>Rule</th>
<th>Weight</th>
<th>Cut-off point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planar $R^2$</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>Linear $R^2$</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>Lake</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Rock</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Forest track</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Roughness</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>Vegetation</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Co-occurrence between Image &amp; DTM</td>
<td>0.8</td>
<td>3</td>
</tr>
<tr>
<td>Co-occurrence between Image &amp; Image</td>
<td>0.2</td>
<td>4</td>
</tr>
</tbody>
</table>

The final confidence map is shown in Figure 6.52. This clearly shows areas where non-geological features have been found (in pink), i.e., in areas of forest, water, and shadow. Those features showing the highest confidence (in yellows and orange) occur where rocks are exposed at the surface, indicating that the knowledge based rules have been effective in delineating geological and non-geological primitives. A threshold can be applied to this image to produce a final geological set (Figure 6.53). The threshold of 0.3 has been set interactively to remove the least geological primitives, i.e., those within the lake, the shadows, and the forest.

### 6.3 Production of a Geological Model

Whatever the application when mapping structural geology a model of the structure is a fundamental requirement. For instance, in oil exploration structural models are used to identify likely locations for reservoir accumulations of hydrocarbons (Whitten and Brooks 1981, Figure 6.54). This section will describe how a structural model may be created using a combination of manual and automated techniques. The model is initialized by finding the most common dip and strike orientation and creating a linearly dipping structural model. This model is then refined by the inclusion of manually identified faults and automatically derived bedding sequences and folds.
Figure 6.52 - Final confidence map.

Figure 6.53 - Thresholded confidence map (threshold set at 0.3).
Figure 6.54 - Simple structures in which hydrocarbons accumulate. (A) Anticlinal fold; (B) Unconformity; (C) Stratigraphic, by change of facies; (D) Fault.
The structural measurements, which have been derived from the techniques described in this Chapter, may be plotted on a map (e.g., Figure 6.9) and used manually by the geologist to gain an insight into the geological structure of an area. Alternatively, these measurements can be combined using additional automated modelling techniques to produce a 3D map, and a more complete understanding, of the geology. The automated modelling techniques are more likely to produce a more comprehensive understanding of the geology due to the large volume of data to be interpreted. However, the complex nature of geological structures means that an automated expert system can only hope to produce an approximate model of the real-world situation. Such a generalization is in fact what the geologist requires and also what all maps provide, the question is the degree of generalization the user is willing to accept, set against the accuracy of the map. The following sections will describe a series of procedures which are used to derive structural models from the primitive structural data and to display or represent the geological model in a number of different modes according to the geologist's requirements.

6.3.1 The Initial Model

Contouring structural geological data within stereonets is described in section 6.2.1. This can help to describe the geological structure of an area. Figure 6.51 shows a contoured stereonet of structural data derived from band 7 before any object or validation rules have been applied to the data. The figure exhibits the most obvious use of the stereonet which is to show how structural data are often clustered together and how the contouring may be used to derive the most common structural orientation (in this case D/S = 11°/198°). The clustering of the data shows that the geological primitives from which the data were derived belong to the same geological structure. If each data point was randomly located throughout the stereonet this could signify that the structural measurements were grossly in error.

The most common orientation may clearly be identified manually from the stereonet but may also be derived automatically. A program has been designed to produce stereonets. This counts the frequency of the data into a raster grid covering the area of the stereonet. It is therefore a simple procedure to examine each cell of this raster grid to search for the highest frequency and to calculate the orientation of that grid cell. The accuracy of this derived orientation is dependent not only on the accuracy of each original measurement but also on the size of the raster grid. Accuracy may be increased if the number of cells in the grid is increased. A grid size of 50 could be expected to provide an accuracy of between 10° and 1.2° relating to dips of 10° and 80° respectively. On the other hand, a grid size of 5000 would provide accuracies of 0.07° and 0.01° for the same dips, which is beyond the average accuracy possible from the original dip and strike measurements.
A grid size of 1000 approximates this average accuracy \( i.e., \) between 0.3° and 0.06° and has therefore been employed to contour the structural data and also so as not to introduce any further errors into the structural modelling procedure.

The most common orientation, derived from the stereonet, is used to initialize a model of the geology. The model begins existence as an equation of a plane describing the selected orientation and passing through one of the primitives possessing that orientation. This plane then describes the boundary between two unknown geological units and can be used to derive a geological map of the current geological model. Figure 6.55 shows a geological map given the orientation derived from the final automated stereonet \( i.e., \) after the application of knowledge-based rules. The equation of the plane is used to estimate the height value of the geological boundary for each x-y location in the area. If the height value is greater than the corresponding elevation extracted from the DEM then the map is coloured to represent the lower geological unit and vice versa. At this initial stage of the model development it is unimportant which of the primitives representing the 'most common' orientation is used to derive the planar equation. The next stage is to include all of the measurements which have this orientation in order to establish a sequence of bedding units. If each of these primitives are used to calculate separate planes, then a whole series a parallel planes will be produced from which a geological sequence may be extracted. The perpendicular distance between each individual plane can be used to determine the bedding thicknesses of each geological unit. Again at this early stage in the model it is not important to relate these units to any real geological type as the model is simply attempting to portray the structure of the geology.

Figure 6.56 shows the 46 planes that are derived from each of the primitives exhibiting the 'most common' orientation. It can be seen that a number of the planes are extremely close together (as little as 20 cm) and allowing for the errors inherent in the measurement procedures some of these planes can be combined. Each primitive has two associated attributes within its structure relating to the calculated error in both dip and strike. These may be used to construct error planes either side of the derived 'boundary' plane (Figure 6.57). As the error attributes are specified in angular units this results in the error planes not being parallel to the boundary plane. If another boundary plane lies within these error planes, at its own original primitive x-y location, then the two boundary planes may be combined. Figure 6.58 shows a modified sequence of bedding units after a series of combinations have been made according to the associated error attributes. Figure 6.59 shows a planimetric map of these units, again derived using height values calculated from each planar equation.
Stereonet of thresholded primitives (derived from all data sources).

Planimetric map of initial model using the most common orientation (MCO).

Figure 6.55, Creation of the initial geological model.
Figure 6.56 - Bedding thicknesses derived by calculating the perpendicular distance between those primitives having the most common orientation (MCO).
Figure 6.57 - Error planes emanating from the primitive centres. Overlapping planes are used to combine primitives which are considered to belong to identical lithological units.
Figure 6.58 - Revised bedding thicknesses after combination of similar planes.

Figure 6.59 - Geological model using above bedding sequence.
6.3.2 Building a More Complex Model

The initial model described above could represent a simple linearly dipping geological structure and such structures do occur throughout the world. However, the vast majority of even these most simple of structures will include some minor folding or faulting or geological units which peter out and which are therefore not parallel to other units. In most cases it is necessary to include further geological structural information into the model. This section will describe ways in which folds and faults can be introduced into the structural model.

Figure 6.60 shows a stereonet of structural measurements derived from the slope data after application of the object identification and validation rules. This stereonet shows two concentrations of measurements. These concentrations may be considered as part of a folded structure which is represented by the great circle passing through them. The two concentrations represent structural measurements taken from the limbs of the folded structure - if the apex of the fold is relatively sharp or angular, then few measurements will be found with orientations different from those of the limbs. The distance of the great circle away from the centre point of the stereonet indicates the degree to which the axis of the fold dips from the horizontal. The most common orientations can be derived for each of the limbs and combined with corresponding primitives to calculate the equation of the curved surface for each geological unit. The major problem with this technique is to find primitives which represent the same geological unit. Another problem is that the stereonet may be representative of just one fold or a number of similar folds (Figure 6.61).

In an ideal scene, with spectrally distinguishable geological units, an obvious way to ascertain which primitives are associated with the same geological unit, would be to investigate and match the spectral properties of the image to each side of the primitives. Most scenes, however, do not exhibit such ideal properties and most geological units are either spectrally similar or as in this study area covered in vegetation, making the spectral comparison very difficult. An alternative method of matching the primitives is to attempt to match the thicknesses of bedding units. Figure 6.62 shows the local bedding thicknesses derived for those primitives representing the most common orientations on both limbs of the fold. It can be seen that it is virtually impossible to match any of the individual units in this way due to the obvious differences between the derived sequences. These differences are probably due to a number of factors, including:

- the errors inherent in the dip and strike measurements,
Figure 6.60 - Stereonet showing two concentrations, indicating a folded structure.
Figure 6.61 - Different local fold scenarios - illustrating the difficulty in matching bedding thicknesses.
Figure 6.62 - Comparison of bedding sequences on either side of a fold.
• the non-parallel nature of some of the geological units (particularly the dolerite sills which are irregular in their extent),
• the fact that a number of faults are likely to separate local geological sequences, thus offsetting the bedding thicknesses and confusing the match,
• the presence of a number of similar folds which will also confuse the sequence, and
• the fact that some boundaries may not have been identified.

Most of these points can be partly accounted for by obtaining bedding thickness sequences over local areas rather than the whole scene. Of course an additional problem is then to choose appropriate areas which are not dissected by faults or multiple folds. The automatic segmentation of the scene into sub-areas according to fold axes is impossible before the structural model has been made. At present the only sub-area designation that can be made is to segment the area using an interpreted system of structural units. Figure 6.63 illustrates a manually identified set of units which are used to segment the study area. These are combined to produce a set of complete polygons; artificial line segments are added at the outer limits of the study area to facilitate this.

By using these segmented structural units, folds may be approximated by neighbouring linearly dipping models. Planimetric maps may then be made by checking each pixel in the output map against the list of interpreted faults to determine which structural unit it belongs to. The equation of each lithological boundary is then checked for that pixel to determine which units occur at the surface (Figure 6.64). A series of geological maps has been produced in this way using the automatically identified primitives and those identified manually. These maps are produced from:-

- Model 1 - automatically identified primitives
- Model 2 - a manually selected subset of the above
- Model 3 - manually identified primitives
- Model 4 - a combination of manually and automatically identified points

These models are illustrated in Figures 6.65 and 6.66, and may be compared to a simplified version of the published geological map shown in Figure 6.67. Model 1 provides an accurate representation of the geology in regions B and H with the planimetric map closely matching the published map. Other areas are poor, particularly region D which contains a large number of primitives but has a most-common-orientation (MCO) out by approximately 30° in strike. Tighter confidence levels are probably required in these regions in addition to further geological knowledge-based

11 N.B. Each of these locations is shown in Figure 6.63.
Figure 6.63, Manual segmentation of the study area into structural units.
Figure 6.64 - Solving each planar equation for each pixel in the image to determine which geological layer is exposed at the surface.
Model 1 - Derived from automatically identified primitives.

Model 2 - Derived from manually selected automatic primitives.

Figure 6.65, Models 1 and 2.
Model 3 - Derived from manually identified primitives.

Model 4 - Combined inputs of models 2 and 3.

Figure 6.66, Models 3 and 4.
Figure 6.67 - A simplified version of the published 1:50,000 geological map, with the manually interpreted structural segments. A dolerite intrusion occurs at X.
rules to identify those primitives which fit in better with the neighbouring structural segments. Model 2 shows a great improvement in the poorer regions of model 1. Region D now matches the published map if the dolerite intrusions at X are ignored; such intrusions have not yet been incorporated into the modelling procedure, but it can be seen that these intrusions roughly follow the surrounding sedimentary structure. Similarly, region C has igneous intrusions, but the published structure is still reflected in the results. This region also has a fold towards the north, but it was impractical to segment this region further due to the lack of primitives identified here. It has not been possible to match bedding sequences in each structural segment and so any displacements across faults is impossible to determine. Therefore, the apparent displacement shown in this planimetric map does not refer to any real or modelled shift. It is also apparent from this model that the DEM affords a far greater level of detail in the surface expression of the bedding than the 1:50,000 published map. Models 3 and 4 have produced results very similar in accuracy to model 2, with the exception of region C in model 3 which is markedly different and slightly modified strikes in the remaining regions. These differences are probably due to the lack of fine detail included in the manual interpretation. The eye often smoothes over the fine detail and takes in a broader picture, so the fine detail in the DEM which helps produce more accurate dip and strike results is lost.

