TEXTURE ANALYSIS TECHNIQUES

FOR MULTI-SPECTRAL CLOUD CLASSIFICATION

by

ZHI-QIANG GU

Master of Philosophy

University of Edinburgh

1988

EX BIBL. UNIV. EDINBURGH

ABSTRACT

To find more accurate texture descriptors and to reveal clearly the importance of texture analysis techniques in aiding multi-spectral cloud classification, this study was designed to develop some new textural features and compare the relative merits of these newly-developed features and some widely-used textural features in multi-dimensional feature space. The textural features considered in this study include both spatial and frequency features. Spatial features are mainly those based on spatial grey level difference statistics and circular Moran autocorrelation features, and frequency features are those based on summed energies of polar coordinate Fourier power spectra and entropy-based measures of the spatial distribution of frequency entries in the polar spectra. Some other textural features such as the Roberts gradient measure and local variance were also considered. The work was performed with TIROS-N AVHRR image data acquired in the late spring of 1979 over areas near the British Isles. A cluster analysis was performed using the three-channel spectral images. Class statistics arising from the analysis were iteratively optimized and then used to form a data set to assess the separability of 13 cloud classes which were identified in the analysis, and so to evaluate various textural features. The results from the evaluation, mostly given in terms of the multivariate transformed divergences, are discussed, and corresponding conclusions are drawn to show which individual and group of the textural features considered in this study appear the most suitable for aiding multi-spectral cloud classification in several different cases. Some other problems in achieving high classification accuracies are also tackled.

TABLE OF CONTENTS

1 INTRODUCTION 1

1.1 Background 1
1.2 Statement of Problem 2
1.3 Texture Description 5
1.4 Summary of Chapters 8

2 REVIEW OF PREVIOUS CLOUD CLASSIFICATION STUDIES 10

2.1 Cloud Observations 10
2.2 Nephanalysis 11
2.3 Computer-assisted Cloud Classification 13
2.3.1 Conover's Criteria and Cloud Field Description 13
2.3.2 Texture Analysis Techniques 16
2.4 Summary 25

3 DATA AND PREPROCESSING 29

3.1 TIROS-N Satellite and AVHRR Data 29
3.2 Ingestion of Data 33
3.3 Use of Low-level Image Processing Techniques 34

4 TRAINING STAGE 37

4.1 Introduction 38
4.1.1 Problem in Deriving Class Statistics 38
4.1.2 Cluster Analysis 41
4.2 Iterative Training Algorithm 44
4.2.1 Analysis of Cloud Type 44
4.2.2 Iterative Unsupervised Classification 55

5 TEXTURAL FEATURES 62

5.1 Spatial Features 63
5.1.1 Spatial Grey Level Difference Statistics (SGLDs) 63
5.1.2 Roberts Gradient (R.G.) 67
5.1.3 Autocorrelation Features 69
5.2 Spatial Frequency Features 72
5.2.1 Spectral Analysis 72
5.2.2 Normalized Average Amplitude Fourier Power Spectrum (NAAs) 76
5.2.3 Entropy-based Measures in Polar Frequency Domain (POLENTs) 83

6 EVALUATION OF TEXTURAL FEATURES APPLIED TO AVHRR SATELLITE IMAGERY 89

6.1 Introduction to the Texture Analysis Algorithm 90
6.1.1 Calculation of Spectral and Textural Features 90
6.1.2 Feature Selection and Assessment of Class Separability 91
6.2 Separability Indices 94
6.2.1 Fisher Distance 95
6.2.2 The Transformed Divergence 104

6.3 Evaluation of Textural Features 109
6.3.1 Spatial Features 110
6.3.2 Spatial Frequency Features 130
6.3.3 Intercomparison of the Textural Features 142
6.4 Summary and Discussion 153

7 SUMMARY **157**

7.1 Aims of the Research 157
7.2 Contributions of the Research 159
7.3 Proposals For Future Research 161

I ACKNOWLEDGEMENTS **163**

II REFERENCES **164**

CHAPTER 1

INTRODUCTION

1.1. Background

Clouds are the most distinctive feature of the Earth atmosphere as seen from space. They contribute to both the planetary reflectivity and emitted terrestrial radiation, and indicate several atmospheric processes on synoptic scale or smaller scales. They have traditionally been characterized by defining the fractional cloud cover and by subjective verbal descriptions of cloud shapes that were defined by L.Howard in 1803. These schemes are inappropriate in defining large-scale cloud field properties such as are required, for example, in dynamical models of the atmosphere (Chin and Jau, 1987). Many factors such as cloud size, shape, texture and internal variability need to be characterized in terms of a limited number of parameters that describe these morphological characteristics.

Satellite data give a unique perspective view of the Earth's atmosphere. A vast number of TIROS-N/NOAA images have been accumulated since the launching of TIROS-1. There is a considerable amount of information contained in these pictures which has unfortunately not been fully utilized. One of the major reasons for this is the difficulty of quantifying the information contained in the pictures in a meaningful form. Extracting cloud field information as well as other information from these digital images by image interpretation techniques is a quantitative approach to the problem.

Cloud mapping is a direct application of remote sensing to meteorology and climatology. It arose from the need for cloud classification. Operational classification products derived from digital images have been used for various purposes. Among the most important approaches are that of using classified

pictures to guide forecasters when interpreting satellite images, deriving cloud and precipitation fields for use in numerical forecasting models (e.g., Desbois *et al.*, 1982), and extracting a limited amount of information from satellite pictures that can be archived for climatological purposes.

1.2. Statement of Problem

Generally, cloud classification algorithms fall into two broad categories. One uses radiative transfer theory, in which cloud field properties are calculated from a model of the atmosphere, surface and clouds. The other applies statistical pattern recognition techniques, which is the concern of this study and is discussed further in the following paragraphs. Moreover, some researchers have also used a bi-spectral threshold method or a hybrid bi-spectral threshold method, which incorporates a radiative transfer model (see Rossow *et al.* 1985 for general discussion).

Statistical pattern recognition algorithms consist of several procedures, which include the choice of features and the choice of decision strategies. The term feature in this thesis is used to refer mainly to one of the measurements from satellite radiance images (with some exceptions, which will be noted when used). The former as the concern of this study is primarily designed to find more accurate features from available spectral information for different cloud classes of interest (e.g., Booth, 1973; Harris and Barrett, 1978; Bunting, 1980; Ebert, 1987), and the latter to find optimal statistical classifiers (e.g., Darling and Joseph, 1968).

In remote sensing, there are basically two kinds of features: spectral features and textural features. Spectral features comprise a set of values for the

reflectivity and emissivity of objects such as clouds measured over specific wavelength intervals. These features have been widely used to understand and characterize the appearance of clouds on remotely sensed imagery. Parikh (1977, 1978), Desbois *et al.* (1982) and Liljas (1984, 1986) employed these respectively in terms of radiance histogram statistics in their classification studies using multi-spectral threshold methods or statistical histogram analysis. A comprehensive discussion was given by Rossow *et al.* (1985) of an intercomparison of some previous studies in the International Satellite Cloud Climatology Project (ISCCP). All these studies showed the problem of the inadequate specification of cloud properties by spectral features in multi-spectral feature space. These problems are in essence attributed to the fact that the spectral features are dependent on the spectral distribution of physical properties of clouds, radiant flux onto clouds, geometric relationships between incident energy and sensor angle-of-view and atmospheric effects. The improvement of understanding of clouds in general and of satellite observations of clouds in particular, which was gained from previous studies, leads naturally to the study of what other cloud properties can be observed from space, or what can be retrieved from currently available spectral information. The former is beyond the scope of this study. The latter, i.e., information retrieval, is the major concern. Since multi-spectral information is not adequate, accurate identification of cloud types in the atmosphere requires the use of other elements of image interpretation in addition to the spectral response.

The question of what other features in satellite images may aid cloud field interpretation has increasingly drawn attention in the last two decades. To interpret the images manually, six qualitative criteria were suggested by Conover (1962, 1963) for trained meteorologists to produce cloud charts. These criteria are brightness, texture, size, shape, organization and shadow effects. Using the first two is the most attractive idea for the quantitative interpretation.

Among the techniques of image interpretation, texture analysis is an important aspect. Texture is often defined as a structure composed of a large number of more or less ordered similar elements or patterns without one of these drawing special attention (Van Gool *et al*, 1985). The analysis proves to be successful to varying degrees in many applications, especially in applications of remote sensing.

The texture of various objects such as clouds can be characterized by different mathematical descriptors in either spatial or spatial frequency domain. However, which of these specifications the characteristics of cloud fields conform best to remains unknown so far, since cloud fields tend to be considered randomly distributed due to the complexity of their visual patterns. In fact, they are not truly a purely stochastic process, and they are also not strictly ordered arrays of identical subpatterns. Satellite images reveal that cloud fields, even at the mesoscale (100km x 100km areas), are not always organized in a random fashion (Garand and Weinman, 1986). They actually lie between deterministic and stochastic models. Thus, it is believed by people working on cloud classification, such as Liljas (1986) and Ebert (1987), that most of their patterns may be described by a set of spatial statistics that are extracted from a large ensemble of local picture properties. An example of how clouds are verbally described by texture specification is given in section 1.3. However, it has proved difficult to find an accurate and robust measure which may convert intuitive ideas of texture into a mathematical form (Harris, 1982).

In practice, researchers such as Barrett & Grant (1978) and Liljas (1984) showed the necessity and feasibility of using texture analysis to characterize cloud morphologies because they found that many clouds such as convective clouds and some other highly variable (in space) clouds cannot be separated by their multi-spectral signatures. Further research on cloud property spatial statistics was also recommended by the ISCCP (Rossow *et al*, 1985). Some

research studies in cloud classification, e.g., investigating the simple statistics "local variance" by Harris & Barrett (1978) and Karlsson (1987B), and applying grey level difference statistics by Ebert (1987), have all achieved some success. Therefore, from either the theoretical or experimental point of view, the quantification of cloud fields is feasible.

The objective of this study is to investigate the merits of both improved spatial-frequency-domain features and spatial-domain features, to compare them with the best of commonly-used textural features, and then to show their applicability in further cloud classification studies.

1.3. Texture Description

As described above, this study is concerned mainly with the analysis of cloud field images observed by TIROS-N satellite, and the extraction of some quantitative parameters that may describe the scene pattern for cloud type identification.

Textural information is the major concern of the study. To show how textural clouds can be well described by texture specification, an example is quoted from the handbook written by Anderson and Veltishchev (1973) to form a table illustrating qualitatively the characterization of a specific cloud class - Cumulus. In fact, many of cloud patterns are much more complex than this specific example and difficult to describe verbally.

The illustration is given in Table 1.1, and some explanations are given below.

Cumuliform clouds are one of the important cloud types which are indicative of atmospheric convective activity. This type of cloud has been investigated by

researchers either using mathematical models (e.g., Krishnamurti, 1975) to find the relationships of "open cell" or "closed" cellular cloud patterns to large-scale sinking or rising motion, or using stochastic process (e.g., Garand and Weinman, 1986) to model cloud field morphology. The difficulties in characterizing the cloud type are obvious, since the appearance of the cloud pattern is extremely changeable (Wallace, 1977). While the clouds are young and growing they contain much fine detail and their boundaries are sharp, appearing as a typical cauliflower. At this stage they consist mainly of liquid water. Then, the clouds will become increasingly diffuse as the concentrations of ice particles increases.

In visible images, Cumulus appears as a cloud field made up of groups of irregularly shaped elements of varying size. These cloud fields are frequently organized into lines, or cellular patterns. Some detailed textural descriptions of these two patterns are shown in Table 1.1. In infrared images, the Cumulus appears moderately grey but the appearance can indicate its height.

Substantial changes in the appearance and diffuse patterns discussed above often cause the problem that many cumuliform clouds cannot be separated by multi-spectral features. Moreover, the temperature of young Cumulus cloud tops is quite close to the Earth's surface temperature causing this type of cloud to be difficult to detect.

From Table 1.1, we are more confident that cumuliform clouds, and other clouds as well, can be described qualitatively using somewhat definite specifications. This forms the basis of the present study, which aims to find some suitable textural features which show quantitatively a recognizable pattern for some cloud morphological type.

Name of Regions	Cumulus(cell-shaped)	Cumulus(line-shaped)
Texture elements	small cells	lines (narrow)
Textural element size	1. 10 - 30 km across cell; 2. 2 - 10 km across individual clouds.	varying size
Spatial Relationships between elements	1. randomly distributed (cellular); 2. pattern depends on weather but similar all the time; 3. partially covered.	1. dense; 2. roughly parallel lines.
Boundaries of elements	irregularly shaped(cell).	edge fuzzy
Geometric description of elements	1. highly textural; 2. cell-shaped; 3. randomly distributed.	1. highly textural; 2. line-shaped; 3. linear and directional.
Expected contrast	edge gradient is relatively small.	1. high contrast in visible; 2. low contrast in infrared.

Table 1.1 Cloud texture description

1.4. Summary of Chapters

A brief summary of each chapter is given below.

Chapter 1 is an introduction to the thesis. It describes the background to this study and outlines the particular problem tackled. An example of how a cloud type is qualitatively specified by texture terminology is also given in this chapter.

Chapter 2 surveys the literature in the field covered by the thesis and gives a historical review of previous work in cloud classification.

Chapter 3 gives an introduction to AVHRR satellite radiance imagery and describes the choice of data set and a series of preprocessing steps: data ingestion, calibration and normalization as well as the use of low-level image processing techniques for the training stage.

Chapter 4 describes some practical problems tackled in the training stage and the fundamentals of cluster analysis algorithm. An iterative training procedure, in which the cluster analysis is a major part, is then illustrated with examples.

Chapter 5 describes the five groups of textural features used in this study. The modification of some textural features is also discussed in this chapter.

Chapter 6 evaluates the different groups of textural features applied to TIROS-N AVHRR data by a number of experiments using a multivariate divergence measure and shows which individual and group appear the most suitable for aiding multi-spectral cloud classification based on the separability index. The comparisons between feature pairs for each group and the comparison between the one-dimensional Fisher distance and the divergence measure are also given in this chapter.

Chapter 7 summarizes the thesis and proposes some areas for future research and development.

CHAPTER 2

REVIEW OF PREVIOUS CLOUD CLASSIFICATION STUDIES

The importance of cloud observations in weather monitoring and forecasting has been briefly described in Chapter 1 as a background. To understand fully the usefulness of the observations in meteorology and climatology and then the application of computer-assisted interpretation techniques applied to satellite sensing, a historical review is given in this chapter of satellite observations of clouds and the development of cloud classification techniques.

2.1. Cloud Observations

The observation of clouds is very important to meteorological and climatological studies. The importance can be seen from the following two aspects.

Firstly, the strong interaction of clouds with terrestrial infrared radiation greatly influences the climate. There are two effects that clouds have on the planetary radiation budget. One is that they reflect back into space some of the incoming solar short-wave radiation incident upon them due to their high albedo, and the other is that they absorb and radiate planetary long-wave radiation, thus contributing to the greenhouse effect. It has been widely recognized that the radiative properties of clouds need to be better known and understood (GARP/JOC, 1978). It is generally believed that for low- and middle-level clouds the cloud albedo effect dominates, whereas for cirrus clouds the cloud greenhouse effect dominates. Which effect takes overall precedence will depend on the amount, type and height of cloud present. Thus,

to know fully the effects of clouds on the global radiation budget it is necessary at least to know the amount, type and height of clouds on a global scale. In addition, cloud shape, fractional cover and the cloud field morphology must all be specified if people try to adequately compute the radiative heating rate profiles in cloud filled atmospheres (Harshvardhan, 1982; Welch and Wielicke, 1984).

Secondly, the appearance of clouds can clearly reveal important atmospheric processes. In satellite images, each cloud type has a characteristic pattern and brightness or temperature. Thus, by analysing the type of clouds, meteorologists can locate fronts, squall lines, jet streams, troughs and ridges, and further they can assess the stage and trend of development in mid-latitude and tropical storm systems. In many cases, it is also possible to infer the presence of turbulence, orientation of surface and upper-level winds, and the atmospheric stability (Colwell, 1983).

2.2. Nephanalysis

As discussed above and indicated by Barrett (1974), the most important cloud information in satellite observations is the form and distribution of the cloud. The most widely-used way of reducing the satellite recorded cloudiness is in nephanalyses. This simplified representation of cloud field information depicts the dominant cloud types and distributions, as well as other information for short-term weather forecasting.

Nephanalyses are often made from two major data sources: surface observations and satellite observations. In surface observations, three principal properties, cloud amount, cloud type, cloud height, are estimated. They have

been mapped on man-made nephanalyses for many years with other cloud properties. However, because surface observers only see the sides and bases of clouds, they often overestimate the amount of the cloud. The estimation is also subjective. Surface observations of the other two cloud properties are also poor, especially during night time. Therefore, nephanalyses made from surface observations are not adequate but can be used when people compare cloud climatologies derived from different source data.

Meteorological satellites provide a synoptic view of the atmosphere, and so update the surface-based nephanalyses. Since a visual summary of cloud information contained in satellite images is often adequate for many climatological and meteorological purposes, these images have been interpreted by human analysts in operational services to produce man-made satellite nephanalyses. Observations marked on these nephanalyses are cloud amount, cloud-top heights, and cloud types, precipitation, weather. But, since there is no standard procedure for such image interpretation, these nephanalyses are often too subjective or too dependent upon each interpreter's experience. Therefore, many experiments have been designed to estimate objectively cloud amount, either using spectral radiance data in some specific wavelength intervals (Blankenship, 1962), or using more sophisticated histogram weighting procedure (Reynolds and Vonder Haar, 1977). Cloud-top heights can also be readily derived from infrared cloud images by combining the calculation of the equivalent black-body temperature of cloud-top surfaces and a knowledge of the atmosphere's temperature structure. The knowledge of the structure may be obtained from radiosonde observations. Actually, the information about cloud heights is embodied in a first-order spectral feature, e.g., the brightness temperature of image pixels. The information is reflected by the "local mean" in this study, which is calculated from a grid of radiance data.

The estimation of cloud type has met many difficulties due to the high

variability of atmospheric data so that no procedure for analysing cloud type has been widely accepted. The next section gives a broad review of the procedures used in the past for cloud type identification.

2.3. Computer-assisted Cloud Classification

2.3.1. Conover's Criteria and Cloud Field Description

An automatic algorithm for cloud type identification requires definitive mathematical descriptions of cloud types. In fact, these descriptions do not currently exist and the problem has been investigated by researchers in the past. Six criteria, suggested by Conover (1962, 1963), are widely accepted by satellite meteorologists to assess cloud types (e.g., Anderson *et al.*, 1969), which are brightness, texture, pattern, shadow effects, size, shape. Although they are very qualitative and can only be subjectively used, they really provide guidelines for further quantitative studies.

In a satellite image, the value of the first element of the six, brightness, is proportional to the intensity of radiation received within some specific wavelength regions from the underlying atmosphere and surface. The grey level of each image element, i.e., pixel, is the brightness value of such a resolution area. Meteorological satellites primarily sense in visible and infrared wavelength regions. In this study, another spectral region, called Near-Infrared region, was also employed. In the visible region, the brightness values depend on illumination, scene geometry (incident, viewing and relative angles) and cloud reflectivity. In the infrared region, the brightness refers to the intensity of emitted radiation from clouds and surfaces, and thus is given another name, brightness temperature (dependent on the temperature and emissivity). The

cloud reflectivity is, in turn, related to the thickness, droplet size distribution, liquid water content of clouds, solar elevation and the characteristics of the clouds' upper surface. It varies directly with the thickness and liquid content, and inversely with the dropsize. It has been well-recognized that the brightness is a complex element, and quite often, it is meaningful only in a comparative sense (Barrett and Martin, 1981). The two infrared spectral regions usually give more information concerning cloud height and amount.

In pattern recognition theory, pattern is considered as the spatial arrangement of natural or constructed objects. For example, the repetition of certain general forms or relationships can give these objects a pattern that aids a corresponding manual interpretation. Cloud elements visible in a satellite image are usually either randomly distributed or organized into mesoscale forms. These arrangements can be considered as a cloud pattern. Spatial features in cloud patterns have been extensively used, for example, to subjectively determine the circulation fields over land and sea, and to explain current or impending changes such as the development of waves along a front in response to an approaching upper tropospheric vorticity maximum (Colwell, 1983). Unfortunately, there is no appropriate mathematical description for these patterns so far. The exact characteristics depend strongly on the field of applications.

Shadow effects are an important criterion for human analysts interpreting outdoor scenes. The shape or outline of a shadow offers a profile view of objects. In meteorological cases, these effects usually reveal a three-dimensional cloud structure as they can show the relative heights of different cloud layers. But shadows also have some detrimental effects on image interpretation. The obvious problem is that natural objects within shadows reflect little light and so are difficult to recognize. Because of this problem, some clouds of small size or complex structure such as convective clouds may

be misinterpreted. Thus, the shadow effects can only be used in manual interpretation by trained meteorologists.

Size and shape are the geometric specifications of objects. The size of an object is considered in the context of a picture scale. Shape refers to the general form or outline of individual objects. The size and shape of some objects are sometimes so distinctive that their images can be interpreted by these criteria. Unfortunately, all cloud categories vary greatly in size and shape. Their patterns are often arranged in straight or curved bands, or appear ragged and diffuse. Thus, there have been as yet few attempts in using these two criteria in cloud classification.

Comparatively, the criterion, texture, appears as the most attractive among the other five besides brightness as it has been widely accepted in applying remote sensing to many other areas such as geography and geology. It is defined as a structure composed of a large number of more or less ordered similar elements or patterns without one of these drawing special attentions (Van Gool *et al*, 1985). It is also considered as the frequency of grey level change. In essence, it is a product of the individual shape, size, pattern, shadow and brightness. As the scale of an image is reduced, the texture of any given object will become finer and finally disappear. This fact forms one of the final proposals of the study discussed in Chapter 7.

Cloud fields can be verbally described by the criterion texture. For example, a cloud area can be considered to have low texture when its top surface has a smooth appearance with little variation in its thickness. Similarly, a cloud area can be considered highly textured when its top surface has a rough appearance or appears ragged. An example of using texture terminology to qualitatively describe Cumulus is given in Section 1.3.

However, it has been shown by many researchers such as Harris & Barrett

(1978), Parikh (1978), Parikh & Rosenfeld (1978), that it is difficult to find an accurate measure which can be used to convert an intuitive idea of the criterion, texture, into a mathematical form. In practice, many textural features have been designed and tested by other researchers on standard texture patterns, or for cloud classification studies, and so may be considered as potential candidates for cloud classification. Section 2.3.2 aims to give a historical review of these previous cloud classification studies using various mathematical specifications of cloud texture.

2.3.2. Texture Analysis Techniques

The definition of a texture has been given in Section 2.3.1. The definition looks rather qualitative. A clearer description of a texture is given by defining its two major dimensions: coarseness and directionality. By this definition, a number of textural feature candidates may be used to characterize these two dimensions. For example, a popular way of measuring coarseness is to see it as a spatial frequency distribution, which is based on the assumption that fine textures are rich in high frequencies while coarse textures are rich in low frequencies. The other dimension, directionality, can be measured by phase information. Moreover, textures with relatively regular structure can be described by primitive extraction methods (Van Gool *et al*, 1985).

Generally speaking, there are two categories of texture analysis approaches: statistical and structural. The former is applicable to fine textures which have no regular spatial arrangement but obey some statistical laws, and show various quantitative features of interpixel relationships. Many applications (e.g., Ebert, 1987; Harris, 1982; Karlsson, 1987B) have shown that these features are useful for cloud classification. Structural approach deals with some textures based on

the assumption that they are made up of a set of texture elements arranged with some regular placement rules.

The use of cloud classification techniques using textural features for cloud classification dates back to the 1960s when Rosenfeld *et al.* (1965) and Darling & Joseph (1968) applied feature extraction techniques to the visible pictures obtained from the TIROS and NIMBUS satellites. Their experimental results showed the considerable potential of textural features for future use in cloud classification.

When Nimbus-2 was put into operational use in 1966, the application of multi-spectral classification techniques became feasible. A substantial amount of effort using first-order spectral statistics for multi-spectral cloud classification gave many researchers an incentive to applying image analysis techniques for more accurate information retrieval from satellite images. Since then, many research studies have been performed on radiance data obtained from different satellites. Some of these studies are reviewed in this section, together with comments.

2.3.2.1. Structural Approaches

Structural approaches are based on the assumption that texture is made up of primitives which appear in nearly regular repetitive spatial arrangements. As defined by Haralick (1979), a primitive is a connected set of resolution cells characterized by a list of attributes. The simplest of the attributes is the grey level of an image pixel. By defining the grey level as a property, a primitive can be regarded as a maximally connected set of pixels having the same grey level attribute. Thus, primitives can be generated by grouping image pixels or splitting an image on the basis of their individual properties such as grey level or edge

direction. To describe the texture, we must describe the primitives and placement rules (Haralick, 1979).

Some structural approaches have been widely accepted in applications, for example, methods based on placement rules or primitive extraction or both (Conners and Harlow, 1980B; Wang *et al*, 1981), and syntactic approach (Lu and Wu, 1978).

Cloud fields are so diffuse that the assumption for them cannot be made and justified adequately. A typical cloud field, Cumulus clouds appearing relatively structural compared to other cloud classes, is shown in Table 1.1 of Chapter 1. From the table, we may convince ourselves that it is very difficult to find out which attributes can describe primitives within the cloud fields and what placement rules these primitives appear in, at least until now. Thus, structural approaches were not used in this study.

However, we may get some impression from the table that the cloud fields have in fact some structures, for example, at mesoscale (Garand and Weinman, 1986), but are actually difficult to describe in a mathematical form. This leads to the idea of using some structural-statistical approach to quantifying the cloud fields. The hybrid approach was discussed by Haralick (1979), and an experiment was performed by Garand and Weinman (1986) using a stochastic process.

2.3.2.2. Statistical Approaches

Statistical approaches are generally applicable to many textural cases. On a statistical level a texture is defined by a set of statistics which are extracted from a large ensemble of local picture properties (Van Gool *et al*, 1985). Some first-order statistics such as mean and variance may be used to classify a

limited set of textures. Nevertheless, human beings are sensitive to second-order statistics (Julesz, 1975). Examples of such second-order statistics are the grey level co-occurrence probability matrix and the grey level difference statistics (Van Gool *et al*, 1985) (The latter is discussed in Section 5.1.1).

A fundamental problem is often neglected in the initial consideration of applying texture analysis techniques for cloud classification, which is the appropriate scale at which textural features should be measured. The scale is affected by two factors: the resolution level of the observations and the sample window size. The former is related to the extent to which image characteristics can be measured. It is obvious that a texture with fine structure can become coarser if we look at it with greater magnification, i.e., with increased resolution. Some research studies have been directed towards a resolution-insensitive characteristics of texture, e.g., by Chetverikov (1981) using a "texture anisotropy indicatrix". The sample window size also has a strong effect on the measurement of image characteristics. When textural features are calculated on a small grid of image data they are often not reliable, but if we use a larger grid, we will find that it is not easy to find regions that can be considered uniformly textured. There have also been many research studies to overcome these difficulties, e.g., by Pietkainan and Rosenfeld (1981). The problem is considered in Chapter 4.

There are eight groups of statistical textural features, which have been widely accepted in the measurement and characterization of image textures. These are autocorrelation, optical transforms, digital transforms, textural edgeness, structural elements, spatial grey level co-occurrence probability matrix, grey level run length and autoregression models (Haralick, 1979). The first three of these approaches are related in that they all measure spatial frequency directly or indirectly. Of these eight approaches, four are given a comprehensive review respectively in this section, and were directly used or

modified in this study. These four are autocorrelation, digital transforms, textural edgeness and grey level co-occurrence probability matrix (An alternative of this approach was actually used in this study: See Section 5.1.1 for explanation). The other four were not considered in this study due to the fact that they have not been used so far in application of remote sensing to meteorology and also they rarely appear in other applications.

As indicated above, the measurement of spatial frequency distribution is a straightforward way for texture analysis. Two major dimensions of textures which can be measured by the spatial frequency distribution are their coarseness and directionality. Another is the amount of edge per unit area, since coarse textures have a small number of edges per unit area while fine textures have a large number of edges per unit area (Haralick, 1979). Moreover, the size of primitives is also an important dimension of texture (for the definition of a primitive, see Section 2.3.2.1).

Autocorrelation is a function which measures the size of grey level primitives. As shown by Haralick (1979), grey level primitives of large size are indicative of coarse textures while grey level primitives of small size are indicative of fine textures. In essence, the correlation coefficients of autocorrelation functions measure the linear dependence that one image pixel has on another. The relationship between autocorrelation and Fourier power spectral density functions is well known. Some studies of these functions have been performed in other applications, e.g., by Unser and Coulon (1982) in an automatic visual inspection system of texture.

The application of autocorrelation functions to meteorological satellite images was made by Katz (1965), and more recently, by Bunting and Fournier (1980). In Bunting and Fournier (1980)'s work, they also used several other textural measures to detect small-scale clouds. In their experiments, they

computed the autocorrelation functions for a series of two-dimensional data by simply shifting the series along scan lines by a given lag. The process was carried out at lags 1, 2, 3, ..., 10. The results of the experiments showed that this simple definition of autocorrelation functions gave results highly correlated with those given by an FFT power spectrum, and suggested that this information would appear redundant when FFT measures were used. This conclusion is obvious since the autocorrelation and the FFT power spectrum are the transforms of each other.

There are several digital transform methods: Fourier transforms, Hadamard transforms and Slant transforms etc (A discussion of the first two is given in Rosenfeld & Kak (1982), and the last in Pratt *et al.* (1974)), but, little difference could be found when they were applied to textural classification (Kirvida, 1976). Thus, in many cases, the use of the Fourier transform alone is adequate.

Similarly to autocorrelation functions, the digital Fourier transform measures the regularly repetitive primitives of texture by a number of spatial frequency entries in the power spectrum. The Fourier transform uses the sine-cosine basis set of functions. The new coordinate system generated by the Fourier transform has an interpretation that relates to spatial frequency. This frequency is a useful representation of textures.

Applying the Fourier transform to a given time series data to estimate its spectral density function is named spectral analysis. Spectral analysis is capable of revealing the features of many natural or constructed patterns in practice. In meteorology, a pioneering work was done by Leese and Epstein (1963) using two-dimensional spectral analysis to quantify satellite cloud images in order to show atmospheric waves based on satellite observations of vertical motions and displacement of the air evidenced by cellular convective clouds. Their experiment showed that the technique could not only separate the

periodic features from a substantial amount of noise but also reveal patterns which tended to be obscured by dominant features. This enabled one to differentiate different types of cloud patterns.

In 1973, Booth applied a similar two-dimensional spectral analysis to the high resolution scanning radiometer data from the polar orbiting NOAA-1 environmental satellite. It was found in his experiment that when used separately, the spatial frequency features in infrared imagery provided better overall classification accuracy, and the overall classification accuracy approaches both visible and infrared data classification accuracy in all clouds except Cirrus. The detection of Cirrus required dual-channel data. Bunting and Fournier (1980) applied a similar approach to the Defense Meteorological Satellite Program (DMSP) fine mode satellite data, and showed that multi-spectral classification accuracies could not be significantly improved when two-dimensional Fourier transform features were added. Lately, this approach was examined by Ebert (1987) using a one-dimensional Fisher distance measure. This approach and its extension used in this study are discussed in Chapter 5.

Different from the spatial frequency approach is another called edgeness. By this approach, texture can be conceived (e.g., by Rosenfeld and Thurston (1971)) in terms of edgeness per unit area. An edge passing through an image pixel can be detected by comparing the values for some local properties obtained in pairs of nonoverlapping neighbourhoods bordering the pixel (Haralick, 1979). As shown by Rosenfeld and Thurston (1971), visible edges in an image are often associated with abrupt changes in average grey level. Such changes can vary greatly in degree of size; at one extreme, two adjacent points having different grey levels define a micro-edge, while at the other extreme, two large coarsely textural regions having average grey levels can define a sharp edge if the transition between them is abrupt. In order to avoid detecting such micro-edges, A window size has to be selected, large enough with respect to

the extent to which image characteristics can be measured. One of the local properties is the nonlinear Roberts Gradient, which is defined as the sum of the absolute value of the difference between diagonally opposite neighbouring pixels. Thus, a textural feature can be obtained by calculating firstly the gradient and then the average value of the gradient in a sample window. Other linear edge operators also exist such as the compass gradient masks (Pratt, 1978).

The application of the Roberts Gradient as a textural feature for cloud classification was first performed by Parikh and Ball (1980). In their experiment, they used the textural feature for SMS-1 (Synchronous Meteorological Satellite-1) satellite data to classify the cloud images into five cloud classes. Several problems were outlined, mainly that the differentiation of low-level clouds form middle-level clouds is often not possible; high-level clouds with significantly lower clouds and Cumulonimbus are difficult to differentiate; High clouds can be confused with low clouds because of high transmissivity of the cloud layers, e.g., thin cirrus layers. Wu *et al.* (1985) and Ebert (1987) also used the Roberts gradient for the determination of rainfall rates from GOES satellites, and for cloud classification in polar regions, respectively.

Perhaps, the most widely-used statistical group of textural features for natural scene analysis is the spatial grey level co-occurrence probability matrix. This method is based on the estimation of the second-order joint conditional probability density functions, $f(\mathbf{i},\mathbf{j}/\mathbf{d},\theta)$. Each $f(\mathbf{i},\mathbf{j}/\mathbf{d},\theta)$ is the probability of going from grey level **i** to grey level **j**, given the intersample spacing is **d** in the direction given by the angle θ. This group of textural features is second-order for visual discrimination, which is compatible with the way the human visual system differentiates textural objects (Julesz, 1975). It is actually a context-related process, as dependences between pixels are taken into account explicitly. It characterizes primarily the spatial interrelationships of individual pixels in an image.

To avoid a number of problems encountered when calculating the spatial grey level co-occurrence probability matrix, an alternative was given by Weszka *et al.* (1976), called spatial grey level histogram difference statistics. These two are actually comparable in performance (Weszka *et al.*, 1976). The textural features derived from this group of statistics measure factors such as the amount of local variation within a grid of data and the overall homogeneity within it. From the group of statistics, a number of textural features can be defined. Four of them have been popularly used in experiments such as land and cloud mapping (Weszka *et al.*, 1976; Ebert, 1987) in the past. These are Mean, Angular Second Moment (ASM), Contrast (CON), and Entropy (ENT). A detailed mathematical description of these statistics is given in Chapter 5.

Parikh (1977) applied some of these spatial textural features to NOAA-1 data (approximately 4 n.mi. resolution at nadir in infrared images, and 2 n.mi. resolution at nadir in visible images) for a comparative study of cloud classification techniques. In her study, she used the one-dimensional Fisher distance to determine the best choice of distance between neighbouring pixels, direction between neighbouring pixels, and various textural features. A comparison of the Fisher distance values was based on neighbouring pixels at distance 1, distance 2 (separated by one pixel), distance 3 and distance 4. The results showed that distance 1 was preferable. Among these textural features, infrared entropy proved superior for the separation of low clouds from Cumulonimbus, and visible entropy and ASM were the best textural features for discrimination of Cirrus from other high cloud categories.

Seddon (1983) used the same approach, and demonstrated the ability of the visible left-diagonal ASM to distinguish between Cirrostratus and Cirrocumulus with Meteosat imagery. Ebert (1987) showed the potential of maximum entropy in both visible and infrared images for cloud classification studies in polar regions.