Models 2, 3, and 4 are far more successful than model 1, showing that some form of manual interpretation or detailed geological knowledge is required to map the structure in this area. Furthermore, the manual results on their own, have been shown to be insufficient in this area, unless each feature is painstakingly traced through each pixel of the data. Therefore, a combination of techniques provide the best method for the mapping of structural geology for the data used in this study.

6.3.3 Map Products

Sections 6.3.1 and 6.3.2 have described methods of creating planimetric maps of the structural model which are all that is required in many applications. In other applications, however, geologists require further map products to aid their interpretations and decision making (Hobbs et al. 1976, Barnes 1981). Such products include:-

- a perspective view of the planimetric map,
- block diagrams of the geology,
- a cross section through the geology,
- a series of fence sections through the geology,
- a bedding sequence at a point location,
• a 3D representation of a particular geological unit, e.g., an ore body (Figure 6.68 - after Raper 1989), and
• an extension of a unit above the surface, for visualization of the structure.

The first of these additional products can be created simply by overlaying the planimetric map onto a perspective view of the elevation data. The remaining products require a further step in the map production process which includes a depth or height parameter. A voxel\(^{12}\) version of the planimetric map can be made by calculating a map for each plane in the block (Figure 6.69); in this case the height condition is not taken from the DEM but from the elevation of each plane within the block. Most of the remaining map products can be derived in a similar manner by solving the equations of the lithological boundaries for a particular condition (i.e., a chosen plane, in the case of a cross-section - see Figure 6.70). A voxel block is shown in Figure 6.71 where the surface is displayed as a pale green colour, the area above surface as grey, and the various sub-surface geological units coloured individually. The block is displayed within a commercial package, called Application Visualization System - AVS (Advanced Visual Systems Inc. 1992), which allows the user to display the block from any angle. At present the user can see very little of the structural model, but AVS has a facility whereby the user can specify that all pixels exhibiting a particular digital number can be made transparent or translucent. Figure 6.72 shows the same block diagram with the above ground pixels made transparent and the surface pixels given an opacity of 0.1 (where 0.0 is transparent and 1.0 is opaque). This feature allows the user to see much more of the structure and of the surface, and is therefore of much greater value.

The ability to change the opacity of a geological unit also helps to produce the map products vi) and vii) in the above list. Displaying one particular geological unit can be made possible by making the other units transparent. This can be useful if the geologist is interested in the shape of a reservoir or cap rock, or in the shape of an irregular ore body. The extension of a geological unit above the surface can be achieved by using the surface equations above the surface and then making this unit translucent (Figure 6.73).

6.3.4 Improvements for the geological modelling procedure

At present the modelling procedure provides a first approximation to the actual geological structure. Improvements are required to obtain more detail and accuracy within the model. Such improvements might incorporate more geological knowledge into the procedures and a better representation of the individual surfaces.

\(^{12}\) A voxel is a volume element which has three dimensions and is comparable to a pixel, or picture element, which has only two dimensions.
Figure 6.68 - A 3D voxel representation of an ore body (after Raper 1989).
Figure 6.69 - Solving each planar equation for each pixel at each depth in the output volume to determine which geological layer is present.
Figure 6.70 - A cross-section derived from the 3D model.

Figure 6.71 - A solid 3D block diagram showing the geological structure.
Figure 6.72 - Block diagram with surface visible through transparent 'soil' layer.

Figure 6.73 - Geological unit projected into the above ground space for visualization purposes.
Knowledge-based rules could be designed to implement geological interpretative skills. For instance, if two identified planes or surfaces converge, then an unconformity could be present or simply a petering out of a lithological unit (Figure 6.74). The former could affect the structure on a much wider scale than the latter which is often more local in nature. Similarly, if a certain kind of folding structure or faulting orientation is discovered then these features may be repeated elsewhere in the scene. Such knowledge could then be used to help interpret other structural information.

It has been noted that the use of polynomial surfaces can be inaccurate at distances away from the primitive location. Therefore, the modelling is likely to be improved by the incorporation of an improved representation of each surface. There are several ways of representing 3D surfaces, including triangulated networks (TINs), 3D points, voxel surfaces, or as planar equations. The problem arises in areas between the known points where an interpolation must be made. At present polynomial equations have been used, but a more accurate interpolation may be provided by TINs, standard interpolations of 3D points (see Chapter 5), or by incorporating surface spline functions (McLaren and Kennie 1989). It is hoped that future work in this area could study these alternatives.

A quite different, and possibly more successful, method of creating the model might be to invoke a type of region growing technique, whereby the most common orientations are used as a starting points. The successively closest orientations could then be incorporated into a modified version of the model. This could have the advantage of easily including minor local perturbations into the structure but could be weak where there are few structural primitives present.

6.4 Conclusions

This chapter has shown that accurate dip and strike values may be extracted from DEMs and remotely sensed images. The accuracy of each measurement can be estimated using the known 3D inaccuracies of the input data and likely errors suggested by measuring the curvature of each primitive feature. The average measured inaccuracy for the estimates in dip is 0.27°. This is sufficient when compared to the quoted general accuracy of field measurements, i.e., ±1° (Barnes 1981). Of course other inaccuracies occur simply because of the misrepresentation of the geological features. This could be due to the identification techniques used or the partial inadequacies of the data used (i.e., vegetation cover, poor quality DEM, and low spectral and spatial resolution). To investigate these possibilities the techniques described in this thesis need to be tested on a number of different study areas varying in scale, climatological environment, and structural type.
Figure 6.74 - Geological phenomena which can confuse structural interpretations.
Several knowledge-based rules have been designed to separate the geological primitives from those representing non-geological features. The rules fall into two categories, those which attempt to identify specific objects, and those which attempt to validate a feature by comparison with knowledge about the surface or with similar features identified from similar data sources. The rules are combined to give a confidence statistic for each primitive indicating the likelihood that the primitive represents a geological feature. These techniques successfully identify the most likely geological features. However, the weighting factors have been set using knowledge of the area and trial and error techniques. An improved method would be to incorporate a learning technique into these procedures, whereby the weighting factors could be modified according to new evidence and facts learned (Forsyth 1989, Beale and Jackson 1990). Several of the rules designed for this project are, to a certain degree, specific to the Llyn Cowlyd area. These require modification for general application in other areas. Again, this could be incorporated into a learning process to investigate which methods and rules work best in different environments.

The structural model that has been derived for the Llyn Cowlyd study area provides a good first approximation to the actual geology of the area and compares favourably with the geological map. The detail that can be achieved in certain parts of the study area far exceeds the detail shown on the published geological map but is poor in other areas. Improvements to the model could be made by the incorporation of more complex geological structures, such as different types of folding, irregular bedding units, and inclusions of igneous intrusions. These additional features may be facilitated by an alternative representation of surfaces and interpolation between them, such as TINs or splined surfaces. Also the incorporation of geological 'knowledge' would assist the creation of the structural model. This knowledge, as with the rules designed in this chapter, would need to be defined so that it is generally applicable in other study areas and could encompass any number of structural environments.

A system of processing is also required which, at any stage during the creation of the model, could return to the images or DEM and search for additional information, or return to the primitive database and modify the weighting factors to include or remove features from the model. Such a system is an expert system. The design and planned implementation of an expert system for mapping geological structures is discussed in the following chapter.
Chapter 7 - Progress Towards a Structural Mapping Expert System

An expert system is a computer system that can perform above, at, or near, the level of a human expert (Harmon and King 1985). Within the scope of research described in this thesis, an expert system should aim to provide a full understanding of a scene (e.g., in terms of geology, vegetation, shading, and geomorphology), a 3-D structural model of the geology, and be able to gain experience from each new area encountered. This is obviously a tall order. The following chapter describes a proposed expert system for structural mapping and demonstrates the progress this research has made towards that goal.

Several computer vision expert systems were briefly introduced in Chapter 3. Further general expert systems will be introduced here, with specific examples of how these may be applied to geological structural mapping. Included throughout the chapter are brief descriptions of parts of the proposed system which have been implemented and/or evaluated.

A preliminary introduction to the proposed expert system is followed by a more detailed discussion of the benefits and problems associated with various designs and implementations of expert systems. The basic components of an expert system are outlined, together with a discussion of different control strategies, methods of computer learning, and suitable user-interfaces. In this context, the current implementation of the proposed system is detailed, and initial results and evaluations are discussed. The scope for further work in the development of this system is discussed in the conclusion to the chapter.

7.1 The Proposed Expert System

An expert system is proposed for the full or semi automated mapping of geological structures from remotely sensed images and digital elevation data. The system, which is
currently under development, is shown graphically in Figure 7.1. It comprises a number of segments which provide tools for an 'inference engine' (whether or not this 'engine' has artificial or human control). These segments include:-

- a database of input parameters,
- a database of derived products,
- a library of knowledge-based rules,
- a library of data processing techniques,
- a library of default and modified variables,
- a results database, and
- a history of learned facts.

Current and future implementations for each of these segments will be discussed later in the chapter, in addition to control of the expert system afforded by the inference engine.

In brief, Figure 7.1 illustrates the mechanism behind the expert system. Input to the system is provided by remotely sensed images, digital elevation data, a set of input parameters describing the input data, and by input from the user via 'conversations' with the expert system and via graphical input. Results of initial processing of the input data, using low-level information extraction techniques (described in Chapters 4, 5, and 6), are stored in the results database. Control of the expert system is then assumed by the inference engine. This engine attempts to make sense of the results in terms of creating possible structural models from the data. This is achieved through the application of knowledge-based rules on expert geological knowledge and by a learning process. Such processes are facilitated by a number of tools within the expert system, including a library of data processing techniques, a library of variables, a results database, and a sophisticated user-interface. Following interpretation of the initial results, the engine may decide that further processing is required in certain parts of the study area and that, as a result, further knowledge-based rules should be applied. This process is repeated until the engine decides either that an accurate structural model has been identified or that no new information can be gathered from the input data provided. Finally, the system produces a structural model (or series of possible models) of the geology in whichever format the user requires (see section 6.3.3). The learned facts are stored and may be used again given new data for the same area or a totally different area.

### 7.2 Benefits of Expert Systems

Chapters 4 and 5 offer techniques for obtaining structural measurements from remotely sensed images either automatically or manually. However, if each method is performed
Figure 7.1 - Schematic flow chart of the proposed expert system
individually the path to obtaining the required measurements is tortuous, involving many iterations of certain techniques before a final result can be determined. One solution is to create a batch process, which automatically executes each of the programs in sequence to produce the desired results. Although simple to implement, it is difficult to encode flow control and adaptability into such techniques. Thus, the degree of success will depend on the applicability of the batch program to the particular land cover and geomorphological features present in the study area. An alternative would be to design a number of batch processing programs, each differing slightly, each appropriate to different geological conditions. Clearly, this removes many of the advantages of automation and would be impractical. Ideally, then, automated processing requires some form of flexible control which mimics the human decision-making process. One way to achieve this is through use of an expert system (Harmon and King 1985, Graham and Jones 1988, Lucas and Van der Gaag 1991). The expert system should not only mimic an expert in the field but should also provide a tool to assist the expert, i.e., not simply replace certain aspects of an expert's job but also to benefit other aspects of their role (Sproull 1985).

Automated techniques have a number of advantages over manual techniques and it is these properties which should be exploited to the full within an expert system. The advantages of automation within the proposed expert system include:-

- speed of performing analysis (and in particular of dull and repetitive jobs),
- accuracy and repeatability,
- objectivity; if the system is properly configured, it should be objective (although it is acknowledged that the system is set up by humans who are, by nature, subjective), and
- the ability to carry out an extremely large number of processes, directed towards any part of the image (Nazif and Levine 1984).

Other more general benefits of expert systems, according to Graham and Jones (1988), include:-

- Experts retire, taking their knowledge with them.
- Experts may be in short supply.
- Humans need sleep and can fall ill.
- Humans are sometimes forgetful or inconsistent.
- Experts can be impatient if required to repeat themselves in training situations.
- Experts may be better employed on the more difficult cases.
The complexity of the tasks set as the aims of this thesis, the large volumes of data resulting from the processing stages, and ambiguity in mapping the geology of some areas, makes this study a good candidate for implementation of an expert system.

7.3 Knowledge-based Expert Systems

An expert system has been defined by Feigenbaum and McCorduck (1983) as:-

"... an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution. Knowledge necessary to perform at such a level, plus the inference procedures used, can be thought of as a model of the expertise of the best practitioners in the field.

The knowledge of an expert system consists of facts and heuristics. The 'facts' constitute a body of information that is widely shared, publicly available, and generally agreed upon by experts in a field. The 'heuristics' are mostly private, little-discussed rules of good judgement (rules of plausible reasoning, rules of good guessing) that characterize expert-level decision making in the field. The performance level of an expert system is primarily a function of the size and the quality of a knowledge base it possesses."

Expert systems, then, are essentially based on three components; the underlying environment (including computer hardware, software and data), the representation of knowledge, and the control or application of that knowledge.

A number of classic expert systems were developed in the 1960's and 1970's which have provided the techniques and building blocks for most of today's expert systems. These include MYCIN, a medical expert system designed to give expert advice regarding identification and treatment of infectious diseases (Buchanan and Shortliffe 1984), DENDRAL, which attempts to predict the molecular structure of an unknown molecule given a spectroscopic analysis (Lindsay et al. 1980), and PROSPECTOR, which provides consultation to geologists in the early stages of investigating a site for ore-grade deposits (Duda and Reboh 1984). PROSPECTOR is a top-down system incorporating five models of possible mineralization (Table 7.1). The system asks the
user certain questions to satisfy a number of assertions that make up the model. These assertions are nodes within a semantic network (see Chapter 6) and typical assertions include:

- "There is pervasively biotized hornblende"
- "There is alteration favourable for the potassic zone of a porphyry copper deposit"

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of assertions</th>
<th>No. of rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koroko-type massive sulphide</td>
<td>39</td>
<td>34</td>
</tr>
<tr>
<td>Mississippi Valley-type lead-zinc</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td>Type-A porphyry copper</td>
<td>187</td>
<td>91</td>
</tr>
<tr>
<td>Komatitic nickel sulphide</td>
<td>75</td>
<td>49</td>
</tr>
<tr>
<td>Roll-front sandstone uranium</td>
<td>212</td>
<td>133</td>
</tr>
</tbody>
</table>

Table 7.1 - Five models comprising PROSPECTOR (from Harmon and King 1985)

The system constantly explains its thinking to the user during questioning and allows the user to determine why the system is progressing along certain lines. The following is an example dialogue:-

1 - I am considering the possibility of a (Type-A porphyry copper deposit - PCDA) in the target area.

The following questions are intended to determine the nature of the regional environment.

2 - To what degree do you believe that:
   (there are granitic intrusives in the region) ?

5

8 - To what degree do you believe that:
   (igneous rocks in the regions have a porphyritic texture) ?

why
I am trying to establish whether some of the intrusive rocks in your area have textures suggestive of a hypabyssal to subvolcanic environment. ....

The user answers with values between -5 and 5 depending on their degree of certainty regarding individual facts. Rules are then used to combine these data using Boolean logic and Bayesian probability to provide certainty factors for each assertion and in turn for each mineralization model (Katz 1991).

PROSPECTOR was originally designed as a research project but for test purposes was employed in a real exploration study. This resulted in the discovery of large molybdenum deposits in both Canada and the United States. Unfortunately, since the early success further deposits have not been found.

7.4 Components of an Expert System

The basic components of an expert system are illustrated in Figure 7.2 (from Harmon and King 1985). Expert systems usually consist of a knowledge-base, an inference engine, and a user-interface. Harmon and King (1985) also add two subsystems, a knowledge acquisition subsystem and an explanation subsystem. Ranzinger and Ranzinger (1984) further specify a method base and a data base containing tools for the job and input data, intermediate results, and outputs respectively. The SPAM system of McKeown et al. (1985) similarly includes an image/map database and suite of image processing tools. Clark (1990) and Hart (1990) include machine learning as an important component of an expert system, which could possibly be incorporated into the knowledge acquisition subsystem of Harmon and King. The proposed system therefore contains the standard components of general expert systems. The remainder of this section will describe possible implementations of these components.

Construction of an expert system includes stages comparable to classical systems analysis (Robinson and Frank 1987). These include identification, conceptualization, prototyping, creating user-interfaces, testing and redefinition, and knowledge-base maintenance (Bobrow et al. 1986). Robinson and Frank (1987) further point out that while ordinary computer programs represent knowledge on two levels, those of data and program, expert systems organize knowledge on three levels, those of facts, rules, and inference. Inference or reasoning is defined as the gaining of new information from the available knowledge and facts (Lucas and Van der Gaag 1991). It is the separation of knowledge and inference that is essential to the design of an expert system (Duda and Gaschnig 1981). Modification or update of the system is then simplified as new knowledge is acquired (Ripple and Ulshoefer 1987).
Figure 7.2 - The standard architecture of a knowledge-based expert system (after Harmon and King 1985).
Knowledge and inference are therefore the most important parts of an expert system. The representation and implementation of knowledge-based rules have been discussed in Chapter 6. Acquisition of knowledge for an expert system is crucial to its success (Lucas and Van der Gaag 1991) and this is introduced in the next section. It will be followed by a description of a number of possible implementations of inference engines and finally by a review of learning strategies for expert systems.

7.4.1 Knowledge Acquisition

The success of an expert system in achieving its goals is heavily dependent on the knowledge it contains and the relevance of this knowledge to the application in hand (Graham and Jones 1988). The study of knowledge-base design is termed knowledge engineering (Harmon and King 1985) and is a non-trivial task. As Lucas and Van der Gaag (1991) summarize, to achieve a performance comparable to human experts, even within a restricted application, an expert system requires large amounts of knowledge.

Knowledge may be thought of in two different ways, that which is derived from well known facts, acquired from books or learnt during formal education, and that which is derived from experience. The latter type of knowledge is termed heuristics (Graham and Jones 1988). Figure 7.3 illustrates the typical development of knowledge throughout life and shows how heuristics become more important throughout a professional career. When acquiring knowledge for an expert system one should not simply rely on facts that can be gleaned from published material, but should also include heuristics gained from a number of experts in the field.

Robinson and Frank (1987) suggest that knowledge should be initially acquired from one source, an expert in the field, and the expert system designed around this knowledge. This should then be followed by testing of the system by several other experts to help identify idiosyncrasies and determine alternative problem solving styles. This enables knowledge-design documents to be drawn up which may then be circulated to further experts in the application field and related fields for further criticism and comment.

Knowledge may also be acquired interactively from the user of the system. McKeown and Harvey (1985) describe an aerial image interpretation system which automatically generates rules from knowledge gained from the user. A user specifies such parameters as typical road widths and spatial context of different objects in a scene. The rule generator then creates rules using these parameters which can then be applied to other scenes. This interactive acquisition is a form of learning for the expert system and is, in some ways, comparable to the induced automated learning described in section 7.4.3.
Figure 7.3 - General pattern of professional development (after Harmon and King 1985).
7.4.2 Control Strategies for the Application of Knowledge-Based Rules - the Inference Engine

A simple form of inference was introduced in Chapter 6, using Shortliffe's inference rules to combine primitive identification rules. Such inferences can be applied using Boolean logic or alternatively using fuzzy uncertainties, as applied in Chapter 6. These inferences can be defined as bottom-up or data-driven inferences as they are driven by products derived from the input data. Such strategies may be repeated until it is no longer possible to derive new information from the data (Lucas and Van der Gaag 1991). An alternative strategy is that of top-down or goal-directed processing, which generates a hierarchy of sub-goals until they can be reached using the input data. These two schemes need not be exclusive and can be combined to give a bi-directional flow control (Nicolin and Gabler 1987, Matsuyama 1987). For instance, in the case of structural mapping a bottom up process can be used initially to produce a first approximation structural model, which is then followed by a top-down procedure whereby the model guides further processing.

Another way of conceptualizing these control strategies is to conceive a 'state-space' containing a large number of alternatives where each might lead to a solution (Robinson and Frank 1987). The space may be searched for a particular path which solves the problem. For example Figure 7.4 shows backward chaining through a search tree to prove H, given eight rules. Note that there are two solutions to arrive at H, one much shorter than the other, which highlights another control problem, i.e., whether the rules should be considered in a depth-first or breadth-first manner. A depth-first approach would initially consider rule one and try to solve the problem along that branch of the tree before considering other branches. A breadth first approach would treat each branch equally and consider each level of the tree structure in turn (Graham and Jones 1988). Each approach would fair better under different circumstances depending on the number of branches, the number of levels, and the path of the solution. Other more intelligent search techniques include costs for each node in the path, in order that some paths may be less costly to explore and may reach the goal sooner (Nilsson 1980).

The correct choice of problem-solving steps is essential to the performance of the expert system (Hayes-Roth and Hewett 1988). Hayes-Roth and Hewett (1988) point out that an intelligent problem solver will always be adjusting its strategy appropriate to the current state of the problem-solving process, and to take advantage of the resources available to it. In other words it should act opportunistically and be capable of reasoning about its actions.
Figure 7.4 - Backward chaining through a search tree.
Hallam (1990) argues that 'blackboard' architecture offers the best potential for realizing such a system. A blackboard system consists of a number of internal experts each contributing their own expertise to the problem at hand (Figure 7.5, Hallam 1990). The system partitions the problem into loosely coupled sub-tasks, or areas of specialization, for which each 'expert' has responsibility. Each expert may add or remove contributions to the blackboard as and when required as the current understanding of the problem changes. The advantages of such a systems include:

- "...it allows expertise to be brought to bear on the problem in a manner and order determined by the progress of the problem solving process and the expertise available, i.e., problem solving proceeds opportunistically;
- it is easy to add new experts, even during the problem solving process, since all the information required by the experts can be found on the blackboard;
- different aspects of the problem can be considered simultaneously by different groups of experts using different areas of the blackboard;
- since all interactions between experts take place through the blackboard, complete control of the problem-solving process can be exercised solely by determining the expert's order of access." (Hallam 1990).

In the case of the proposed structural mapping expert system, the internal experts would include one with full knowledge and access to the results database, another expert controlling the library of data processing tools, a subsequent expert controlling knowledge relating to the identification of vegetation or anthropogenic features in the scene, and others controlling geological features, geomorphological analysis, drainage network analysis, spectral identification, ancillary data, determination of dip and strike, and many different geological modelling strategies (for example, sedimentary, metamorphic, igneous, or mixed geological environments). With the task of structural mapping involving so many diverse fields of expertise, the modular approach of the blackboard system would appear to have many advantages. Development is also made easier as each 'expert' may be designed separately and linkage with other expert systems in related fields is simplified.

7.4.3 Machine Learning

A principal requirement for learning in an expert system is a result of the so-called Feigenbaum bottleneck, related to the difficulty in acquiring knowledge from an expert, particularly at the level of detail required for coding into an expert system (Clark 1990). Learning could also be considered a primary characteristic of intelligence and an essential part of Michie's (1986) eventual goal for artificial intelligence, that of
Figure 7.5 - Components of a blackboard system (after Hallam 1990)
superarcticulacy, or the ability of a machine to acquire knowledge and to form theories, previously unknown, and to communicate these to people.