2.4. Summary

From the review of texture analysis techniques in section 2.3.2, it was found that there are a number of problems which exist in cloud classification studies and which strongly affect the subject's development. Some major problems, which have been realized so far, are outlined as follows:

1. As pointed out by Harris (1982), there is a lack of consistency between the available techniques for either cloud classification or feature extraction and for the presentation of final results. No agreement exists to indicate what constitutes a cloud classification scheme. Thus, it is difficult to compare the final results.

2. There is no standard data base that people may use for the assessment of cloud classification accuracies and the evaluation of either spectral or textural features.

3. The choice of cloud classes is often arbitrary when a supervised classification scheme is used. It depends upon the application of interest, expertise of human analyst and archived data sets. This also makes it difficult to compare the results of cloud classification and feature extraction.

In addition, there are a number of specific fundamental problems in applying texture analysis to cloud classification. These problems also appear as obstacles to the progress of texture analysis studies for cloud classification:

1. It has been shown that textural features can aid multi-spectral cloud classification since multi-spectral features are not adequate in representing some of highly textural cloud classes. As texture analysis techniques are often developed on a standard data base (e.g., Brodatz, 1966) as an independent discipline and textural features representing cloud classes have been always selected from studies of the discipline, verbal cloud descriptions such as the subjective classification scheme used by satellite meteorologist tend not to be taken into account.

2. There is no agreement on what scale is appropriate at which cloud textural features should be calculated. Scale selection often depends upon the subjective judgement of cloud coverage or the knowledge of horizontal distribution of cloud cover adopted from other meteorological and climatological studies, as well as computational convenience.

3. There is also no agreement on what separability measure should be used in cloud classification studies and how the usefulness of textural features can be assessed. This further causes difficulties in the comparison of feature extraction results. As discussed above, Parikh (1977, 1978), Seddon (1983), Wu *et al.* (1985) and Ebert (1987) all used the one-dimensional Fisher distance as the separability measure, which does not take into account the intercorrelation of spectral channels and textural features. A detailed evaluation of the separability measure is given in Section 6.2.1.

4. The previous research studies were all concentrated on

creating new textural features (e.g., Harris and Barrett, 1978; Bunting and Fournier, 1980), or selecting some optimal textural features from existing textural features individually (e.g., Wu *et al.*, 1985; Ebert, 1987) to aid multi-spectral cloud classification. No effort has been made so far to exploit the spatial information reflected by an appropriate combination of some existing textural features such as spatial grey level difference statistics for aiding multi-spectral cloud classification. In fact, some existing textural features, although not outstanding in performance for improving multi-spectral cloud classification accuracy when used alone, may provide a substantial amount of additional discriminating information and become powerful when they are appropriately combined and jointly used in multi-dimensional feature space. Unfortunately, the additional discriminating information has not been considered so far for cloud classification. The information that these features contain may be specially helpful in enhancing the distinction between some hard-to-separate cloud class pairs, or may be used in such a way that some textural features which contain the same information can be discarded.

It is certain that the solution of these problems would make a large step towards the maturity of cloud classification. Thus, only some of the problems can be tackled in this study. The objective of the study is mainly to assess the relative merits of two newly-developed groups of statistical textural features, and compare them with the well-known textural features described in Section 2.3.2 on the basis of a multivariate separability index to give a suggestion of

which textural features appear the most useful for cloud classification. Meanwhile, the multivariate separability index, the transformed divergence, is compared with the one-dimensional Fisher distance. The problem with the arbitrary choice of cloud classes is to be inherently avoided to some extent by using a cluster analysis. The investigation of appropriate combinations of some textural features is also carried out in this study to give a basic understanding of the usefulness of the feature combination for cloud classification.

CHAPTER 3

DATA AND PREPROCESSING

In this study different texture analysis techniques were developed and assessed using satellite imagery collected by the Advanced Very High Resolution Radiometer (AVHRR) instrument on board the TIROS-N polar orbiter. Some brief description of the AVHRR instrument and the data is given in Section 3.1, and several preprocessing procedures concerning data handling, calibration and enhancement are given in the other two sections.

3.1. TIROS-N Satellite and AVHRR Data

The TIROS-N satellite is a third-generation environmental satellite carrying instruments to support both day-to-day enviromental monitoring and global research programmes. It flies in a circular, polar, Sun-synchronous orbit with repeat cycles of about 10 days. It passes over an area near the British Isles twice every twenty-four hours, once northbound and once southbound. The satellite is operated by the National Environmental Satellite Service (NESDIS) of the NOAA.

The AVHRR instrument on board the TIROS-N satellite contains a four-channel radiometer, which are sensitive to visible, near-infrared and infrared radiation. The resolution (or equivalently, instantaneous field of view (IFOV)) at nadir is 1.1 km.

A general requirement for the image data used in this study is that the images are multi-spectral and the selected parts of them include as many cloud and surface classes of interest as possible, since the purpose of this study is to

assess the contributions of various texture analysis techniques in differentiating between cloud classes, especially those that are hard to separate.

Another requirement is the necessarily high resolution of the remotely sensed images, especially in our case, i.e., interpreting cloud texture patterns. Bunting and Fournier (1980) showed that satellite data of resolution 0.6 nautical mile (approximately 1.11 km) are adequate for demonstrating the feasibility of spectral analysis (time series) techniques and estimating the performance of an automatic classifier for cloud types. Seddon (1983) gave a similar result. The AVHRR data of resolution 1.1 km satisfy the requirement of Bunting and Fournier (1980).

The AVHRR data used in this study, provided by the Satellite Data Receiving Station at Dundee, are recorded on computer compatible tapes (CCTs). Each scanline is comprised in turn of 2048 image pixels. These pixels are represented on the CCTs as 10-bit digital radiance counts with values from 0 to 1023. Calibration coefficients are included in each scanline. A detailed description of the TIROS-N satellite, AVHRR instrument, data extraction and calibration can be found in Schwalb (1979) and Lauritson *et al.* (1979).

Two 2048x2048-pixel AVHRR images of three spectral channels (visible, Channel 1, near-infrared, Channel 3, and infrared, Channel 4) were used here, which cover a large area (0° to 20°W and 45° to 60°N) near the British Isles. Figures 3.1 and 3.2 show these two images in the infrared channel, which were acquired at 14 54 GMT on May 13th, 1979 and at 14 43 GMT on May 14th, 1979, respectively.

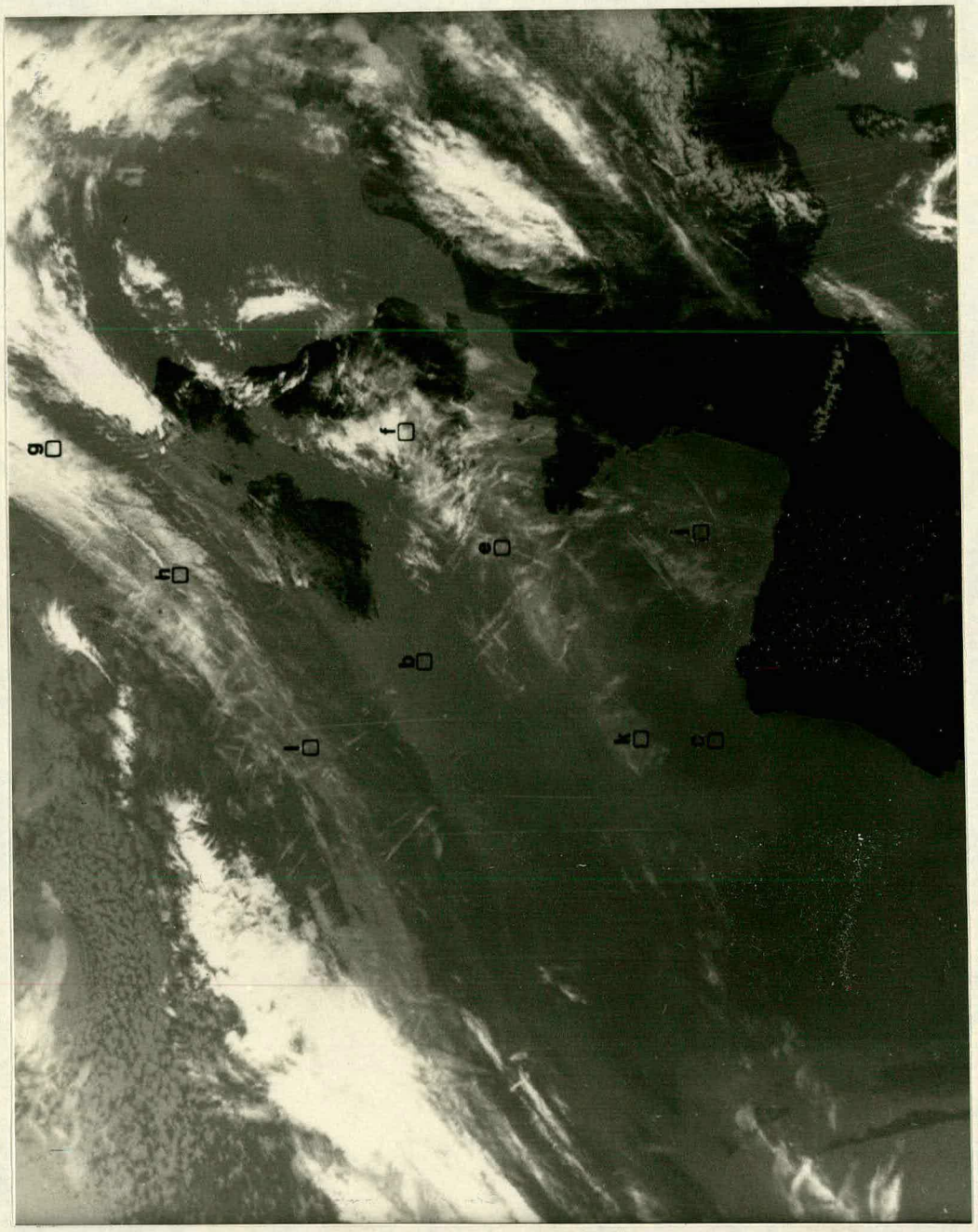


Figure 3.1 The infrared TIROS-N image of 14 54 GMT 13 May 1979 covering a large area near the British Isles. Each area marked by a letter (e.g., b, c) in the figure indicates one typical sample for the cloud class. Refer forward to Page 45 for cloud categories represented by these letters.

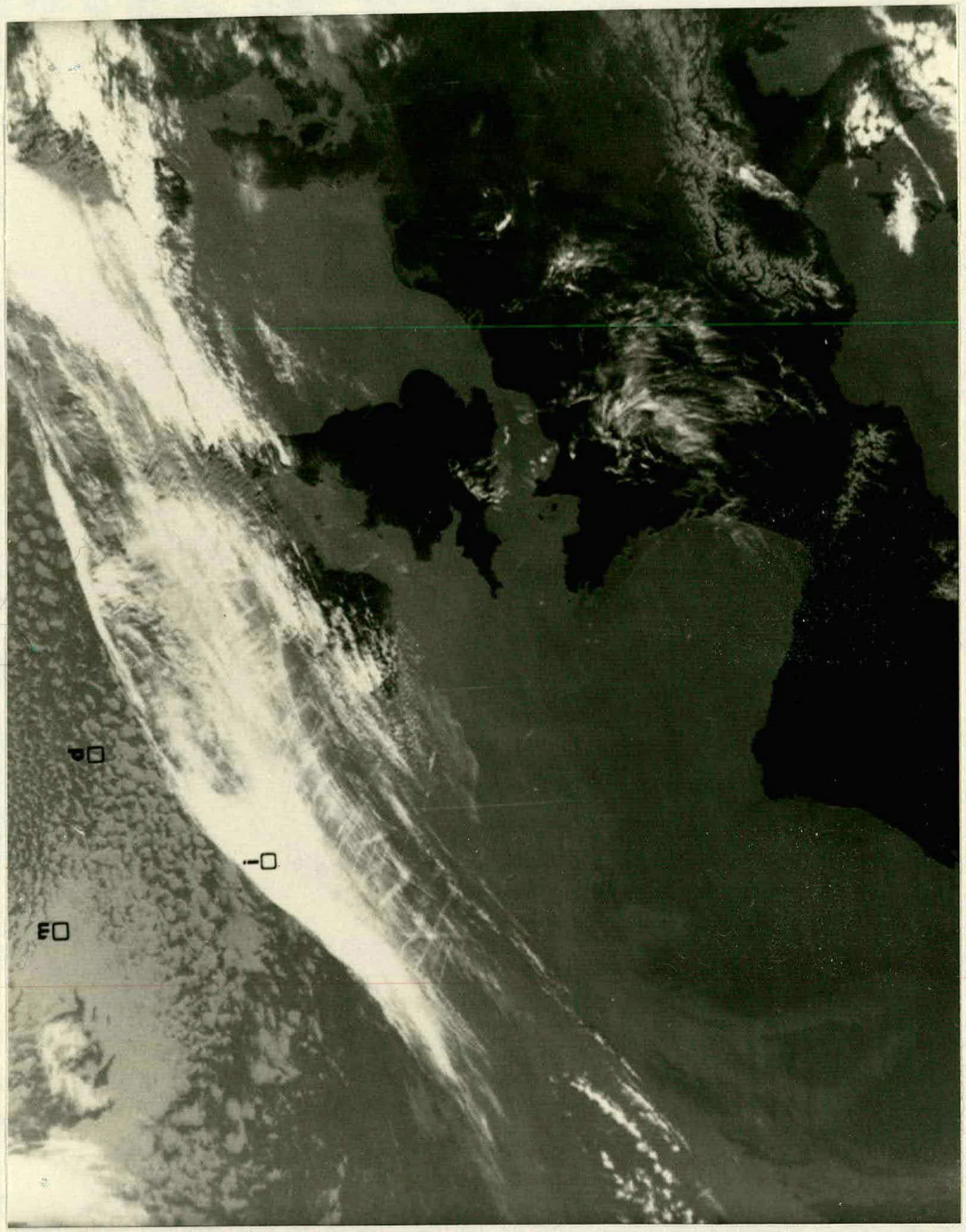


Figure 3.2 The infrared TIROS-N image of 14 43 GMT 14 May 1979 covering a large area near the British Isles. Each area marked by a letter (e.g., d, m) in the figure indicates one typical sample for the cloud class. Refer forward to Page 45 for cloud categories represented by these letters.

3.2. Ingestion of Data

The ingestion of AVHRR raw data from CCTs is an essential step in data preprocessing. This algorithm, together with data calibration and normalization, is implemented by the AVHRR data processing programs of the British Meteorological Office. The programs were slightly altered for use on the Edinburgh University Computing Service (EUCS) VAX 8500 computer.

The program AVHIN reads an AVHRR raw data tape and extracts the pre-specified areas of the required channels. The 10-bit images for each channel (up to five) are first written to data files. The first record written is the standard image header. The program allows the images to be sampled before they are written to disc. The image pixels are arranged in such a way that the first element of each scanline contains the most westerly pixel for an ascending pass and the most easterly pixel for a descending pass. The areas within the input images can be selected either by specifying starting lines, starting columns, number of lines and number of columns, or by specifying the latitudes and longitudes of the area centres as well as the sizes of the areas in lines and columns. For display and later use, the dynamic range of the acquired images is compressed from 10 to 8 bits by excluding the two least significant bits.

The programs AVHCAL and RTOALB were finally not used here to calibrate the AVHRR data. These procedures were skipped since the relative values of the data is all that are required for our purpose, i.e., comparing various texture analysis techniques. The normalization of visible radiance due to varying sun elevation angles was also omitted following an oral and subsequent written discussion with Karlsson (1987B) in which he pointed out that the normalization did not seem to appear helpful because the reflectivities of cloud surfaces are markedly non-Lambertian, and the variation depends very much on object classes. Instead, we chose satellite observations from two days which were

acquired at nearly the same time each day, and the training sampling was restricted to a narrow band of about 1200 km wide along the track of the raster data.

3.3. Use of Low-level Image Processing Techniques

The image processing and analysis algorithms of this study basically include four procedures: (1) data ingestion, (2) data preprocessing, (3) data analysis, and (4) data storage. The detailed description of each procedure can be found in Richards (1986) or Colwell (1983). The three procedures used in this study, except the third, are discussed in this chapter, and they are often referred to as low-level image processing techniques because they are mainly used to handle the data sets and improve the data quality, and their success depends only on the nature of the data while the third procedure data analysis is fundamentally knowledge-based. The third procedure constitutes a major part of this study. However, the other three procedures also play an essential role throughout. The first of these three procedures, data ingestion, was described in Section 3.2. The others are briefly described in this section.

Image enhancement techniques are an important class of the second procedure. Such techniques include geometric correction and grey level modification.

The former is used for correcting geometrical distortions present in a satellite image. These distortions include perspective distortion, which results when an image is taken from an oblique viewing angle; and barrel distortion, due to the limitation of optical imaging or electronic scanning systems.

Grey level modification is used for reducing pictorial degradations: noise,

attenuations, and low contrast between the image pixels of interest and background. By means of these techniques, the interesting part of the pictorial information can be illustrated in a more pronounced way and the background information can be suppressed or at least deemphasized. The use of these techniques varies with the particular interest of a special user.

A detailed mathematical treatment of these techniques is given by Rosenfeld & Kak (1982), and Rosenfeld (1976).

The geometrical distortion of AVHRR images, considered in this study, is perspective distortion. It is mainly caused by the scanning process and the Earth's curvature because the 2048x2048-pixel images cover an area of over 2300 km square. With an increase in the scanner's oblique viewing angle along scan lines from the satellite sub-point, the resolution of the images, 1.1 km at subtrack, starts to degrade towards the sides of the images, and eventually degrades to about 3.8 km at the farthest point of the scanlines.

Since a sufficient amount of AVHRR radiance data were available, no efforts were made to correct these distortions on both sides of the TIROS-N satellite track but the training sampling was restricted to an area of about 2300x1200 km along the track of the raster data.

The pictorial degradation of AVHRR images, considered in this study, is the very low visual contrast between the image pixels showing different cloud layers and surfaces. This problem necessitates the application of image enhancement techniques for visual evaluation of the radiance images. Histogram equalization technique was used in this study during the training stage (see Chapter 4). The algorithm is implemented using the image processing program library SPIDER which is written in FORTRAN (SPIDER working group, 1983). Two examples of the resultant images from the histogram equalization are shown in Figures 4.1 and 4.2 of Chapter 4. The impact of these low-level image

processing techniques on image classification needs to be further investigated.

For easy access, efficient storage and transfer, AVHRR images were stored in the Edinburgh University Computer Science Department standard image format IFF (run-length coding) and transferred in hexdecimal data form between different computer systems.

CHAPTER 4

TRAINING STAGE

The training stage is a critical step in computer-assisted cloud image interpretation. The general requirement for this stage is that it provide a reasonably valid data base for the interpretation procedure. The success of the interpretation relies directly on the quality of class statistics generated and delineated from this stage. It is a time-consuming iterative process involving a substantial amount of subjective judgement by a human analyst.

The training stage began with the careful selection of training fields that were representative examples of each cloud category. It required a thorough knowledge of the spectral characteristics of typical cloud classes being analysed (see Anderson and Veltishchev 1973 for discussion). These characteristics were evaluated with the help of an interactive image processing system, and by consulting a reference data source given by Liljas (1984) and surface weather charts.

There are a number of practical problems in deriving cloud class statistics. One of the major problems is that the determination of class classes tends to be too dependent on the application of interest, and the judgement and expertise of the analyst. Another is the appropriateness of the horizontal scale for calculating spectral and in particular, textural features. These two problems and the corresponding strategies in this study are explained in Section 4.1.1.

Some experiments performed in this study, including an interactive analysis of cloud type and iterative clustering procedures to maximize the separability of cloud classes, are described in Section 4.2.

4.1. Introduction

4.1.1. Problem in Deriving Class Statistics

4.1.1.1. Choice of Cloud Classes

The inherent properties of different cloud classes are characterized by their spectral and textural characteristics. In applying statistical pattern recognition theory to satellite radiance data, a basic underlying premise is that these classes are indeed separable by these characteristics. This assumption allows us to compare different texture analysis techniques as aids to multi-spectral cloud classification.

Since the choice of cloud classes is dependent on the application of interest, the expertise of analyst and the archived data set, there is often a tendency to define the classes of interest that people want to map and then to force classes which are identified visually from the radiance data set to match the classes of interest. This often leads to some serious problems in the cloud classification and results in substantial classification errors. For example, in assessing the separability of cloud classes in multi-spectral feature space, the assumption that cloud classes have multivariate normal probability functions is often made. The choice of cloud classes based on the application and human skill rather than on the natural groupings of the lattice data often leads to the problem that this assumption is violated. When such a situation arises, these offending classes have to be split into subclasses of which each is unimodal and has an approximately normal probability function.

In this case, an effective alternative is usually made of clustering the radiance data based on multi-spectral features, thus obtaining "spectral maps" in which the spectrally distinct classes are displayed, and then using reference data to identify the cloud types in each class and to derive the statistics

quantifying their characteristics. By this approach, people can easily determine if some classes of interest are spectrally similar, or conversely, if several spectral classes actually represent a single class of interest.

Moreover, some of the other practical problems may also be solved or at least alleviated by using such a cluster analysis.

In analysing a satellite image, people often have little idea of the number of distinct unimodal groups that the image falls into in multi-dimensional spectral space. Clustering procedures tend to enable the inherent data structures of the image to be determined.

As cloud fields are often highly textural, some clouds actually exist but may not be visually resolved. In this case, the cluster analysis can be helpful in the multi-spectral feature space to identify the existence of these classes.

Therefore, to some extent, an arbitrary choice of cloud classes can be avoided so that more consistent classification results may be produced.

The cluster analysis algorithm used in this study is described in Section 4.1.2, and the application of this analysis algorithm for cloud type identification is discussed with two examples in Section 4.2.1.

The choice of cloud classes also relied upon the previous studies of Liljas (1984) and Karlsson (1987A,B) where some of these cloud classes were found difficult to separate in their multi-spectral signature space.

4.1.1.2. Choice of Sample Size

To gather samples for the training procedure, the two AVHRR radiance images (which were shown in Figures 3.1 and 3.2 of Chapter 3; see Section 3.1

for explanation) were divided into arrays of cells, of which each is 32 pixels square, corresponding to approximately 35x35 km areas at subtrack. By this specification, the basic spectral feature MEAN, i.e., the "local mean" of such a cell of radiance data, is in essence the average brightness of the cell.

The determination of the cell size is rather subjective, since there have been no consistent criteria for cloud texture analysis so far. It is our belief that the cell size, i.e., 35x35 km, is small enough to contain appropriate homogeneity within each cell and large enough so that the structure of some clouds such as Cumulus is apparent. More importantly, the size is a power of two so that fast Fourier transform can be applied to calculating frequency features (Booth, 1973). The cell size for calculating other textural features may be more flexible (see, for example, Parikh 1980).

4.1.1.3. Optimization of Class Statistics

In this study, cloud samples were gathered by dividing the two radiance images into sixteen 512x512-pixel subimages. Clustering procedures were operated on each of these subimages in the three-channel (visible, near-infrared and infrared) feature space. Cloud samples (each is 32 pixels square) arising from the clustering procedures were carefully selected and subjectively assigned to appropriate classes. Among the selected samples (649 in total), some might be subjectively misclassified, and some might include the boundaries between two or more different cloud classes. Moreover, some samples of a specific cloud class or surface class had significantly different spectral characteristics due to latitudinal or longitudinal difference but they were obviously of the same class by visually examining the corresponding images and so these samples were finally assigned to that single class. Thus, the spectral variability of the

statistics for each cloud class was considerably large.

To reduce the bias and variance of the class statistics, i.e., to maximize the separability of the 13 cloud classes, a cluster analysis algorithm was used for unsupervised classification to relocate iteratively these samples so that the separability of these classes could be maximized. The results of the relocation were given in a classification matrix (or called table). In this matrix, "predicted assignment" class samples, which were subjectively assigned prior to the unsupervised classification, are placed versus "actual assignment" class samples, which were objectively assigned by the unsupervised classification. The numbers along the main diagonal axis of the matrix represent the consistent results given by the two assignments, and the others represent the number of class samples which were misclassified by the subjective assignment (see Table 4.4 for example).

4.1.2. Cluster Analysis

There are a number of cluster analysis algorithms available for use in remote sensing. The application of some of these methods to mete orological images has been performed by Desbois *et al.* (1982), Seddon & Hunt (1985) and Pairman & Kittler (1986) for isolating cloud fields.

The iterative clustering algorithm, which was used in this study, is implemented by Wishart (1978) in the package CLUSTAN 2.1. The algorithm includes the following set of basic steps:

1. A given number MINC, the minimum number of class clusters required, is chosen. Meanwhile, the number (k) of initial clusters is defined in a classification array (ICLA(I), I=1, N), where N is the number of samples to be clustered and k

must be smaller than N.

2. A set of the samples is then chosen for each class by relating the k initial clusters to the N samples in the classification array. For each of the initial clusters, the centre of gravity and the variance are calculated from the set of the chosen samples.

3. Relocate the N samples by assigning them to the classes of the closest mean. The assignment is made on the basis of the Euclidean distance measure (The code of similarity coefficient is 24). During each relocation scan, the centroids of clusters are recomputed to account for this stage at the time that the switch occurs. The population is repeatedly scanned until no samples are relocated during one full scan, when a local optimum solution for k clusters in terms of the dissimilarity functions will have been obtained.

4. The dissimilarities (Euclidean distance) between all pairs of classes are computed and those two clusters which are most similar are fused, thereby reducing the classification to (k-1) clusters.

5. The relocation phase is repeated to obtain a local optimum for the (k-1) clusters, and then the fusion phase is repeated to yield (k-2) clusters. In this way, local optimum solutions for all clusters from k to MINC are obtained.

If MINC is not less than k, no fusion occurs and the relocation procedure will only optimise the previous classification. In this case, this iterative clustering technique is essentially the ISODATA algorithm (Richards, 1986). The approach was used for unsupervised classification to optimize the class statistics (see Section 4.1.1.3).

The clustering criterion used in this cluster analysis is the sum of squared error (SSE) measure (see Everitt 1974 or Wishart 1978). Other quality-of-clustering measures also exist. One popular one is to derive a

"within-cluster scatter matrix" by determining an average covariance matrix of the clusters, and a "between-cluster scatter matrix" by taking the difference between the means of clusters and the global mean of the data (see, for example, Seddon 1983). These two measures can be combined into a single figure. Richards (1986) indicates that these measures are essentially the same as the sum of squared error criterion.

The standardization of the image data was performed in this training stage before the raster data were classified. The reason that this is necessary is that the dissimilarity function Euclidean distance used on the raw data is very unsatisfactory since it is badly affected by changing the scale of a variable (Everitt, 1974), and similarity values will be biased towards those variables which have large variances (Wishart, 1978).

The choice of the two parameters k and MINC depends mainly on the empirical evaluation of different cloud fields to be clustered. The number k of initial clusters is usually chosen conservatively higher than we expect. Since we did not know exactly how many cloud classes actually exist in an image to be clustered and also did not know how many spectral class clusters could be formed prior to the cluster analysis, a number of experiments were carried out to find an optimal MINC value in order to ensure that the segments obtained through the clustering procedures contain appropriate homogeneity, and to ensure that the groupings are as natural as possible so that the segments are most appropriately classified.

The analysis *posterior* to the clustering procedures was also an important step in finding spectrally separable cloud classes within the sixteen 512x512-pixel training areas during the training stage although it is rather subjective. This process was mainly a cloud type analysis similar to that of Desbois *et al.* (1982) in the analysis of METEOSAT imagery for automatic

classification of high-level clouds. The corresponding statistics, mean, standard deviation and percent coverage of a specific cloud type over a clustered subimage were obtained from such an analysis. With knowledge of the spectral characteristics of cloud fields (Refer to Anderson and Veltishchev 1973), we could see that some cloud clusters were so close together that they actually represented an unnecessary thresholding of radiance data, and thus they could be merged. It could also be found sometimes that after clustering procedures some assigned class clusters had a substantially large amount of within-class variability which is reflected by their standard deviations. These elongated clusters in multi-spectral feature space were often split when their standard deviations were larger than an empirical threshold.

4.2. Iterative Training Algorithm

4.2.1. Analysis of Cloud Type

Clustering procedures for analysing cloud type were applied to the three-channel AVHRR image data, since it has been shown by Liljas (1984, 1986) that three-channel AVHRR data is effective for the separation of a considerable number of cloud classes. The data consist of sixteen 512x512-pixel subimages, each containing 256 field-based average samples (32x32 pixels for each field sample). The results of the cloud type analysis on these subimages were given in terms of percent coverage, mean and standard deviation of a specific cloud class. An analysis of cloud fraction within the individual samples can be performed later by some other approaches such as those given by Saunders and Kriebel (1987).

As discussed in Section 4.1.1.1 and 4.1.2, the number of spectrally separable

cloud class clusters within each 512x512-pixel subimage was subjectively determined through a number of experiments with the help of the cluster analysis. These experiments were generally based on the criterion that clusters which are similar in their spectral characteristics were merged while clusters with a large amount of within-class variability were split. In order to match the class clusters obtained from the clustering procedures to the cloud classes of our interest which appeared hard-to-separate in the Karlsson (1987A) and Liljas (1984)'s studies, some trade-offs were made between the parameter MINC (see Section 4.1.2) and the subjective assignment of cloud categories.

In these experiments, the category indices of class clusters obtained from the clustering procedures performed on each subimage of 256 samples were assigned by the CLUSTAN programme in order so that they have only relative meanings for each subimage (see Tables 4.1 and 4.2 for example). And later, these class clusters from different subimages were subjectively merged and eventually re-assigned into 13 cloud classes, each with varying numbers of training samples. These are:

Cloud Types	Number of Samples
Land	41
Stratus	75
Sea surface	91
Cumulus	69
Thick cirrus over low cloud	77
Altostratus over middle frontal cloud	26
Thick cirrostratus	42
Thin cirrostratus over middle frontal cloud	33
Cumulonimbus	57
Thin cirrus with open water	54
Thin cirrus over low cloud	16
Thin cirrus over middle frontal cloud	56
Stratocumulus	12
in total	649

The applications of such a cloud type analysis are illustrated by two

examples as follows to show the analysis stages involved as well as the corresponding problems tackled.

Figures 4.1A and 4.1B are one of the sixteen 512x512-pixel subimages in the visible and infrared channels respectively, showing the diffuse pattern of Cirrus over the underlying water surface. It is located approximately on the right-bottom part of the central 1024x1024-pixel area in Figure 3.1 of Chapter 3, which covers a large area near the British Isles from 0° and 10°W and 43° to 48°N.

It should be noted that the original concurrent visible and infrared subimages are so faint that cirriform clouds cannot be visually identified. Figures 4.1A and 4.1B are those which have been improved by the histogram equalization technique to reveal clearly the relative brightness of different objects within the subimages. To relate these images to their classification map (see Figure 4.3 or Figure 4.4) for explanation, the two subimages shown in Figures 4.1A and 4.1B were vertically reversed.

The initial clustering analysis was performed several times, with MINC ranging from five to eight. The results, when MINC is five, constitute a classification map (see Figure 4.3 with the corresponding class statistics given in Table 4.1). Comparing this figure with Figures 4.1A and 4.1B, we can see that the major parts of Land surface (class 3) and Sea fog/Stratus (class 16) are well revealed in the classification map, whereas cirrus clouds are embedded in class 7, a mixture of Cirrus and water surface in fact. Class 1 and Class 6 were not subjectively assigned here to any class since they are actually either boundaries of clouds or mixture of clouds, Land and sea surfaces.

The reason that cirrus clouds were not differentiated in the cluster analysis when the parameter MINC was five is due to the very small contrast between this cloud class and its underlying background. By increasing gradually the MINC

value, we found that Class 7 could be resolved and eventually split into several subclasses. The experiments with the different MINC values show that when MINC is eight, Class 7 in Table 4.1 is split into Open water (class 7), Thin cirrus over open water (class 8) and Very thin cirrus over open water (class 9). The corresponding map in this case is given in Figure 4.4 and Table 4.2. This also indicates the potential of the objective classification approach for detecting some clouds which are visually unresolved.

Land surface (class 3) in Figure 4.3 was also split by the clustering procedures with MINC = 8 into two subclasses with 8% and 7% coverage, Class 3 and Class 4 in Figure 4.4. Class 16, Sea fog/Stratus, remains unchanged.

Class 1 and Class 6 were again not assigned, showing nearly the same properties as in Figure 4.3.

Some of the assigned classes were then merged based on the values of the distances between the centres of gravity of these cloud classes and the similarity of their appearances and positions in Figures 4.1A and 4.1B: Class 3 and Class 4 were merged into one class, Land surface, and similarly, Class 8 and Class 9 were merged into Thin cirrus over open water.

Finally, the 32x32-pixel sample areas forming the definitive cloud classes, i.e., Land surface (class 3 and 4), Open water (class 7), Thin cirrus over open water (class 8 and 9) and Sea fog/Stratus (class 16), were used as training samples to assess the texture analysis techniques.

The second example is a complex frontal case containing a number of cloud classes over areas to the west of Ireland (see Figures 4.2A and 4.2B): frontal clouds lie diagonally across the subimage, and Sea fog/Stratus lie at the right-top and convective clouds within a large area below the frontal clouds.

It should be noted again that the visible and infrared subimages (Figures

4.2A and 4.2B) have been enhanced by the histogram equalization technique. As Figures 4.1A and 4.1B, these two subimages were also vertically reversed.

The iterative clustering analysis was again performed iteratively also with the parameter MINC starting at five. The global optimum based on the values of the distances and the visual evaluation of Figures 4.2A and 4.2B was achieved when MINC was eight. The classification map and corresponding statistics are given in Figure 4.5 and Table 4.3.

Comparing Figure 4.5 with Figures 4.2A and 4.2B, we can see that several cloud fields are well outlined by the clustering procedures. Thus, Class 6 and Class 16 in Table 4.3 with a total of 20% coverage were jointly assigned to Thin cirrus over middle cloud. Class 3 with 11% coverage, mostly appearing on the right-top, was assigned to Sea fog/Stratus. Class 7 with 6% coverage, appearing along the frontal band, was considered as Cirrostratus.

Some class clusters were not assigned, since their patterns are too complex and diffuse to be representative of any typical cloud class. By comparing Figure 4.5 with Figures 4.2A and 4.2B, we can see that Class 1 and Class 18 contain convective clouds, and Stratus and Sea surface (Actually, they were all assigned as Cumulus in the nephanalysis); Class 5 is mainly the boundary of Cirrostratus (class 7); Class 17 is Thick stratus mixed with high-level cirrus. These classes were not assigned and so not used later as training samples.

The above two examples illustrate a general procedure in analysing each of the sixteen subimages. During the training stage, the analysis segmented the sixteen subimages individually and then gathered the varying numbers of training samples from the isolated areas in these subimages for some definite cloud classes which were identified during the analysis procedures. With suitable training samples, the total number of classes, 13 in this study, was determined from the subsequent overall analysis. Then, the training samples

from the different subimages were gathered together for each cloud class. It was noted that the final step would lead to considerably larger within-class variations for each class than existed in each single subimage. To ensure that further iterative optimization of the class statistics converged reasonably quickly, we carefully selected the optimum parameters MINC and also the training samples in each subimage through the classification maps in such a way that each class had relatively small within-class variations.

From the cluster analysis, we obtained the training statistics with a high level of confidence since the clustering results provided very good guidelines by outlining cloud fields present in each subimage for cloud type identification. The cluster analysis is in essence a histogram partitioning method, thus the probability that the resultant class statistics have statistical distributions closely approximated to the normal distributions is high. This finally led to the optimal class statistics of the three-channel spectral features so that the error in image interpretation by the divergence analysis (see Section 6.1.2) could be expected to be reduced.

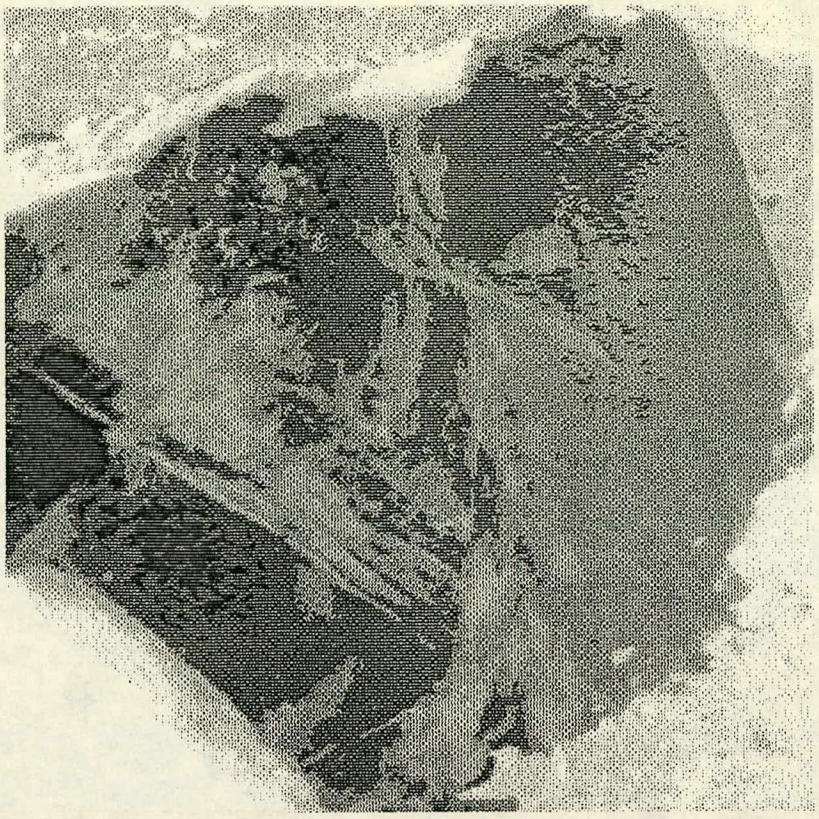


Figure 4.1A An example of the visible subimages showing the Cirrus clouds

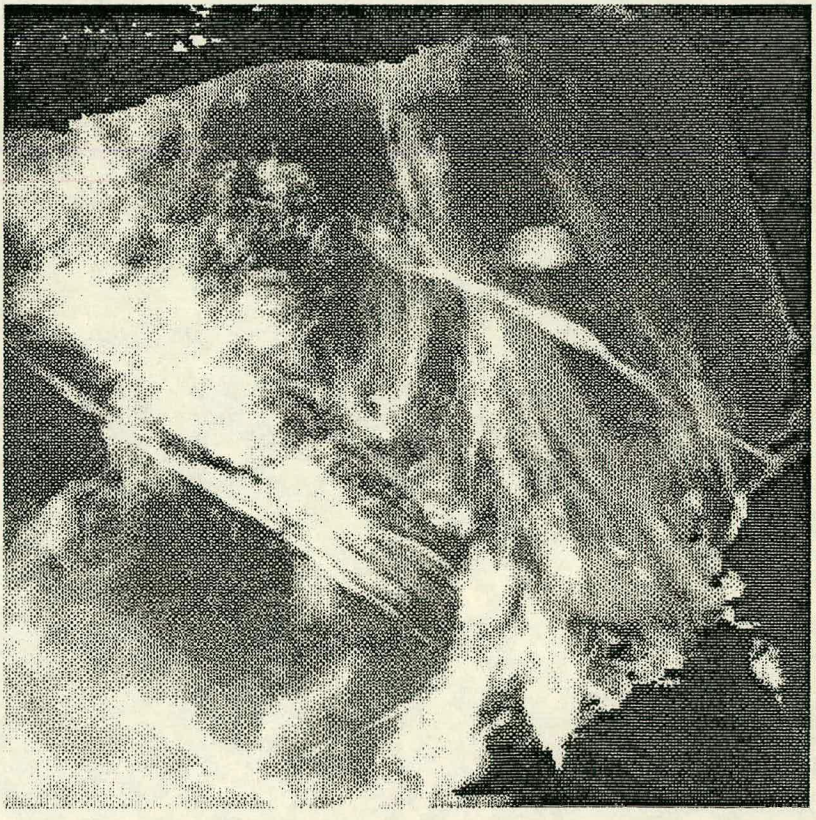


Figure 4.1B An example of the infrared subimages showing the Cirrus clouds

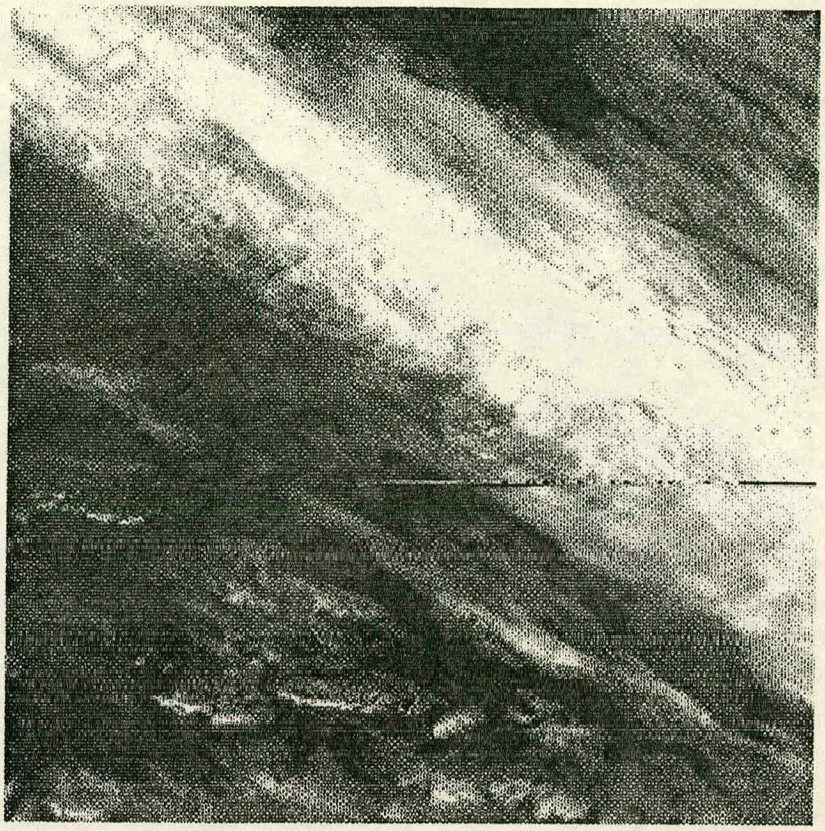


Figure 4.2A An example of the visible subimages showing a complex cloud field in frontal situation

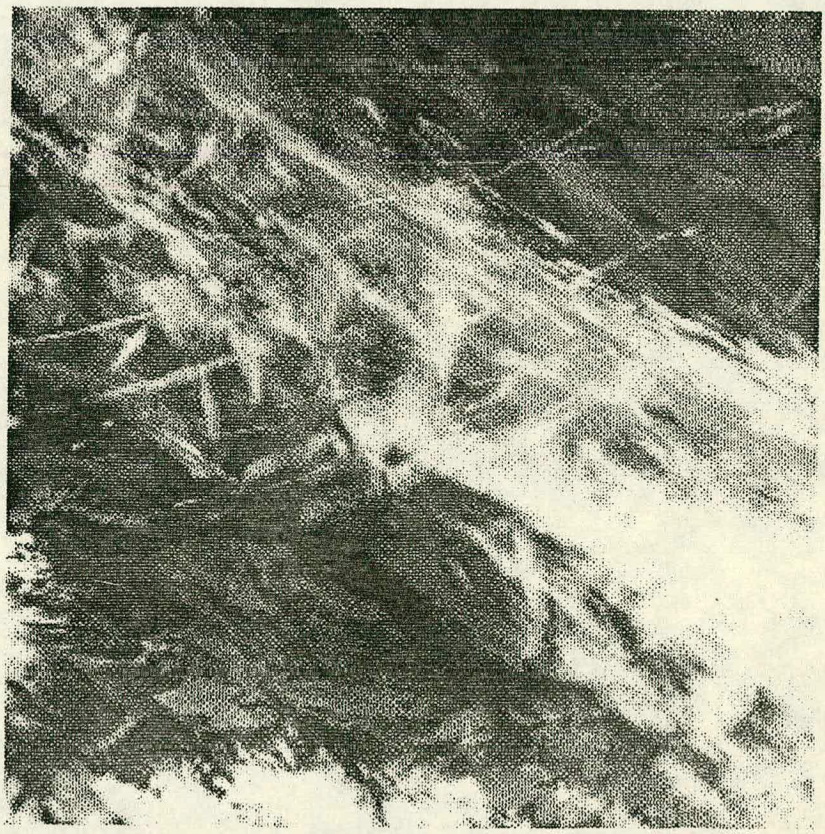


Figure 4.2B An example of the infrared subimages showing a complex cloud field in frontal situation

EX BIBL. UNIV. EDINBURGH

3	3	3	3	3	3	3	3	3	6	6	6	6	6	3	3
3	3	3	3	6	6	1	1	1	7	7	7	7	7	3	3
6	1	1	6	7	7	7	7	7	7	7	7	7	7	6	3
1	1	7	7	7	7	7	7	7	7	7	7	7	7	6	3
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	3
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	3
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	3
7	7	7	7	7	7	7	7	7	7	7	7	7	7	6	3
1	7	7	7	7	7	7	7	7	7	7	7	7	7	6	3
16	1	1	7	7	7	7	7	7	7	7	7	7	7	6	3
16	16	16	1	7	7	7	7	7	7	7	7	7	6	3	1
16	16	16	1	1	7	7	7	7	7	7	6	3	3	3	3
16	16	16	16	16	1	7	7	7	6	3	3	3	3	3	3

Figure 4.3 Classification results of the subimage in Figure 4.1 (Refer to text for explanation and cloud category numbers)*

Class	Percent Coverage	Visible Average (S.D.)	Infrared Average (S.D.)	Near-infrared Average (S.D.)
1	5	40.16 5.61	132.96 4.27	138.88 10.64
3	14	36.96 4.65	108.48 5.00	125.77 10.17
6	7	27.38 4.90	122.20 3.88	158.50 11.72
7	67	16.26 1.49	132.17 3.94	187.48 3.26
16	4	67.78 6.72	137.74 1.99	127.99 4.92

Table 4.1 The table gives the percent coverage of the subimage by each class, the centre of gravity and standard deviation of each class. Units given in the table are digital counts of the AVHRR radiometers. (Refer to text for cloud category numbers)*

*Cloud category numbers in the above figure and table were assigned by the cluster analysis program, so that they have only relative meanings for each 512x512-pixel subimage.

4	4	4	3	4	4	3	3	3	3	6	6	6	6	4	4
4	4	3	3	6	6	1	1	1	7	7	7	7	7	4	4
3	1	1	6	7	7	7	7	7	7	7	7	7	7	6	4
1	1	9	7	7	9	7	7	9	7	7	7	7	7	6	4
9	9	9	9	7	7	7	9	9	7	9	7	7	7	7	4
9	8	8	8	9	7	7	7	9	9	9	7	7	7	7	3
7	8	8	8	8	9	7	9	9	9	9	9	7	7	7	6
7	7	8	8	8	9	9	9	7	9	9	7	7	7	7	6
7	7	8	8	8	9	9	9	7	9	9	9	7	9	7	6
7	7	9	9	8	8	8	9	9	9	9	9	9	7	7	3
9	9	7	7	7	9	8	9	9	9	9	9	9	7	6	3
1	6	7	7	7	7	8	9	7	8	9	9	9	9	6	3
16	1	1	7	7	7	9	7	7	8	8	9	7	7	3	3
16	16	16	1	9	9	9	7	9	8	8	9	6	3	3	1
16	16	16	1	1	8	9	9	8	9	9	6	3	3	4	3
16	16	16	16	16	1	9	8	8	6	4	3	4	3	4	3

Figure 4.4 Classification results of the subimage in Figure 4.1 (Refer to text for explanation and cloud category numbers)*

Class	Percent Coverage	Visible Average (S.D.)	Infrared Average (S.D.)	Near-infrared Average (S.D.)
1	5	40.16 5.61	132.96 4.27	138.88 10.64
3	8	37.64 4.77	113.70 3.58	134.04 7.73
4	7	35.37 4.15	104.40 2.90	119.37 8.70
6	7	25.60 3.87	123.81 4.29	163.37 9.11
7	30	15.59 1.02	128.95 1.57	186.54 2.60
8	10	17.59 0.92	139.32 2.17	189.84 1.77
9	26	16.29 1.15	133.06 1.32	188.13 2.38
16	4	67.78 6.72	137.74 1.99	127.99 4.92

Table 4.2 The table gives the percent coverage of the subimage by each class, the centre of gravity and standard deviation of each class. Units given in the table are digital counts of the AVHRR radiometers. (Refer to text for cloud category numbers)*

*Refer back to the notes for Figure 4.3 and Table 4.1.

17	16	16	17	16	3	3	1	1	1	1	1	3	3	3	3
16	16	16	16	17	17	17	17	3	1	1	1	3	3	3	3
16	16	16	16	16	17	17	17	17	17	3	1	1	3	3	3
1	3	17	16	16	16	16	17	17	17	17	17	3	3	3	3
1	1	3	16	16	16	16	16	16	17	16	17	16	3	3	3
1	1	18	18	18	3	6	6	6	6	16	16	17	16	3	3
1	1	18	18	18	18	3	16	6	6	6	6	6	6	17	17
1	1	3	1	18	18	18	18	18	16	6	6	6	6	6	6
1	1	1	18	1	1	18	5	5	5	6	6	16	16	16	6
1	1	1	1	1	1	1	18	18	5	7	7	7	6	6	6
18	1	1	1	1	1	1	1	1	18	5	5	7	7	7	7
18	1	1	1	1	1	1	1	1	1	18	18	5	7	7	7
5	1	1	1	18	1	1	1	1	1	18	5	18	5	7	7
18	1	1	18	18	18	18	18	1	1	1	18	5	5	5	5
1	1	1	1	18	18	18	1	18	18	18	18	18	18	5	5
18	5	5	7	7	7	7	5	7	5	18	18	18	18	5	5

Figure 4.5 Classification results of the subimage in Figure 4.2 (Refer to text for explanation and cloud category numbers)*.

Class	Percent Coverage	Visible Average (S.D.)	Infrared Average (S.D.)	Near-infrared Average (S.D.)
1	26	28.33 5.92	139.13 1.79	163.26 6.30
3	11	58.79 11.29	139.61 2.62	148.95 8.11
5	8	46.61 12.83	155.06 2.97	181.79 6.81
6	8	108.87 8.49	152.79 3.42	157.73 9.73
7	6	78.06 18.71	171.73 6.13	198.55 7.65
16	12	100.42 10.52	145.50 3.06	143.67 5.56
17	8	94.72 13.19	142.54 3.37	126.19 5.82
18	17	34.24 10.27	145.30 2.60	171.82 5.75

Table 4.3 The table gives the percent coverage of the subimage by each class, the centre of gravity and standard deviation of each class. Units given in the table are digital counts of the AVHRR radiometers. (Refer to text for cloud category numbers)*
*Refer back to the notes for Figure 4.3 and Table 4.1.

4.2.2. Iterative Unsupervised Classification

The training samples obtained through the cluster analysis in Section 4.2.1 were then used as input for the iterative unsupervised classification algorithm. The reason for using such an algorithm was briefly given in Section 4.1.1.3. The cluster analysis routines were employed again in this algorithm. As opposed to those used in the earlier cloud type analysis, the routines had no fusion process in the Relocation procedure as the parameter MINC was set equal to k in this algorithm, i.e., the subsequent relocation procedure only optimizes the previous classification. The optimization of the training statistics is characterized by the relative classification accuracy (predicted assignment versus actual assignment) in terms of percentage (see Section 4.1.1.3 for details).

The unsupervised classification was performed iteratively on the class statistics (see Section 4.2.1). After each iteration, the resultant classification table (see Table 4.4 for example) was examined. Some "misclassified" samples were relocated subjectively when they were found wrongly assigned by referring them to the corresponding subimages, classification maps (see Figure 4.3 for example) and surface weather charts, and others were left unchanged. The altered class statistics were then used as the "predicted" class samples and led to a new classification table in the next iteration. Three iterations raised the relative accuracy of the initial model from 71.3% to 89.7%, and then the accuracy stabilized to some extent, i.e., the "misclassified" samples could not be altered further based on our visual interpretation. The final classification table is given in Table 4.4.

From the table, we can see that seventeen of the Stratus (class 2) samples move to Thin cirrus over low cloud (class 11) and two to Stratocumulus (class 13) during the final iteration; eleven of the Cumulus (class 4) samples are misclassified as Stratocumulus (class 13) and six as Altostratus over middle

frontal cloud (class 6); eleven of the Thick cirrus over low cloud (class 5) samples are relocated to Cumulus (class 4) and four to Stratocumulus (class 13); three of the Thin cirrus over low cloud (class 11) samples are re-assigned to Stratocumulus (class 13). The table shows the similarity between these class pairs, and the results are consistent with our knowledge of cloud field interpretation.

The class statistics obtained from only the two-day (13th, 14th May 1979) satellite observations and optimized in the above procedures, although they did not contain enough within-class variation for each cloud or surface class, reveal reasonably well the relative spectral characteristics of the 13 cloud classes. To show the difference between these classes clearly, three coincident plots were formed from the class statistics (see Figures 4.6, 4.7 and 4.8), representing the means and standard deviations of these three-channel spectral features MEANs for each class. It should be noted that only one standard deviation of each of the three spectral features is plotted for each class. In these figures, the horizontal axis represents the digital radiance levels (up to 255). It can be seen from these figures that the spectral signatures of the 13 classes are all within an interval of about 130 radiance levels rather than spread along the whole axis. This indicates that the contrast of these radiance images is relatively small. Class pairs which are hard to separate in the three channels, e.g., those pairs (which will appear in Table 6.4 of Chapter 6) to which we were interested in applying texture analysis techniques, are well reflected in the figures: these class pairs considered in Table 6.4 all overlap either within one standard deviation (shown in these three figures) or within two standard deviations from the mean in all three spectral features.

Since the unsupervised classification was performed only using the three-channel radiances, it can be expected that the involvement of additional "bands", i.e., textural features in this study, may contribute to the separation of

these hard-to-separate cloud class pairs. This is what Chapter 6 mainly aims to reveal. The class statistics gathered and optimized in the training stage of this chapter form the basis of the study.

	Predicted												
Actual	1	2	3	4	5	6	7	8	9	10	11	12	13
1	41	0	0	0	0	0	0	0	0	0	0	0	0
2	0	53	0	0	0	0	0	0	0	0	0	0	0
3	0	0	91	0	0	0	0	0	0	0	0	0	0
4	0	0	0	49	11	0	0	0	0	0	0	0	0
5	0	3	0	1	61	0	0	0	0	0	0	1	0
6	0	0	0	6	0	26	0	0	0	0	0	0	0
7	0	0	0	0	0	0	42	0	1	0	0	0	0
8	0	0	0	0	1	0	0	33	4	0	0	0	0
9	0	0	0	0	0	0	0	0	52	0	0	0	0
10	0	0	0	1	0	0	0	0	0	54	0	0	0
11	0	17	0	1	0	0	0	0	0	0	13	0	0
12	0	0	0	0	0	0	0	0	0	0	0	55	0
13	0	2	0	11	4	0	0	0	0	0	3	0	12

Accuracy = 89.7%

Table 4.4 Classification table of "Predicted assignment" versus "Actual assignment" training samples.

1-Land
2-Stratus
3-Open water
4-Cumulus
5-Thick cirrus over low cloud
6-Altostratus over middle frontal cloud
7-Thick cirrostratus
8-Thin cirrostratus over middle frontal cloud
9-Cumulonimbus
10-Thin cirrus with open water
11-Thin cirrus over low cloud
12-Thin cirrus over middle frontal cloud
13-Stratocumulus

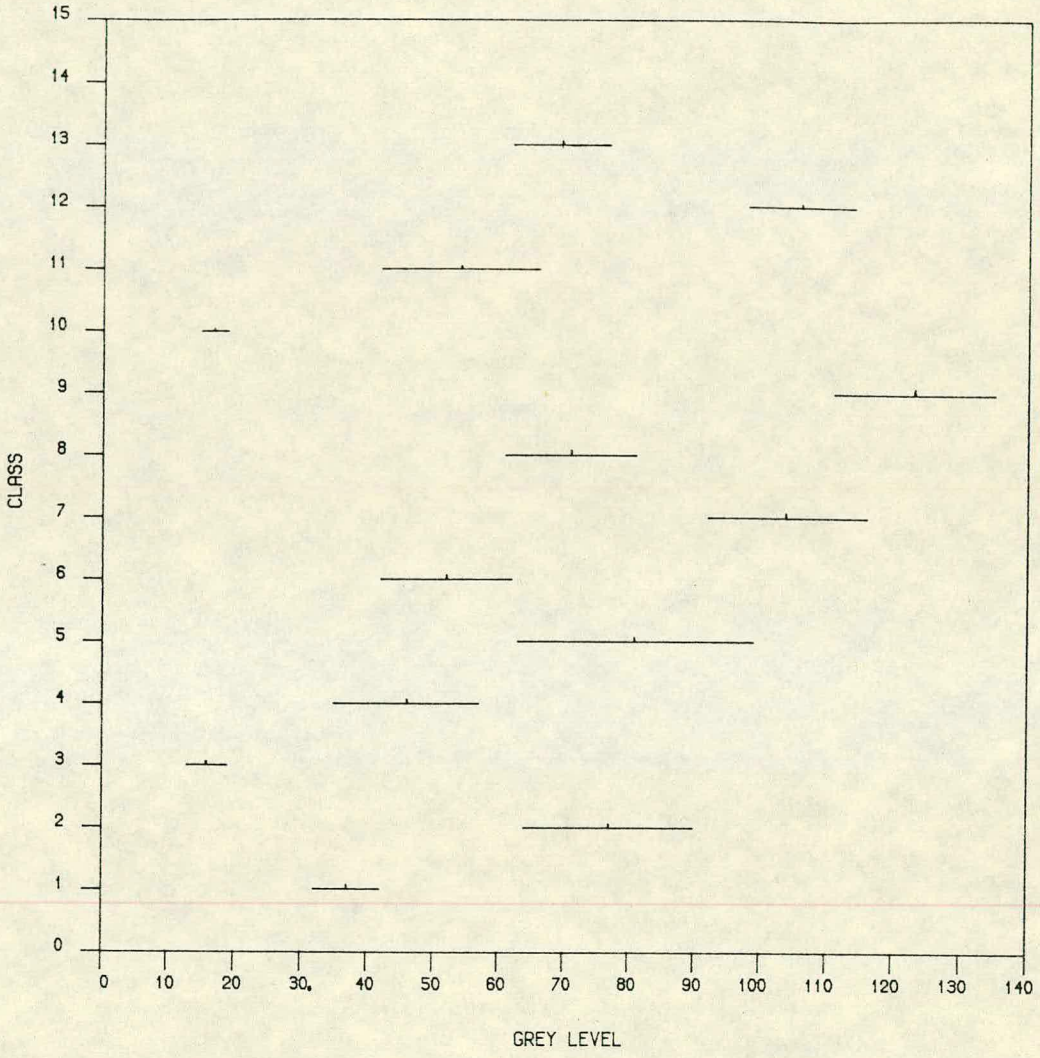


Figure 4.6 Coincident plots of class means and standard deviations of the visible spectral feature MEAN used in this study

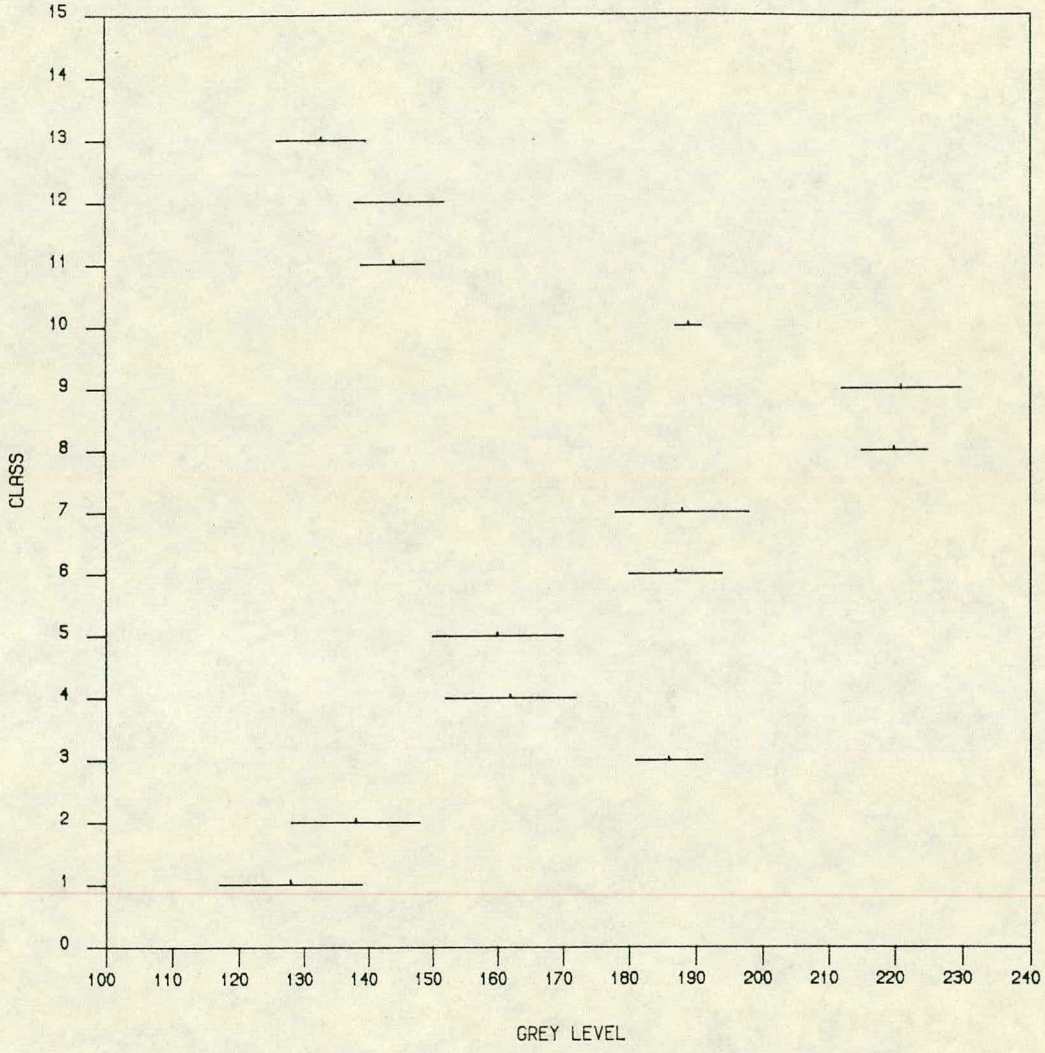


Figure 4.7 Coincident plots of class means and standard deviations of the near-infrared spectral feature MEAN used in this study.

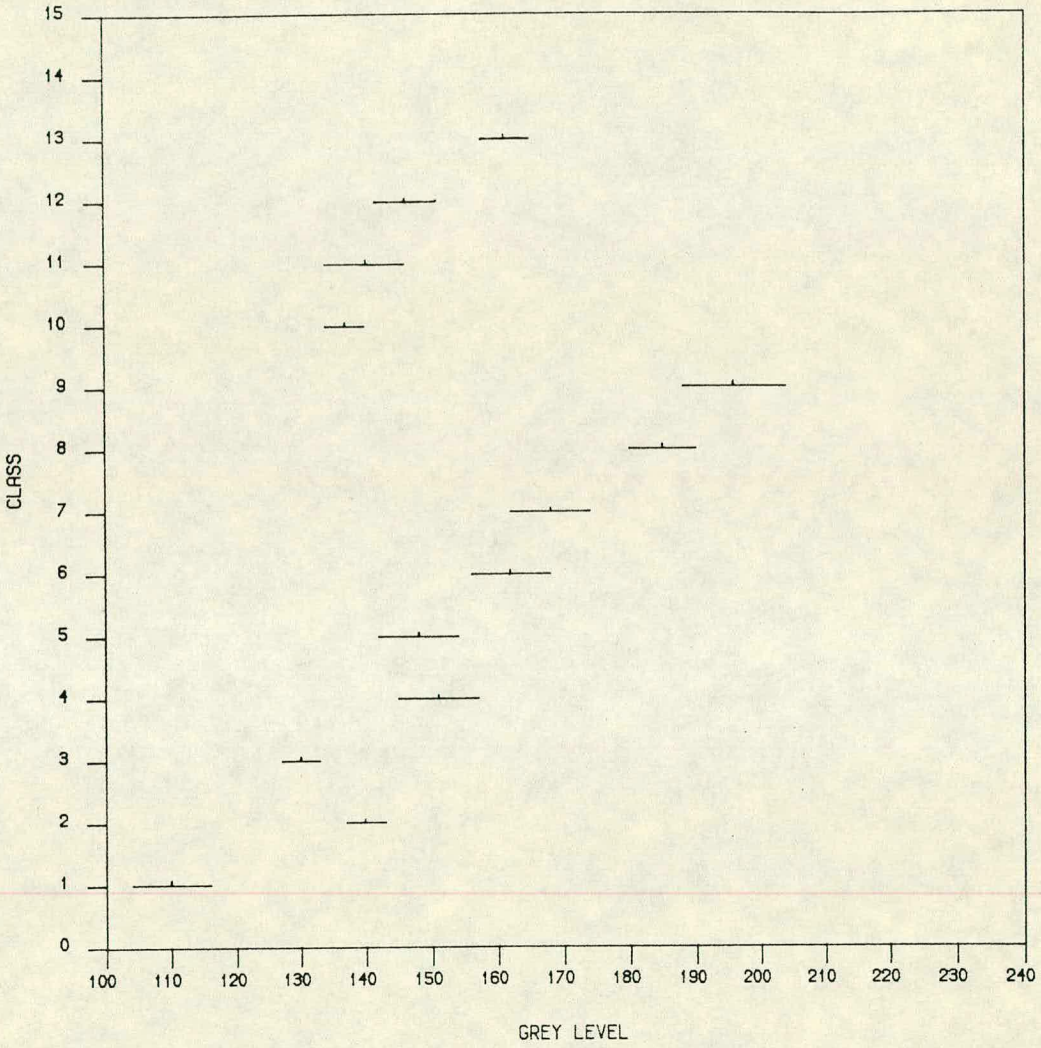


Figure 4.8 Coincident plots of class means and standard deviations of the infrared spectral feature MEAN used in this study

CHAPTER 5

TEXTURAL FEATURES

As described in Chapter 2, visual identification of cloud types by human analysts, e.g., to create nephanalysis, is based on a small number of important attributes. By analogy we may expect that an automatic classification of cloud types may also be done on the basis of only a few significant features characterizing the classes of cloud patterns, preferably those that can be used for cloud type identification by humans. However, in practice, we often have only indirect observations of properties of cloud classes, in which the discriminating information is embodied in a complex manner and so the measurements which are important for cloud classification are not apparent. Also, the verbal description of cloud types used by human analysts is difficult to be converted into a mathematical form.

Although there have been many textural features available for use, most of these features are application-dependent. It is difficult to find any guidelines about what features are optimal for a specific cloud classification study. In this study, textural features were selected mainly from the four groups of statistical features which were briefly discussed in Section 2.3.2. Most of these have been used in the earlier classification studies outlined in that section. Some were modified into more appropriate representations of cloud fields. A detailed mathematical explanation of the features used in this study is given in the next two sections on spatial features and spatial frequency features. Autocorrelation functions are related to spatial features although they in essence measure the spatial frequency information of texture indirectly (Haralick, 1979). The simple textural feature, local variance, evaluated in Section 6.3.1.3, is not discussed in this chapter.

In the experiments described in Chapter 6, the values of all these textural features were calculated over a 32x32-pixel grid using two-channel (visible and infrared) radiance data. The choice of the sample size was explained in Section 4.1.

5.1. Spatial Features

5.1.1. Spatial Grey Level Difference Statistics (SGLDs)

The spatial grey-level difference statistics are probably the best feature set in current use for producing computer-assisted nephanalysis. As shown in the study of Ebert (1987), three textural features, out of eight spectral and textural features which were ranked as the most useful among 66 features (inclusive of 36 textural features) in four channels, are all SGLDs. In other studies such as by Parikh (1977) and Wu *et al.* (1985), SGLD features were either their unique or major choice of textural features for cloud classification studies. This set of the second-order statistics has been extensively used by these researchers to analyse remotely sensed satellite data. The major reasons for this may be attributed to their compatibility with human vision (see Section 2.3.2.2) and the ease of computation. The latter is particularly important when any textural feature set is applied to the analysis of satellite imagery.

A part of the study was designed to assess the SGLD features in multi-dimensional feature space, and then compare their relative merits with other newly-developed feature sets (see other sections of this chapter). The specification of spatial grey level difference statistics is the same as that used by Wu *et al.* (1985). The brief mathematical descriptions follow.

Let {G(n,m)} be the satellite image data within a sample area, e.g., 32 x 32

pixels, then for any displacement δ = (Δn, Δm), we can define the grey level difference in a 3x3-pixel grid (see Figure 5.1) as

$$G_{\delta,\theta}(n,m) = |G(n,m) - G(n+\Delta n, m+\Delta m)| \tag{5.1}$$

Where Δn and Δm are integers, $\delta = (\Delta n^2 + \Delta m^2)^{1/2}$ and $\theta = \tan^{-1}(\Delta m/\Delta n)$. Possible choices of Δn, Δm in the vector δ are $-d \leq \Delta n, \Delta m \leq +d$ so that the direction θ in which the pixel pairs are measured may be 0°, 45°, 90° and 145°. The grey level difference usually ranges from 0 to 255. Figure 5.1 illustrates a neighbourhood of image pixels.

G(m-1,n-1)	G(m-1,n)	G(m-1,n+1)
G(m,n-1)	G(m,n)	G(m,n+1)
G(m+1,n-1)	G(m+1,n)	G(m+1,n+1)

Figure 5.1 A neighbourhood of image pixels.

Grey level differences are produced by selecting a pair of values δ and θ, and finding the $G_{\delta,\theta}(n,m)$ over a sample area of the image. Many other grey level differences are possible using other choices for the δ and θ.

From the grey level difference of a sample area, a probability density function $f_{\delta,\theta}(i)$ associated with a possible value $G_{\delta,\theta}(n,m)$ can be estimated, i.e.,

$$f_{\delta,\theta}(i) = \text{Prob}(G_{\delta,\theta}(n,m) = i) \tag{5.2}$$

The probability density function provides a measure of the scale size of cloud field characteristics. If a cloud area is smooth across its whole horizontal scale and δ is small compared to the cloud texture element size, then the pairs of the image pixels at separation δ should usually have similar grey levels. In

this case, the probability that the grey level difference $\{G_{\delta,\theta}(n,m)\}$ is zero will be high. Conversely, if the cloud area appears inhomogeneous, compared with the separation δ comparable to the element size, the pairs of the image pixels should often have quite different grey levels. In this case, the probability that the grey level difference $\{G_{\delta,\theta}(n,m)\}$ is not zero will be high, i.e., the values in $f_{\delta,\theta}(i)$ should be more spread out. Since such a density function is associated with the angle of the pixel pairs, the directional information is also reflected.