Machine learning falls into two main categories, that of induced learning from a set of training examples, and that of neurocomputing which attempts to model directly the thought processes of the human brain. Induced learning basically takes a set of results and backtracks to discover what steps were involved in their derivation. Figure 7.6 shows an example of induced learning for the recognition of mammals (taken from Clark 1990). This example illustrates several limitations of this form of learning. First, the quality of learnt facts is heavily dependent on the training examples given. Second, it is also dependent on the applicability of the knowledge to the input data (Dreyfus and Dreyfus 1986). For instance, is it reasonable to assume that all medium sized mammals are lions? This is clearly an extreme example, and induction learning, if designed effectively, can overcome such difficulties to produce the required results. One advantage over neurocomputing is that induced learning is relatively easy to implement.

There are many tasks at which humans excel which defy explanation. In many pattern recognition tasks we just 'see' the solution (Hart 1990), and subsequent analysis of the task is inadequate. A method which tries to simulate rather than emulate the process of the human mind may therefore learn more readily in such circumstances. Neurocomputing is based on an analogy to the physiology of the brain. There are approximately $10^{11}$ neurons in the brain, between which there is a very high connectivity; any neuron can be connected up to $10^5$ other neurons. Activity within the brain is considered highly parallel in nature which appears to give the brain its power. Neural networks are the computer simulation of this principle.

A neural network consists of simple processing units, nodes, which are connected by links (Figure 7.7). A series of layers of nodes may be created (Figure 7.8) through which a network of paths can be made. Each connection has a strength associated with it and each node has a threshold value $\theta$, representing its bias towards an 'on' or 'off' state (i.e., TRUE or FALSE). The state of the node is determined by $\theta$ and the states of all nodes feeding into it:

$$ n = \left( \sum n_i l_i + \theta \right) $$

(7.1)

where $n_i$ is the state of node $i$ which feeds in and $l_i$ represents the links between them (Hart 1990).

Current values of nodes within the network provide the short-term memory of the system. Learning is afforded by changing these values until the network can perform
Figure 7.6 - An example of rule induction for classifying mammals.
Figure 7.7 - A node is connected to several other nodes by links. Its state depends on the values of the other nodes and on the connective strengths and thresholds (after Hart 1990).

Figure 7.8 - A multi-layer neural network (after Hart 1990).
the designated task and match the example training set (Clark and Cañas 1993). Knowledge or long-term memory is therefore the weights assigned to contributing nodes.

Clark and Cañas (1993) report considerable success in identifying components of mixed spectral signatures using neural networks. However, success can be dependent on the initial guess given to the network. Neural networks hold much hope for the future of expert systems and machine learning, but at present induced learning is arguably easier to implement and to design around the concept of interest.

7.4.4 Intelligent User-interface

So that a user may obtain the best results from the expert system there must be a clear, user-friendly interface, through which the user and computer interact (Long and Whitefield 1989, Turk 1990, Edwards 1991). Figure 7.9 (adapted from Bass and Coutaz 1988) illustrates the processes involved in human-computer interaction (HCI). The user of the system must have a mental model of the processes involved in the system in order to understand the results. Conversely, the system should have a model of the user and the user's capabilities in order to elucidate its knowledge and results in the most appropriate manner. Users of an expert system will have different levels of skill in both computing and the expert domain of the system. Inexperienced users will require guidance through the system, while others may wish to take more control (Ripple and Ulshoefer 1987). Ideally then, the HCI should have selectable levels or be dynamically adaptable to the user's requirements (van der Veer 1989). The optimization of this arrangement is termed 'cognitive ergonomics'.

Edwards (1991) lists a number of factors which should be considered when designing a HCI, including:

- "the degree of expertise of the prospective user;
- the amount and nature of direct user interaction required;
- the amount of textual information required;
- the importance of static or animated graphics;
- the degree of menu-type interaction necessary;
- the amount of concurrent help facilities required during operation;
- the appropriateness of windows, icons and mouse-style operation (Wimps);
- the importance of making explicit the inference process and data being used."

1 Turk (1990) provides an excellent review of current research in cognitive ergonomics, including the sharing of cognitive responsibility and the differences in human cognitive ability.
system designer's mental models of the system and users

user's mental model of the system

system's model of the user

Figure 7.9 - System and user models (after Bass and Coutaz 1988)
Some of these points have been discussed above, but the remainder show the importance of textual information about the processes involved and the graphical user-interface (GUI) between system and user. A help facility recently incorporated into many systems is the 'Hypertext system', which provides contextual help in both text and graphical form (Watt and Van den Berg 1989). Furthermore, it allows the user to navigate through different help screens in an application oriented manner.

Windows, menus, icons, and other graphical widgets (Figure 7.10) provide an easy to use interface allowing the user to readily access information which might otherwise be difficult and time consuming to obtain via numerous keyboard commands. Similarly, graphical displays allow simple visual appreciation of results more clearly than a table of potentially less understandable data. Figure 7.11 illustrates an example of this where geological primitives complete with symbols indicating error-of-fit are overlaid on a perspective view of the terrain. It is conceivable that with future advances in computer technology such graphical interfaces could be controlled in a virtual reality environment (Parsons 1993, Raper et al. 1993) enabling the user to gain a more complete appraisal of the data and results.

7.5 Current Implementation of the Proposed System

The proposed system described in section 7.1 combined with certain learning and inference control techniques described in the previous sections, provides the 'ideal' for a geological structural mapping expert system. The present system combines structural mapping methods described in Chapters 4, 5 and 6, with a number of graphical interface developments and learning techniques, all within the framework of an expert system. Control of the inference engine is primarily through manual interaction, although some automated control of the modelling procedure has been developed (see section 6.3). Any references to the inference engine, throughout this section, therefore refer mainly to manual control at present. However, the reader should bear in mind the proposed automation of this component of the system.

This section will briefly describe some of the segments of the expert system which have been developed, including:

- a database of input parameters,
- a database of derived products,
- a library of knowledge-based rules,
- a library of data processing techniques,
- a library of default and modified variables,
Figure 7.10 - Windows features, used to design user-friendly interfaces
Figure 7.11 - Errors of fit for individual vertices on a primitive
(size of dot indicates magnitude of error)
• a results database,
• a history of learned facts,
• a GUI, and
• an inference engine.

7.5.1 Database of input parameters

A database of input parameters holds information describing properties of the input data required in later processing. Incorporated in the database are names of files containing remotely sensed images and elevation data, and units relating to these data; i.e., radiance measured in \( \text{W}^{-7}\text{m}^{-2}\mu\text{m}^{-1}\text{sr}^{-1} \) and height measured in metres. Related to these parameters are the ground resolution of both data sets, any known errors in the geometric location of the data, and wavelength range of each of the input images. Each parameter is used in subsequent processing. For example, geometric parameters are used in the calculation of dip and strike values, while wavelength ranges are used to specify which wavebands are to be utilized in calculating the band ratios and within several of the knowledge-based rules (e.g., to describe the spectral response of water). Spectral responses used within these rules may be substituted by the optional inclusion of a classified image within the input data. Information describing each class is included within the input database.

Other parameters included in this database are details of the location of the study area (i.e., latitude and longitude values) and the date and time of day of image acquisition. These parameters are used in the Lambertian shading process to simulate the natural shading of the study area, and could also be used to remove effects of shading in remotely sensed images. Latitude and longitude values are also available to link input data with other geolocated datasets, such as geophysical or geochemical data, providing additional information to the mapping procedure.

At a later stage in the development of fully automated functionality where the system has learned certain facts from previous areas (for example, that certain techniques are more successful in a particular environment e.g., vegetated, highland areas), this input database may also include more detailed information concerning the study area. This may include items such as characteristics of the geological environment present, including geological structure, a brief description of geomorphology (e.g., glaciation, recent upheaval etc.), or the type of vegetation present.
7.5.2 Library of Knowledge-Based Rules

The library of knowledge-based rules holds programs which execute the object-specific and validation knowledge-based rules described in Chapter 6. The 'library' is more conceptual than tangible. It is a collection of programs, written in a standard format, each called in the same way by the inference engine. The library includes an index comprising a file of program names, a list of required inputs and a series of keywords describing the function or application of the rule (Figure 7.12). These keywords enable the inference engine to interact with the library and to search for appropriate rules to perform certain tasks.

Interaction occurs between the library of knowledge-based rules and the results database, the database of default variables, and the database of derived products. Interaction with the results database is via the inference engine, as the engine may wish to control which primitives are processed, rather than processing the entire database. For instance, the engine may decide that the lake identification rule should only be applied to primitives within a particular area, e.g., low-lying land.

7.5.3 Library of Data Processing Tools

The library of data processing tools is designed in a similar manner to the library of knowledge-based rules. It provides the complete range of data processing techniques required within the system. For example, it includes image processing programs to perform edge detection, spectral enhancement, texture, segmentation, Lambertian shading, and various image management routines, such as sub-image extraction and image header manipulation. Additional data processing routines include thresholding, line thinning, line extraction and a program called surface which fits surfaces to the primitive data in order to derive dip and strike measurements. Although many of these programs have been written as part of this thesis, others have been developed by colleagues or are part of a commercial image processing package, HIPS (Landy et al. 1982). As a result, certain tools cannot at present be called in a standard fashion. The index to the library describes the program options and uses, but programs within the library do require some standardization to facilitate easier control from the inference engine.

Several image processing tools are used to create the database of derived products (Table 7.2), which includes a number of data sets derived from both the input remotely sensed and elevation data. For example, a Sobel filter is used to create slope and aspect images while a band ratio is used to derive a vegetation index. The only routine within the library which outputs data to the results database is surface which provides raw
Figure 7.12 - Library index for object-specific rules
geological primitive data. In controlling all inputs to the data processing tools, the inference engine may selectively apply tools to certain areas of the study area or to certain primitives in the results database.

<table>
<thead>
<tr>
<th>Tools used</th>
<th>Derived Products</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>convolve</em>^2 Sobel filter.</td>
<td>Slope</td>
</tr>
<tr>
<td><em>convolve</em> Sobel filter.</td>
<td>Aspect</td>
</tr>
<tr>
<td><em>raindrop</em> drainage simulator.</td>
<td>Drainage</td>
</tr>
<tr>
<td><em>texture_co</em>^3 Haralick’s co-occurence matrix texture operators.</td>
<td>Geomorphological Texture</td>
</tr>
<tr>
<td><em>fcalcframe</em>^4 general image arithmetic program.</td>
<td>Band Ratios</td>
</tr>
<tr>
<td><em>pca</em>^5 principal components analysis</td>
<td>Principal Components</td>
</tr>
<tr>
<td><em>fcalcframe</em> general image arithmetic program.</td>
<td>Vegetation Index</td>
</tr>
<tr>
<td><em>xclass</em> interactive maximum likelihood classifier</td>
<td>Land Cover Classification</td>
</tr>
</tbody>
</table>

Table 7.2 - List of derived products

7.5.4 Database of Default Variables

A database of default variables houses a list of initial values for all parameters used within the knowledge-based rules and data processing techniques. This allows variation by the inference engine (whether the engine is controlled artificially or by the user). A list of default variables and their initial values is given in Figure 7.19. The database consists of two files, one containing default values for each parameter and the other containing current values. Default values have been determined from experience during the evaluation phase of this study, and found to be the most generally applicable values. Future testing of the system may find that different defaults should be used or that alternative sets of defaults should be used in different geological environments.

A number of parameters are initialized from the input database, such as the ground resolution of the data and which bands to use in order to create the vegetation index. Other parameters include variables describing the object to be identified. For instance,

---

2 Written by Philip Lewis, Remote Sensing Unit, Department of Geography, University College London.
3 Written by Stuart Barr, Remote Sensing Unit, Department of Geography, University College London.
4 HIPS program (Landy et al. 1982).
5 Written by James Pearson, Department of Photogrammetry and Surveying, University College London.
the lake identification rule uses parameters defining the 'levelness' (i.e., variation in height along the primitive) of the lake edge and the relative importance of the levelness to the spectral response of water in the three spectral wavelengths used (see section 6.2.1). Each parameter may be modified by the inference engine to adapt to different circumstances in the modelling procedure or different features in the input data. Alternatively, this could be done interactively using a set of slider bars to alter the values.