From any of these probability functions, four textural features are usually defined. They are as follows:

Mean:

$$\text{Mean}(\delta,\theta) = \sum_{i=0}^{N_G-1} i \left(\frac{f_{\delta,\theta}(i)}{N} \right) \tag{5.3}$$

Contrast:

$$\text{CON}(\delta,\theta) = \sum_{i=0}^{N_G-1} i^2 \left(\frac{f_{\delta,\theta}(i)}{N} \right) \tag{5.4}$$

Angular Second Moment:

$$\text{ASM}(\delta,\theta) = \sum_{i=0}^{N_G-1} \left(\frac{f_{\delta,\theta}(i)}{N} \right)^2 \tag{5.5}$$

Entropy:

$$ENT(\delta,\theta) = -\sum_{i=0}^{N_G-1} \frac{f_{\delta,\theta}(i)}{N} \log_{10}\left(\frac{f_{\delta,\theta}(i)}{N}\right) \qquad (5.6)$$

where N_G is the total number of grey levels, e.g., usually 256, and N is the total number of pixel pairs in a sample area separated by distance δ in direction θ.

The MEAN(δ,θ) is the average of the grey level differences. It is small when $\{f_{\delta,\theta}(i)\}$ are concentrated near the origin and large when they are concentrated far from the origin. The contrast CON(δ,θ) is a measure of the relative amount of the local variation within a sample area. Large values of these two textural features correspond to structured cloud fields such as Cumulus.

The angular second moment ASM(δ,θ) is a measure of the homogeneity of a sample area. It is smallest when the $\{f_{\delta,\theta}(i)\}$ are all as equal as possible and large when some values are high and others low, e.g., when the values are concentrated near the origin. The entropy, ENT(δ,θ), is a measure of the absence of the distinct structure of grey level differences over distance δ and angle θ, i.e., it is largest for equal $\{f_{\delta,\theta}(i)\}$ and small when they are very unequal (Refer to Section 5.2.3 for a general discussion of "entropy" property). Small values of this feature correspond to smooth or highly structured cloud fields.

By adopting this approach, the spatial information of a cloud field can be extracted from various angles θ when the distance δ is determined. To reduce the number of the textural features, we usually combine the information into some composite form. In this study, the mean and maximum values of these textural features over four angles are calculated to produce eight textural features as follows:

mean MEAN	max MEAN
mean CON	max CON
mean ASM	max ASM
mean ENT	max ENT

The choice of an appropriate distance δ for cloud classification was initially investigated by Parikh (1977). She concluded from a comparative study of different distances that the the distance 1 was preferable for revealing the characteristics of a cloud field. Wu *et al.* (1985) gave a similar conclusion that the textural features which were defined in terms of the smallest pixel separation were the most important. The effect of different distances was also assessed in this study by using the most useful feature, mean MEAN (see Section 6.3.1.1 for explanation), of the eight SGLD features (see Section 6.3.3 for assessment).

The evaluation of the eight SGLD textural features is described in Section 6.3, together with a comprehensive comparison with other textural features.

5.1.2. Roberts Gradient (R.G.)

As discussed in Section 2.3.2.2, one of the dimensions of texture is the amount of edge per unit area. Coarse textures have a small number of edges per unit area while fine textures have a large number of edges per unit area.

Textural edge is considered as a more or less abrupt change in texture, indicating the end of one region and the beginning of another. Edges through a neighbourhood (of 2x2 pixels) can be detected by comparing the values of some local properties obtained in pairs of nonoverlapping neighbourhoods bordering that neighbourhood. The relative orientation of the neighbourhoods determines the direction of the edges that will be detected, and the size of the neighbourhoods determines the widths of the edges (see Rosenfeld and Kak 1982 for discussion of textural edge).

An important local property is Gradient. There are a number of gradient

measures available for use. One of these is the quick Roberts gradient (R.G.), which was suggested by Rosenfeld and Thurston (1971). The Roberts gradient (R.G.) is defined as the sum of the absolute value of grey level differences between diagonally opposite neighbouring image pixels, i.e.,

$$R.G. = |G(m,n) - G(m+\delta,n+\delta)| + |G(m+\delta,n) - G(m,n+\delta)| \quad (5.7)$$

where $\{G(m,n)\}$ is the two-dimensional image data. The R.G. is calculated within an image block shown in Figure 5.2. In fact, the differences in any pairs of perpendicular directions can be used to compute the gradient. The differences in Equation (5.7) are symmetrical about the interpolated pixel at $(m+\delta/2,n+\delta/2)$ (see Figure 5.2), and thus the R.G. is an approximation to the continuous gradient at that pixel, rather than at the pixel G(m,n).

$G(m,n)$	$G(m,n+\delta)$
$G(m+\delta,n)$	$G(m+\delta,n+\delta)$

Figure 5.2 Block of A neighbourhood of pixels.

The distance δ indicates the extent of the block. For micro-edges, small neighbourhoods (δ) should be used while for macro-edges, large neighbourhoods (δ) should be used. In this study, the distance δ was chosen as one. The results of the experiments are given in Section 6.3.1.2.

Since the R.G. used in this study was used as a textural feature for a sample area of 32 x 32 pixels, the appropriate form of the R.G. measure is re-defined as the area-averaged measure given below.

$$R.G. = \frac{1}{(M-\delta)(N-\delta)} \sum_{m=1}^{M-\delta} \sum_{n=1}^{N-\delta} \{|G(m,n)-G(m+\delta,n+\delta)|+|G(m+\delta,n)-G(m,n+\delta)|\} \quad (5.8)$$

where G(m,n) is the same as above; M and N are the size of a sample area.

5.1.3. Autocorrelation Features

As discussed in Section 2.3.2.2, the autocorrelation function is a textural feature which measures the spatial size of grey level primitives in an image. A primitive is defined as a connected set of resolution cells (i.e., image pixels), characterized by a list of attributes (see Section 2.3.2.1). If the grey level primitives in an image are relatively large, then the autocorrelation value will drop off slowly with increasing distance. Conversely, if these primitives are relatively small, the autocorrelation value will drop off quickly. For some textural fields in which the grey level primitives appear periodic, the autocorrelation value will change in a periodic way. One dimension of texture, coarseness, can be well reflected in the autocorrelation changes. Another major dimension, directionality, can also be more or less reflected by the function.

The application of a simple autocorrelation function to meteorological patterns has been performed by Bunting and Fournier (1980), in which they calculated the autocorrelation function for a series of two-dimensional radiance data by simply shifting the series along scan lines from one to ten lags. Their experiments showed that the functions gave similar results to Fourier spectrum features.

A set of spatial autocorrelation functions was used in this study as an alternative to the simple functions, which was recommended by Renshaw (1987), called Circular Moran Autocorrelation Functions. These circular functions are defined to determine whether lattice data are spatially autocorrelated, while the simple autocorrelation functions described above to measure autocorrelation in

stationary time series (Cliff and Ord, 1973). The variable in a time series is influenced only by past values, while for a spatial process dependence extends in all directions. The circular functions also measure the spatial size of grey level primitives in an image but the directional bias is inherently avoided by their specification.

The circular autocorrelation functions evaluate the spatial relationships between image pixels in a way fundamentally different from the simple autocorrelation function. The relationships reflected by these functions are defined by a connection matrix. An illustration is given in Figure 5.3, in which the matrix is organized in a circular manner around the central pixel (with asterisk): the first annular band contains eight pixels while the fourth band has twenty-four pixels. The autocorrelation of the image pixels in such an approximately circular band with the central pixel is evaluated by multiplying each value in the band by the central value and then summing upon shifting the image both horizontally and vertically. These pixel pairs, one within the band and the other at the centre, lie in the directions emanating from the centre. The number of the directions varies with the number of lags, e.g., when the number of lags is one, four directions, 0°, 45°, 90°, 135°, are included. Finally, the sum is divided by the variance (from zero) of the central pixel and the number of the image pixels within the band to normalize the value of the circular autocorrelation reflecting the local property of the image. From the connection matrix, circular Moran autocorrelation functions at different lags can be defined. With $Y_{ij} = G_{ij} - \bar{G}$, where G_{ij} is the grey level of an image pixel at position (i,j) and $\bar{G}$ the grey level mean of the image, the first circular Moran autocorrelation function at lag 1 is defined for an image area of 32x32 pixels as:

$$\text{CIRC } 1 = \sum_{i=5}^{28} \sum_{j=5}^{28} \{Y_{i,j}(Y_{i-1,j-1}+Y_{i-1,j}+Y_{i-1,j+1}+Y_{i,j-1}+Y_{i,j+1}+Y_{i+1,j-1}+Y_{i+1,j}+Y_{i+1,j+1})\}$$

$$\div(8^* \sum_{i=5}^{28} \sum_{j=5}^{28} Y_{i,j}^2) \qquad (5.9)$$

Other circular autocorrelation functions (CIRC 2, CIRC 3 and CIRC 4) at lags 2, 3, 4 are defined in a similar way. Due to the limitation of its size, the connection matrix is operated over a restricted region of an image area. For a sample area of 32x32 pixels, the region within which the circular autocorrelation functions of four lags are calculated ranges from the 5th to the 28th pixel both horizontally and vertically.

The circular autocorrelation functions at four successive lags were used in this study and the results (shown in Section 6.3.1.4) in terms of the transformed divergences (see Section 6.2.2) showed that their discriminating power is equivalent to the commonly-used spatial grey level difference matrix, although their applicability still needs to be verified by other measures, e.g., classification accuracies.

```
      4 4 4 4 4
    4 3 3 3 3 3 4
  4 3 2 2 2 2 2 3 4
  4 3 2 1 1 1 2 3 4
  4 3 2 1 * 1 2 3 4
  4 3 2 1 1 1 2 3 4
  4 3 2 2 2 2 2 3 4
    4 3 3 3 3 3 4
      4 4 4 4 4
```

Figure 5.3 The connection matrix of circular Moran autocorrelation functions.

5.2. Spatial Frequency Features

As shown by Bajcsy and Lieberman (1976), mathematical descriptors derived from spatial frequency domain are useful in characterizing the texture of outdoor scenes. The application of these descriptors to meteorological remote sensing was discussed in Section 2.3.2. The work described in this section was primarily based on that of Bunting and Fournier (1980), firstly using the normalized average amplitude Fourier power spectrum features and further modifying them to the entropy measures in polar frequency domain.

A detailed mathematical treatment of these two groups of the statistics is given respectively in the three following subsections.

5.2.1. Spectral Analysis

The choice of Fourier power spectrum as a cloud field descriptor is based on the assumption that a cloud field can be modelled by a stationary stochastic process. Spatial information in such a process can be simply described by its autocovariance function. An equivalent description of such an autocovariance function is provided by its power spectrum, which is the transform of the autocovariance function. The power spectrum shows how the variance of the stochastic process is distributed with frequency. Thus, the inherent information contained in the spectrum, e.g., size, orientation and location, can be revealed (see Jenkins & Watts 1968, or Oppenheim & Schafer 1975).

The sample autocovariance function can be defined at lag (j,k) for $0 \leq j < m$ and $-n < k < n$ by

$$C_{jk} = (1/mn) \sum_{s=1}^{m-j} \sum_{\Omega_t} G_{s,t}\, G_{s+j,t+k} \qquad (5.10)$$

where G_{st} (s=1, ..., m; t=1, ..., n) is a two-dimensional array of image data which is corrected for their overall mean by setting the mean to zero; Ω_t = {1, ..., n−k; k ≥ 0} and Ω_t = {−k+1, ..., n; k<0}. For j,k ≥ 0, it can be shown that $C_{-j,-k} = C_{j,k}$ and $C_{-j,+k} = C_{j,-k}$, which means $C_{j,k}$ is symmetric about the origin of the coordinate. Here, we only consider j=0,..,m−1.

The interpretation of such an autocovariance function is often difficult, especially in two dimensions. However, the difficulties can be overcome by using the Fourier transform,

$$f(\omega_1,\omega_2) = \sum_{j=-m+1}^{m-1} \sum_{k=-n+1}^{n-1} C_{j,k}\cos(j\omega_1 + k\omega_2) \qquad (5.11)$$

This spectral function, or periodogram, is usually evaluated at frequencies (ω1, ω2) = (2πp/m, 2πq/n), where p = 0, ..., m−1; q = 0, ..., n−1.

In practice, f(ω1,ω2) can also be derived by transforming directly the mean-corrected image data matrix. Thus, the periodogram is written as:

$$I_{p,q} = mn(a_{p,q}^2 + b_{p,q}^2) \qquad (5.12)$$

where p and q are integers, and

$$a_{p,q} + i\, b_{p,q} = 1/mn \sum_{s=1}^{m} \sum_{t=1}^{n} G_{s,t}\exp\{2\pi i\,(ps/m + qt/n)\} \qquad (5.13)$$

The full range of p,q values should be over p = 0, ..., m−1; q = 0, ..., n−1. However, due to the fact that the periodogram is symmetric about the origin, i.e., $I_{m-p,q} = I_{p,n-q}$, we need only consider half of the periodogram. Here, p=0, ..., m/2 and q=−n/2, ..., n/2−1 are chosen.

The periodogram allows us not only to infer some properties such as coarseness but also to distinguish directional and non-directional components of textural patterns. This is due to the fact that the periodogram is invariant with respect to translation in the corresponding spatial patterns, but not with respect to rotation, thus the directionality of an image pattern is preserved in the spectrum (Rosenfeld and Kak, 1982).

However, natural objects such as clouds are always horizontally distributed in random directions. Thus, the directionality of natural scenes, reflected by the spectra of the scenes, can not provide any useful discriminating information. To obtain the spatial information concerning the scales of a textural pattern from its power spectrum, we usually calculate the average spectral energies at the same frequency magnitudes over all directions in the periodogram to form a simplified spectrum. Obviously, a cartesian coordinate is not appropriate since the rings outlined by the same frequency magnitudes on the horizontal and vertical axis of the periodogram are actually squared rather than circular. To solve the problem, we have to transform the periodogram from a cartesian coordinate system (m,n) into a polar coordinate system (r,ψ).

The direct calculation of a polar spectrum through the use of Bessel functions involves a considerable amount of computation time and thus is not suitable for satellite image analysis.

To reduce the calculation time, a fast procedure was adopted from Renshaw and Ford's (1984) work, in which they transformed the periodogram directly from a cartesian coordinate system to a polar coordinate system.

Two spectra, **R**-spectrum and **θ**-spectrum, can be derived from the procedure. First, we scaled the $I_{p,q}$ values to ensure that their average values is unity for each value of $I_{p,q}$ over the range

```
p=0              ;   q=-n/2, ..., -1
p=1, ..., m/2-1;   q=-n/2, ..., n/2-1
p=m/2            ;   q=-n/2, ..., 0.
```

Given that $r \equiv \text{sqrt}(p^2 + q^2)$ and $\theta \equiv \tan^{-1}(p/q)$, the groups of spectral entries within $0 < r \leq 1$, $1 < r \leq 2$, ... and $-5^\circ < \theta \leq 5^\circ$, $5^\circ < \theta \leq 15^\circ$, ..., $165^\circ < \theta \leq 175^\circ$ were then considered by summing all the $I_{p,q}$ within the groups. Finally, we divided the sum of $I_{p,q}$ values in each interval by the number of $I_{p,q}$ values counted within it. Thus, the **R**-spectrum is a plot of frequency elements with approximately the same frequency magnitude ($\text{sqrt}(p^2 + q^2)$), and the **θ**-spectrum is a plot of frequency elements with approximately the same frequency angle ($\tan^{-1}(p/q)$).

The visual interpretation of these two polar spectra is straightforward although some care has to be taken. However, a quantitative decision should be possible using a limited number of parameters, each of which may be a combination of spatial information at a certain number of wavenumbers.

One of the possible approaches to this problem is to convert the polar spectrum into a limited number of parameters, which was used by Bunting and Fournier (1980), and also used by Ebert (1987) in her cloud classification study. The detailed treatment of the approach is given in the following subsection.

The **R**-spectrum is what we call the Normalized Average Amplitude Fourier power spectrum in Section 5.2.2 following the convention of Bunting & Fournier (1980) and Ebert (1987).

5.2.2. Normalized Average Amplitude Fourier Power Spectrum (NAAs)

Features based on the normalized average amplitude Fourier power spectrum (NAAs) were first used when Bunting and Fournier (1980) applied two-dimensional spectral analysis to the detection of small-scale clouds. The main purpose of their approach is to create a suitable feature vector for Bayes decision maker.

To reduce the number of NAA parameters from $r=1/2\ \mathrm{sqrt}(m^2 + n^2)$ the following three parameters are used to characterize the NAA: the average amplitude of the NAA spectrum, the maximum amplitude of the NAA spectrum and the dominant wavenumber corresponding to the maximum amplitude. These were extracted as textural features for cloud classification. The three parameters are abbreviated mean NAA, max NAA and DOM WVN, respectively.

These three textural features were assessed and then used in this study to compare their merits with other groups of textural features as an aid to multi-spectral cloud classification. The results of the experiment are given in Section 6.3.

The usefulness of the NAA power spectrum in polar frequency domain for cloud field interpretation was first tested on several synthetic patterns such as a sine wave pattern, a stripe pattern and a checkerboard pattern with and without random noise, which were designed to simulate cloud edge, cellular and ragged appearance, and then on some typical cloud patterns such as Cumulus (Cu), Cirrus (Ci), Cumulonimbus (Cb) and Stratocumulus (Sc) in both visible and infrared images. Figures 5.4 to 5.7 show respectively the results of the test on four 32x32-pixel sample areas of each of these four cloud classes in visible (A) and infrared (B) images. Each of the four figures contains four curves representing the spectra of the four sample areas. The four sample areas were

randomly selected from 69, 54, 57 and 12 cases, respectively for the four cloud classes. The logarithmic scale was used to conventionally plot the spectra for all wavenumbers. We did not use the average spectra of the four cloud classes for the illustration, as used by Bunting and Fournier (1980), as we intended to show the substantially large within-class variability of them, i.e., the problem of the cloud analysis by the technique.

The relative within-class compactness and between-class separability of the four classes can be seen in these figures: the class pairs, Cu/Ci, Cu/Sc, Ci/Cb in visible spectra and Cu/Ci, Ci/Cb, Ci/Sc in infrared spectra are different to some extent, while the separability between the others are not visually apparent. The separability is reflected mostly at relatively low wavenumbers. It is also obvious that the NAA spectra appear rather smoothed at higher wavenumbers, for example, after wavenumber 12. The quantitative interpretation of the separability of these cloud classes is given in Section 6.3.

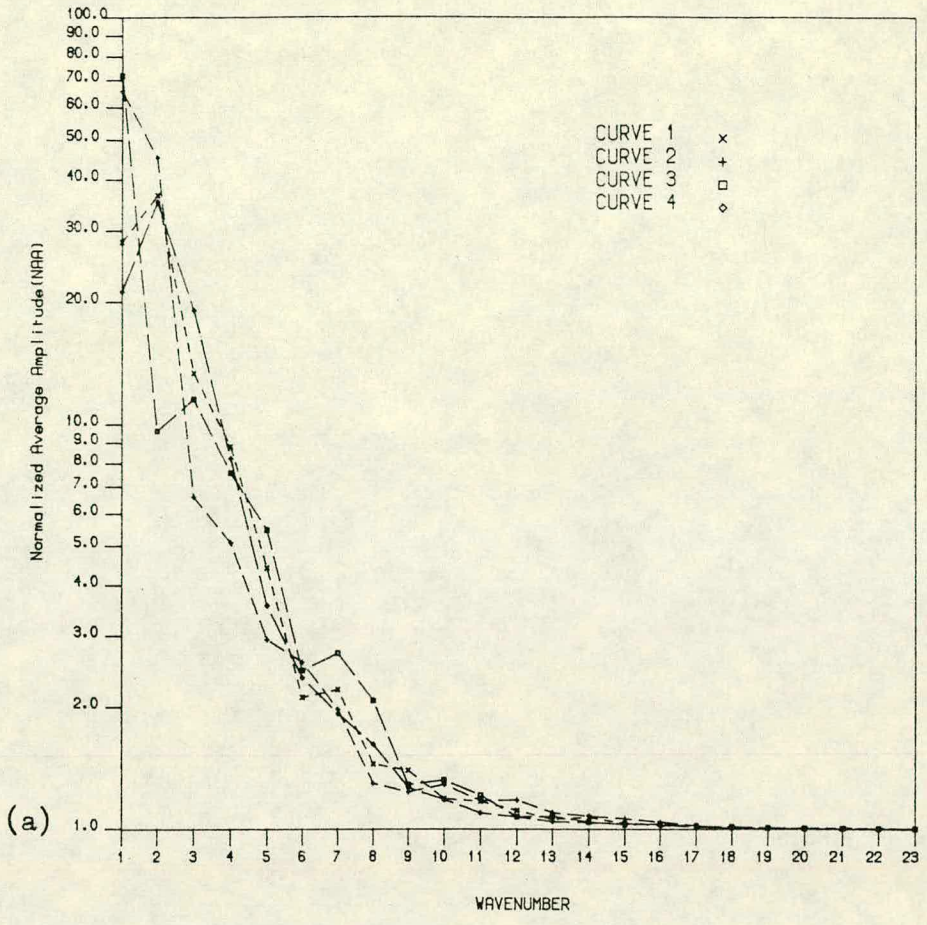
The absence of significant peaks at high wavenumbers should be mostly attributed to the inherent specification of the NAA spectrum, since at these wavenumbers the spectral frequency entries have much smaller energies than those near the origin (see Figures 5.4 - 5.7) but averaged in the same way. The observations in this study are well consistent with those of Bunting and Fournier (1980) using the DMSP fine-mode satellite data.

Another crucial problem with the specification of the NAA spectra, which was considered in this study, is its uncertainty in characterizing the spatial distributions of frequency entries within circular bands. An example for this is that, a single pure sinusoid within a band would be considered the same as a mixture of sinusoids all within the same band since the energy of a pure sinusoid is equal to the average energy of the mixture. This is to say, two textures with even distinct spectra would be judged the same as long as their

average energies within bands were the same.

A careful evaluation of the facts indicates that these two problems are in essence closely related to the inadequacies of the NAA power spectrum in measuring the spatial distributions of frequency entries within circular bands. This leads to the necessity of designing new frequency measures above or independent of the amplitude information contained in the NAA spectrum.

The NAA representation only implicitly contains information about the spatial distributions. As an alternative, entropy measure for the distributions can give more impressive results (see Section 5.2.3 for discussion).



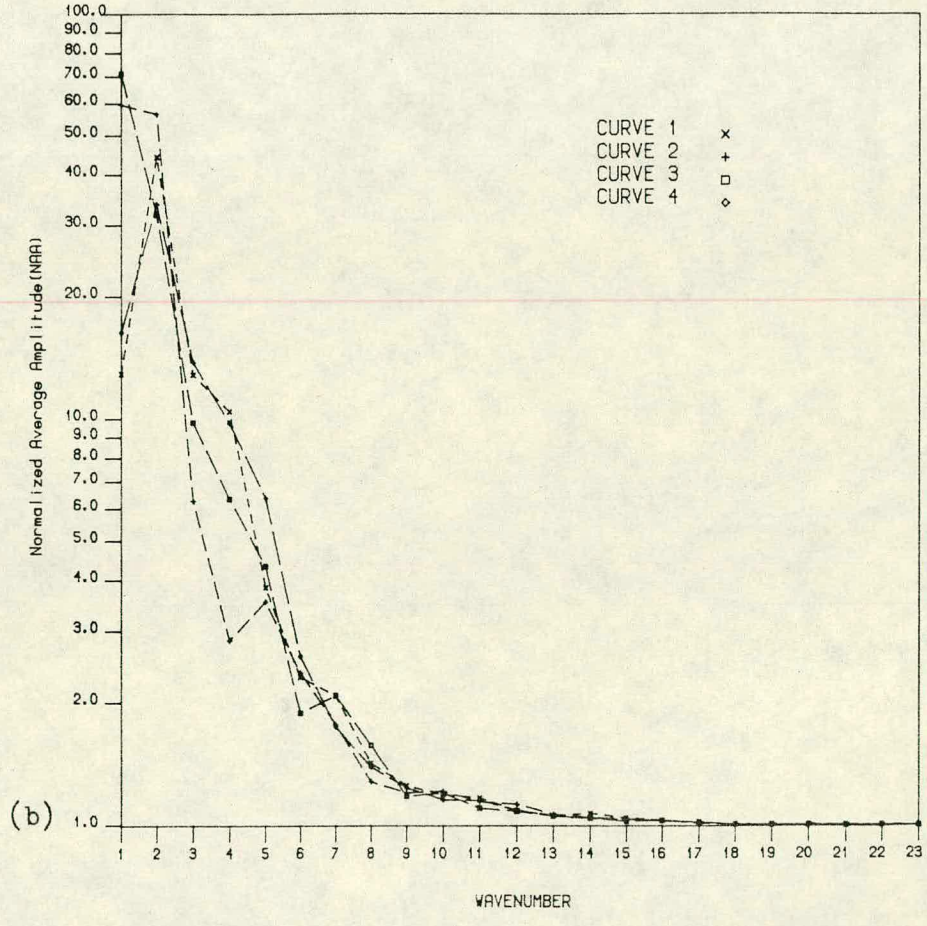
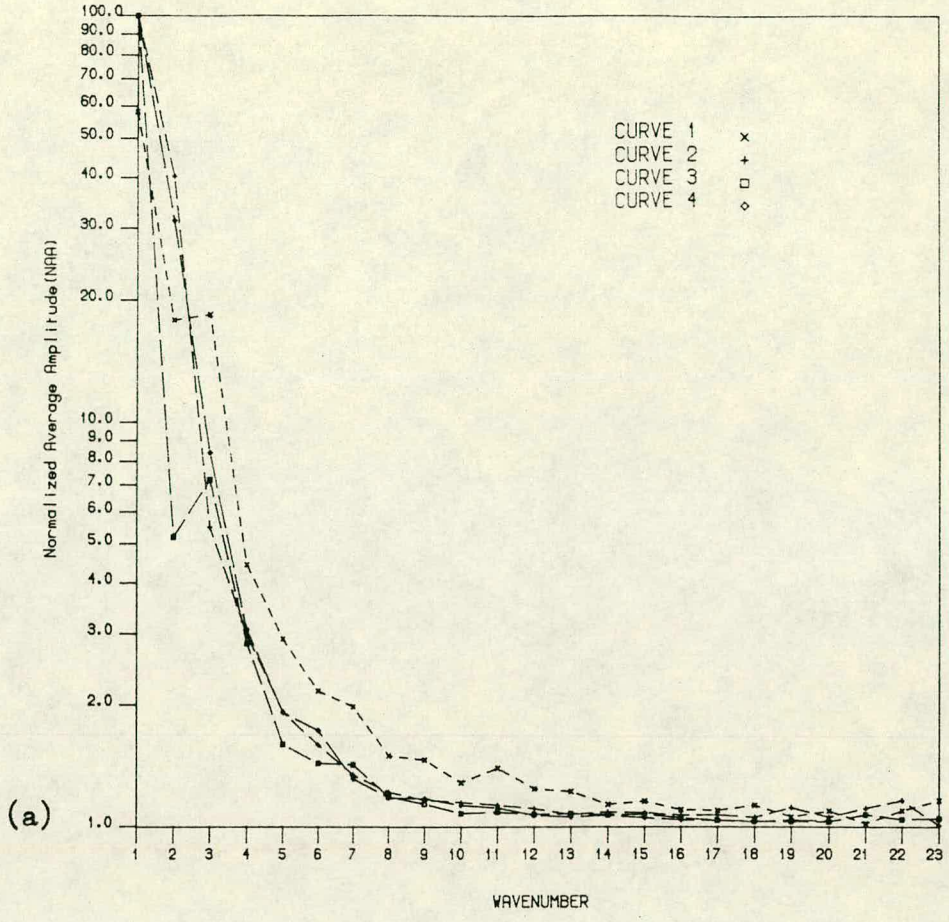


Figure 5.4 Visible (a) and Infrared (b) spectra of four Cumulus samples.



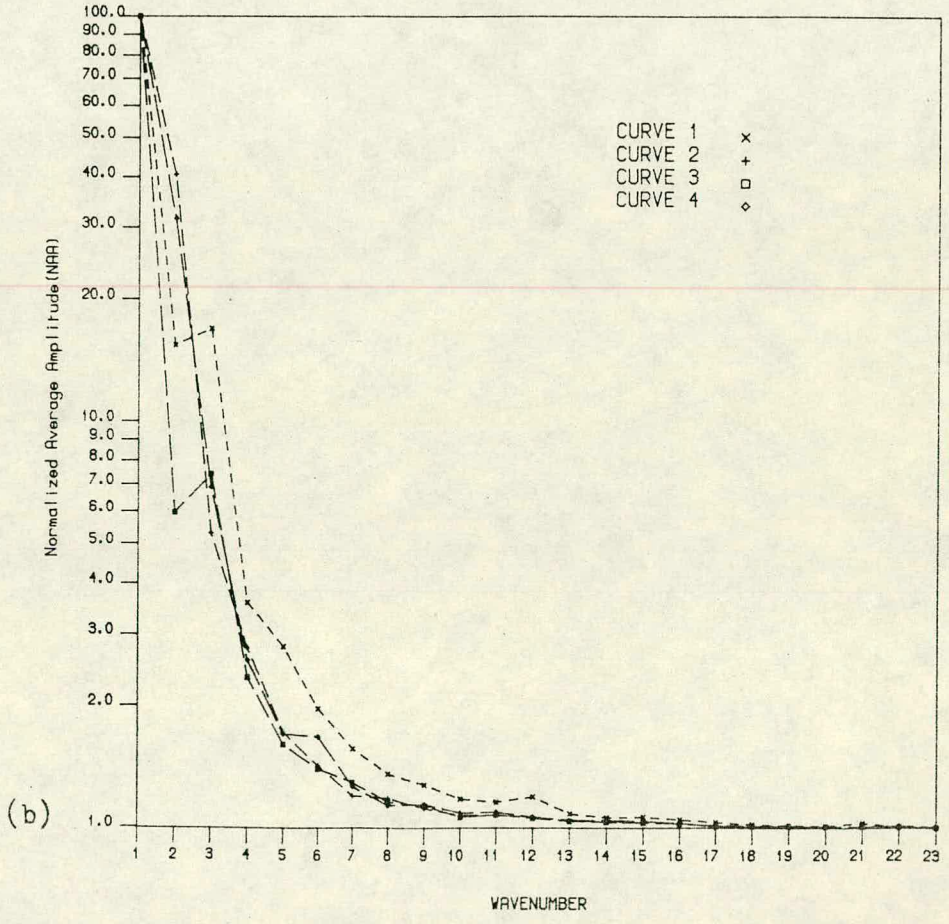
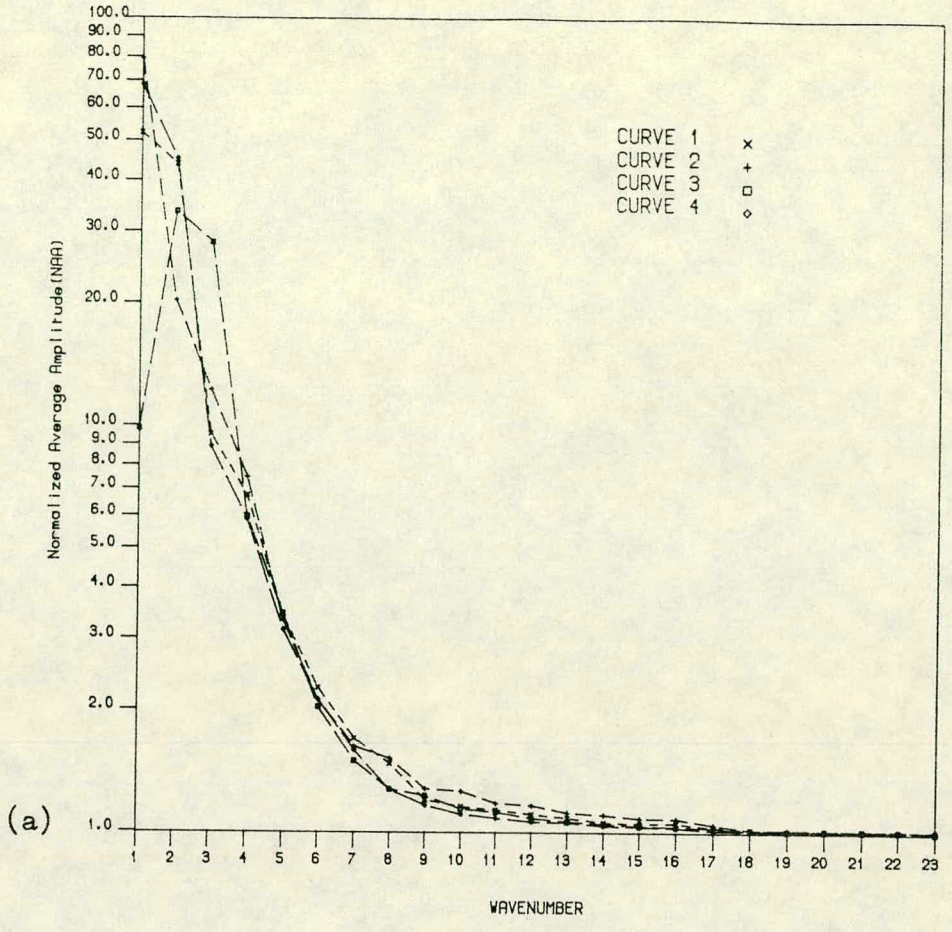


Figure 5.5 Visible (a) and infrared (b) spectra of four Cirrus samples.



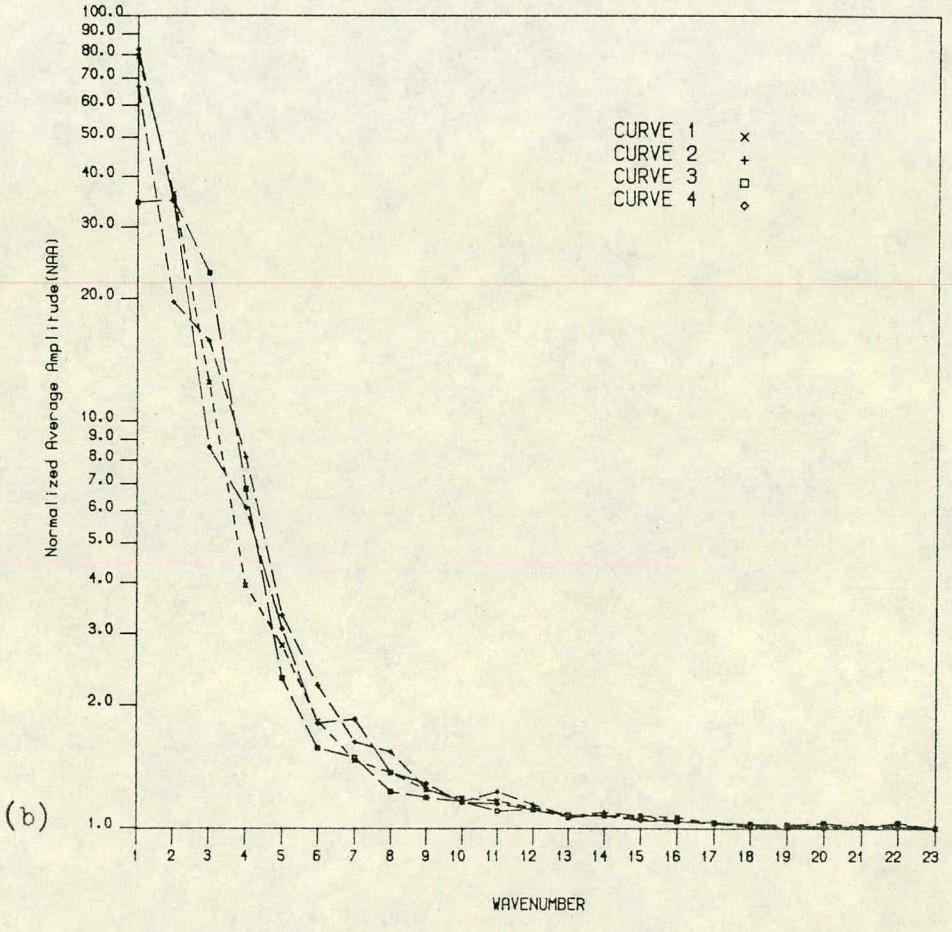
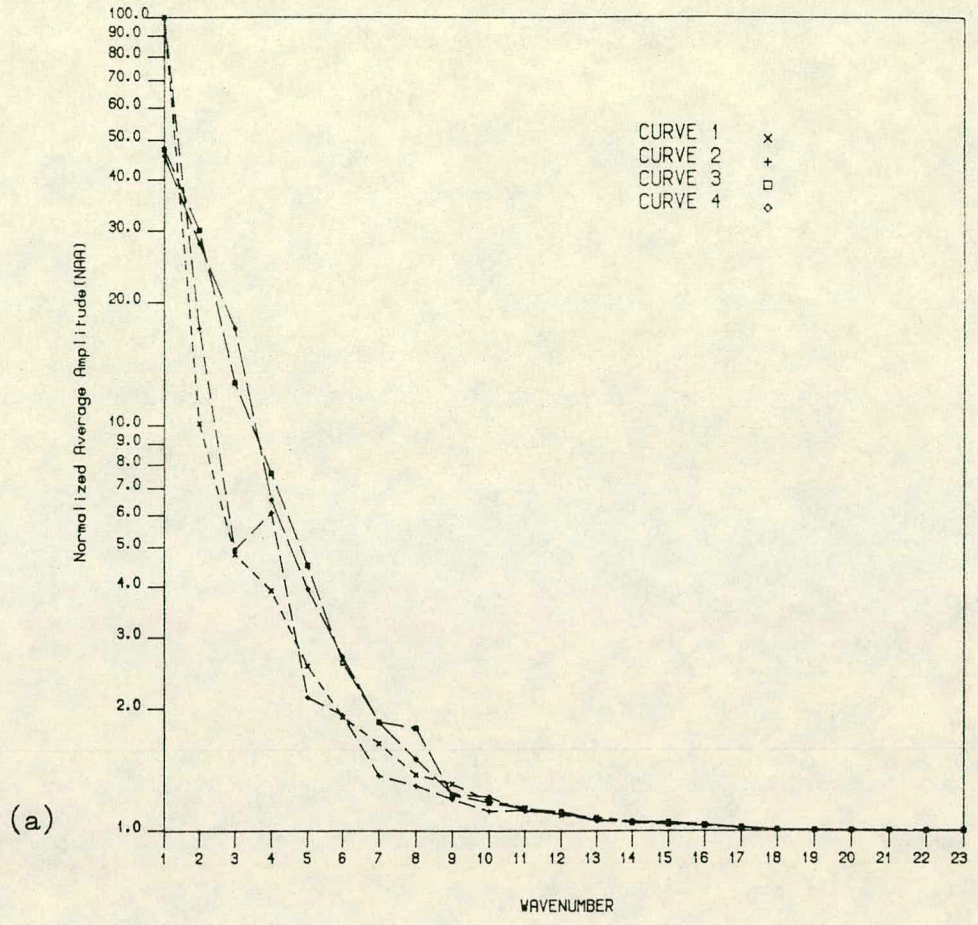


Figure 5.6 Visible (a) and infrared (b) spectra of four Cumulonimbus samples.



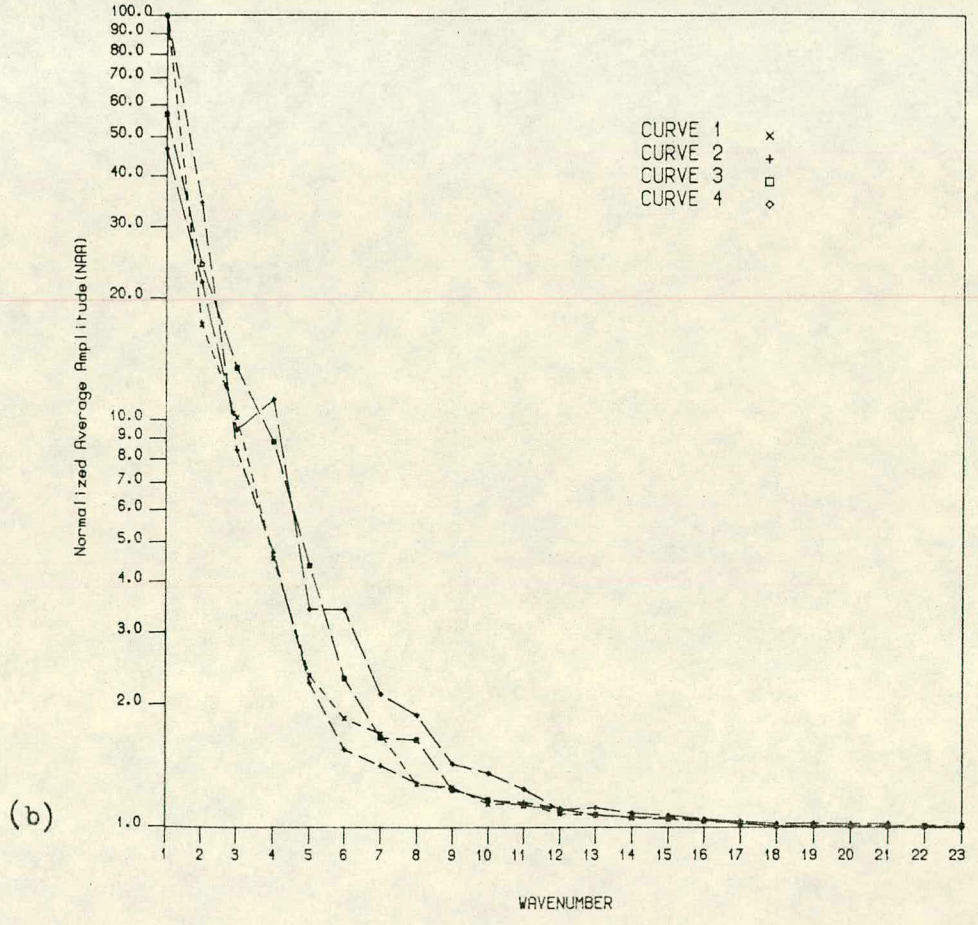


Figure 5.7 Visible (a) and infrared (b) spectra of four Stratocumulus samples.

5.2.3. Entropy-based Measures in Polar Frequency Domain (POLENTs)

Assume that there is a set (R) of m frequency entries $\{x_i, i=1, m\}$ and let the probability that x_i will appear in the set be $P(x_i)$, we can define $I(x_i) = -\log P(x_i)$ as the self-information. The expected value of self-information in this set is called the entropy of the set R, and given by

$$H(R) = -\sum_{i=1}^{m} P(x_i) \log P(x_i)$$

H(R) is a measure of the average amount of information contained in the set, or may be interpreted as a measure of the average *a priori* uncertainty regarding the appearance (see Rosenfeld and Kak 1982 for a general discussion).

It can be verified (see Sun and Shen 1985) that H(R) reaches its maximum when $\{P(x_i)\}$ are all equal. The base of the logarithm in the above equation determines the information unit. The unit of information for base *e* logarithms is called a *nat* (natural unit), while that for base 10 and base 2 logarithms are respectively called the *Hartleys* unit and a *bit* (Bell, 1953). The base 10 was used here as there is no specific requirement for interpreting a power spectrum.

In this study, the entropy-based approach was designed to characterize the absence of structures in cloud patterns by treating frequency entries within each annular band of an image spectrum as a probability distribution rather than summed spectral energies. Thus, the textural features defined by this approach provide additional spatial information which is independent of that contained in the summed energies.

The approach was originally developed by Jernigan and D'astous (1984) in cartesian coordinates. For natural scenes, cartesian coordinates are not

appropriate, as discussed in Section 5.2.1. Thus, the approach was modified in this study so that the entropy feature vector was calculated from polar spectra. An annular band R is defined as a circular band in the spatial frequency domain. The band width may be one wavenumber or more.

The significant improvement brought by such entropy-based measures can be illustrated by the same example in Section 5.2.2 of the mixture of sinusoids versus a pure sinusoid within the same band: by the measures, these two cases are clearly discriminable.

In this algorithm, for each 32x32-pixel image sample, the entropy values within annular bands are calculated from the distribution of spatial frequency entries in polar frequency domain. One entropy value can be found for each circular band so that several values of entropy are available for each image sample. The collection of these entropies then forms a feature vector representing the texture property of the sample. In essence, the entropies measure the spread of spatial frequency entries within circular bands.

The basic principles of the approach are given below.

Let $G_{s,t}$ (s=1, ..., m; t=1, ..., n) be an image array which is corrected for its overall mean by setting the mean to zero. Its two-dimensional discrete Fourier power spectrum $I_{p,q}$ can be obtained through Equation (5.12).

By defining a circular band **R** in the frequency domain, we can calculate the total energy in the band by summing

$$E_R = \sum_{p,q \in R} I_{p,q} \tag{5.14}$$

In fact, the term E_R is equivalent to the NAA spectrum in the band R

described in Section 5.2.2.

The normalized energy at (p,q) can then be expressed as

$$P_{p,q} = \frac{I_{p,q}}{E_R} \tag{5.15}$$

The set $\{P_{p,q} \mid p,q \in R\}$ may be considered as a set of probabilities. It satisfies

$$P_{p,q} \geq 0 \text{ and } \sum_{p,q \in R} P_{p,q} = 1 \tag{5.16}$$

From Equation (5.15), we can then define the entropy measure of the frequency entries within R as

$$H(R) = - \sum_{p,q \in R} P_{pq} \log_{10} P_{pq} \tag{5.17}$$

It can be shown (see Sun and Shen 1985) that this entropy measure in R actually satisfies $0 \leq H(R) \leq H_{max} = \log_{10} N$ where N is the total number of frequency entries in R. The physical properties of the entropy measure are similar to those in other applications of entropy measures and can be interpreted as follows. When $P_{p,q} = 1$ for some (p,q), i.e., the power spectrum of the sample has a single frequency entry in R, H reaches its minimum. Conversely, when $P_{p,q} = 1/N$ for all (p,q) in R, i.e., all the frequency entries in R of the power spectrum have the same amount of energy, H reaches its maximum.

Through a large number of observations, it was found that the frequency

entries at the first nine wavenumbers contained nearly all the spatial information of cloud textures in an image sample. These experiments were performed on 649 samples (32x32-pixel samples for 13 classes) of two spectral channels (see Chapter 4 for detail of the data set). In the experiments, all the frequency entries in the cartesian spectrum of an image sample were first expressed as percentages of the total energy, as the first step of the "censoring" procedure (see below). The frequency entries of the spectrum with values larger than C% (0.39 was chosen for C; see below for explanation) of the total energy were considered as entries containing the spatial information. It is obvious, e.g., from Figures 5.4 to 5.7, that not all the frequency entries in a polar spectrum contain the spatial information, especially those at relatively high wavenumbers. Thus, we intended to reduce the number of wavenumbers. The general criterion for determining an appropriate number is that the spectrum within a circular region outlined by the outermost wavenumber contains as much spatial information as possible. The first nine wavenumbers were found relatively optimal in this study: the probability that the frequency entries at these nine wavenumbers contribute over 97% (inclusive) of the total energy is just larger than 0.90 for the 649 samples in both the two spectral channels. Other choices of the outermost wavenumber are also acceptable, for example, eight and ten. Therefore, a trade-off was made here: only the frequency entries of the first nine wavenumbers were considered for the entropy feature vector.

Another trade-off was made to combine the first nine wavenumbers into three equal-width bands since we intended to compare the POLENT group with other groups of four features such as CIRCs and SGLDs on a group basis. The entropy values for each of these three equal-width bands were averaged to form an entropy feature. Thus, three POLENT features resulted from this specification, which were abbreviated to AVE ENT 1, AVE ENT 2 and AVE ENT 3. These three features, combined with the feature NO DOM (see below), form the

POLENT group.

The parameter NO DOM is the number of dominant frequency entries in a polar power spectrum. A dominant entry is one which has a larger value than some "censoring" value (see below). This parameter was used as an additional textural feature as well as the other three entropy measures.

As discussed above, entropy-based measures in polar frequency domain have the potential of characterizing the spatial distribution of frequency entries. However, in applying these measures to natural scenes such as cloud fields, some serious problems often occur. One of these is that frequency entries with relatively small values tend to contribute considerably to the total entropy within circular bands. This problem greatly reduces the discriminating power of the entropy measures for image analysis. By visually examining the sixteen subimages used in this study, it was found that most of these small entries were actually made up of noise or some ragged cloud structures in the images.

To solve this problem, a "censoring" procedure was adopted from the Renshaw and Ford's (1983, 1984) work to eliminate these small entries before the Fourier power spectrum was transformed from a cartesian coordinate system to a polar coordinate system. Following their work, all the frequency entries in half of a cartesian spectrum were first expressed as percentages of the total variance (or called the total energy). The frequency entries of the spectrum with values less than C% of the total variance, were then replaced by zeros. The value C was also adopted from their work (Renshaw and Ford, 1984) in which C = 0.39 (i.e., 400/1024) was found appropriate for their ecological patterns in such a way that it was just large enough for most background noise to be removed and thus distinct patterns to be revealed, although it might not be optimal for interpreting our cloud patterns. The later experiments showed that the "censoring" procedure with an appropriate C value is essential if these

POLENT features are to be applied to remotely sensed data. The application of these features to characterizing cloud categories and their discriminating potential measured by some separability indices are discussed in Section 6.3.2.2.

CHAPTER 6

EVALUATION OF TEXTURAL FEATURES APPLIED TO AVHRR SATELLITE IMAGERY

As discussed and summarized in Chapter 2, a number of fundamental problems remain in applying remote sensing techniques to cloud classification. One of the problems we have to deal with first is, perhaps, the lack of a mathematical form into which the verbal description of cloud types used by human analysts can be converted. It was discussed in Section 2.3 that the six qualitative criteria are usually used to interpret cloud types but it is very difficult to quantify these criteria, since the study of human vision and the corresponding automatic techniques have not been taken to sufficient depth. However, the discipline of image processing, combined with psychological studies, does provide many texture descriptors in terms of either structural or statistical approaches to characterize the grey level primitive properties of image texture and spatial interrelationships between them. These descriptors may be selected to aid multi-spectral cloud classification. Thus, the problem for those working towards developing a quantitative cloud analysis scheme is how to find the most appropriate descriptors from a large number of available choices.

Another problem is to choose an appropriate separability measure with which to assess the merits of textural features under investigation.

The first problem is mainly considered in Section 6.3. Some introduction to the texture analysis algorithm is given in Section 6.1. The choice of a separability measure is discussed in Section 6.2.

6.1. Introduction to the Texture Analysis Algorithm

The texture analysis algorithm consists of three basic procedures: (1) training stage, (2) calculation of spectral and textural features for all gathered samples, (3) feature selection and assessment of class separability.

The first procedure and a detailed description of various textural features used in this study were given in Chapters 4 and 5, respectively. Other procedures are the major concern of this chapter.

6.1.1. Calculation of Spectral and Textural Features

Besides the textural features (see below), five spectral features (MEAN, MAX, MIN, MAX/MIN, RANGE) (see Table 6.1B) were also incorporated in this algorithm. These five features were adopted from the Ebert's (1987) work in order to show the dominating importance of some of the spectral features for cloud classification.

Two general types of textural features, spatial features and spatial frequency features, were computed in this study. The textural features incorporated in this algorithm include the twenty-one features (2th, 7-26th in Table 6.1B), including both types in two AVHRR spectral channels: visible and infrared channel. Except the simple textural feature S.D. (i.e., local variance), all the textural features were fully described in Chapter 5. Textural information in Channel 3 (near-infrared) was not used due to the highly noisy appearance of the AVHRR images in this channel. Each textural feature acts as an additional "channel" to the original three-channel spectral feature space and its value was calculated from an ensemble of 649 radiance sample areas. These 649 sample areas were obtained

from different 3-channel image segments of the 13 cloud classes (see Table 6.1A), which were isolated by a cluster analysis (see Chapter 4 for details). The size of each sample area is a 32 x 32-pixel grid, representing about 35 x 35 km square area at the satellite sub-track. The reason for this choice is given in Section 4.1.

Comparatively, the calculation of all the spatial features, even the sophisticated second-order grey level difference statistics (SGLDs), took much less CPU time than that of the spatial frequency features (0.04 CPU second for calculating 4 SGLD features of each sample area; 0.21 CPU second for calculating 3 NAA features and 0.22 CPU second for calculating 4 POLENT features of each sample area on Hitachi NAS VL80 machine). This indicates their efficiency in computation.

6.1.2. Feature Selection and Assessment of Class Separability

The general objective of a feature selection procedure is to find an optimal combination of spectral features and textural features which will enhance the separability of the cloud classes of interest. The procedure provides an indication of the discriminating potential of features selected, and consequently, about the highest achievable performance of a decision making system for a given representation space. Thus, it is often considered that a further role of the feature selection procedure is to establish whether it is necessary to seek additional measurements which will contain discriminating information allowing the improvement of the performance of the system, and meanwhile, the assessment of separability of some hard-to-separate class pairs.

Cloud Types	Number of Samples
1. Land	41
2. Stratus	75
3. Sea surface	91
4. Cumulus	69
5. Thick cirrus over low cloud	77
6. Altostratus over middle frontal cloud	26
7. Thick cirrostratus	42
8. Thin cirrostratus over middle frontal cloud	33
9. Cumulonimbus	57
10. Thin cirrus with open water	54
11. Thin cirrus over low cloud	16
12. Thin cirrus over middle frontal cloud	56
13. Stratocumulus	12
in total	649

Table 6.1A Thirteen cloud classes (i.e., 78 cloud class pairs) used in this study.
(Note that this table is quoted from Section 4.2.1)

MEAN	max ASM
S.D.	max ENT
MAX	mean NA
MIN	max NAA
MAX/MIN	DOM WVN
RANGE	AVE ENT 1
R.G.	AVE ENT 2
mean MEAN	AVE ENT 3
mean CON	NO DOM
mean ASM	CIRC 1
mean ENT	CIRC 2
max MEAN	CIRC 3
max CON	CIRC 4

Table 6.1B Twenty-six spectral (1st, 3-6th) and textural (2th, 7-26th) features used in this study. (Textural features were calculated from visible and infrared channels only.)

The performance of a classification system is often estimated in terms of error probability of classification, but it is shown by Swain and Davis (1978) that computing the error probability is often not feasible. Therefore, an alternative predictor of the performance is often used instead, called a statistical separability index. It is related to the error probability: the greater the statistical separability of class pairs, the smaller the error probability of classification.

There have been several statistical separability indices, which may be used for remote sensing purpose. The most widely-used index in cloud classification studies is the one-dimensional (1-D) Fisher distance. It has been employed by researchers such as Parikh (1977, 1978), Seddon (1983) and Wu *et al.* (1985), and proved to be a simple and computationally efficient measure. However, This index has many limitations so that we are often not content with the results derived by using it. One of the most crucial limitations is, perhaps, that its characterization of class separability is performed in one dimension only. Due to the limitation, features can only be assessed individually for cloud class pairs but their composite function for cloud classification cannot be revealed. In fact, features selected through a feature selection procedure by the index are often not optimal in multi-dimensional feature space, and conversely, features which rank highly in multi-dimensional feature space often cannot be selected by the index (The problem will be illustrated in Section 6.2.1). This limitation is especially significant in our case, because multi-spectral features are dominant in performance for cloud classification and so textural features, which are often not comparable to the spectral features, can only be expected to aid these spectral features in multi-dimensional feature space.

The accurate assessment of various textural features in multi-dimensional feature space requires a multivariate separability measure, which takes into account the distributions of class clusters. Based on this idea, a commonly-used pairwise separability index, the transformed divergence, was finally used in this

study. This index is described in Section 6.2.2, to highlight the roles of various textural features in aiding multi-spectral cloud classification. It can show not only the overall usefulness of a textural feature in addition to some basic spectral features for all class pairs, but also show its importance in separating some class pairs of interest.

To compare the relative merits of various textural features by means of the multivariate separability index, several experiments (see Section 6.3) were designed to carry out a comparative study based on the class statistics (see Chapter 4 for details of the statistics). These single- or group-based feature comparisons were made in several different cases such as, in visible channel alone, in infrared channel alone, or in both channels for these features. The contributions of primitive size variations to the separability of various cloud class pairs, reflected by feature pairs in which each feature was calculated at a different distance, were also assessed. The results of these comparisons are given in Section 6.3.

6.2. Separability Indices

AVHRR radiance images used in this study, like other remotely sensed data, involve a pixel having its characteristics recorded over a number of spectral channels plus, in this study, textural "channels" (textural features). A cloud class, represented by a group of such pixels, may also be characterized over the same channels. The difference between these two cases is that the former is specified as a single point in the multi-dimensional feature space while the latter is specified as a cluster in the space. The representation of the cluster is specified by a joint probability distribution, which leads to a parametric approach of pattern recognition. The mean vector in the multi-dimensional feature space

locates the single peak of an unimodal probability density distribution while the covariance matrix outlines the region of influence of the probability density distribution in each dimension. Unfortunately, the distributions of different classes tend to overlap each other, as the distributions are often broad and not sharply defined. To quantitatively reveal the separability of these class distributions requires a definite measure. A separability index is often used to assess the effects of spectral channels as well as textural features for class separation in multi-dimensional feature space. It can be defined as a function of separation between class probability distributions in either one-dimensional or multi-dimensional feature space (Swain and Davis, 1978).

In the following two subsections, two separability indices, the 1-D Fisher distance and the multivariate transformed divergence, are discussed.

6.2.1. Fisher Distance

Fisher distance is a measure of class separability (Duda and Hart, 1973). It is converted from a Fisher linear discriminant function and can be defined as

$$J_{ij} = \frac{|\mu_i - \mu_j|}{\sigma_i + \sigma_j} \qquad (6.1)$$

where μ_i and σ_i are respectively the mean and standard deviation of one feature or channel for the samples of Class Ci.

The distance value J_{ij} is associated with an error probability of classification. To illustrate this we consider two normal distributions shown in Figure 6.1, one for each cloud class. The positions of the means μ_i and μ_j lead to the gross separability of the two classes, while different spreads in the training class

statistics lead to different spreads in the likelihood distributions. If these two distributions overlap in the 1-D feature space as shown in the figure so that the distance value J_{ij} is small, the error probability of classification will be high. This distance J_{ij}, although considered only in one dimension, takes into account two statistical factors, mean and variance. The overlapping of any two cloud classes can be attributed to two reasons: their mean values are similar, and the variance value of either one or both is large. In these two cases, the distance values appear small.

By using this measure for each feature, the separabilities of class pairs can be revealed in a single table, e.g., Table 7.3 of Seddon (1983). The table is useful in showing the relative usefulness of a single feature for distinguishing between different class pairs.

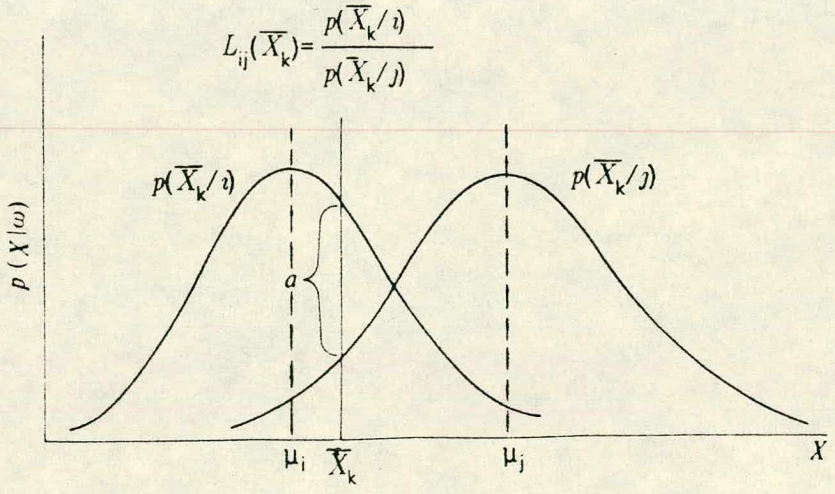


Figure 6.1 Likelihood distribution for two classes in one-dimensional feature space

Following the technique used by Ebert (1987) we initially applied the Fisher distance for feature selection, and then the results from this procedure were compared to the later results given by the transformed divergence analysis. Some results arising from the former procedure are given in Tables 6.2A and 6.2B. Table 6.2A shows the 13 most useful features (the feature category numbers are given in Table 6.2B) for differentiating Class C_i from Class C_j and Table 6.2B shows the relative usefulness of these features in one dimension, which is characterized by the number (in parenthesis) of each feature ranking first for the 78 class pairs.

These 13 features can then be placed in order of importance as follows:

- MEAN 3
- MEAN 1
- MEAN 4
- MAX 3
- MIN 4
- max ENT 1
- mean CON 4
- MAX 1
- MIN 1
- mean MEAN 1
- RANGE 1
- MAX/MIN 1
- mean NAA 1

To reveal any relationship between these 13 features, the correlation coefficients between them were analyzed on the basis of the sample statistics (see Chapter 4 for details). The correlation coefficients are listed in Table 6.3.

C_i \ C_j	1	2	3	4	5	6	7	8	9	10	11	12	13
1	*												
2	3	*											
3	2	1	*										
4	3	7	11	*									
5	9	2	1	5	*								
6	3	2	3	2	2	*							
7	3	3	1	1	3	1	*						
8	3	3	3	2	2	2	2	*					
9	13	3	8	8	9	2	9	4	*				
10	2	1	3	4	1	1	1	3	8	*			
11	3	10	2	6	2	2	2	2	2	2	*		
12	1	1	1	5	1	1	2	2	2	1	1	*	
13	13	13	11	2	12	2	2	2	2	8	12	12	*

Table 6.2A Most useful features (as measured by one-dimensional Fisher distance) for distinguishing between class pairs C_k and C_j. (Refer back to Table 6.1A for cloud category numbers)

1 = MEAN 1	(16)	8 = MAX 3	(4)
2 = MEAN 3	(26)	9 = MIN 4	(3)
3 = MEAN 4	(14)	10 = mean NAA 1	(1)
4 = MAX 1	(2)	11 = mean MEAN 1	(2)
5 = MIN 1	(2)	12 = max ENT 1	(3)
6 = RANGE 1	(1)	13 = mean CON 4	(3)
7 = MAX/MIN 1	(1)		

Table 6.2B Feature category

Note: 1. The suffices of these features indicate the channels from which they are calculated.
2. The number of times each feature appearing in Table 6.2a is indicated in parenthesis.

F_j \ F_k	1	2	3	4	5	6	7	8	9	10	11	12	13
1	*												
2	0.05	*											
3	0.65	0.65	*										
4	0.51	0.02	0.38	*									
5	0.90	0.12	0.57	0.28	*								
6	0.18	-.02	0.18	0.93	-.01	*							
7	0.07	0.03	0.05	0.83	-.13	0.91	*						
8	0.10	0.89	0.66	0.11	0.09	0.09	0.06	*					
9	0.65	0.64	0.94	0.33	0.63	0.10	0.02	0.56	*				
10	0.12	0.08	0.26	-.16	0.09	-.20	-.31	0.11	0.19	*			
11	0.73	-.16	0.51	0.55	0.47	0.39	0.11	0.01	0.43	0.12	*		
12	0.49	-.26	0.37	0.44	0.21	0.38	0.06	-.05	0.25	0.21	0.92	*	
13	0.64	0.65	0.99	0.38	0.57	0.18	0.05	0.67	0.92	0.25	0.50	0.36	*

Table 6.3 Correlation coefficients of the 13 most useful features F_k versus F_j for all the cloud classes.
Refer back to Table 6.2B for feature numbers.

To reduce the redundancy between these features, any feature with a correlation coefficient with a higher-ranking feature exceeded 0.80 was discarded. From this procedure, seven out of the thirteen features were left. These are:

- MEAN 1
- MEAN 3
- MEAN 4
- MAX 1
- MAX/MIN 1
- mean NAA 1
- max ENT 1

From the above results, we see that the number of features which can be selected by this procedure is fairly small (seven in this case), and among these only a few textural features are left (two in this case). Comparatively, the occurrence of the three basic spectral features, MEANs, is very high (see Table 6.2B). Some of the others such as MIN 1, MAX 3, MIN 4 and mean CON 4 are highly correlated to these three spectral features (see Table 6.3) and so discarded by the subsequent correlation analysis. Moreover, some features such as the spectral features: MAX 1, RANGE 1 and MAX/MIN 1, and the textural features: max ENT 1 and mean MEAN 1, although not highly correlated to these three spectral features, are highly intercorrelated and so most of them were also discarded. The features, which were finally selected (seven in this case), are all which ranked first for differentiating between class pairs C_i and C_j in 1-D feature space.

By comparing the two remaining textural features with the results of the multivariate divergence analysis (shown in Table 6.18), the significant difference of feature selection between in one dimension only and in multi-dimensional feature space can be observed: the higher of the two textural features, max ENT 1, appears behind mean MEAN 4, S.D. 4 and mean MEAN 1; the other one, mean NAA 1, is far behind many other textural features. Thus, these two textural features are not optimal in multi-dimensional feature space, i.e., the three-channel spectral features MEANs, when combined with these two textural features, cannot be expected to achieve the highest classification accuracies. This indicates the necessity of using a multivariate separability index for assessing the auxiliary function of each textural feature in aiding multi-spectral cloud classification.

It has been shown in Chapter 4 that the three-channel spectral features MEANs alone were not adequate in separating between some highly textural cloud classes and so the involvement of additional features was necessary. The

results from the 1-D feature selection indicates further the necessity by showing quantitatively the dominant importance of these spectral features for cloud classification as well as their inadequacies (see Table 6.4). The separability of 11 cloud class pairs was considered in Table 6.4. These cloud classes were mostly those reported as hard-to-separate by Liljas (1984) and Karlsson (1987A). From the table, it can be seen that all the distance values $\{J_{ij}\}$ of the 11 class pairs are below 2.85, and most of the features ranking first are the spectral features MEANs in different spectral channels. This means that the error probability of cloud classification is still high and so needs to be reduced. Therefore, appropriate features, either spectral or textural features, are strongly needed to enhance the separability. To find such textural features is the concern of this study.

The Fisher distance was used in this study as a one-dimensional approach to select the most useful spectral and textural features for the 78 class pairs (see Tables 6.2A and 6.2B), and meanwhile to show the order of the within-group textural features (6 groups considered in this study: SGLDs, R.G., S.D., NAAs, POLENTs and CIRCs) for comparison with that ranked by the multivariate transformed divergence (see Section 6.3). But it is difficult to use this method to compare the order of the textural features in different groups. To give an overall view of the order of all the twenty-six features on basis of a unique criterion, a variance analysis was used instead in this study. The 26 features in the two spectral channels were ranked into order according to the F-ratio values. The results are shown in Table 6.5 to give an overall view of the relative importance of these 26 features in one dimension.

	Class 2	Class 3	Class 4	Class 5
C5	1.1085 (MEAN 3)		1.6763 (MIN 1)	
C6			1.4556 (MEAN 3)	1.6458 (MEAN 3)
C7				1.6950 (MEAN 4)
C10		1.2547 (MEAN 4)		
C11	1.1730 (mean NAA 1)		1.3897 (RANGE 1)	
C12	1.4030 (MEAN 1)			
C13			1.6890 (MEAN 3)	2.8357 (max ENT 1)

Table 6.4 Fisher distance values J_{ij} between cloud class pairs of interest for spectral and textural features ranking first among 57 features in three channels (Refer back to Table 6.1A for cloud category number; only spectral features (1st, 3-6th) in Table 6.1B were used in near-infrared channel). In the table, the feature names are given in parenthesis below the distance values J_{ij}, and the suffix of each feature name indicates the channel in which the feature was calculated.

	Visible		Infrared	
Order	Feature	F-Ratio	Feature	F-Ratio
1	MEAN	588.9	MEAN	975.3
2	mean CON	318.6	mean CON	766.3
3	mean ASM	285.7	max CON	505.7
4	max CON	279.3	MIN	479.5
5	max ASM	265.3	MAX	311.9
6	mean ENT	215.4	mean MEAN	183.6
7	max ENT	209.6	max MEAN	152.9
8	mean MEAN	195.2	max ENT	70.5
9	MIN	194.7	mean ENT	66.6
10	max MEAN	191.7	max ASM	65.8
11	S.D.	78.8	mean ASM	65.7
12	MAX	31.3	MAX/MIN	42.7
13	R.G.	23.5	RANGE	41.0
14	CIRC 1	21.9	S.D.	40.4
15	CIRC 2	15.5	CIRC 1	13.6
16	mean NAA	14.4	CIRC 2	10.9
17	CIRC 3	13.4	R.G.	10.0
18	RANGE	12.7	CIRC 3	9.1
19	CIRC 4	12.4	CIRC 4	8.0
20	NO DOM	11.9	mean NAA	6.2
21	max NAA	11.6	AVE ENT 3	5.2
22	AVE ENT 2	9.7	NO DOM	4.5
23	AVE ENT 3	7.4	AVE ENT 2	4.5
24	MAX/MIN	4.7	max NAA	4.4
25	AVE ENT 1	3.0	AVE ENT 1	3.0
26	DOM WVN	2.0	DOM WVN	2.0

Table 6.5 The rank of 26 features derived from a variance analysis.

6.2.2. The Transformed Divergence

There are a number of multivariate separability indices in common use. One of these, called the transformed divergence (Swain and Davis, 1978), was chosen for this comparative study. The primary reason for this choice is that it has been verified in many cases to be superior to other indices as a measure of the multi-dimensional separability of probability distributions, and it has been extensively used in land-use classification studies (e.g., Boyd *et al.*, 1983).

To provide a basic understanding of the multivariate separability index, some mathematical explanation is given below. A more detailed treatment of the index can be found in either Swain and Davis (1978) or Thomas *et al.* (1987).