7.5.5 Results Database

The results database retains structural results derived from data processing and knowledge-based rules. As described in section 6.1.2 these results contain not only the structural measurements themselves but also the co-ordinates of each point along the line and other associated attribute data (Table 6.1). The database is manifested as a flat file comprising columns of attribute data for each primitive. Each entry also includes the map co-ordinates of every vertex in the primitive, allowing access to the database on a geographical basis.

7.5.6 History of Learned Facts

Chapter 6 described many data processing techniques whose function it is to derive different geological structural measurements from both the spectral data and the elevation data. Knowledge-based rules were applied to remove non-geological primitives and to condense the remainder of the results and extract only those exhibiting the highest confidences. An alternative approach is to introduce a learning facility into the system. This attempts to learn the best techniques to use in different geographical areas. For example, in a flat, highly vegetated area, the most successful techniques to use, in terms of number of accurate geological primitives derived, might be a Canny edge filter applied to a principal components image. Learning methods could help to reduce redundancy in the results since only the best image processing techniques are used for each area, rather than an exhaustive evaluation of all possible techniques many of which may produce very similar results.

Early attempts at a learning system have been made based on the induced learning methods described in section 7.4.3. A number of experiments has been performed which demonstrate the usefulness of the technique. These experiments assessed the success of various input data layers (e.g., spectral waveband, band ratios, DEM) from which the primitives are derived. These images are assessed to determine the most appropriate methods for different terrain types. Terrain types have been classified using four descriptors; geomorphological texture, height, slope, and the vegetation index. The
success of each image has been measured in terms of the number of geological primitives obtained from each image. Other measures could be used, such as the number of primitives used in the final geological model, but for evaluation purposes the preferred measure is sufficient.

Geomorphological and land cover descriptors are first classified into broad categories in order that simple assessments can be made. To limit the number of possible terrain types, each of the descriptors used - slope, height, texture, and the vegetation index, has been divided into four broad categories, resulting in a total of 256 possible combinations. Figure 7.13 illustrates the resulting images. Each image has been median filtered (Schowengerdt 1983) so as to remove very small areas of any class.

Figure 7.14 indicates the number of primitives identified using each input layer for the ten largest classes in the study area. The largest class represents 14.2% of the study area and is depicted along the bottom of the graph. The greatest number of primitives for this class was derived from the fourth principal component and from the Lambertian shaded image with an illumination azimuth of 135°. Over all ten terrain classes, the latter data layer and the third principal component identify the most primitives. The relative success of other data layers for different classes can also be seen from the Figure 7.14. It would be fairly straightforward to specify these conclusions as rules in further processing by simply stating that only certain data layers should be used in specified terrain. However, there are still a number of uncertainties regarding these conclusions or learned facts. These include uncertainties over whether the terrain classes have been defined appropriately to reflect identifiable terrain on the ground. For example, the ten largest classes all fall within the zero slope class (0 - 22.5°); perhaps the slope classes should be further subdivided or the aspect of the slope included in the terrain description to better classify the area. Second, the measure of success could be improved by counting only those primitives having a high confidence of representing geological features. Third, there is no assessment of the study area as a whole, these terrain classes may be identified in other areas but the geology and vegetation cover may be totally different, resulting in different techniques being required. These uncertainties illustrate difficulties in defining appropriate tests and measures to achieve the desired results; i.e., clear rules that can be used in further processing. As Blicher (1985) points out, "finding the right way to state a problem turns out to be the cornerstone of the problem". Clearly further work is required in this area, but this evaluation serves to show possible gains which induced learning may provide.
Figure 7.13, Four terrain classes used in the induced learning example.
Figure 7.14 - Number of primitives found for each terrain class
Graphical User-interface

Several advances have been made in creating a suitable graphical user-interface to the mapping system. As the system predominantly employs manual control, this interface is geared towards access and manipulation of data for manual interpretation. This section will briefly describe and illustrate some of these features, which are incorporated into a program called tracer^, written to assist the manual mapping of structural geological features. In summary, these features include:-

- **Display input data** (Figure 7.15)
- **Edit primitive features** (Figure 7.16)
  - Draw new manually identified geological primitives.
  - Edit those features identified previously by automated processes to provide first guess to manual mapping or to correct automated results.
  - Delete unwanted features.
  - Add attribute information to primitives, including assigned confidence and colour code (which could define lithological type or bedding/fault if required).
- **Display attribute data** - Several of these attributes, namely those relating to the surface fits, are updated each time the user edits a primitive. Others attributes must be updated outside the GUI at present.
  - Display as textual information (Figure 7.17).
  - Display attributes as graphical symbols. For example by displaying the residuals of a planar fit for each vertex on a primitive the user can identify which vertices do not fit very well and adjust their positions accordingly (Figure 7.11).
  - Display dip and strike symbols (Figure 7.18). Standard dip and strike symbols on geological maps are used to give the reader an understanding of the geological structure. Here the symbols provide the same function in addition to permitting the user to identify those features which do not fit well into the local structure.
- **Modify system-wide parameters** (Figure 7.19) - A separate GUI (params) has been developed to enable users to modify system-wide parameters and to provide a brief description of how each parameter is used within the system.

---

^Tracer has been written to take advantage of the OpenWindows windows environment. Openwindows is based on the X protocol which means that tracer may be displayed on any X-based terminal, including all Unix workstations and PC's.
Figure 7.15 - Interactive display of input data (DEM on left and best ITI on right).
Figure 7.16 - Interactive editing of primitives. Colours are used to represent different lithologies (N.B. orange is used to represent faulting).
<table>
<thead>
<tr>
<th>Vector ID:</th>
<th>55</th>
<th>Curvature:</th>
<th>0.510910</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colour Code:</td>
<td>199</td>
<td>Lake prob.:</td>
<td>0.000000</td>
</tr>
<tr>
<td>Num. of Pts:</td>
<td>19</td>
<td>Forest prob.:</td>
<td>0.9435023</td>
</tr>
<tr>
<td>Planar R2:</td>
<td>0.9808894</td>
<td>Frack prob.:</td>
<td>0.345599</td>
</tr>
<tr>
<td>Linear R2:</td>
<td>0.959943</td>
<td>Rock prob.:</td>
<td>0.8315036</td>
</tr>
<tr>
<td>Curve R2:</td>
<td>0.998170</td>
<td>Veg prob.:</td>
<td>1.000000</td>
</tr>
<tr>
<td>Dip:</td>
<td>19.275335</td>
<td>Roughness:</td>
<td>1.000000</td>
</tr>
<tr>
<td>Strike:</td>
<td>145.848963</td>
<td>Overall Prob.:</td>
<td>0.9908946</td>
</tr>
<tr>
<td>Dip error:</td>
<td>0.385242</td>
<td>Origin:</td>
<td>Image 2</td>
</tr>
<tr>
<td>Strike error:</td>
<td>1.050461</td>
<td>Link string:</td>
<td>Image 2</td>
</tr>
</tbody>
</table>

Figure 7.17 - Attribute information displayed interactively.
Figure 7.18 - Display of dip and strike values.
Figure 7.19 - GUI allowing modification of parameters
• Modify the number of initial techniques and data layers - params also provides a GUI to change those techniques and data layers used in the initial extraction of primitives. At present, during initial testing of the system, the default setting includes all techniques and data inputs. In future these will be reduced to reflect facts learned by the system and its users.

• Manipulate images and DEM.
  • Interactive contrast stretch of images to enhance small changes in DN.
  • Creation of perspective views, overlaying images onto the DEM (Figure 7.20). Users may rotate a wireframe model of the DEM in real-time (Figure 7.21) to design the desired view before the actual perspective image is created. Increasing CPU speeds of computers should mean that the actual view can be rotated in real-time.
  • Real-time creation of Lambertian shaded images (Figure 7.22). As the user changes azimuth and zenith sliders in the menu, so the shading changes on the screen. This enables the user to choose the most effective enhancement of the DEM.
  • Interactive assessment of the drainage network. The user may click on any pixel within the DEM to initialize the simulation of a stream flowing from that point (Figure 7.23). These streams can then be used to interpret the geological environment in the area (Argialas et al. 1988).

7.5.8 Control of the Expert System

The data processing steps from input data to the production of a geological model have been discussed in detail in Chapters 4, 5, and 6. These steps will be briefly summarized here to describe the control of the expert system (shown diagrammatically in Figure 7.24).

Each input data layer, including derived products, is processed in turn to derive a set of primitives. These processes are not performed on the entire image, partly due to limitations in computer memory but also because some techniques perform better in certain terrain and these sub-areas give a first approximation to such terrain units. Each sub-area overlaps by a user defined amount (default = 20%) in order that edges crossing boundaries are not overlooked. Ideally, sub-areas could be set to irregular shapes which more closely relate to local terrain and/or local geological structure.
Figure 7.20 - Interactive display of perspective views.
Figure 7.21 - GUI allowing real-time manipulation of a wireframe perspective view
Figure 7.22 - GUI allowing real-time manipulation of Lambertian shading
Figure 7.23 - Display of single or multiple simulated streams
Figure 7.24 - Current processing stages within the expert system.
Sub-areas are then processed using several edge detection routines. Primitives derived from the resulting edge images are further processed using thresholding, thinning, and line extraction techniques. Dip and strike values are determined for each primitive using surface fitting algorithms. Object specific and validation rules are then applied to the primitives to ascertain which have the highest confidence of being geologically relevant. This sequence is repeated until all the input data layers have been employed. Results are stored in the results database and overall confidences thresholded to provide primitives for the geological model.

The geological model is progressively refined from an initial linearly dipping model to include more complex faulting and folding. A stereonet is used to identify the most common orientation for the linear model. User defined faults are then added to segment the area into structural units and the model is updated by adding separate linear or folded models into each segment. This is as far as the system has been developed.

Inference comes from controlling this sequence of events, to obtain the best results for each part of the study area, and through incorporation of geological knowledge to combine structural information to produce a reasonable geological model. Validation and combination rules, and manipulation of the stereonet results, are the first steps in this direction. Machine learning will also help to define more procedures within the inference engine.

The main features which are at present missing from the system include:-

- Identification of areas in the image which require further processing or the use of different algorithms. Such areas may have produced few primitives and the inference engine may deduce that certain directional filters may be needed to enhance geological features of a certain dip and strike (gathered from the current geological model).

- Knowledge of more geological structures and environments. At present the system assumes that all geological units are parallel throughout an area and that the structure can be represented by linear sequences or simple folds segmented by vertical faults. Structures are more complex than this and may include features such as unconformities, thrusts, nappes, box folds, slump structures. Metamorphic structures are often very complex and difficult to model while igneous intrusions may cut straight through sedimentary structures or modify structures locally (Hobbs et al. 1976). Each of these factors needs to be recognized by the inference engine in order that the system can be applied to different geological environments.
- An adaptable control interface, allowing a range of set-ups from fully automated control to manual control. Different users demand different interfaces dependent on their level of expertise in geology, remote sensing, and artificial intelligence.
- A conversational interface which allows the user to guide the system and the system to explain any internal reasoning and processing steps.

7.6 Conclusions and Future Requirements

The geological structural mapping procedures developed in this thesis may be incorporated into an expert system. Inclusion of such artificial intelligence techniques provides a number of advantages, including execution of onerous tasks, speed, accuracy, objectivity, and creation and continual updating of a pool of expert knowledge. The present developments have been designed with a proposed expert system in mind, resulting in the completion of several segments (Figure 7.25). In fact, many of these segments will never be 'completed' as such, because they may always be improved upon, just as the human brain may be, through the incorporation of new knowledge and techniques.