Given a set of q_i image pixels representing i-th cloud class $\{x_{km}, k=1, ..., q_i; m=1, ..., n\}$ in an n-dimensional feature space, we can express the statistics of the k-th sample point (pixel) of the data set in matrix form as

$$\bar{X}_k = \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{km} \\ \vdots \\ x_{kn} \end{bmatrix} \qquad \bar{\mu}_i = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{im} \\ \vdots \\ \mu_{in} \end{bmatrix}$$

and

$$\bar{V}_i = \begin{bmatrix} v_{i11} & v_{i12} & \cdots & v_{i1n} \\ v_{i21} & v_{i22} & \cdots & v_{i2n} \\ \vdots & \vdots & v_{ilm} & \vdots \\ v_{in1} & v_{in2} & \cdots & v_{inn} \end{bmatrix} \tag{6.2}$$

where μ_{im} and V_{ilm} are given by

$$\mu_{im} = \frac{1}{q_i} \sum_{k=1}^{q_i} x_{km}, \quad m = 1, 2, ..., n \tag{6.3}$$

$$v_{ilm} = \frac{1}{q_i - 1} \sum_{k=1}^{q_i} (x_{kl} - \mu_{il})(x_{km} - \mu_{im}), \quad l,m = 1, 2, \ldots, n \tag{6.4}$$

and q_i is the total number of image pixels in the class-**i** sample set. The probability of a point $\bar{X}_k$, taken over all n dimensions, being assigned to class **i** is $P(\bar{X}_k/i)$:

$$p(\bar{X}_k|i) = (2\pi)^{-n/2} |\bar{V}_i|^{-1/2} \exp\{-\tfrac{1}{2} (\bar{X}_k - \bar{\mu}_i)^T \bar{V}_i^{-1} (\bar{X}_k - \bar{\mu}_i)\} \tag{6.5}$$

Assume that the *a priori* probabilities for all classes are all equal, we can define a likelihood function as

$$L_i(\bar{X}_k) = P(\bar{X}_k/i), \qquad i=1, \ldots, \ldots, r \tag{6.6}$$

Referring back to Figure 6.1 as an illustration, we can see that for a given point $\bar{X}_k$ the separability of the class pairs (i,j) depends on the relative values of the class likelihood distribution at that point. We can define this relation at the point $\bar{X}_k$ as a ratio:

$$L_{ij}(\bar{X}_k) = L_i(\bar{X}_k)/L_j(\bar{X}_k) \tag{6.7}$$

where $L_{ij}(\bar{X}_k)$ is the likelihood ratio between the two classes **i** and **j** at $\bar{X}_k$. If $L_i(\bar{X}_k)$ is larger than $L_j(\bar{X}_k)$ at the point, $\bar{X}_k$ will be assigned to the **i**-th class.

The likelihood ratio $L_{ij}(\bar{X}_k)$ can be re-defined as the logarithmic likelihood ratio $L_{ij}'(\bar{X}_k)$ (see below) so that it can be used mathematically in an easier manner. When Equation (6.5) is incorporated, $L_{ij}'(\bar{X}_k)$ can be expressed as:

$$\begin{aligned} L_{ij}'(\bar{X}_k) &= \ln L_i(\bar{X}_k) - \ln L_j(\bar{X}_k) \\ &= \tfrac{1}{2} \ln |\bar{V}_j|/|\bar{V}_i| + \tfrac{1}{2} (\bar{X}_k - \bar{\mu}_j)^T \bar{V}_j^{-1} (\bar{X}_k - \bar{\mu}_j) \\ &\quad -\tfrac{1}{2} (\bar{X}_k - \bar{\mu}_i)^T \bar{V}_i^{-1} (\bar{X}_k - \bar{\mu}_i) \end{aligned} \tag{6.8}$$

Through a number of the mathematical operations, the multivariate divergence index can be derived and expressed as:

$$D_{ij} = \frac{1}{2} \mathrm{tr}(\bar{V}_i - \bar{V}_j)(\bar{V}_j^{-1} - \bar{V}_i^{-1}) + \frac{1}{2} \mathrm{tr}(\bar{V}_i^{-1} + \bar{V}_j^{-1}) * (\bar{\mu}_i - \bar{\mu}_j)(\bar{\mu}_i - \bar{\mu}_j)^T \quad (6.9)$$

where **tr** stands for the trace of a matrix. In Equation (6.9), the first term gives the difference between the respective covariance matrices over an n-dimensional feature space while the second term is in essence the average of Mahalanobis distance between class means. D_{ij} is never zero unless both means and covariance matrices are identical.

The explanation of Equation (6.9) leads to five mathematical properties of the divergence D_{ij}, which are discussed fully by Swain and Davis (1978) and are given briefly as follows:

1. $D_{ii} = 0$

The divergence of one likelihood distribution relative to itself is zero, i.e., when the two classes are identical.

2. $D_{ij} > 0$

For two different likelihood functions the divergence is always larger than zero.

3. $D_{ij} = D_{ji}$

Divergence is symmetrical between two classes over the same n-dimensional feature space.

4. $D_{ij}(x_1, x_2, ..., x_n) = \sum D_{ij}(x_k)$

If the n-dimensions in a feature space are statistically independent, then the overall divergence in a multi-dimensional feature space is equal to the sum of the divergences of all individual dimensions.

5. $D_{ij}(x_1, x_2, ..., x_n, x_{n+1}) \geq D_{ij}(x_1, x_2, ..., x_n)$

For two different classes, the addition of any extra dimension never decreases the class separability.

Some practical problems remain in using this divergence measure. An

obvious problem is that the divergence D_{ij} is not saturated even when two classes **i** and **j** are well separated in feature space and the probability of correct classification stabilizes towards 100%. This leads to some difficulties in interpreting these D_{ij} values.

One solution is to transform the divergence into a saturation function. For this purpose, Swain and Davis (1978) define a pairwise transformed divergence,

$$^{T}D_{ij} = a\,[1 - \exp(-D_{ij}/b)] \tag{6.10}$$

where the constants **a** and **b** are chosen to suit the values of $^{T}D_{ij}$ for the computation process and the desired saturation value.

Another major problem is often met when more than two classes are dealt with. To solve it, an average weighted pairwise divergence $^{T}D_{ave}$ may be used for the total of r classes:

$$^{T}D_{ave} = \sum_{i=1}^{r} \sum_{j=1}^{r} p(i)\; p(j)\; {}^{T}D_{ij} \tag{6.11}$$

The interclass divergence is weighted by the *a priori* probabilities, p(i) and p(j), of the presence of two classes C_i and C_j. The probabilities p(i) and p(j) were assumed to be equal in this study as in land-use classification studies. They are important in practice in a statistical decision maker where they act as weights. However, due to their high dependence on weather changes, the estimation of the *a priori* probabilities is impossible. This averaged divergence helps us to select an optimal feature set for the differentiation of r classes.

Similarly, $^{T}D_{ave}{}^{*}$ is defined as an average weighted pairwise divergence for the least separable pairs of r classes with transformed divergences $\{^{T}D_{ij}\}$ less

than 1500 (when **a** in Equation (6.10) is 2000), since we are often more interested in finding a best feature set that can achieve the best separability for the least separable class pairs but ${}^{T}D_{ave}$ could not well reflect the separability of these class pairs. The evaluation of the $\{{}^{T}D_{ij}\}$ values for different textural features in this study showed that class pairs which were well separated contributed the large ${}^{T}D_{ij}$ values (mostly larger than 1800 when **a** is 2000) to the average pairwise divergence ${}^{T}D_{ave}$ and these values inappropriately weighted the average divergence ${}^{T}D_{ave}$ so that the separability of other class pairs which were not well separated could not be well reflected. This led to the alternative choice ${}^{T}D_{ave}{}^{*}$. The threshold 1500 was decided based on the evaluation of the $\{{}^{T}D_{ij}\}$ values for thirty-two textural features, which were finally used in this study (see Table 6.18), in addition to the three basic spectral features MEANs. When the amplitude **a** in Equation (6.10) is 2000, the frequency distribution of the $\{{}^{T}D_{ij}\}$ values (${}^{T}D_{ij}$ versus Number of Class pairs C_i and C_j having the value) is mainly concentrated within two intervals: one above 1900 and the other below 1000. For a group of textural features, the probabilities that ${}^{T}D_{ij}$ is above 1900 and that ${}^{T}D_{ij}$ is below 1000 were about 0.30 and 0.50, respectively. The valley of the probability distribution appeared between 1300 and 1700. For a single textural feature with the three basic spectral features, the probabilities that ${}^{T}D_{ij}$ is above 1900 and that ${}^{T}D_{ij}$ is below 1000 were about 0.15 and 0.70, respectively. The valley appeared between 1100 and 1700. 1500 was chosen here as a trade-off.

The three separability measures, ${}^{T}D_{ij}$, ${}^{T}D_{ave}$ and ${}^{T}D_{ave}{}^{*}$ were all used in this study to give different judgement criteria for measuring the separability between two classes C_i and C_j, and the separability of either the overall 78 class pairs or the least separable pairs of these class pairs. ${}^{T}D_{ave}{}^{*}$ was used in this study as a major criterion, while the joint use of ${}^{T}D_{ave}$ and ${}^{T}D_{ave}{}^{*}$ was to avoid misinterpretation of class separability by using ${}^{T}D_{ave}{}^{*}$ alone. Their applications

to different cases are illustrated in Section 6.3.

As seen from Equation (6.10), two parameters have to be decided before the equation can be used to assess the separability of class pairs for a given set of spectral and textural features. In this study, just following Thomas *et al.* (1987), we derived these two parameters by solving Equation (6.10) for $D_{ij} = {}^{T}D_{ij} = 1800$ with **a** = 2000. Thus, we obtained **b** = 781.73.

6.3. Evaluation of Textural Features

This section contains the major results of the experiments carried out throughout the study. These experiments were performed on the basis of the cloud class statistics, which arose mainly from a cluster analysis and subsequent subjective judgements (see Chapter 4 for details), to compare the relative merits of both spatial and spatial frequency features (see Chapter 5 for details) in aiding conventional multi-spectral cloud classification.

These experiments include mainly the comparisons within and between either spatial features or spatial frequency features in addition to the three basic spectral features - visible MEAN, near-infrared and infrared MEAN. To assess further the discriminating potential of SGLD, CIRC and POLENT features, the comparisons between feature pairs for each of these three groups were also performed to reveal the importance of size variations that these features involve for cloud classification.

All the comparisons were made in terms of both the transformed divergence of all the 78 class pairs (${}^{T}D_{ave}$) and the transformed divergence of the least separable class pairs (${}^{T}D_{ave}$*) (see Section 6.2.2).

The simple textural feature, local variance, was also assessed here in the two spectral channels to compare its merit with the merits of the other textural features considered in this study.

6.3.1. Spatial Features

6.3.1.1. Spatial Grey Level Difference Statistics (SGLDs)

It was shown in Section 6.2.1 that the feature selection performed in one dimension cannot reveal the auxiliary contributions of various textural features in aiding multi-spectral cloud classification. To highlight the importance of various textural features for the classification, these textural features have to be assessed on an individual or group basis with spectral features in multi-dimensional feature space by using a multivariate separability measure.

The assessment of the SGLD features within the group in this study was made on both the bases using the transformed divergences (see Section 6.2.2). The purpose of this experiment is to compare the relative merit of each SGLD feature individually as an aid to multi-spectral cloud classification (i.e., the contribution of each SGLD feature with the three basic spectral features to the separability of the 78 class pairs), and to show the contribution of the SGLD features to the separation of several hard-to-separate class pairs (i.e., the contributions of the SGLD group with the three basic spectral features in the two channels to the separability of several hard-to-separate class pairs).

The relative importance of the eight SGLD features in the two spectral channels was initially sorted by the Fisher distance measure as follows:

Feature	Order	
	VIS	IR4
mean MEAN	1	3
mean CON	4	1
mean ASM	8	6
mean ENT	6	8
max MEAN	5	4
max CON	3	2
max ASM	7	7
max ENT	2	5

As we intended to compare the SGLD features with the entropy-based features and Moran circular autocorrelation features on a group basis (see Section 6.3.3), the first four SGLD features for each channel were selected: mean MEAN, max ENT, max CON and mean CON for the visible channel; mean CON, max CON, mean MEAN and max MEAN for the infrared channel. These eight SGLD features with the three basic spectral ones MEANs were then evaluated by the transformed divergences and the results are shown in Table 6.6. The results of the evaluation by the Fisher distance measure (see Section 6.2.1) are also included in this table to give the relative order of the importance of these eight SGLD features in one dimension.

From Table 6.6, we can see that the relative order of various SGLD features sorted out by the two transformed divergences is different. Referring back to the inherent specification of the two measures and different SGLD textural features, we may conclude that the different order should be attributed to the sensitivity of these textural features to different class pairs, i.e., cloud field structures. For example, the feature max MEAN 4, ranked first by the $^{T}D_{ave}$, was ranked 4th by the $^{T}D_{ave}$*. The ranking shows implicitly that there are several well-separated class pairs, e.g., Altostratus (class 6) and Thin cirrus with open water (class 10), of which the separability benefits with the textural feature max MEAN 4.

However, we were more interested in the least separable class pairs to which different textural features can be applied. Thus, the second divergence measure ${}^{T}D_{ave}{}^{*}$ was chosen for the major use in this study, while the ${}^{T}D_{ave}$ also provided some important guidelines to avoid misinterpretation of class separability by the ${}^{T}D_{ave}{}^{*}$ alone. It is desirable that the values of both the terms change simultaneously but often it may not be so.

By the ${}^{T}D_{ave}{}^{*}$, mean MEAN 4 was considered as the most useful in four dimensions for multi-spectral cloud classification. It is interesting to note that three out of the top four SGLD features, sorted out by this measure, are mean MEAN 4, mean MEAN 1 and max MEAN 4, i.e., the MEAN of the spatial grey level difference histogram of a cloud image contains the most valuable spatial information in addition to the basic radiative information. The MEAN is also computationally the cheapest of the SGLD features, and thus was used later for the assessment of feature combination (This will be described in Section 6.3.3).

Another interesting result is that mean CON 4, the basic feature measuring the local variation of a cloud field in the SGLDs, was ranked first by the Fisher distance measure in one dimension but appeared 6th in the feature selection procedure by using the ${}^{T}D_{ave}{}^{*}$. This result verifies the problems of the 1-D Fisher distance measure used to select textural features which are to be used in multi-dimensional feature space.

To reveal how much the eight SGLD textural features contribute to the separation of several hard-to-separate class pairs (11 in this case), we may assess these features individually or on a group basis. In this study, we chose the latter for compactness and simplicity (For a similar purpose, the other textural features in the following subsections were assessed in the same way). These SGLD features, together with the three basic spectral features, were assessed by the transformed divergence ${}^{T}D_{ave}$ in three different cases: four

visible SGLD features, four infrared SGLD features and both, respectively.

The group-based comparison of the SGLD features in this study was mainly motivated by the work of Weszka *et al.* (1976), in which they found that substantially better classification accuracies could be obtained when they used pairs of feature values with the Fisher linear discrimination function. The motivation was also based on the qualitative interpretation of the SGLD, CIRC and POLENT features (see Sections 6.3.1.4 and 6.3.2.2 for interpretation of the CIRC and POLENT features). Compared with their work, this comparative study was based on the premise that the number of samples is large enough and thus the high-order separability index, the transformed divergence, is appropriate as it can take the intercorrelation of features into account.

The results of the comparison constitute Table 6.7.

A few comments can be made about the results shown in Table 6.7A. From the table, we may conclude that the suitability of visible and infrared SGLD textural features for various class pairs depends on physical cases, i.e., the spatial properties of cloud fields. For instance, infrared SGLD textural features are more useful for the separation between Cumulus (class 4) and Altostratus (class 6), Cumulus and Stratocumulus (class 13), while visible SGLD textural features more helpful for that between Thin cirrus (class 10) and Sea surface (class 3). These results are consistent with those obtained by Booth (1973). In general, textural information reflected by the infrared group-based SGLD features is more useful for differentiating these 11 hard-to-separate class pairs. This conclusion is consistent with that given by the traditional measure $^{T}D_{ave}$* (see Table 6.7B): group-based infrared SGLD features are more useful for separating the least separable pairs of the 13 cloud classes.

From Table 6.7A, we also see that substantially better separability of some class pairs was achieved when the eight SGLD features in both visible and

infrared channels were involved. The most significant improvement is the separability between Cumulus (class 4) and Stratocumulus (class 13). Just as multi-spectral features are important in cloud classification, textural features in multi-spectral channels also play an important role.

From Table 6.6, we find that the best-performing features of the SGLDs in the visible and infrared channels achieved the overall separability ($^{T}D_{ave}$*) 448.9 and 455.4, respectively. Table 6.7B shows that significantly better separability of the least separable class pairs was achieved when the four SGLD features in the single channel were combined: at least 12% improvement was achieved in each channel. The relative importance of the feature combination for different groups of textural features is assessed in Section 6.3.3.

Feature	$^{T}D_{ave}$	$^{T}D_{ave}$*	Order by $^{T}D_{ave}$	Order by $^{T}D_{ave}$*	Order by Fisher D.
mean MEAN 1	838.9	448.9	3	2	3
mean CON 1	825.5	362.8	5	8	5
max CON 1	817.8	405.5	6	7	4
max ENT 1	832.2	444.5	4	3	2
mean MEAN 4	847.9	455.4	2	1	7
mean CON 4	766.7	410.0	8	6	1
max MEAN 4	854.8	439.3	1	4	8
max CON 4	770.9	414.9	7	5	6
MEAN 1,3,4	728.9	409.2			

Table 6.6 The eight SGLD textural features ranked by two transformed divergence measures, Fisher distance and variance analysis.

Notes: 1. "$^{T}D_{ave}$" stands for the average of the overall divergences of all the 78 class pairs;
2. "$^{T}D_{ave}$*" stands for the average of the divergences of those class pairs with transformed divergences less than 1500;
3. "MEAN 1,3,4" stands for visible, near-infrared and infrared feature MEAN.

Table 6.7A Separability of 11 cloud class pairs measured by transformed divergence where the eight SGLD features (Refer to text for explanation) were used in visible, infrared and both channels in addition to the 3 basic spectral features MEANs (The results of the former two cases are given in parenthesis). Refer back to Table 6.1a for cloud category numbers.

	Class 2	Class 3	Class 4	Class 5
C5	84.0 (50.0,56.0)		549.8 (202.2,165.8)	
C6			445.6 (97.9,115.3)	172.4 (110.0,90.7)
C7				114.5 (84.7,81.0)
C10		198.1 (113.2,69.2)		
C11	367.6 (124.8,75.7)		404.4 (156.4,110.0)	
C12	129.8 (94.0,62.3)			
C13			1997.9 (316.6,686.8)	1999.0 (1251.0,1405.9)

$T_{D_{ave}}$ in visible channel: 236.4
$T_{D_{ave}}$ in infrared channel: 265.3
$T_{D_{ave}}$ in both channels: 587.6

Table 6.7B Overall separability of thirteen cloud classes in two channels.

Feature	$T_{D_{ave}}$	$T_{D_{ave}}$*
SGLDs 1	1046.8	502.8
SGLDs 4	951.9	521.7
SGLDs 1&4	1287.9	702.6

6.3.1.2. Roberts Gradient (R.G.)

As discussed in Section 5.1.2, this textural feature is defined to measure the amount of edge per unit area. Its performance to interpreting meteorological satellite imagery has been measured by Parikh & Ball (1980), Wu *et al.* (1985) and Ebert (1987). The usefulness of this feature was shown to be not comparable to the SGLD features (see Section 6.3.1.1), and thus it was always discarded in their feature selection stage. In this study, ranked by the one-way variance analysis, it appeared 13th in the visible channel and 17th in the infrared channel among the twenty-six features (see Table 6.5). However, it was our belief that the potential of this feature has not been sufficiently assessed, especially in more than one dimension.

Similar to Section 6.3.1.1, one of the experiments in this study was designed to assess the relative merit of the R.G. in aiding multi-spectral cloud classification. The assessment of the feature was made by the transformed divergences in three cases as in Section 6.3.1.1. The results are given in Table 6.8.

As expected, the separability of different class pairs varies between the two spectral channels. The class pairs, Cumulus (class 4) and Altostratus (class 6), and Cumulus and Stratocumulus (class 13), are more separated in the visible channel, while the class pair, Thin cirrus (class 10) and Sea surface (class 3), is more separable in the infrared channel. The two channels have more or less the same discriminating power for most of the other class pairs. In general, the R.G. feature in the visible channel leads to an overall better separability of the 11 hard-to-separate class pairs. This conclusion corresponds to that given by the two traditional measures $^{T}D_{ave}$ and $^{T}D_{ave}$* (see Table 6.8B): visible R.G. feature is more useful for separating between the 13 cloud classes and the least separable pairs of these classes.

Table 6.8A Separability of 11 cloud class pairs measured by transformed divergence where the R.G. feature was used in visible, infrared and both channels in addition to the 3 basic spectral features MEANs (The results of the former two cases are given in parenthesis). Refer back to Table 6.1A for cloud category numbers.

	Class 2	Class 3	Class 4	Class 5
C5	37.4 (34.7,36.9)		100.5 (96.2,91.4)	
C6			81.2 (78.0,51.1)	64.6 (57.4,61.5)
C7				64.9 (63.2,63.1)
C10		48.1 (33.5,47.2)		
C11	46.0 (39.5,36.0)		80.5 (72.2,57.6)	
C12	54.4 (50.0,54.1)			
C13			292.9 (277.4,199.1)	1266.6 (1259.5,1097.1)

$T_{D_{ave}}$ in visible channel: 187.4
$T_{D_{ave}}$ in infrared channel: 163.3
$T_{D_{ave}}$ in both channels: 194.3

Table 6.8B Overall separability of thirteen cloud classes in two channels

Feature	$T_{D_{ave}}$	$T_{D_{ave}}$*
R.G. 1	783.1	424.4
R.G. 4	771.3	373.5
R.G. 1&4	821.1	405.8

6.3.1.3. Local Variance

The "local variance" (the standard deviation S.D. was used instead in this study) of a grid of satellite radiance data is a simple statistic of texture. It measures the spread of grey level values around the mean in a sample area. The application of this feature to remote sensing is extensive in either cloud classification (e.g., Karlsson, 1987B; Ebert, 1987) or land-use classification (e.g., Alm, 1985). The major benefit of this feature, as realized in the past, is its efficiency in computation.

However, due to the simplicity of such a local variance measure, it cannot be used to separate some object classes which are similar in their patchiness (Alm, 1985). In this study, it was ranked 11th in the visible channel and 14th in the infrared channel among the twenty-six features by the one-way variance analysis.

A more precise assessment was then performed with the transformed divergence in multi-dimensional feature space in the three same cases as used in the previous sections. The results of the assessment are given in Table 6.9.

From Table 6.9A, we can see that, in common with other textural features, the separability of various class pairs varies with the two channels from which the features were calculated. The class pair, Cumulus (class 4) and Altostratus (class 6), is more separable by this feature in the visible channel, while the class pairs, Cumulus (class 4) and Stratocumulus (class 13), and Thin cirrus (class 10) and Sea surface (class 3), are better separated in the infrared channel. In general, the S.D. in the infrared channel leads to a slightly better separability of the 11 hard-to-separate class pairs than in the visible channel. This conclusion corresponds to that given by the traditional measures $^{T}D_{ave}$ and $^{T}D_{ave}$* (see Table 6.9B): infrared S.D. feature is more useful for differentiating the 13 cloud classes and the least separable pairs of these classes.

The performance of both channels together is better than that of the equivalent for the R.G. feature (see Section 6.3.1.2). From Table 6.9A, it can be seen that the distinction between some class pairs such as Cumulus (class 4) and Thick cirrus over low cloud (class 5), and between Cumulus and Stratocumulus (class 13) was greatly enhanced.

Table 6.9A Separability of 11 cloud class pairs measured by transformed divergence where the S.D. feature was used in visible, infrared and both channels in addition to the 3 basic spectral features MEANs (The results of the former two cases are given in parenthesis). Refer back to Table 6.1A for cloud category numbers.

	Class 2	Class 3	Class 4	Class 5
C5	42.9 (35.3,38.6)		341.0 (191.0,143.8)	
C6			109.1 (88.8,65.7)	63.5 (60.5,57.7)
C7				90.7 (85.5,66.9)
C10		49.1 (39.9,43.0)		
C11	89.5 (49.5,61.5)		167.6 (80.4,108.1)	
C12	60.2 (52.6,56.7)			
C13			336.5 (200.6,297.0)	1142.3 (1024.7,1092.7)

$T_{D_{ave}}$ in visible channel: 173.5
$T_{D_{ave}}$ in infrared channel: 184.7
$T_{D_{ave}}$ in both channels: 226.6

Table 6.9B Overall separability of thirteen cloud classes in two channels

Feature	$T_{D_{ave}}$	$T_{D_{ave}}$*
S.D. 1	809.6	413.3
S.D. 4	847.9	452.0
S.D. 1&4	905.3	465.0

6.3.1.4. Circular Moran Autocorrelation Features (CIRCs)

As opposed to the digital transform approach, autocorrelation functions measure spatial frequency information in an indirect way. By these features, grey level primitives of larger size are indicative of coarser textures while grey level primitives of smaller size are indicative of finer textures. Although it is closely related to its power spectral density function, an autocorrelation function is sometimes preferable because the function performed in the spatial domain directly reflects the size change of grey level primitives and also it is computationally cheaper. A modified version, circular Moran autocorrelation functions (CIRCs), which can take the field properties of texture into account, was used in this study (Refer back to Section 5.1.3).

The CIRCs values, calculated from the raw samples of the 13 cloud classes, are given in Table 6.10. An verbal interpretation of some of these values follows.

Land (class 1) and Sea surface (class 3) appear relatively smooth in the visible images. The image samples of these two classes consist mostly of constant grey level values plus some additive noise. The "noise" of land surfaces results mainly from large differences in reflectance of topography, soil and vegetation types, which all have different physical and chemical properties. The "noise" of sea surfaces comes also from a difference from pixel to pixel. Because the noisy pixels are mostly uncorrelated, the autocorrelation features at four lags (CIRC 1-4) for these two surface classes have relatively low values (0.73 and 0.60 at most, respectively) compared to those for cloud classes. Thick cirrostratus (class 7) has a similar result in the visible images, although it is more structured than the two surface classes. In the infrared images, the autocorrelation values of these three classes at the four lags are all greater than those in the corresponding visible images, i.e., the uncorrelated within-class variations of these four classes are reduced in the infrared images. This

corresponds well to our knowledge of visual interpretation.

Cumulus (class 4) and Stratocumulus (class 13) are relatively structured clouds according to the moderate CIRCs values in the two spectral channels, but their grey level primitives are relatively small compared to other clouds: their autocorrelation values drop off more quickly (from 0.88 to 0.44 and from 0.87 to 0.41 in the visible images, respectively; from 0.90 to 0.49 and from 0.85 to 0.39 in the infrared channel, respectively) than those of other clouds with distance from one to four. The appearance of Stratocumulus is smoother than that of Cumulus in the infrared images, which is reflected by the smaller absolute values of the four autocorrelation functions.

The autocorrelation values at lag 1 (CIRC 1) for Cirrus either with or without other underlying clouds (classes 5, 10 and 12) in the visible images, except Thin cirrus over low cloud (class 11), are all smaller than those of Cumulus (class 4) and Stratocumulus (class 13), i.e., these three cloud classes are less structured than Cumulus and Stratocumulus in the visible images at the finest detail. It is important to note that the relative drops from CIRC 1 to CIRC 4 of these clouds in the visible images, including Thin cirrus over low cloud (class 11), are significantly smaller than those of Cumulus (class 4) and Stratocumulus (class 13). This should be attributed to the fact that the Cirrus in each visible sample of these classes is fairly faint and thus the background of the sample dominates the circular autocorrelation values so that the size of the dominating grey level primitives in the sample is in fact relatively large. In the infrared channel, the outline of Cirrus either with or without other underlying clouds is clearer than that in the visible images so that the autocorrelation values at the four lags (CIRC 1-4) of these cloud classes, except Thin cirrus over low cloud (class 11), are greater.

It is interesting to observe that Thin cirrus over low cloud (class 11) and

Thin cirrus over middle frontal cloud (class 12) in the visible and infrared images have either relatively large (e.g., 0.93 and 0.92 for CIRC 1 in the two channels, respectively) or relatively small autocorrelation values (e.g., 0.81 and 0.82 for CIRC 1 in the two channels, respectively) at the four lags. By analysing the corresponding image samples, it was found that the high values should be attributed to the fact that the structured pattern of low cloud as a background in the mixture of class 11 was caused by low-level vertical wind shear, and the low values attributed to the fact that the contrast between thin Cirrus and middle cloud in both the visible and infrared images is so small that thin Cirrus is hardly discernible. From the study of these satellite observations, we may conclude that the quantitative interpretation of thin Cirrus by the autocorrelation functions (and so Fourier power spectra as well) tends to be difficult, because the high transparency of thin Cirrus results in a great variability of its properties in combination with the background.

As in previous sections, the following experiments were designed to assess the relative merits of these four circular autocorrelation features on an individual and a group basis using the transformed divergences. The results of their individual assessment are given in Table 6.11. Interestingly, from this table, we find that the ranks of all the eight CIRC features sorted out by the two divergences are the same.

From Table 6.11, we find that the top two of the most useful autocorrelation features are CIRC 1 (infrared) and CIRC 1 (visible). This indicates that the nearest circular ring from the centre is preferable for the quantitative interpretation. This is coincident with the Parikh (1977)'s conclusion that the spatial textural features, MEAN, CONTRAST, ASM, ENTROPY, calculated at distance 1, were the best for her cloud classification study. Moreover, we also find that the three other most useful features are all from the infrared channel. This indicates that the individuals of infrared CIRC features contain more

valuable spatial information than those of visible CIRC features for separating between the 13 cloud classes and the least separable pairs of these classes.

There is considerably high correlation between the four CIRC features in each channel: at least four of the six feature pairs in each channel are highly correlated (the correlation coefficients of these four pairs, CIRC 1/2, 2/3, 2/4 and 3/4, are all over 0.90). However, since the divergence measure theoretically takes the intercorrelation between the features into account, the group-based approach for feature assessment was still employed here for these autocorrelation features in three different cases: four visible CIRC features, four infrared CIRC features and both. The results of the assessment are shown in Table 6.12A. The table shows that the suitability of visible and infrared CIRC features for various class pairs depends on the spatial properties of cloud fields, e.g., visible features are more useful for separating between Cumulus (class 4) and Stratocumulus (class 13), and between Thin cirrus (class 10) and Sea surface (class 3), while infrared features more useful for separating between Cumulus and Altostratus (class 6). It is interesting to note that the visible value $^{T}D_{4,13}$ (1074.6) for Cumulus and Stratocumulus is 3.6 times greater than that (298.3) measured in the infrared channel. In general, the group-based visible CIRC features are better than the group-based infrared features in separating the 11 hard-to-separate class pairs, especially in separating Stratocumulus from the other cloud classes. It is interesting to note that the group-based assessment by the $^{T}D_{ave}$* for the least separable pairs of the 13 cloud classes seems to give an opposite conclusion (see Table 6.12B), although the difference between the $^{T}D_{ave}$* (VIS) and the $^{T}D_{ave}$* (IR4) is smaller than the 3% of the $^{T}D_{ave}$* (VIS). This problem should be attributed to the shortcoming of the inherent definition of the $^{T}D_{ave}$* or the non-optimal choice of the threshold for the $^{T}D_{ave}$* (1500 was chosen in this study; see Section 6.2.2 for explanation). In fact, the substantial increase of the separability of some class pairs in the

visible channel, e.g., Thick cirrus over low cloud (class 5) and Stratocumulus (class 13), although neglected by the ${}^{T}D_{ave}*$, is reflected by the ${}^{T}D_{ave}$ (1038.8), indicating the greater importance of the visible channel for the group-based CIRC features.

From Table 6.11, we find that the best-performing features of the CIRCs in the visible and infrared channels achieved the overall separability 409.4 and 412.2 (${}^{T}D_{ave}*$), respectively. Table 6.12B shows that substantially better separability of the least separable class pairs was achieved when the four CIRC features in one channel were combined: 33.8% and 36.7% improvement was obtained in the visible and infrared channels, respectively. The relative importance of this feature combination compared to other groups of textural features is emphasized in Section 6.3.3.

Table 6.10 Means and standard deviations (in parenthesis) of the two-channel circular Moran autocorrelation measures for each class. Refer back to Table 6.1a for cloud category numbers.

	VISIBLE				INFRARED			
	CIRC 1	CIRC 2	CIRC 3	CIRC 4	CIRC 1	CIRC 2	CIRC 3	CIRC 4
1	0.73	0.49	0.37	0.30	0.87	0.74	0.64	0.57
	(0.11)	(0.17)	(0.19)	(0.19)	(0.08)	(0.15)	(0.18)	(0.21)
2	0.83	0.65	0.53	0.45	0.80	0.68	0.59	0.52
	(0.09)	(0.15)	(0.17)	(0.18)	(0.15)	(0.16)	(0.17)	(0.19)
3	0.60	0.53	0.45	0.40	0.84	0.73	0.65	0.57
	(0.32)	(0.27)	(0.28)	(0.25)	(0.12)	(0.14)	(0.15)	(0.17)
4	0.88	0.70	0.55	0.44	0.90	0.74	0.61	0.49
	(0.05)	(0.11)	(0.14)	(0.16)	(0.04)	(0.10)	(0.15)	(0.17)
5	0.86	0.71	0.62	0.54	0.91	0.79	0.68	0.59
	(0.09)	(0.14)	(0.17)	(0.19)	(0.07)	(0.11)	(0.14)	(0.16)
6	0.91	0.80	0.69	0.61	0.92	0.80	0.68	0.58
	(0.05)	(0.12)	(0.16)	(0.18)	(0.03)	(0.06)	(0.09)	(0.11)
7	0.78	0.61	0.49	0.40	0.93	0.81	0.71	0.61
	(0.26)	(0.24)	(0.23)	(0.22)	(0.04)	(0.10)	(0.14)	(0.17)
8	0.93	0.83	0.74	0.66	0.94	0.85	0.77	0.69
	(0.04)	(0.10)	(0.14)	(0.17)	(0.03)	(0.07)	(0.11)	(0.14)
9	0.89	0.73	0.62	0.53	0.91	0.81	0.73	0.65
	(0.07)	(0.15)	(0.18)	(0.21)	(0.06)	(0.09)	(0.13)	(0.15)
10	0.82	0.68	0.57	0.49	0.91	0.79	0.69	0.59
	(0.11)	(0.14)	(0.16)	(0.17)	(0.04)	(0.08)	(0.11)	(0.13)
11	0.93	0.86	0.80	0.74	0.92	0.82	0.73	0.65
	(0.03)	(0.07)	(0.10)	(0.13)	(0.05)	(0.09)	(0.12)	(0.15)
12	0.81	0.62	0.50	0.42	0.82	0.67	0.56	0.48
	(0.12)	(0.16)	(0.17)	(0.18)	(0.15)	(0.16)	(0.16)	(0.16)
13	0.87	0.67	0.52	0.41	0.85	0.64	0.49	0.39
	(0.04)	(0.09)	(0.13)	(0.15)	(0.06)	(0.12)	(0.15)	(0.16)

Table 6.11 The four circular Moran autocorrelation features ranked by two transformed divergences.

Feature	$T_{D_{ave}}$	$T_{D_{ave}}$*	Order by $T_{D_{ave}}$	Order by $T_{D_{ave}}$*
CIRC 1(VIS)	764.6	409.4	2	2
CIRC 2(VIS)	752.0	395.2	6	6
CIRC 3(VIS)	751.3	393.7	7	7
CIRC 4(VIS)	751.0	393.3	8	8
CIRC 1(IR4)	765.3	412.2	1	1
CIRC 2(IR4)	760.8	405.6	3	3
CIRC 3(IR4)	760.4	404.8	5	5
CIRC 4(IR4)	760.4	405.0	4	4
MEAN 1,3,4	728.9	409.2		

Note: Refer back to Table 6.6 for explanation of MEAN 1,3,4.

Table 6.12A Separability of 11 cloud class pairs measured by transformed divergence where the four CIRC features were used in addition to the 3 basic spectral features MEANs in visible, infrared and both channels, respectively (The results of the former two cases are given in parenthesis. Refer back to Table 6.1a for cloud category numbers.

	Class 2	Class 3	Class 4	Class 5
C5	81.7 (49.2,66.1)		397.3 (181.7,131.0)	
C6			135.8 (67.6,86.0)	456.4 (155.0,247.9)
C7				289.8 (122.4,207.2)
C10		496.2 (135.6,110.6)		
C11	1337.8 (131.2,274.3)		487.8 (111.5,93.0)	
C12	86.2 (67.4,62.3)			
C13			1366.8 (1074.6,298.3)	1999.0 (1997.3,1228.7)

$T_{D_{ave}}$ in visible channel: 372.1
$T_{D_{ave}}$ in infrared channel:255.0
$T_{D_{ave}}$ in both channels: 648.6

Table 6.12B Overall separability of thirteen cloud classes in two channels.

Feature	$T_{D_{ave}}$	$T_{D_{ave}}$*
CIRCs(VIS)	1038.8	547.7
CIRCs(IR4)	934.8	563.5
CIRCs(both)	1320.0	753.2

6.3.2. Spatial Frequency Features

6.3.2.1. Normalized Average Amplitude Fourier Power Spectrum (NAAs)

As discussed in Section 5.2.2, the normalized average amplitude Fourier power spectrum (also called **R**-spectrum) was characterized by the following three composite features, mean NAA, max NAA, and DOM WVN, as used by Bunting & Fournier (1980) and Ebert (1987). These features were designed to measure the coarseness of textural cloud fields. Although the measurement of spatial frequency distribution is intuitively straightforward for texture analysis, whether the specification of these features is necessarily optimal still remains in doubt. The ranks of the three features, sorted by the variance analysis (see Table 6.5) give a general view of the relative merits of the features. From the table, it was found that the mean NAA appeared ahead of the other two in both channels. The feature DOM WVN was observed to be of no value by examining its values calculated from the 649 raw samples of the 13 classes, and this was verified by the F-test: the F-ratio (2.00) of this feature is smaller than 2.184 ($F_{0.01}(12,\infty)$), i.e., the difference between the 13 cloud classes by this feature is not significant at 1% level.

The further evaluation of the other two NAA features was performed with the transformed divergences. Table 6.13 shows the order of these two features ranked by the ${}^{T}D_{ave}$ and ${}^{T}D_{ave}$*.

Again, we chose the ${}^{T}D_{ave}$* as the major separability measure. From Table 6.13, we find that max NAA 1 was the most useful for separating the least separable class pairs of the 13 classes. It is interesting to see that the two best-performing features are all from the visible channel. This indicates that the visible frequency features contain the most valuable spatial information for aiding multi-spectral cloud classification. The results given by the ${}^{T}D_{ave}$ are consistent with those given by the variance analysis (see above): mean NAA

performed better than max NAA for separating between the 13 cloud classes in both channels.

Feature	$^{T}D_{ave}$	$^{T}D_{ave}$*	Order by $^{T}D_{ave}$	Order by $^{T}D_{ave}$*
mean NAA 1	752.8	396.6	2	2
max NAA 1	744.2	425.7	4	1
mean NAA 4	752.9	394.8	1	3
max NAA 4	750.7	393.1	3	4
MEAN 1,3,4	728.9	409.2		

Table 6.13 The three NAA features ranked by two transformed divergences.
Notes: 1. The suffices, 1 and 4, of the feature names in this table denote the visible and infrared channels, respectively.
2. Refer back to Table 6.6 for explanation of MEAN 1,3,4.

As before, we then used the group-based approach for feature assessment to reveal the contributions of the NAA features in three different cases: two visible NAA features, two infrared NAA features and both. The results of the assessment are given in Table 6.14A. The same conclusion as in the other cases described above may be drawn from the table that the separability of different class pairs varies between the two spectral channels. Although the separability of most of the 11 class pairs in the table is relatively small, the values $\{^{T}D_{ij}\}$ of these class pairs still show the relative merits of the NAA features: visible NAA features provide more spatial information for differentiating between Thin cirrus (class 10) and Sea surface (class 3), and between Cumulus (class 4) and Stratocumulus (class 13), while infrared NAA features are slightly better suited for differentiating between Cumulus and Altostratus (class 6). In general, the visible spatial information reflected by the NAA features is more useful for differentiating these hard-to-separate class pairs. This conclusion is

consistent with those given by the ${}^{T}D_{ave}$ and ${}^{T}D_{ave}$* (see Table 6.14B).

From Table 6.13, we have seen that the best-performing features of the NAAs in the visible and infrared channels achieved the overall separability 425.7 and 394.8 (${}^{T}D_{ave}$*), respectively. Table 6.14B shows that slightly better separability of the overall 78 class pairs was achieved in the visible channel but the separability of the least separable class pairs appears to be reduced from 425.7 to 418.8 when the two NAA features in the same channel were combined. Although the decrease in the separability should be in fact attributed to the inherent specification of the separability measure ${}^{T}D_{ave}$*, this fact indicates that the performance improvement of the NAAs in the visible channel cannot benefit much from the feature combination. The performance of the infrared NAAs was slightly improved when the features were considered in combination. The comparison of the relative importance of the NAA feature combination with others is given in Section 6.3.3.

Table 6.14A Separability of 11 cloud class pairs measured by transformed divergence where the two NAA features were used in addition to the 3 basic spectral features in visible, infrared and both channels (The results of the former two cases are given in parenthesis). Refer back to Table 6.1A for cloud category numbers.

	Class 2	Class 3	Class 4	Class 5
C5	47.2 (36.4,44.5)		105.7 (92.3,90.3)	
C6			63.7 (50.9,52.5)	75.1 (59.4,65.4)
C7				85.8 (78.6,67.8)
C10		61.5 (39.2,33.7)		
C11	75.8 (57.6,43.6)		113.5 (97.5,61.0)	
C12	56.7 (51.8,53.5)			
C13			308.8 (246.5,205.6)	1599.4 (1323.3,1056.7)

$T_{D_{ave}}$ in visible channel: 193.9
$T_{D_{ave}}$ in infrared channel:161.3
$T_{D_{ave}}$ in both channels: 235.7

Table 6.14B Overall separability of thirteen cloud classes in two channels

Feature	$T_{D_{ave}}$	$T_{D_{ave}}$*
NAAs 1	779.5	418.8
NAAs 4	767.0	412.0
NAAs 1&4	825.4	431.5

6.3.2.2. Entropy-based Measure in Polar Frequency Domain (POLENTs)

The property of entropy-based measures in the polar frequency domain as an alternative to the normalized average power spectrum was one of the major concerns in this study. Referring back to the inherent property of entropy, the entropy measures were designed to characterize the absence of frequency pattern structures by measuring the probability distribution of each annular band within an image spectrum rather than summed spectral energies (see Section 5.2.3).

The entropy values, calculated from the raw 649 samples of the 13 cloud classes, are given in Table 6.15. Some qualitative interpretation of cloud fields can be made from the table. It was noted that the feature, AVE ENT 3 had a surprisingly large standard deviation for all the 13 classes and so its reliability was degraded. This led further to some interpretation errors by the transformed divergence.

The first class, Land surface, contains a considerable amount of variation during day-time in the visible images, which is reflected by the relatively large NO DOM (29.4). The relatively large AVE ENT 3 (12.8) shows that its fine pattern in the visible images is rather unstructured. The corresponding infrared images give a similar but less significant result. For Sea surface (class 3), the value of the feature NO DOM (18.3) in the visible images and the value (19.0) in the infrared images are significantly smaller than those of Land surface. This indicates that Sea surface appears more homogeneous than Land Surface in either the visible or infrared images. The relatively moderate values of the feature AVE ENT 3 in both the visible and infrared images implies that its patterns is quite fuzzy.

The texture pattern of Cumulus (class 4) is fairly well reflected by the POLENT features in both channels. The relatively large value (23.4) of the visible

feature NO DOM shows that Cumulus is highly textured, and the relatively small value (2.1) of the visible feature AVE ENT 3 indicates that its textural pattern appears rather structured (or regularly arranged). Cumulus in the infrared images gives a similar result. Compared to the Cumulus pattern, the highly textural pattern of Stratocumulus (class 13) appears less structured in either the visible or the infrared images.

The appearance of Cumulonimbus (class 9) is different in the visible and infrared images. The relatively large visible AVE ENT 3 (3.5) shows that the top surfaces of this class are rather unstructured, while the relatively small infrared AVE ENT 3 (0.5) indicates that the thermal structures of the top surfaces appears somewhat regularly arranged within about 35 km square areas. Thick cirrostratus (class 7) shows a similar but slightly less textural pattern.

The somewhat structured pattern of Thin cirrus over low cloud (class 11) and the fuzzy pattern of Thin cirrus over middle frontal cloud (class 12) are well reflected by the visible AVE EVT 3 values (0.9 and 5.9, respectively). Their patterns in the infrared images are less apparent than in the visible images by the feature AVE ENT 3. The relatively small and relatively large NO DOM values of these two classes in the visible and infrared images (17.1 and 17.7 for Class 11 in the two channels, respectively; 24.8 and 21.3 for Class 12 in the two channels, respectively) show respectively that the texture of Thin cirrus over low cloud is coarse and the texture of Thin cirrus over middle frontal cloud is rather fine. This confirms again that the high transparency of thin Cirrus results in a great variability of its properties in combination with the background (Refer back to Section 6.3.1.4 for interpretation of CIRC values).

As the CIRC features, the qualitative interpretation of the POLENT features depends on appropriate feature pair(s), e.g., AVE ENT 3 and NO DOM in this case. However, this feature pair may not be an optimal set for a quantitative

interpretation, i.e., it may not lead to the maximum separability of the 13 cloud classes. Therefore, this also leads further to the necessity of investigating optimal textural feature pairs for multi-spectral cloud classification (We initially intended to reveal the effect of the size sensitivity of various textural features on the separability of the 13 classes). The results of the investigation are given in Section 6.3.3.

The problems with this entropy-based approach are obvious. One of the problems is that the four POLENT features cannot directly reflect the spatial information contained in the absolute values of frequency entries with values larger than a "censoring" value. This problem may be solved or alleviated by combining them with NAA features. Moreover, the substantial standard deviations of the POLENT features, especially the third feature AVE ENT 3, certainly degrade the reliability of the quantitative interpretation. A possible approach to alleviating the problem is to apply spectrum smoothing technique (see Renshaw and Ford 1983 for example). It is also possible to improve the entropy-based approach by pre-thresholding or smoothing the training images. These works are beyond the scope of this comparative study.

The following experiments were designed to assess the relative merits of these four entropy features using the transformed divergences. The results are given in Tables 6.16 and 6.17.

We chose the ${}^{T}D_{ave}^{*}$ as the major separability measure. From Table 6.16, we note that the visible feature NO DOM appears the best for separating the least separable pairs of the 13 cloud classes. Referring back to the definition of NO DOM, this result confirms implicitly the importance of the "censoring" procedure. It is also interesting to see that the four best features for separating the least separable class pairs are all from the visible channel. This indicates that the individuals of the visible POLENT features contain more valuable spatial

information than those of the infrared POLENT features for separating the least separable pairs of the 13 cloud classes.

The group-based assessment approach was also employed for the entropy features to reveal the contributions of these features in three different cases: four visible entropy features (POLENTs 1), four infrared entropy features (POLENTs 4) and both (POLENTs 1&4). The results of the assessment are shown in Table 6.17A. Once again, we may conclude from the table that the suitability of the visible and infrared entropy features for various class pairs depends on the spatial properties of cloud fields, e.g., the visible entropy features are more useful for differentiating between Cumulus (class 4) and Stratocumulus (class 13), while the infrared entropy features are more helpful for differentiating between Cumulus and Altostratus (class 6), and between Thin cirrus (class 10) and Sea surface (class 3). In general, the group-based infrared POLENT features are nearly equivalent to the group-based visible POLENT features in separating the 11 hard-to-separate class pairs. They are also nearly equivalent in the other two cases reflected by the ${}^{T}D_{ave}$ and ${}^{T}D_{ave}$* in Table 6.17B.

Interestingly, the separability of the 11 hard-to-separate class pairs was significantly improved when the entropy features in both the channels were involved, especially for some class pairs such as Cumulus (class 4) and Stratocumulus (class 13), for which the dual-channel divergence value is 1734.6, seven times greater than the better (246.6) of the two single channels.

From Table 6.16, we find that the best-performing features of the POLENTs in the visible and infrared channels achieved the overall separability 423.6 and 395.4 (${}^{T}D_{ave}$*), respectively. Table 6.17B shows that better separability of the least separable class pairs was achieved when the four features of the POLENTs in the same spectral channel were combined: 8.2% and 13.8% improvement was achieved in the visible and infrared channels, respectively. The relative

importance of this feature combination compared to others is assessed in Section 6.3.3.

It should be stressed that the accurate analysis of the ERRORs made in these experiments is not possible. Since the involvement of an additional feature never decreases the original divergence value (see Section 6.2.2), the group-based assessment may also lead to some doubt. However, from the relative increase of the $\{^{T}D_{ij}\}$ values of the various class pairs considered, some tentative conclusions can be drawn with confidence.

As indicated in Table 6.15, the values of the POLENT features were derived at the "censoring" (C) value 0.39 (i.e., 400/1024). The C value was adopted directly from the Renshaw and Ford's (1984) work for this comparative study. The usefulness of the "censoring" procedure with such a C value for aiding multi-spectral cloud classification was verified by comparing the overall separability ($^{T}D_{ave}$) of the 13 cloud classes and the separability ($^{T}D_{ave}$*) of the least separable pairs of these classes when the POLENT features in both visible and infrared channels with C = 0.39 were used with those when the POLENT features with C = 0.00 were used:

C	$^{T}D_{ave}$	$^{T}D_{ave}$*
0.00	869.1304	401.2771
0.39	972.0176	488.8618

The results show that significantly better separabilities in terms of the $^{T}D_{ave}$ and $^{T}D_{ave}$* were achieved when the "censoring" procedure with C = 0.39 was used, although the C value might not be optimal for our case. This reveals that the "censoring" procedure is essential for applying the POLENT features to interpreting the AVHRR radiance images, and meanwhile suggests implicitly that the discriminating potential of the POLENT features may be increased by selecting a more appropriate C value.

Table 6.15 Means and standard deviations (in parenthesis) of the two-channel entropy-based measures for each class. Refer back to Table 6.1A for cloud category numbers.

	VISIBLE				INFRARED			
	AVE ENT1	AVE ENT2	AVE ENT3	NO DOM	AVE ENT1	AVE ENT2	AVE ENT3	NO DOM
1	73.5	50.3	12.8	29.4	67.2	33.9	2.1	21.1
	(13.1)	(16.1)	(12.7)	(7.5)	(17.5)	(16.8)	(4.3)	(5.2)
2	62.3	38.7	6.3	22.9	66.3	36.3	4.9	22.1
	(18.9)	(13.8)	(9.7)	(5.6)	(17.9)	(18.0)	(8.5)	(6.9)
3	59.9	27.0	3.3	18.3	65.4	25.9	1.8	19.0
	(21.3)	(17.8)	(7.9)	(6.9)	(16.8)	(17.6)	(5.5)	(5.5)
4	68.6	39.3	2.1	23.4	69.6	31.3	1.0	20.0
	(15.2)	(15.3)	(3.9)	(5.5)	(14.3)	(14.0)	(2.4)	(5.1)
5	68.2	36.4	3.6	22.2	69.3	26.6	0.8	19.1
	(15.0)	(16.1)	(7.2)	(5.6)	(15.2)	(13.5)	(3.2)	(4.2)
6	67.0	29.6	1.3	19.1	71.0	28.4	1.3	19.9
	(15.1)	(17.1)	(4.2)	(5.0)	(13.3)	(11.8)	(4.1)	(3.6)
7	61.3	36.7	3.4	21.4	69.0	29.7	0.7	19.2
	(23.1)	(17.5)	(5.8)	(8.3)	(16.2)	(12.5)	(2.9)	(4.5)
8	70.4	23.2	0.2	17.9	67.8	24.2	0.2	18.0
	(14.7)	(16.3)	(1.4)	(4.7)	(13.5)	(14.1)	(1.3)	(4.2)
9	66.9	36.0	3.5	22.7	68.9	30.4	0.5	19.6
	(16.1)	(16.1)	(7.2)	(5.6)	(13.9)	(14.1)	(2.0)	(4.4)
10	69.7	33.7	2.3	21.1	67.1	27.8	0.3	18.6
	(14.8)	(16.0)	(6.4)	(5.9)	(14.9)	(12.6)	(1.5)	(4.1)
11	67.1	22.0	0.9	17.1	62.4	23.2	1.5	17.7
	(14.8)	(10.1)	(2.6)	(3.1)	(19.5)	(15.7)	(3.1)	(5.3)
12	64.5	42.9	5.9	24.8	64.4	33.1	4.2	21.3
	(14.3)	(17.4)	(8.9)	(6.3)	(16.8)	(15.0)	(8.3)	(6.0)
13	70.7	43.5	3.0	24.9	70.3	46.7	3.2	26.1
	(14.4)	(13.3)	(3.6)	(4.5)	(18.3)	(14.3)	(4.2)	(5.8)

Note: 1. All the POLENT values were derived at the "censoring" value 0.39(i.e., 400/1024) and then normalized by 100 to have a maximum possible value of 100. (Refer to Section 6.3.2.2 for explanation).
2. Refer back to Table 6.1A for cloud category numbers.

Table 6.16 The four polar entropy features ranked by two transformed divergences.

Feature	$T_{D_{ave}}$	$T_{D_{ave}}$*	Order by $T_{D_{ave}}$	Order by $T_{D_{ave}}$*
AVE ENT1(VIS)	741.0	403.4	7	3
AVE ENT2(VIS)	747.1	412.1	5	2
AVE ENT3(VIS)	748.5	396.5	4	4
NO DOM(VIS)	741.5	423.6	6	1
AVE ENT1(IR4)	736.8	380.2	8	8
AVE ENT2(IR4)	750.3	395.4	3	5
AVE ENT3(IR4)	754.0	395.1	1	6
NO DOM(IR4)	752.4	393.2	2	7
MEAN 1,3,4	728.9	409.2		

Note: refer back to Table 6.6 for explanation of MEAN 1,3,4.

Table 6.17A Separability of 11 cloud class pairs measured by transformed divergence where the four entropy-based measures were used in addition to the 3 basic spectral features MEANs in visible, infrared and both channels, respectively (The results of the former two cases are given in parenthesis). Refer back to Table 6.1a for cloud category numbers.

	Class 2	Class 3	Class 4	Class 5
C5	57.8 (39.3,47.8)		114.4 (98.8,93.7)	
C6			81.9 (52.6,63.4)	84.8 (64.8,64.1)
C7				98.1 (81.9,70.5)
C10		96.3 (39.1,50.2)		
C11	170.8 (81.7,63.7)		290.4 (110.4,92.8)	
C12	63.3 (54.4,53.9)			
C13			1734.6 (246.6,237.3)	1999.0 (1068.9,1175.3)

$T_{D_{ave}}$ in visible channel: 176.2
$T_{D_{ave}}$ in infrared channel:182.9
$T_{D_{ave}}$ in both channels: 435.6

Table 6.17B Overall separability of thirteen cloud classes in two channels.

Feature	$T_{D_{ave}}$	$T_{D_{ave}}$*
POLENTs 1	805.7	458.2
POLENTs 4	799.2	450.0
POLENTs 1&4	972.0	488.9

6.3.3. Intercomparison of the Textural Features

The above within-group comparisons of the textural features have revealed the relative discriminating potential of these features on an individual and also a group basis (There were only three groups of the features for these comparisons, which are calculated respectively from the visible, infrared and dual channels). The experiments described in this subsection were designed to compare the relative merits of these features on a between-group basis (i.e., between various groups of the textural features calculated from the same channel). The investigation of optimal feature pairs for the three groups CIRCs, SGLDs and POLENTs is also described in this subsection.

A. Comparison between individual textural features

Table 6.18 shows the ranks of the thirty-two textural features in the visible and infrared channels, sorted out by the transformed divergence of the least separable pairs of the 13 cloud classes ($^{T}D_{ave}$*). The values of the transformed divergence were calculated in a four-dimensional feature space using the three basic spectral features MEANs and the respective textural features. From the table, we see that the infrared features mean MEAN and S.D. appear the most useful, i.e., these two features give the lowest error probability of cloud classification when they are used with the basic three-channel spectral features. The visible mean MEAN and max ENT appeared among the four most useful textural features when assessed by the 1-D Fisher distance (see Table 6.2B), and also rank third and fourth respectively in the multi-dimensional feature space. It is interesting to note that the infrared mean CON and the visible mean NAA were among these four features but rank thirteenth and twentieth in the multi-dimensional feature space.

As shown in Table 6.18, the infrared feature S.D. in the infrared channel appears the second best textural feature for discriminating the least separable pairs of the 13 cloud classes (while it is ranked third by the $^{T}D_{ave}$). This result indicates that this simple statistic as a texture descriptor is not only computationally efficient, as realized in the past, but also highly powerful in aiding multi-spectral cloud classification.

From the between-group point of view, the individuals of the spatial features including the SGLDs, R.G., S.D. and CIRCs are more useful than those of the spatial frequency features (NAAs, POLENTs): the spatial features accounting for 62.5 percent of the overall textural features considered in this study have 80 percent of the high-ranking (in the range of the first fifteen) features.

From the table, we can also see that the CIRC features are not comparable in performance to most of the SGLD features on an individual basis. It is interesting to note that the POLENT feature NO DOM in the visible channel appears relatively high-ranking as the best of the POLENT group. This fact confirms implicitly the usefulness of this newly-defined frequency feature and importance of the "censoring" procedure.

B. Comparison between feature groups

To show the relative importance of each group of textural features, the group-based combinations of these features in three different cases: in the visible, infrared and both channels, were assessed by the divergence analysis. The relative discriminating potential of each combination was first evaluated based on the two divergences of the overall 13 cloud classes, and then on the separability of the 11 hard-to-separate pairs of these classes. The comparative results are given respectively in Tables 6.19 and 6.20.

A practical problem in this comparative study is that the number of textural features in various groups used here is different. The POLENT features and CIRC features were all designed as four-feature groups. The SGLD features were reduced into four by a feature selection procedure (see Section 6.3.1.1). Some groups such as the S.D., R.G. and NAAs have only one or two features. Although this problem makes the comparison between all the six groups not feasible, it is still of value to compare between some groups of various group sizes, e.g., between the NAAs and local variance.

From Tables 6.19 and 6.20, some tentative conclusions can be drawn as follows:

1. In the visible channel, CIRC features perform about as well as SGLD features in separating the 78 pairs of the 13 cloud classes, and do somewhat better than SGLD features in separating the least separable pairs of these classes, and do much better in separating the 11 hard-to-separate class pairs.

In this channel, CIRC features do much better than POLENT features in either case. SGLD features perform much better than POLENT features either in separating all the 78 class pairs or in separating the 11 hard-to-separate class pairs, and do somewhat better in separating the least separable pairs of the 13 cloud classes.

In the infrared channel, CIRC features perform about as well as SGLD features in separating the 78 pairs of the 13 cloud classes and do somewhat better than SGLD features in separating the least separable pairs of these classes, but somewhat worse in separating the 11 hard-to-separate class pairs.

In this channel, both CIRC features and SGLD features perform much better than POLENT features in either case.

The good performance of the visible CIRC features suggests that the textures of the cloud samples may be more appropriately modelled statistically in the spatial domain, e.g., as random fields with autocorrelations, rather than in the spatial frequency domain.

2. In the visible channel, the simple statistic S.D. does about as well as R.G. feature either in separating the 78 pairs of the 13 cloud classes or in separating the least separable pairs of these classes, but does somewhat worse in separating the 11 hard-to-separate class pairs. Interestingly, infrared S.D. performs significantly better than infrared R.G. in separating the 78 class pairs and much better both in separating the least separable pairs of the 13 cloud classes and in separating the 11 class pairs. The relative importance of infrared feature S.D. in aiding multi-spectral cloud classification suggests that it may replace some sophisticated textural features.

Compared with the performance of R.G. feature, the relative weakness of S.D. feature in the visible channel, shown by Tables 6.19 and 6.20, indicates that it may not adequately distinguish cloud fields with a large within-class variability. In this case, a bit more sophisticated textural feature is necessarily needed.

3. In the visible channel, NAA features, which include two components, do even slightly worse than the simple feature S.D. in separating the 78 class pairs and do about as well as the S.D. feature in separating the least separable pairs of the 13 cloud classes, but significantly better in separating the 11 hard-to-separate class pairs. It is interesting to note that infrared feature S.D. does significantly better than infrared NAA features in either case, and even somewhat better than infrared POLENT features (including four components) in separating all the 78 class pairs.

C. Comparison using feature pairs

The experiments described above have shown that significantly better separability of various class pairs can be achieved in most cases when textural features are considered in group-based combination. The reason can be explained, as Weszka *et al.* (1976) pointed out for other spatial features, that they contain the information concerning the size variations of cloud fields.

To reveal further the composite properties of CIRC features, various pairs of the features were studied. Table 6.21 shows the performance of all the possible combinations of the four CIRC features in each channel. Comparing this table with Table 6.11, we see that significantly better separability of both cases, the 78 pairs of the 13 cloud classes and the least separable pairs of these classes, was obtained in either channel when we used the feature pairs (as the cases of the feature groups discussed above): in the visible channel, ${}^{T}D_{ave}$ increases from 752.0 (CIRC 2) to 854.1 (CIRC 1,2) and ${}^{T}D_{ave}$* from 409.4 (CIRC 1) to 463.3 (CIRC 1,3); in the infrared channel, ${}^{T}D_{ave}$ increases from 765.3 (CIRC 1) to 856.5 (CIRC 1,2) and ${}^{T}D_{ave}$* from 412.2 (CIRC 1) to 506.9 (CIRC 1,2).

In either visible or infrared channel, the best pairs of CIRC features always involve small sizes, i.e., one CIRC feature at distance 1, which implies that coarseness, as one of the two major texture dimensions, is important for cloud classification. It is also noted that the class separabilities (${}^{T}D_{ave}$ and ${}^{T}D_{ave}$*) by the six infrared feature pairs are all better than those by the six visible feature pairs to varying degrees, respectively.

SGLD feature pairs were also assessed in this study to investigate the importance of using SGLD feature pairs which are calculated at different distances. The most useful feature mean MEAN was chosen for the comparison since this feature in either the visible or the infrared channel appeared the best of the SGLD features according to Table 6.18. Four successive distances starting

from one were considered here for the single feature mean MEAN. The results of the assessment are shown in Table 6.22. Comparing Table 6.6 with Table 6.22, we see that somewhat better separability of both cases, the 78 pairs of the 13 cloud classes and the least separable pairs of these classes, was obtained in either channel when we used the feature pairs: in the visible channel, $^{T}D_{ave}$ increases from 838.9 (mean MEAN at distance 1, shown in Table 6.6) to 881.4 (mean MEANs at distances 1,3, shown in Table 6.22) and $^{T}D_{ave}$* increases from 448.9 (mean MEAN 1) to 477.2 (mean MEANs 1,2); in the infrared channel, $^{T}D_{ave}$ increases from 847.9 (mean MEAN 1) to 905.3 (mean MEANs 1,4) and $^{T}D_{ave}$* from 455.4 (mean MEAN 1) to 496.0 (mean MEANs 1,4).

Similarly, in either visible or infrared channel, the best pairs of SGLD features always involve small sizes, i.e., one single mean MEAN at distance 1. From Table 6.22, we also find that except the feature pair mean MEAN 1,2, the class separabilities by the other five infrared feature pairs are all better than those by the other five visible feature pairs to varying degrees, respectively.

Comparing the improvement of class separability (discussed above in terms of $^{T}D_{ave}$ and $^{T}D_{ave}$*) produced by CIRC feature pairs and that produced by SGLD feature pairs, we may conclude that CIRC features have more size sensitivity which we can exploit further to increase the discriminating potential of CIRC features by finding optimal feature pairs or feature groups.

In comparison, POLENT feature pairs do not perform as well as CIRC and SGLD feature pairs (see Tables 6.16 and 6.23): in the visible channel, there is small increase of $^{T}D_{ave}$ from 748.5 (AVE ENT 3) to 767.1 (AVE ENTs 2,3) and even small decrease of $^{T}D_{ave}$* from 423.6 (NO DOM) to 418.4 (AVE ENTs 2,3) (due to the problem with the inherent specification of $^{T}D_{ave}$* measure); in the infrared channel, there is small increase of $^{T}D_{ave}$ from 754.0 (AVE ENT 3) to 771.1 (AVE ENTs 2,3) and increase of $^{T}D_{ave}$* from 395.4 (AVE ENT 2) to 415.6 (AVE ENTs 2,3).

The relatively small improvement of class separability (in terms of $^{T}D_{ave}$ and $^{T}D_{ave}$*) compared to that produced by CIRC and SGLD features indicates that POLENT features have less size sensitivity than CIRC and SGLD features.

It is interesting to note from Table 6.23 that the best pairs of POLENT features always include the feature AVE ENT 3 in either of the two channels. The feature AVE ENT 3 corresponds to the highest frequency band (three wavenumbers inclusive), revealing the finest details of cloud fields.

Comparing the results in this part (Part C) with those given in Part B, we can conclude that the optimal combination of existing textural features calculated at different distances may provide more discriminating information than the joint use of textural features, which are individually selected from existing textural features without taking the size sensitivity of the features into consideration, in multi-dimensional feature space. This fact has been illustrated above by comparing CIRC and SGLD features: CIRC features contributed more to the separation of the least separable pairs of the 13 cloud classes in both channels, and that of the 11 hard-to-separate class pairs in the visible channel when used on a group basis (see Tables 6.19 and 6.20); the discriminating power of CIRC features increased faster than mean MEAN, the best of SGLD features, when they were used on a pair basis. This suggests a new approach to feature selection for future cloud classification study.

Table 6.18 Ranks of textural features by transformed divergence of the least separable class pairs.

Order	Feature	$^{T}D_{ave}$*
1	mean MEAN 4	455.4
2	S.D. 4	452.0
3	mean MEAN 1	448.9
4	max ENT 1	444.5
5	max MEAN 4	439.3
6	max NAA 1	425.7
7	R.G. 1	424.4
8	NO DOM (VIS)	423.6
9	max CON 4	414.9
10	S.D. 1	413.3
11	CIRC 1 (IR4)	412.2
12	AVE ENT 2(VIS)	412.1
13	mean CON 4	410.0
14	CIRC 1 (VIS)	409.4
15	CIRC 2 (IR4)	405.6
16	max CON 1	405.5
17	CIRC 4 (IR4)	405.0
18	CIRC 3 (IR4)	404.8
19	AVE ENT 1 (VIS)	403.4
20	mean NAA 1	396.6
21	AVE ENT 3 (VIS)	396.5
22	AVE ENT 2 (IR4)	395.4
23	CIRC 2 (VIS)	395.2
24	AVE ENT 3 (IR4)	395.1
25	mean NAA 4	394.8
26	CIRC 3 (VIS)	393.7
27	CIRC 4 (VIS)	393.3
28	NO DOM (IR4)	393.2
29	max NAA 4	393.1
30	AVE ENT 1 (IR4)	380.2
31	R.G. 4	373.5
32	mean CON 1	362.8

Note: Except the channels from which the POLENT and CIRC features are calculated are noted with extra abbreviated words in parenthesis, the channels concerned with the other features are noted with numerical sufficies (1=visible, 4=infrared).

Table 6.19 Separability of the overall thirteen cloud classes when the textural features are combined on group basis.

Feature Group	Visible $T_{D_{ave}}$	Visible $T_{D_{ave}}$*	Infrared $T_{D_{ave}}$	Infrared $T_{D_{ave}}$*	Visible & Infrared $T_{D_{ave}}$	Visible & Infrared $T_{D_{ave}}$*
SGLDs (4)	1046.8	502.8	951.9	521.7	1287.9	702.6
R.G. (1)	783.1	424.4	771.3	373.5	821.1	405.8
S.D. (1)	809.6	413.3	847.9	452.0	905.3	465.0
NAAs (2)	779.5	418.8	767.0	412.0	825.4	431.5
POLENTs (4)	805.7	458.2	799.2	450.0	994.9	450.6
CIRCs (4)	1038.8	547.7	934.8	563.5	1320.0	753.2

Note: 1. Separability of the overall thirteen cloud classes when the three basic spectral features, MEANs (visible, near-infrared, infrared), were used alone:

$T_{D_{ave}}$: 728.9

$T_{D_{ave}}$*: 409.2

2. The number of features in each group is given in parenthesis adjacent to the group name.

Table 6.20 Separability of 11 cloud class pairs measured by transformed divergence when the textural features are combined on a group basis.

Feature Group	Visible $T_{D_{ave}}$	Infrared $T_{D_{ave}}$*	Visible & Infrared $T_{D_{ave}}$
SGLDs (4)	236.4	265.3	587.6
R.G. (1)	187.4	163.2	194.3
S.D. (1)	173.5	184.7	226.6
NAAs (2)	193.9	161.3	235.7
POLENTs (4)	176.2	182.9	435.6
CIRCs (4)	372.1	255.0	648.6

Note: 1. Separability of 11 hard-to-separate cloud class pairs when the three basic spectral features, MEANs (visible, near-infrared, and infrared) were used alone:

$T_{D_{ave}}$: 147.6

2. The number of features in each group is given in parenthesis adjacent to the group name.

Table 6.21 Separability of thirteen cloud classes when the CIRC feature pairs were used in addition to the 3 basic spectral features MEANs.

Feature Pair	Visible $T_{D_{ave}}$	Visible $T_{D_{ave}}$*	Infrared $T_{D_{ave}}$	Infrared $T_{D_{ave}}$*
CIRC 1,2	854.1	457.5	856.5	506.9
CIRC 1,3	823.5	463.3	833.9	495.8
CIRC 1,4	813.3	451.1	818.8	476.9
CIRC 2,3	781.0	431.1	798.0	450.2
CIRC 2,4	777.8	427.6	792.5	443.9
CIRC 3,4	773.1	421.7	787.2	437.2

Table 6.22 Separability of thirteen cloud classes when mean MEAN feature pairs of different distances (1 to 4) were used in addition to the 3 basic spectral features MEANs.

Distance Pair	Visible $T_{D_{ave}}$	Visible $T_{D_{ave}}$*	Infrared $T_{D_{ave}}$	Infrared $T_{D_{ave}}$*
1,2	881.0	477.2	884.6	475.8
1,3	881.4	454.9	894.6	484.2
1,4	879.4	449.5	905.3	496.0
2,3	862.5	431.0	894.2	463.4
2,4	864.9	430.0	904.1	475.3
3,4	860.6	426.7	905.2	478.3

Table 6.23 Separability of thirteen cloud classes when the POLENT feature pairs were used in addition to the 3 basic spectral features MEANs.

Feature Pair	Visible $T_{D_{ave}}$	Visible $T_{D_{ave}}$*	Infrared $T_{D_{ave}}$	Infrared $T_{D_{ave}}$*
AVE ENT1,AVE ENT2	763.2	409.7	758.9	404.2
AVE ENT1,AVE ENT3	765.0	412.8	760.2	402.5
AVE ENT1,NO DOM	755.9	401.2	759.9	402.2
AVE ENT2,AVE ENT3	767.1	418.4	771.1	415.6
AVE ENT2,NO DOM	762.6	410.7	767.3	411.4
AVE ENT3,NO DOM	760.2	410.2	765.4	407.4

6.4. Summary and Discussion

This chapter evaluates comprehensively the effect of the different texture analysis techniques, of which some are newly developed, on interpreting AVHRR satellite radiance imagery, and compares the commonly-used 1-D Fisher distance with the multivariate separability index - the transformed divergence. Some problems discussed in Sections 2.4 were tackled in this study. Several conclusions are drawn from this study as follows about which techniques appear the best approach to aiding multi-spectral cloud classification as well as the consideration of the problems.