The system, as it stands, could arguably be described as an expert system (Harmon and King 1985), as it makes use of artificial intelligence methods. However, only when all proposed segments have been developed will it become a fully operational system. There are two major segments that require further work:-

- a conversational interface and
- the control of the inference engine.

The proposed interface is planned to include hypertext style help and information which is context sensitive. It will also allow graphical, as well as textual information to be displayed. A blackboard-style control system is proposed for the inference engine. The application is well suited to this style of control, as there are a number of sub-experts that can be defined within the engine, each responsible for certain tasks within the system (proposed sub-systems are depicted in Figure 7.26).

There are a number of other ways in which the system could be improved, including further processing techniques and data:-

- *Additional data.* Any *a priori* knowledge of an area should be used within the system. There is no point in re-inventing the wheel, if
Figure 7.25 - The 'completed', started, and untouched parts of the proposed expert system.
Figure 7.26 - Possible sub-systems for the proposed blackboard architecture.
valuable information exists then it should be used. Readily available data may include the following.

- Geophysical data such as seismic, aeromagnetic, and gravity data can provide valuable information on the sub-surface structure of the geology (Khan 1976, Barr 1990) and help to further constrain the geological model. The interpretation of such data is not straightforward and the system requires additional knowledge-based rules for processing and inference.

- Important stratigraphic and mineral information may be gained from a study of geochemical data acquired from soil, rock, or stream samples (Rose et al. 1979).

- Borehole data and old mine records may provide structural and lithological information at depth, where the extrapolation of the surface model may be poor.

- Field mapping may have been carried out at reconnaissance scales.

- **Further processing techniques.**
  - 'Geologically' intelligent line following algorithms which follow a feature preferably in a direction related to the current structural understanding of the geology.
  - 'Geologically' intelligent region growing algorithms, in a similar vein to the above.
  - Spectral mapping using high spectral resolution data to identify lithological units.
  - Pruning, which removes short branches from a skeleton (thinned) structure.

- **Further interactive components.**
  - Editing a wireframe display of the current geological model, enabling the user to interact with that part of the inference engine.
  - Manual interpretation of the stereonet plot.
  - Interactive fly-by facility allowing a roam of the study area in perspective view mode.

The expert system will never be 'finished', as new techniques and knowledge will always become available. The system proposed in this thesis, however, does include much of the latest thinking in remote sensing and artificial intelligence and has provided a framework for future expansion.
8.1 Requirements for an Automated Geological Mapping System

As world resources become more scarce the requirement for accurate and cost-effective methods of geological mapping increases. This thesis has introduced a number of techniques designed to improve the mapping of geological structures from remotely sensed images and digital elevation data, which are both reasonably cheap data sources when compared to traditional ground-based surveys. These improvements have enabled quantitative, automated modelling of structures from these data, which until recently has not been possible.

With the exception of mixture modelling techniques, very few previous geological remote sensing applications have attempted to produce quantitative results, and processing techniques often require qualitative interpretation of each new scene. Although many of the image enhancement methods used in geological applications are successful in highlighting useful information, the results tend to be scene specific and there are no fixed rules as to which technique works best in a given area or why. One aim of the work started in this thesis is to create a geological mapping system which can be used for any area, regardless of the geological/geomorphological environment or vegetation cover. It has therefore been important to investigate each of these enhancement methods with respect to the study area chosen in this thesis, and in future, to other widely varied sites.

Automation is introduced into the mapping system for a number of reasons. First, increasing data volumes of remotely sensed images mean that it is becoming less practical to interpret such data manually. Similarly, the multitude of processing techniques available for geological applications and general image vision techniques fit easily into an automated environment. Automated techniques may also be more objective than manual interpretation, which may vary considerably between interpreters. Finally, 'knowledge' may also be incorporated into the application of these methods. 'Knowledge' may represent geological knowledge about a scene, knowledge about
processing techniques, knowledge about the input data, knowledge of elements in the scene, and/or knowledge used to create a structural model from individual dip and strike values.

The automated geological mapping system developed in this research has not been attempted before as it has been considered too difficult a task. This thesis brings together some of the latest thoughts and technology in a number of varied disciplines, including geology, remote sensing, image vision, geomorphology and DEMs, 3D computer modelling and graphics, knowledge engineering, expert system design, and GUIs. The work also highlights a number of important areas requiring further development to produce a fully functional geological mapping system.

8.2 Evaluation of Data and Techniques to Extract Geological Structural Measurements

Estimates of geological structural parameters may be derived from a combination of remotely sensed images and elevation data. Either data set may be used to define the 2D planimetric extent of a lithological boundary, while the elevation data provides the third dimension. A least-squares approach is used to fit a plane to each 3D vector and estimates of dip and strike are derived from the equation of the plane. It is important that the two data sets are accurately geo-located and that the DEM describes the topography accurately and at a resolution comparable to that of the image data.

The remotely sensed data were acquired by an airborne scanner. Consequently, the images were affected by distortions introduced by the motion of the plane, in addition to the distortion produced by the mountainous terrain and scanning properties of the sensor. The images were geometrically corrected using a polynomial warping techniques. The resulting root-mean-square error of the warp was 9.5 metres with a maximum error of 18.7 metres (section 4.1.3). This is a poor fit and has undoubtedly introduced errors into the estimation of dip and strike.

A DEM was created by interpolating digitized contours onto a regular grid. A number of interpolation methods were evaluated; the most accurate was judged to be the kriging method. The DEM was interpolated to a resolution of 5 metres to match the remotely sensed images, but it is apparent that the DEM does not contain the fine detail apparent in the image data (section 5.1.2.6). This is due to the scale of the original topographic map (1:10,000) and the smoothing introduced by the interpolation algorithms.
Many image enhancement and processing techniques have been developed in the fields of remote sensing and image vision. However, there are no widely accepted views as to which method(s) should be employed in which circumstances. Therefore, it has been important to investigate a range of procedures both to enhance remotely sensed images for geological interpretation and to extract relevant geological information from the data (section 4.2). Similar evaluations have been made for the enhancement of elevation data and the extraction of geological information (section 5.2.1).

Of the enhancement and processing techniques evaluated no one method stood out as highlighting or extracting the most geological structural information. Although there were many notable differences between the results, most techniques contributed at least some new information, however small. As a result, few methods could be rejected as being of no use to the geological mapping system, and even those that were of little use for this study area may prove to be more important for other sites. This highlights the need for expert system control of the mapping system, so that automated learning techniques may be used to better determine which techniques to use and when. For the Snowdonia study area investigated in this thesis, the most successful techniques for enhancing the remotely sensed images included PCA and ITI images (section 4.2.1), and for enhancing the elevation data, the slope, change of slope, and the Lambertian shading techniques were optimum (section 5.2.3).

Geological primitives have been extracted from the enhanced data using a combination of manual and automated techniques. The automated procedure incorporates edge detection, thresholding, line thinning, and vector extraction. The most important part of this procedure is the edge extraction and a number of different algorithms were evaluated. The Canny optimal edge detector was found to produce the best edges, in terms of accuracy, continuity, width and number (section 4.2.4).

Estimates of dip and strike have been derived by fitting planes to 3D primitives. The accuracy of this method was evaluated by applying the technique to known mathematical models. The sensitivity of this technique to errors in the input data was evaluated with respect to geometric errors determined from the geometric correction of the image data and the creation of the DEM. Evaluation against the mathematically produced elevation models showed that the accuracy of the dip and strike estimates is strongly related to the curvature of the primitive. Evaluation of the image and elevation results have shown that errors introduced by inaccuracies of these data produce average errors in dip of less than 1°, which is greater than the accuracy needed for oil exploration (Berger et al. 1992) (section 6.1.3).
8.3 Development of an Expert System

A large amount of primitive data may be extracted using both manual and automated techniques. This information represents both geological and non-geological information (e.g., vegetation and anthropogenic features). The likelihood that any of these primitives represent a geological feature has been determined using a series of knowledge-based rules. For example, rules have been designed to determine whether a primitive represents a lake edge by examining the height differences along the vector and the spectral properties to either side of the feature. Similarly, primitives identified from the image data are compared against those derived from the elevation data; due to the premise that most geological features in this area have some geomorphological expression, those that coincide are assigned a higher confidence of representing a geological feature than those that do not coincide. The former example has a negative contribution to the geological likelihood while the latter has a positive effect. Each of the rules has been combined using the Shortliffe method (Shortliffe 1976), which allows a weighting to be applied to a rule, according to its value in defining whether or not the primitive represents a geological feature (section 6.2.1).

An expert system for the manual and automated mapping of geological structures has been proposed, and initiated, as part of this thesis (Chapter 7). Expert systems can provide a number of advantages in that they can incorporate the knowledge of a number of experts in different fields, perform dull and mundane jobs quickly, and provide consistent and objective results. The system proposed in this research incorporates all of the techniques described above within a framework used to derive a geological structural model. The inference engine of the expert system controls processing depending on the current state of the model.

An initial structural model is created from the dip and strike data extracted from the image and elevation data. The dip and strike data are plotted onto a stereonet which is then contoured to find the most common orientation. This orientation is then used to create a simple linearly dipping structural model. Bedding thicknesses can be derived by calculating the perpendicular distance between those primitives exhibiting the most common orientation. Further stereonets have been used for sub-areas (defined manually at this stage) to investigate any folding in these areas (section 6.3.2).

In the future, the expert system could be used to process the input data further, either to help prove the current model, or to investigate geographical areas that might provide additional information. For example, the inference engine would direct the image processing tools to where it expected to find an outcrop, given the current structural model, bedding sequence and local topography. It might be that a slight change in one
An initial evaluation of learning techniques has been made to enable the system to learn continuously which techniques work best in different scenarios. This will help the inference engine to decide which procedure to use. Furthermore, it is planned to redesign the inference engine using a blackboard system approach so that several 'experts' can contribute to any decision-making process using their particular field of expertise (section 7.4.2).

8.4 Problems Encountered During this Study

A number of problems have been encountered within this study. These have been related to the field site chosen and the input data used. Although the Snowdonia study area exhibits a strong relationship between topography and geology, the relationship is confused and masked in places by the effects of glaciation. Moreover, the geological structure is not simple, with dolerite intrusions, and folding and faulting complicating the picture. Furthermore, the natural vegetation cover and anthropogenic features, such as the forest and reservoir, hide the underlying geology. The site was therefore not an ideal study area for the initial testing of the mapping system. However, it did provide a number of challenges which needed solving; for instance, the object-specific knowledge-based rules that were developed to identify lakes, forests, and forest tracks. Other environments will require additional identification rules. It is proposed that each of these should be designed in such a way that they will be generally applicable to as wide a range of circumstances as possible. A more appropriate initial study area might have been a semi-arid or arid environment with a simple sedimentary geological structure that is reflected quite clearly in the topography. Unfortunately, suitable data for such an area was available for this thesis. Future work will include a wide variety of climate, structures, and topography.

Further problems were discovered with the input data. First, the geometric correction of the remotely sensed images was poor due to the instability of the airborne platform. This could, in part, be overcome by employing a different correction technique. Allison and Muller (1992) have shown that more accurate results can be obtained by matching the images to ortho-aerial photographs using area-based correlation techniques. The ortho-photographs are, in turn, created through stereo-matching of the aerial photographs to produce a DEM, followed by topographic correction of the photography. Second, it was discovered that the elevation data did not include the same level of detail as that shown in the image data. Again, this could be improved upon using different
DEM creation techniques. For example, Allison and Muller (1992) have shown that application of stereo-matching techniques to aerial photographs can produce a DEM with a greater level of detail than that afforded by the interpolation of digitized 1:10,000 contours.

8.5 Areas Requiring Further Work

The geological mapping system that has been developed as part of this thesis has shown that quantitative estimates of geological structural parameters can be extracted from remotely sensed and elevation data, and that these may be combined to produce a simple structural model of the geology. Further work is required to enhance the functionality and performance of the mapping system so that more complex and accurate models can be achieved. Future work (see section 7.6) should concentrate on developing techniques in the following areas:

- improved data processing techniques for the extraction of line information from the image and elevation data
- additional geological structural knowledge for the creation of more complex models
- incorporation of additional data sets, e.g., map data, geophysical and geochemical data, hyperspectral images, borehole data, mine records, and field surveys.
- implementation of a blackboard-type inference engine (Hallam 1990)
- implementation of a learning sub-system (Clark 1990)
- improvement of an intelligent graphical user interface (Edwards 1991)

8.6 General Conclusions

The work described in this thesis has illustrated that geological structural models may be derived both manually and automatically from a combination of remotely sensed images and DEMs. The resulting structural models are good but not perfect. However, the results are considered to be very encouraging considering the problems encountered in the Snowdonia study area (i.e., glaciation, vegetation cover, intrusion of igneous rocks, complexities in the geological structure and errors inherent in the input data). The geological mapping system developed here offers tools to the exploration geologist which will hopefully reduce exploration costs and allow more accurate structural mapping from remotely sensed data than has hitherto been possible.
In the longer term it is hoped that this work will form the basis of more quantitative rather than qualitative geological applications for remotely sensed data. The facility to retrieve quantitative structural parameters automatically and remotely would undoubtedly benefit many studies, including oil and mineral exploration, hydrology, and environmental mapping.

Looking further the work constitutes an essential building block for a fully automated image understanding system (Muller 1988). This would be a system which not only ‘understands’ geological features in an image, but all phenomena within the scene. Wilkinson and Fisher (1984) point to a time when computer systems may eventually out-perform humans in terms of image understanding. A better and more realistic goal might be to provide systems which assist human interpreters rather than replace them. This thesis is one more step in this direction.

At present the techniques developed in this thesis can produce an accurate structural model of linearly dipping sequences and it can map simple folds and faults by segmenting the area into regions that approximate linearly dipping structures. With the addition of further image processing techniques and knowledge-based rules it is likely that the system could map more complex sedimentary structures such as domes and basins, nappes, thrusts, and unconformities. It should also be possible to map igneous intrusions such as sills, dykes, and batholiths; however, the more convolute and complex folds of slump structures and metamorphic rocks may prove more elusive. Extrapolating the more complex structures beneath the surface will be more difficult than for simple structures and may require other ancillary information such as borehole or geophysical data.

The majority of oil and mineral exploration studies, using remotely sensed data, have concentrated on the mapping of lineaments to infer structure. The work of this thesis moves away from this trend and concentrates on quantitative mapping of geological structures. Hopefully, this work will initiate further research into, and the practical implementation of these techniques.
This glossary has been compiled from the following sources:


**Acid rock**
An igneous rock with 10% or more free quartz.

**Albedo**
(1) The ratio of the amount of electromagnetic radiation reflected by a body to the amount incident upon it, often expressed as a percentage, e.g., the albedo of the Earth is 34%. (2) The reflectivity of a body compared to that of a perfectly diffusing surface at the same distance from the Sun, and normal to the incident radiation.

**Algorithm**
(1) A fixed step-by-step procedure to accomplish a given result; usually a simplified procedure for solving a complex problem; also a full statement of a finite number of steps. (2) A computer-oriented procedure for resolving a problem.

**Altitude**
Height above a datum; the datum is usually mean sea level.

**Angle of incidence**
(1) The angle between the direction of incoming EMR and the normal to the intercepting surface; (2) In SLAR systems this is the angle between the vertical and a line connecting antenna and target.

**ASCII**
American Standard Code for Information Interchange.

**Ash**
The unconsolidated fine-grained material formed as a result of volcanic explosions.

**ATM**
Airborne Thematic Mapper. An airborne scanner designed by Daedalus Inc.

**Axial plane**
The plane joining the hingelines in adjacent folded surfaces.

**Azimuth**
The geographical orientation of a line given as an angle measured clockwise from north.
<table>
<thead>
<tr>
<th><strong>Band</strong></th>
<th>(1) A selection of wavelengths. (2) Frequency band. (3) Absorption band. (4) A group of tracks on a magnetic drum. (5) A range of radar frequencies, such as X-band, Q-band, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Batch processing</strong></td>
<td>The method of data processing in which data and programs are entered into a computer which then carries out the entire processing operation with no further instructions.</td>
</tr>
<tr>
<td><strong>BGS</strong></td>
<td>British Geological Survey.</td>
</tr>
<tr>
<td><strong>Bit</strong></td>
<td>(1) An abbreviation of binary digit. (2) A single character of a language employing only two distinct kinds of characters.</td>
</tr>
<tr>
<td><strong>Brightness</strong></td>
<td>(1) The attribute of visual perception in accordance with which an area appears to emit more or less light. (2) Luminance. (3) The luminous flux emitted or reflected per unit projected area per unit solid angle. The unit of brightness, the lambert, is defined as brightness of a surface which emits or reflects one/π lumen per square centimetre per steradian.</td>
</tr>
<tr>
<td><strong>Byte</strong></td>
<td>A group of eight bits of digital data.</td>
</tr>
<tr>
<td><strong>Calibration</strong></td>
<td>The act or process of comparing certain specific measurements in an instrument with a standard.</td>
</tr>
<tr>
<td><strong>Class</strong></td>
<td>A surface characteristic type that is of interest to the investigator, such as a forest by type and condition, or water by sediment load.</td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td>The process of assigning individual pixels of a multispectral image to certain categories, generally on the basis of spectral reflectance characteristics.</td>
</tr>
<tr>
<td><strong>Colour</strong></td>
<td>The property of an object which is dependent on the wavelength of the light it reflects or, in the case of a luminescent body, the wavelengths of light that it emits. If, in either case, this light is of a single wavelength, the colour seen is a pure spectral colour; but if light of two or more wavelengths is emitted, the colour will be mixed. White light is a balanced mixture of all the visible spectral colours.</td>
</tr>
<tr>
<td><strong>Colour composite (multiband photography)</strong></td>
<td>A colour picture produced by assigning a colour to a particular spectral band. In Landsat, blue is ordinarily assigned to MSS band 4 (0.5-0.6μm), green to band 5 (0.6-0.7 μm), and red to band 7 (0.8-1.1 μm), to form a picture closely approximating a colour infrared photograph.</td>
</tr>
<tr>
<td><strong>Contrast stretching</strong></td>
<td>Improving the contrast of images by digital processing. The original range of digital values is expanded to utilise the full contrast range of the recording film of display device.</td>
</tr>
</tbody>
</table>
Control, ground

(1) Control obtained by ground surveys as distinguished form control obtained by photogrammetric methods; may be for horizontal or vertical control, or both. (2) Ground (in-situ) observation to aid in interpretation of remote sensor data.

Covariance

The measure of how two variables change in elation to each other (covariability). If larger values of Y tend to be associated with larger values of X, the covariance will be positive. If larger values of Y tend to be associated with smaller values of X, the covariance will be negative. When there is no particular association between X and Y, the covariance value will approach zero.

CPU

Central Processing Unit - units used to measure execution time for computer programs.

Cursor

Aiming device, such as a lens with cross hairs, on a digitizer or an interactive computer display.

DEM

Digital Elevation Model - a raster array of height values.

Diffuse sky radiation

Solar radiation reaching the Earth's surface after having been scattered from the direct solar beam by molecules or suspensoids in the atmosphere. Also called skylight, diffuse skylight, sky radiation.

Digitization

The process of converting an image recorded originally on photographic material into numerical format or the vectorization of map data into numerical format.

Display

An output device that produces a visible representation of a data set for quick visual access; usually the primary hardware component is a cathode ray tube.

DN

Digital Number. The value that is recorded by the remote sensing scanner. It is a uncalibrated figure and therefore does not have any units.

Dolerite

A medium-grained basic hypabyssal igneous rock, mineralogically and chemically the same as gabbro and basalt.

Edge

The boundary of an object in a photograph or image, usually characterised by a rather drastic change in the grey shade value from the immediate interior of the boundary to the immediate exterior of the boundary.

Edge enhancement

The use of analytical techniques to emphasise transition in imagery.

Electromagnetic radiation (EMR)

Energy propagated through space or material media in the form of an advancing interaction between electric...
and magnetic fields. The term radiation, alone, is commonly used for this type of energy, although it actually has a broader meaning. Also called electromagnetic energy.

**Electromagnetic spectrum** The ordered array of known electromagnetic radiations extending from the shortest cosmic rays, through gamma rays, X-rays, ultraviolet radiation, visible radiation, infrared-radiation, and including microwave and all other wavelengths of radio energy.

**Elevation** (1) Vertical distance from the datum, usually mean sea level, to a point or object on the Earth's surface. Not to be confused with altitude, which refers to points or objects above the Earth's surface. (2 architectural) An orthographic projection of any object into a vertical plane.

**Emissivity** The ratio of the radiance given off by a surface to the radiation given off by a blackbody at the same temperature; a blackbody has an emissivity of 1, other objects between 0 and 1.

**False colour** The use of one colour to represent another; for example, the use of red emission to represent infrared light in the colour infrared film.

**Field of view** The solid angle through which an instrument is sensitive to radiation. Owing to various effects, diffractions, etc., the edges are not sharp. In practice they are defined as the "half-power" points, i.e., the angle outwards from the optical axis, at which the energy sensed by the radiometer drops to half its on-axis value.

**Filtering** In analysis, the removal of certain spectral or spatial frequencies to highlight features in the remaining image.

**GCP** Ground control point. A geographical feature of known location that is recognisable on images and can be used to determine geometrical corrections.

**Geomorphology** The description and interpretation of land forms.

**Histogram** The graphical display of a set of data which shows the frequency of occurrence (along the vertical axis) of individual measurements or values (along the horizontal axis); a frequency distribution.

**Hue** That attribution of a colour by virtue of which it differs from grey of the same brilliance, and which allows it to be classed as red, yellow, green, blue or immediate shades of these colours.

**Illumination** The intensity of light striking a unit surface is known as the specific illumination or luminous flux. It varies directly with the intensity of the light source and inversely as the square of the distance between the illuminated surface and the source. It is measured in a unit called the lux. The total illumination is obtained by multiplying the specific illumination by the area of
the surface covered by the light. The unit of total illumination is the lumen.

**Image**

(1) The counterpart of an object produced by the reflection or refraction of light when focused by a lens or mirror. (2) The recorded representation (commonly as a photo-image) of an object produced by optical, electro-optical, optical mechanical, or electronic means. It is generally used when the EMR emitted or reflected from a scene is not directly recorded on film.

**Image enhancement** Any one of a group of operations that improve the detectability of targets or categories. These operations include, but are not limited to, image compression, image smoothing, and image sharpening.

**Image processing** Encompasses all the various operations that can be applied to photographic or image data. These include but are not limited to, image compression, image restoration, image enhancement, processing, quantization, spatial filtering and other image pattern recognition techniques.

**Infrared** Pertaining to energy in the 0.7-100μm wavelength region of the electromagnetic spectrum. For remote sensing, the infrared wavelengths are often subdivided into near infrared (0.1-1.3μm), middle infrared (1.3-.3.0μm), and far infrared (7.0 - 15.0μm). Far infrared is sometimes referred to as thermal or emissive infrared.

**Irradiance** The measure, in power units, of radiant flux incident upon a surface. It has the dimensions of energy per unit time (i.e., watts).

**Lambertian surface** An ideal, perfectly diffusing surface, which reflects energy equally in all directions.

**Map** A representation in a plane surface, at an established scale, of the physical features (natural, artificial or both) of a part of the Earth's surface, with the means of orientation indicated.

**Maximum likelihood rule** A statistical decision criterion to assist in the classification of overlapping signatures; pixels are assigned to the class of highest probability.

**Microwave** Electromagnetic radiation having wavelengths between 1m and 1mm or 300 - 0.3 GHz in frequency, bounded on the short wavelength side by the far infrared (at 1mm) and on the long wavelength side by very high-frequency radio waves. Passive systems operating at these wavelengths are sometimes called microwave systems. Active systems are called radar, although the literal definition of radar requires a distance measuring capability not always included in active systems. The exact limits of the microwave region are not defined.
<table>
<thead>
<tr>
<th><strong>MSS</strong></th>
<th>Multi-Spectral Scanner. A sensor on board the Landsat satellites.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multispectral</strong></td>
<td>Generally used for remote sensing in two or more spectral bands, such as visible and IR.</td>
</tr>
<tr>
<td><strong>Nadir</strong></td>
<td>(1) That point on the celestial sphere vertically below the observer, or 180° from the zenith. (2) That point on the ground vertically beneath the perspective centre of the camera lens.</td>
</tr>
<tr>
<td><strong>Orbit</strong></td>
<td>(1) The path of a body or particle under the influence of a gravitational or other force. For instance, the orbit of a celestial body is its path relative to another body around which it revolves. (2) To go around the Earth or another body in an orbit.</td>
</tr>
<tr>
<td><strong>Parallax</strong></td>
<td>The apparent change in the position of one object, or point, with respect to another, when viewed from different angles. As applied to aerial photos, the term refers to the apparent displacement of two points along the same vertical line when viewed from a point (the exposure station) not on the same vertical line.</td>
</tr>
<tr>
<td><strong>Perspective</strong></td>
<td>Representation, on a plane or curve surface, of natural objects as they appear to the eye.</td>
</tr>
<tr>
<td><strong>Photograph</strong></td>
<td>A picture formed by the action of light on a base material coated with a sensitised solution that is chemically treated to fix the image points at the desired density. Usually now taken to mean the direct action of EMR on the sensitised material.</td>
</tr>
<tr>
<td><strong>Pitch</strong></td>
<td>Rotation of an aircraft about the horizontal axis normal to its longitudinal axis, which causes a nose-up nose-down attitude.</td>
</tr>
<tr>
<td><strong>Processing</strong></td>
<td>The manipulation of data by means of computer or other device.</td>
</tr>
<tr>
<td><strong>Pixel</strong></td>
<td>A picture element. A single digital measurement (Colwell, 1983).</td>
</tr>
<tr>
<td><strong>Radar, synthetic aperture (SAR)</strong></td>
<td>A radar in which a synthetically long apparent or effective aperture is constructed by integrating multiple returns from the same ground cell, taking advantage of the Doppler effect to produce a phase history film or tape that may be optically or digitally processed to reproduce an image.</td>
</tr>
<tr>
<td><strong>Radiant flux</strong></td>
<td>The time rate of the flow of radiant energy; radiant power (Colwell, 1983).</td>
</tr>
<tr>
<td><strong>Radiance</strong></td>
<td>Radiant flux density per solid angle (W cm(^{-2}) sr(^{-1}))</td>
</tr>
<tr>
<td><strong>Radiation</strong></td>
<td>The emission and propagation of energy through space or through a material medium in the form of waves (Colwell, 1983).</td>
</tr>
</tbody>
</table>
Raster
Two dimensional matrices of contiguous pixel values (Lillesand and Kiefer, 1979).

Reflectance
The ratio of the radiant energy reflected by a body to that incident upon it. The suffix(-ance) implies a property of that particular specimen surface.

Remote sensing
In the broadest sense, the measurement or acquisition of information of some property of an object or phenomenon, by a recording device that is not in physical or intimate contact with the object or phenomenon under study; e.g., the utilization at a distance (as from an aircraft, spacecraft or ship) of any device and its attendant display for gathering information pertinent to the environment, such as measurements of force fields, electromagnetic radiation or acoustic energy. The technique employs such devices as the camera, lasers, and radio frequency receivers, radar systems, sonar, seismographs, gravimeters, magnetometers, and scintillation contours.

Resolution
The ability of an entire remote sensor system, including lens, antennae, display, exposure, processing and other factors, to render a sharply defined image. It may be expressed as line pairs per millimetre or metre, or in many other ways. In radar, resolution usually applies to the effective beam-width and range measurement width, often defined as the half-power points. For infrared line scanners the resolution may be expressed in terms of temperature or other physical property being measured.

Roll
Rotation of an aircraft about the longitudinal axis to cause a wing-up or wing-down attitude.

Roughness
For radar images, this term describes the average vertical relief of small-scale irregularities of the terrain surface.

Satellite
An attendant body that revolves about another body, the primary; especially in the solar system, a secondary body, or moon, that revolves about a planet. A man-made object that revolves about a spatial body.

Saturation
Degree of intensity difference between a colour and an achromatic light-source colour of the same brightness.

Scale
The ratio of a distance on a photograph or map to its corresponding distance on the ground. The scale of a photograph varies from point to point because of displacements caused by tilt and relief, but is usually taken as f/H, where f is the principal distance (focal length) of the camera and H is the height of the camera above mean ground elevation. Scale may be expressed as a ratio 1:24,000; a representative fraction, 1/24,000; or an equivalence, 1 in. = 2,000ft.
**Scene**  
In a passive remote sensing system, everything occurring spatially or temporally before the sensor, including the Earth's surface, the energy source, and the atmosphere, that the energy passes through as it travels from its source to the Earth and from the Earth to the sensor.

**Sensor**  
Any device that gathers energy, EMR or other, converts it into a signal and presents it in a form suitable for obtaining information about the environment.

**Smoothing**  
The averaging of densities in adjacent areas to produce more gradual transitions.

**Software**  
The computer programs that drive the hardware components of a data processing system; includes system monitoring programs, programming language processors, data handling utilities, and data analysis programs.

**Spatial filter**  
An image transformation, usually a one-to-one operator used to lessen noise or enhance certain characteristics of the image. For any particular (x,y) co-ordinate on the transformed image, the spatial filter assigns a grey shade on the basis of the grey shades of a particular spatial pattern near the co-ordinates.

**Spectral signature**  
Quantitative measurement of the properties of an object at one or several wavelength intervals.

**Spectrometer**  
A device to measure the spectral distribution of EMR. This may be achieved by a dispersive prism, grating, or circular interference filter with a detector placed behind a slit.

**Supervised classification**  
A computer implemented process through which each measurement vector is assigned to a class according to a specified decision rule, where the possible classes have been defined on the basis of representative training samples of known identity.

**Swath width**  
The overall plane angle or linear ground distance covered by a multispectral scanner in the across-track direction.

**Synoptic view**  
The ability to see or otherwise measure widely dispersed areas at the same time and under the same conditions; e.g., the overall view of a large portion of the Earth's surface which can be obtained from satellite altitudes.

**Texture**  
In an image, the frequency of change and the arrangement of tones.

**Thermal infrared**  
The preferred term for the middle wavelength range of the IR region, extending roughly from 3μm at the end of the near infrared, to about 15 or 20μm, where the far infrared begins. In
practise the limits represent the envelope of energy emitted by the Earth behaving as a grey body with a surface temperature around 27°C.

**Training areas**  The data samples of known identity used to determine decision boundaries in the measurement of feature space prior to classification of the overall set of data vectors from a scene.

**Visible wavelengths**  The radiation range in which the human eye is sensitive, approximately 0.4-0.7μm.

**Wavelength (symbol \( \lambda \))**  Wavelength=velocity/frequency. In general, the mean distance between maxima (or minima) of a roughly periodic pattern. Specifically, the least distance between particles moving in the same phase of oscillation in a wave disturbance. Optical and IR wavelengths are measured in nanometres (10^{-9}m), micrometres (10^{-6}m), and Angstroms (10^{-10}m).

**Yaw**  Rotation of an aircraft about its vertical axis, causing the longitudinal axis to deviate from the flight line.

**Zenith**  The point in the celestial sphere that is exactly overhead; opposed to nadir.
Thanks to Dr. Mike Barnsley for his supervision, patience, support, and discussions on the relative merits of Arsenal and Aston Villa. I am indebted to my wife Andrea for her continued enthusiasm, encouragement and assistance and latterly to my son Aiden for not minding his Dad being at work on weekends and evenings. Thanks to the rest of my family for putting up with me during this long task. Thanks also to my colleagues at UCL and Plymouth University, Dave Allison, Stuart Barr, Bill Campbell, Eddie Carter, Tim Day, Lewis, Alex McManus, Pete Miller, Prof. Peter Muller, Richard Morris, James Pearson, Dr. Allison Reid, Jon Rogers, and Paul Schooling, for technical support, help, and welcome diversions. I am grateful to the Natural Environment Research Council for provision of the Daedalus ATM images through grant number GR3/7020. Thanks to Sean Nicholson and NERC Computing Services in Plymouth for allowing me to use their computing equipment. Finally, thankyou to the Forestry Commission at Gwydyr Forest for letting me trample all over their land.
References


Almeida-Filho, R. and R.M.G. Castello Branco, 1992, Location of kimberlites using Landsat Thematic Mapper images and aerial photographs: the Redondão diatreme,


Barr, M.W.C., 1990, Pennine exploration dataset re-examined, In *Remote Sensing : an*
operational technology for the mining and petroleum industries, Institute of Mining and Metallurgy, London, 29-31 October, 3-14.


Blicher, A.P., 1985, Edge detection and geometric methods in computer vision, Technical Report No. STAN-CS-85-1041, Department of Computer Science, Stanford University, California, CA 94305, USA.


Campbell, W., 1990, Personal Communication, Department of Geography, University College London.


Clark, C. and T. Cañas, 1993, Spectral Identification by Artificial Neural Network,


Justice, C.O. and B.N. Holben, 1979, Examination of Lambertian and non-Lambertian models for simulating the topographic effect on remotely sensed data, *NASA TM 80557*, Goddard Space Flight Center, Greenbelt, Maryland, USA.


Krishnamurthy, J., P. Manavalan, and V. Saivasan, 1992, Application of digital


Sensing, 30, 103-109.


Podwysocki, M.H., D.B. Segal, and M.J. Abrams, 1983, Use of multispectral scanner images for assessment of hydrothermal alteration in the Marysvale, Utah, mining area,
Economic Geology, 78, 675-687.


Rothery, D.A., 1987a, Improved discrimination of rock units using Landsat Thematic


Anthology, NASA SP-465, National Aeronautics and Space Administration.


Swain, C.J., 1976, A FORTRAN IV program for interpolating irregularly spaced data using the difference equations for minimum curvature, Computers and Geoscience, 1, 231-240.


Sept., University of Edinburgh.


Wilson, A.K., 1985, Calibration of ATM data, In *Proceedings of the NERC 1985*


Geosciences, **18**, 1035-1045.


Appendix A - Calculating Dip and Strike Values from Best Fit Coefficients

The least squares fitting technique provides the coefficients a, b, and c to the following formula:

\[ z = a.x + b.y + c \]

where x, y, and z represent the three-dimensional co-ordinate system.

The coefficients may be used to derive the dip and strike of the fitted plane using simple trigonometry.

If the intersection of the dip line with the z axis is assumed to have a value of 1, then the coefficients a and b represent those lines along the x and y axes shown above. The line x can then be calculated using Pythagoras' Theorem, i.e., \( \text{sqrt}(a^2 + b^2) \). The value for dip is therefore calculated by:

\[ \text{dip} = \tan^{-1} \left( \text{sqrt} \left( a^2 + b^2 \right) \right) \]

The angle alpha may be calculated using:

\[ \text{alpha} = \tan^{-1} \left( \frac{b}{a} \right) \]

The quadrant of the strike is determined from the signs of a and b.

A point worthy of note, in relation to the relevance of the dip and strike results, is the comparison with apparent dip often measured falsely in the field. The apparent dip is the dip seen from a distant when not lined up with the strike of the geology. The results derived here attempt to overcome this by fitting a plane through the data. However, apparent dips may result when the points of the primitive lie along a straight line in 3-D space.