1. The experimental results (e.g., Tables 6.19 and 6.20) from this chapter show the general usefulness of the different texture analysis techniques for aiding multi-spectral cloud classification. Among these features, the spatial features perform substantially better than the spatial frequency features. The spatial information contained in the frequency features is not as effective as that in the spatial features.

2. Features based on spatial grey level difference statistics (SGLDs) appear the best when used individually. However, when used in group-based combination (4 features per group), the newly-developed circular Moran autocorrelation features (CIRCs) appear the best. The reason is due to the different sensitivity of CIRC and SGLD features to the spatial properties of cloud fields. The further investigation on the sensitivity of these two groups of textural features shows that SGLD features have less size sensitivity than CIRC features, i.e., SGLD features are less potential than CIRC

features in enhancing the separability of the cloud classes considered in this study when they are evaluated to find optimal feature pairs or feature groups.

3. Taking the size sensitivity of various textural features into account may greatly aid the selection of an optimal feature set.

4. The specification of the currently-used features based on normalized average amplitude Fourier power spectrum (NAAs) is not appropriate, which is probably one of the major reasons for the poor performance. In fact, the fine detail reflected at high wavenumbers should be emphasized rather than smoothed by defining mean NAA. The further investigation of appropriate frequency features from polar coordinate Fourier spectra is promising, since the newly-developed visible feature NO DOM has ranked relatively high in this comparative study.

5. The newly-developed entropy-based measures (POLENTs) do not perform as well as expected, and their performance is about equivalent (somewhat better in most cases) to the NAA features but somewhat outstanding for separating the 11 hard-to-separate pairs of the 13 cloud classes when these measures calculated from both visible and infrared channels are combined on a group basis. However, a further investigation is still necessary, e.g., to select optimal wavenumbers at which the entropy measures are calculated as well as appropriate "censoring" values. A "censoring" procedure has proved to be essential for using the

entropy-based features in this study, since it can be used to clean up small frequency entries which represent random noise or ragged cloud structures in satellite imagery.

6. The simple statistic "local variance" in the infrared channel appears potentially the second most useful textural feature but the performance is not reflected by either the 1-D Fisher distance or variance analysis. It provides the computationally cheapest statistical measure for aiding multi-spectral cloud classification.

7. Generally, the suitability of visible and infrared channels for calculating textural features varies with the spatial properties of cloud fields.

8. The evaluation of class separability by the 1-D Fisher distance tends to be fallacious and textural features selected by the distance cannot be expected to be optimal in aiding multi-spectral cloud classification, since the statistical distributions of class clusters in multi-dimensional feature space are not fully taken into account.

A number of conclusions have been drawn from this comparative study. However, due to some limitations discussed below, these conclusions should be considered tentative.

One of the major limitations is the insufficient variation included for each cloud class, since the satellite radiance data were gathered from satellite observations on only two days. Therefore, the conclusions drawn in this study will not be universally applicable.

To reveal the contributions of various kinds of textural features for some specific pairs of cloud classes, we evaluated the separability of these class pairs based on each group of the features rather than on individual features. The selection of the group size, four, depended mainly on the pre-specified number of the CIRC features. The choice of both the group size and the group-based feature combination might not be optimal. This is due to the inherent purpose of this comparative study, which made the investigation of these choices not in a certain depth. This may lead to some doubts about the results obtained in this study.

The textural features used in this study were not necessarily the best of each group of statistical textural features (Eight groups were discussed in Section 2.3.2.2 and four of these groups were considered in this study). Most of these features were adopted from other cloud classification studies such as by Parikh (1977), Wu *et al.* (1985) and Ebert (1987), while the entropy-based features (POLENTs) and circular Moran autocorrelation functions (CIRCs) were developed and first used in this study. The design of the POLENT features needs to be quantitatively assessed further, e.g., by determining appropriate frequency intervals or weighting high-frequency components. The specification of the NAA features should also be modified since it seriously degraded the spatial information at high wavenumbers.

In spite of these inadequacies, the conclusions drawn from this study are still of value. It is expected by the author that these experimental results may lead to a further understanding of quantifying cloud fields by texture descriptors and the optimal choices of the textural features, and may encourage a further investigation of texture analysis techniques for aiding multi-spectral cloud classification.

CHAPTER 7

SUMMARY

This thesis has been concerned with three areas of research: texture analysis techniques, the problem of gathering a set of optimal cloud class statistics and a feature selection procedure. These three areas cover a broad field of image processing techniques, statistical pattern recognition techniques and their applications for quantitative remote sensing in meteorology. The background of the three areas and the problems that were tackled in this study have been discussed throughout the previous six chapters. The experiments carried out in this study and the corresponding conclusions were mainly described in Chapters 4 and 6.

This chapter gives a summary of the aims and contributions of this study as well as proposals for future research.

7.1. Aims of the Research

As stated earlier, the general aim of the research was to develop some textural features and to compare these newly-developed features with some widely-used textural features in several different ways. The motivation resulted mainly from a historical review of previous cloud classification studies (Refer back to Chapter 2).

Previous cloud classification studies, e.g., by Parikh (1977), Seddon (1983) and Ebert (1987), although they applied some textural features with multi-spectral features, did not emphasize the importance of these textural features. In fact, most of the textural features that they investigated were

discarded in their feature selection stages. A significant characteristic of their feature selection stages is that they assessed the textural features individually in one dimension by a pairwise separability index. A few spectral features, among all the features that they considered, were dominant in discriminating performance, and so ranked the highest in many cases and were then selected. Unfortunately, the auxiliary function of these textural features in aiding multi-spectral cloud classification was not truly reflected by these previous studies. Moreover, since the features were assessed individually in one dimension, some important properties, e.g., the usefulness of the size variation of a feature pair or group for cloud classification, were also not considered in these studies.

The use of features based on Fourier power spectra is often considered as an important approach to viewing texture in the discipline of image processing. Some frequency features which are specified in the spatial frequency distributions have been given by Bunting and Fournier (1980), and their applicability for cloud classification has been assessed by them and later by Ebert (1987). The unsatisfactory performance of these frequency features, shown by these researchers, necessitates a further investigation for more accurate specification of the Fourier power spectra. One of the problems with the frequency features defined by Bunting and Fournier (1980) was considered in this study, which is their uncertainty in characterizing the spatial distributions of frequency entries within circular bands in the polar frequency domain. A solution by defining a set of entropy-based features in the polar frequency domain was given in this study, and its performance was compared with the other textural features including those frequency features which were defined by Bunting and Fournier.

Since cloud classification studies are still in their infancy, the choice of some important approaches has been often rather arbitrary, e.g., what feature

selection criterion should be used. This study also aimed to tackle some of these specific problems.

7.2. Contributions of the Research

The significance of the research is that it suggests a new approach to evaluating the importance of various textural features in aiding multi-spectral cloud classification, i.e., by emphasizing various textural features in multi-dimensional feature space. By this approach, the auxiliary function of different textural features for aiding multi-spectral cloud classification can be well reflected, and the evaluation of the composite properties of these textural features also becomes possible. The latter suggests further a new approach to future feature selection studies for cloud classification by showing that the appropriate combination of textural features may contribute more to enhancing the separability of hard-to-separate class pairs than the combination without taking size variation of the different textural features into consideration. Moreover, the research experienced again the problem in achieving higher classification accuracies by using a set of new textural features in the frequency domain after the Bunting and Fournier's (1980) study. The unsatisfactory results arising from our study confirm again the inefficiency of using frequency features for cloud classification.

The main conclusions from the research may be summarized as follows.

1. In general, spatial features perform significantly better than frequency features when used on either an individual basis or a group basis. Spatial information contained in frequency features, e.g., size variation, is not effective as that in spatial

features.

2. Compared with circular autocorrelation features, those based on spatial grey level difference statistics do better when used individually. However, when one or more features of the same group, which are calculated at other different distances, are added, the improvement by the addition is greater in relative terms for circular autocorrelation features than those based on spatial grey level difference statistics. This indicates that features based on spatial grey level difference statistics have less discriminating potential than circular autocorrelation features, since they have less size sensitivity.

3. Taking the size sensitivity of various textural features into consideration is important in the selection of an optimal feature set.

4. The newly-developed entropy-based frequency features do somewhat better than frequency features based on summed energies of polar coordinate Fourier power spectra. However, this suggests another potential approach in designing a new frequency feature set, e.g., using entropy-based measures of the spatial distribution of frequency entries in polar power spectra.

5. A "censoring" procedure is essential for the entropy-based frequency features in cleaning up small frequency entries which represent random noise or ragged cloud structures in satellite imagery.

6. The suitability of visible and infrared channels for applying textural features varies with the spatial properties of cloud classes.

7. The simple textural feature "local variance" is powerful in enhancing the separability of some cloud classes with a substantial within-class variability.

7.3. Proposals For Future Research

As discussed in Section 6.4, there are some limitations in evaluating texture analysis techniques for aiding multi-spectral cloud classification in this study. One of these limitations is the insufficient variation included for each cloud class inherent in this comparative study. Another is that some trade-offs were made in using textural features for the evaluation, e.g., in choosing a group size (when groups of textural features were considered in combination) and in choosing the strategy of evaluating various textural features on a group basis rather than on every possible feature. These limitations may be solved or alleviated by taking long-term satellite observations, and selecting one or more features from each group of textural features, which is or are mostly representative of that group.

Besides these, some other proposals can also be made as follows for future research.

One is to investigate the effect of low-level image processing techniques on the interpretation of satellite radiance imagery by texture analysis. As discussed in Sections 6.3.1.4 and 6.3.2.2, the backgrounds of satellite images strongly

affect the values of textural features. Especially for circular autocorrelation functions, random noise or ragged cloud structures in the backgrounds may mislead the interpretation, e.g., for land and sea surface. For entropy-based frequency features, the effect of random noise or ragged cloud structures may be alleviated by an appropriate "censoring" procedure. A possible approach to dealing with this problem is to threshold or smooth the satellite images *prior* to their interpretation by the textural features. In the past, few efforts have been made towards solving this problem.

Although features based on summed energies of Fourier power spectra alone or entropy-based features alone did not seem to be significantly helpful for aiding multi-spectral cloud classification, it is obvious that their usefulness can supplement each other. It can be expected that a texture analysis of cloud type will benefit from a combination of these two groups of frequency-domain textural features, one based on summed energies and the other based on entropy-based measures of the spatial distribution of frequency entries.

Moreover, appropriate spatial scales, e.g., distances at which textural features are calculated and an appropriate sample window size, for applying different texture analysis techniques should be also investigated further.

I. ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors, Dr. C.N.Duncan and Professor P.M.Grant, for their helpful ideas and encouragement throughout this study. I would also like to thank Dr. E.Renshaw of the Statistics Department, University of Edinburgh, for providing his programs for some parts of this study as well as his helpful suggestions. I thank Dr. K.J.Weston for taking the time to help me in identifying cloud types in the satellite radiance images, and Miss M.Mugglestone of the Statistics Department, University of Edinburgh, for our discussions and her help in using some statistical packages.

I would like to thank Dr. K.G.Karlsson, Swedish Meteorological and Hydrological Institute for our oral and written discussions regarding their experience in cloud classification. The help of Dr. E.E.Ebert of the Meteorology Department, University of Wisconsin-Madison, in providing her helpful documents is gratefully appreciated.

My thanks should also go to many people in the Edinburgh University Computing Service (EUCS) for their assistance, especially for allowing me to get easy access to the EUCS VAX 8500 computer, and go to Mr. J.H.Butler of the Computer Science Department, University of Edinburgh, for providing his help in using the utilities of the Computer Science Department Image File Format.

I am indebted to the Chinese State Meteorological Administration and the Chengdu Institute of Meteorology for providing me with the support to study in Britain.

The work was also in part supported by the British Overseas Research Students Awards scheme.

II. REFERENCES

Alm, G., 1985: Land-use classification of aerial photographs using texture and spectral information, Proceedings of 4th Scandinavian Conference on Image Analysis, pp373-382.

Anderson, T.W. and S.L.Sclove, 1978: Statistical Analysis of Data, Houghton Miffin, Boston, 704pp.

Anderson, R.K. and N.F. Veltishchev, 1973: The use of satellite pictures in weather analysis and forecasting, World Meteorological Organization, Technical Note No.124, WMO-No.333.

Bajcsy, R. and L.Lieberman, 1976: Texture gradient as a depth cue, Computer Graphics and Image Processing 5, pp52-67.

Barrett, E.C., 1974: Climatology From Satellites, Methuen, London and New York, 418pp.

Barrett, E.C and C.K.Grant, 1978: An appraisal of Landsat 2 imagery and its application for the design of future meteorological observing systems, J.British.Interpl.Soc., Vol.31, pp3-10.

Barrett, E.C. and D.W.Martin, 1981: The Use of Satellite Data in Rainfall Monitoring, Academic Press, London, New York, 340pp.

Bartlett, M.S., 1975: The Statistical Analysis of Spatial Pattern, Chapman Hall, pp58-59.

Bell, D.A., 1953: Information Theory and Its Engineering Applications, SIR ISAAC Pitman Sons, London, 138pp.

Blankenship, J.R., 1962: An approach to objective nephanalysis from an

Earth-oriented satellite, J.Appl.Meteor., Vol.1, pp581-582.

Booth, A.L., 1973: Objective cloud type classification using visual and infrared satellite data, University of Maryland Institute for Fluid Dynamics and Applied Mathematics Technical Note Bn 768, College Park, Maryland.

Boyd, R.K., J.O.Brumfield and W.J.Campbell, 1983: A comparison of the usefulness of canonical analysis, principal component analysis, and band selection for extraction of features from TMS data for landcover analysis, the 7th International Symposium on Remote Sensing of Environment, Michigan, 10pp.

Brodatz, P., 1966: Textures - A photographic Album for Artists and Designers, Dover, New York.

Bunting, J.T. and R.F.Fournier, 1980: Tests of spectral cloud classification using fine mode satellite imagery, U.S.Air Force Geophysics Lab, AFGL-TR-80-0181.

Chetverikov, D., 1981: Textural anisotropy features for texture analysis, Proc.Int.Conf.on Pattern Recognition and Image Processing, Texas, pp583-588.

Chin, R.T. and J.Y.C.Jau, 1987: The applications of time series models to cloud field morphology analysis, J.Climat.Appl.Meteor., Vol.26, pp363-373.

Cliff, A.D. and J.K.Ord, 1973: Spatial Autocorrelation, Pion, London, 178pp.

Coakley, J.A. and F.P.Bretherton, 1982: Cloud cover from high-resolution scanner data: detecting and allowing for partially filled field of view, J.Geophy.Res., 87, pp4917-4932.

Colwell, R.N.(ed.), 1983: Manual of Remote Sensing (2nd.ed.), I, II, The Sheridan Press.

Conners, R.W. and C.A.Harlow, 1980A: A theoretical comparison of texture algorithms, IEEE trans Patt.Anal.Mach.Intell., Vol.PAMI-2, No.3, pp204-222.

Conners, R.W. and C.A.Harlow, 1980B: Towards a structural textural analyzer based on statistical methods, Computer Graphics and Image Processing 12, pp224-256.

Conover, J.H., 1962: Cloud interpretation from satellite altitudes, AFCRL-62-680, Air Force Cambridge Research Laboratories, M.A., 76pp. (1963, supplement 1, 18pp).

Darling, E.M. and R.D.Joseph, 1968: Pattern recognition from satellite altitudes, IEEE trans Sys.Sci.Cybernet., Vol.SSC-4, No.1, pp38-47.

Davis, L.S., S.A.Johns and J.K.Aggarwal, 1979: Texture analysis using generalized co-occurrence matrices, IEEE trans Patt.Anal.Mach.Intell., Vol.PAMI-1, No.3, pp251-259.

Davis, L.S., M.Clearman and J.K.Aggarwal, 1981: An empirical evalustion of generalized co-occurrence matrices, IEEE trans Patt.Anal.Mach.Intell., Vol.PAMI-3, pp214-221.

Desbois, M., G.Seze and G.Szejwach, 1982: Automatic classification of clouds on Meteosat imagery: Application to high-level clouds, J.Climat.Appl.Meteor., Vol.21, pp401-412.

Desbois, M. and G.Seze, 1984: Use of space and time sampling to produce representative satellite cloud classification, Annales Geophysics, 2, 5, pp599-606.

Devijver, P.A. and J.Kittler, 1982: Pattern Recognition: A Statistical Approach, Englewood Cliffs, New Jersey: Prentice Hall.

Duda, R.O. and P.E.Hart, 1973: Pattern Classification and Scene Analysis, John Wiley & Sons, 482pp.

Ebert, E., 1987: A pattern recognition technique for distinguishing surfaces and

cloud types in the polar regions, J.Climat.Appl.Meteor., Vol.26, No.10, pp1412-1427.

Ebert, E.E., 1988: Classification and analysis of surface and clouds at high latitudes from AVHRR multispectral satellite data, PhD thesis (Scientific Report 8), University of Wisconsin, Madison.

Everitt, B.S., 1974: Cluster Analysis, Heinemann, 122pp.

Garand, L. and J.A.Weinman, 1986: A structural-stochatic model for the analysis and the synthesis of cloud images, J.Climate.Appl.Meteor., Vol.25, pp1052-1068.

GARP/JOC, 1978: Proceedings of the JOC study conference on parameterization of extended cloudiness and radiation for climate models, Oxford (Geneva:World Meteorological Organization), 37pp.

Haralick, R.M., K.Shanmugan, I.Dinstein, 1973: Textural features for image classification, IEEE trans Syst.Man.Cybernet., Vol.SMC-3, No.6, pp610-621.

Haralick, R.M., 1979: Statistial and structural approaches to texture, Proceedings of the IEEE, Vol.67, No.5, pp786-803.

Harris, R. and E.C.Barrett, 1978: Toward an objective nephanalysis, J.Appl.Meteor., Vol.17, pp1258-1266.

Harris, R., 1982: Computer assisted nephanalysis: the story so far, Proceedings of the Remote Sensing and The Atmosphere Annual Technical Conference Held in Liverpool, Remote Sensing Society, 315pp.

Harshvardhan, 1982: The effect of brokeness on cloud-climate sensitivity, J.Atm.Sci., Vol.39, pp1893-1861.

Henderson-sellers, A., 1984: Satellite Sensing of A Cloudy Atmosphere: Observing The Third Planet, Taylor & Francis, London and Philadelphia, 340pp.

Jenkins, G.M. and D.G.Watts, 1968: Spectral Analysis and Its Application, San Francisco: Holden-Day.

Jernigan, M.E. and F.D'astous, 1984: Entropy-based texture analysis in the spatial frequency domain, IEEE trans Patt.Anal.Mach.Intell., Vol.PAMI-6, No.2, pp237-243.

Julesz, B., 1975: Experiments in the visual perception of texture, Scientific American, Vol.232, No.4, pp2-11.

Karlsson, K.G., 1987A: Operational cloud classification - Development and future plans at the Swedish Meteorological and Hydrological Institute, Third AVHRR data user meeting.

Karlsson, K.G., 1987B: personal communication, Swedish Meteorological and Hydrological Institute.

Katz, Y.H., 1965: Pattern recognition of meteorological satellite cloud photography, Proceedings of the Third Symposium on Remote Sensing of the Environment, Ann Arbor, Michigan, pp178-214.

Kirvida, L., 1976: Texture measurements for the automatic classification of imagery, IEEE trans Electromagnet.Compat., Vol.18., pp38-42.

Krishnamurti, R., 1975: On cellular cloud patterns part 1: mathematical model, J.Atm.Sci., Vol.32, pp1353-1363.

Lauritson, L., G.J.Nelson, F.W.Porto, 1979: Data extraction and Calibration of TIROS-N/NOAA radiometers, NOAA Technical Memorandum NESS 107, Washington, 74pp.

Leese, J.A. and E.S.Epstein, 1963: Application of two-dimensional spectral analysis to the quantification of satellite cloud photographs, J.Appl.Meteor., Vol.2, pp629-644.

Liljas, E., 1984: Processed satellite imageries for operational forecasting, Report of SMHI.

Liljas, E., 1986: Use of the AVHRR 3.7 micrometer channel in multispectral cloud classification, Report of SMHI.

Lillesand, T.M. and R.W.Kiefer, 1979: Remote Sensing and Image Interpretation, John Wiley & Sons, New York, 612pp.

Lu, S.Y., and K.S.Fu, 1978: A syntactic approach to texture analysis, Computer Graphics and Image Processing 7, pp303-330.

Minnis, P. and E.F.Harrison, 1984: Diurnal variabilty of regional cloud and clear-sky radiance parameters derived from GOES data I: analysis method, J.Climat.Appl.Meteor., Vol.23, No.7, pp993-1011.

Oppenheim, A.V. and R.W.Schafer, 1975: Digital Signal Processing, Prentice-Hall, New Jersey, 585pp.

Pairman, D. and J.Kittler, 1986: Clustering algorithm for use with images of clouds, Int.J of Remote Sensing, Vol.7, No.7, pp855-866.

Parikh, J., 1977: A comparative study of cloud classification techniques, Remote Sensing of Environment 6, pp67-81.

Parikh, J., 1978: Cloud classification from visible infrared SMS-1 data, Remote Sensing of Environment 7, pp85-92.

Parikh, J. and A.Rosenfeld, 1978: Automatic segmentation and classification of infrared meteorological satellite data, IEEE trans Syst.Man.Cybernet., Vol.SMC-8, No.10, pp736-743.

Parikh, J.A. and J.T.Ball, 1980: Analysis of cloud types and cloud amount during GATE from SMS IR data, Remote Sensing of the Environment 9, pp225-245.

Pietikainen, M. and A.Ronsenfeld, 1981: Image segmentation by texture using pyramid node linking, IEEE trans Syst.Man.Cybernet., Vol.SMC-11, No.11, pp822-825.

Pratt, W.K., L.R.Welch and W.Chen, 1974: Slant transform for image coding, IEEE trans Comm., Vol.Com-22, 1974, pp1075-1093.

Pratt, W.K., 1978: Digital Image Processing, New York Chichester, 750pp.

Pratt, W.K., O.D.Faugeras and A.Gagalowicz, 1981: Application of stochastic texture field models to image processing, Proceedings of the IEEE, Vol.69, No.5, pp54-551.

Renshaw, E. and E.D.Ford, 1983: The interpretation of process from pattern using two-dimensional spectral analysis: method and problems of interpretation, Appl.Statist, 32, No.1, pp51-63.

Renshaw, E. and E.D.Ford, 1984: The description of spatial pattern using two-dimensional spectral analysis, Vegetatio, 56, pp75-85.

Renshaw, E., 1987: personal communication, University of Edinburgh.

Reynolds, D.W. and T.H.Vonder Haar, 1977: A bispectral method for cloud parameter determination, Monthly Weather Review, Vol.105, pp446-457.

Richards, J.A., 1986: Remote Sensing Digital Image Analysis, Springer-Verlag, 281pp.

Rosenfeld, A., C.Fried and J.Orton, 1965: Automatic cloud interpretation, Photogrammetric Engineering, Vol.31, pp991-1002.

Rosenfeld, A. and M.Thurston, 1971: Edge and curve detection for visual scene analysis, IEEE trans Comput., Vol.C-20, pp562-569.

Rosenfeld, A.(ed.), 1976: Digital Picture Analysis, Topics in Applied Physics, 345pp., Springer-Verlag.

Rosenfeld, A. and A.C.Kak, 1982: Digital Picture Processing, Academic Press, New York.

Rossow, W.B., F.Mosher, E.Kinsella, A.Arking, M.Desbois, E.harrison, P.Minnis, E.Ruprecht, G.Seze, C.Simmer, E.Smith, 1985: ISSCP cloud algorithm intercomparison, J.Climat.Appl.Meteor., Vol.24, pp877-903.

Saunders, R.W. and K.T.Kriebel, 1988: An improved method for detecting clear sky and cloudy radiances from AVHRR data, Int.J of Remote Sensing, Vol.9, No.1, pp123-150.

Schowengerdt, D.A., 1983: Techniques For Image Processing and Classification in Remote Sensing, Academic Press, 249pp.

Schwalb, A., 1979: The TIROS/NOAA A-G satellite series, NOAA Technical Memorandum NESS 95, Washington, D.C., 75pp.

Seddon, A.M., 1983: The application of scene analysis techniques to automatic classification of atmospheric data from multispectral satellite data, PhD thesis, University of London, 293pp.

Seddon, A.M. and G.E.Hunt, 1985: Segmentation of clouds using cluster analysis, Int.J of Remote Sensing, Vol.6, NO.5, pp717-731.

Souza, P.D., 1982: Texture Recognition via autoregression, Pattern Recognition, Vol.15, No.6, pp471-475.

Spider Working Group, 1983: Spider User's Manual, AISI MITI Japan.

Sun, Z.K. and C.K.Shen, 1985: Digital Image Processing and Its Applications, Defen.Indust.Publisher (China), 351pp.

Swain, P.H. and S.M. Davis, 1978: Remote Sensing: the Quantitative Approach, New York, Mcgraw-Hill.

Thomas, I.L., V.M.Benning and N.P.Ching, 1987: Classification of Remotely Sensed Images, Adam Hilger, Bristol.

Unser, M. and F.de.Coulon, 1982: Detection of defects by texture monitoring in automatic visual inspection, Proc. of the 2nd Int.Conf.on Robot Vision and Sensory Controls, Germany, pp27-38.

Van Gool, L., P.Dewaele and A.Oosterinck, 1985: Texture analysis anno 1983, Computer Vision, Graphics and Image Processing 29, pp336-357.

Wang, S., D.Velasco, A.Y.Wu and A.Rosenfeld, 1981: Relative effectiveness of selected texture primitive statistics for texture discrimination, IEEE trans Syst.Man.Cybernet., Vol.SMC-11, No.5, pp360-370.

Wallace, J.M. and P.V.Hobbs, 1977: Atmosphere Science, An Introductory Survey, Academic Press, New York, 467pp.

Welch, R.M. and B.A.Wielicki, 1984: Stratocumulus cloud field fluxes: the effect of cloud shape, J.Atm.Sci., Vol.41, pp3085-3103.

Weszka, J., C.Dyer and A.Rosenfeld, 1976: A comparative study of texture measures for terrain classification, IEEE trans Syst.Man.Cybernet., Vol.SMC-6, No.4, pp269-285.

Wickers, A.L. and J.W.Modestino, 1982: A maximum Likelihood approach to texture classification, IEEE trans Patt.Anal.Mach.Intell., Vol.PAMI-4, No.1, pp61-68.

Wielicki, B.A. and R.M.Welch, 1986: Cumulus cloud properties derived using Landsat satellite data, J.Climat.Appl.Meteor., Vol.25, No.3, pp261-275.

Wishart, D., 1978: Clustan User Manual (3rd ed.), Program Library Unit, University

of Edinburgh.

Wu, R., J.A.Weinman and R.T.Chin, 1985: Determination of rainfall rates from GOES satellite images by a pattern recognition technique, J.Atmos.Ocean.Technol., 2, pp314-330.

TEXTURAL AND SPECTRAL FEATURES AS AN AID TO CLOUD CLASSIFICATION

Zhiqiang Gu & C Duncan

Department of Meteorology, University of Edinburgh, U.K.

ABSTRACT

The problem of classifying clouds seen on meteorological satellite images into different types is one which requires the use of textural as well as spectral information. Several textural features are studied to determine their discriminating power across a number of cloud classes including those which have previously been found difficult to separate. Although several features in the spatial frequency domain are tested they are found to be less useful than those in the spatial domain with only one exception.

Keywords: Cloud classification, texture, spatial statistics, spatial frequency, autocorrelation, Fisher distance.

1. INTRODUCTION

A vast number of TIROS-n/NOAA-n images have been accumulated since the launching of TIROS-1. There is a considerable amount of information in these pictures which has not been fully utilized. One of the major reasons for this is the difficulty in quantifying the information contained in the pictures in a meaningful form.

Among the uses of cloud mapping are, classifying pictures to guide forecasters when interpreting satellite images, deriving cloud and precipitation fields for use in numerical forecasting models, and extracting a limited amount of information from satellite pictures that can be archived for climatological purposes.

Spectral features have been widely used in characterizing and understanding the appearance of clouds on remotely sensed imagery (Refs 1-3). However, spectral features alone are inadequate for the specification of all cloud properties. This study investigates some newly-developed textural features to aid cloud classification and compares them with features used in the past.

2. DATA SET

The data selected for this study were two 2048 x 1024 subsets of AVHRR images covering a large area (0° to 20°W and 45°N to 60°N).

The resolution of the satellite image is 1.1 x 1.1 km at nadir. The image subset was chosen to lie along the satellite track in order to avoid the necessity for geometric correction.

The choice of the thirteen cloud classes relied partly upon the natural grouping of clustering procedures and partly upon Ref 2 where some of these cloud classes were found to be overlapping seriously in their multispectral signature space.

The classes used are as follows (numbers indicate the number of samples, 32x32 pixels, in each class - total 649):

a	land (cloud free)	41
b	stratus	75
c	open water (cloud free)	91
d	cumulus (includes cells and lines)	69
e	thick cirrus over low cloud	77
f	altostratus	26
g	thick cirrostratus	42
h	cirrostratus	33
i	cumulonimbus	57
j	cirrus with open water	54
k	thin cirrus over low cloud	16
l	thin cirrus over middle cloud	56
m	stratocululus	12

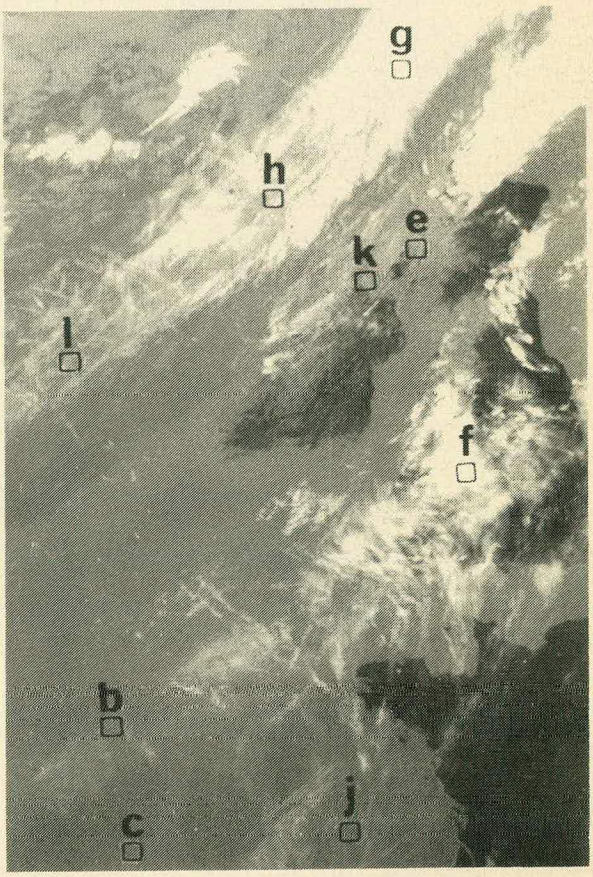


Figure 1. Part of one of the images used showing 9 of the 649 cloud classes. Some classes are difficult to separate in this infrared image.

Proceedings of IGARSS '88 Symposium, Edinburgh, Scotland, 13–16 Sept. 1988. Ref. ESA SP-284 (IEEE 88CH2497-6). Published by ESA Publications Division, August 1988.

3. FEATURES

Twenty six different features were tested to identify their ability to discriminate between the 13 different cloud classes. These 26 features were calculated using visible (CH1), near-infrared (CH3) and infrared (CH4) images to give a total of 58 features. (Only the six spectral features were calculated in the near-infrared channel because noise invalidated textural features.) The features can be separated into four broad classes a) simple statistics, b) spatial grey level difference statistics, c) spatial frequency statistics, and d) autocorrelation statistics. Simple statistics include mean (1), maximum (2), minimum (3), standard deviation (4), range (5), ratio of maximum to minimum (6) and Roberts gradient (7) (Ref 1). Grey level difference statistics include the mean (8-11) and maximum (12-15) values of the four parameters mean, contrast, angular second moment, and entropy (Ref 1). Spatial frequency statistics include the mean (16), maximum (17), and dominant wavenumber (18) of the normalised averaged amplitude of the Fourier power spectrum in polar coordinates. Also included are entropy measures in the spatial frequency domain which are produced by considering annuli in the polar power spectrum as probability distributions. The entropy in three annuli at low (19), medium (20), and high (21) wave numbers are used. The number of spectral components in the entire power spectrum which were greater than a critical 'censoring' value was also used (22). The other four features are circular Moran autocorrelation statistics (Ref 4) in which the usual one dimensional autocorrelation based on lags is replaced by a two dimensional autocorrelation in which the lag is represented by the radius of a circle of points around the reference point. The circles are approximated by points on a regular grid whose distance from the centre is one (23), two (24), three (25) or four (26) grid lengths from the centre.

4. RESULTS

The ability of each of these features to discriminate between the 13 cloud classes has been tested in a number of ways. A simple first step was to calculate the mean and standard deviation of each of the features in two channels, visible and infrared, for every one of the cloud classes. This leads to a matrix, 52x13, which illustrates both the within class and between class variance for every feature, but it is essentially one dimensional in feature space. A useful measure for summarising this large table is the F-ratio (ie the ratio of the between-class variance to the within-class variance). High values of F-ratio indicate good discrimination, but low values are not sufficient evidence to reject features. Table 1 shows the features for which the F-ratio values exceeded 20.

Feature	F-ratio		No. of times ranked first		
	CH1	CH4	CH1	CH3	CH4
1	589 *	975 *	16 *	26 *	14 *
2	79	40			
3	31 *	312	2 *	4	
4	195	480	2 *		3 *
5		43 *	1 *		
6		41	1		
7	23 *				
8	195	184	2		
9	319	766			3
10	286 *	66			
11	215 *	67			
12	191	153 *			
13	279	506			
14	265	66			
15	210	71 *	3 *		
16			1 *		
23	22 *				

Table 1: F-ratio values for selected features for CH1 and CH2, and the number of times these features were the top ranked discriminator using Fisher distance as a measure. Asterisks show those features which remain if all features with correlation >0.8 with higher ranked features are rejected.

The F-ratio test is very general and in no way guarantees that all possible pairs of cloud types are discriminated. Another commonly-used method of checking the separability of different classes is to calculate the Fisher distance between every possible pair of cloud types (78 pairs) for every feature. We can then identify the feature which has the largest Fisher distance for every pair. In this way the best feature for every possible pair is found. Table 1 also shows the number of times each feature had the largest Fisher distance.

5. CONCLUSIONS

A wide range of textural features have been examined to establish their discriminating power for cloud types. In an effort to minimise the number of features used in any cloud type classification the following are recommended using a sample area of 32x32 pixels on NOAA/TIROS images:

- mean radiance in channel 1
- mean radiance in channel 3
- mean radiance in channel 4
- maximum radiance in channel 1
- minimum radiance in channel 1
- minimum radiance in channel 4
- mean normalized average amplitude of the polar spectrum in channel 1
- maximum entropy in the spatial grey level differences in channel 1
- ratio of maximum/minimum radiance in channel 4
- Roberts gradient in channel 1
- maximum mean value of the spatial grey level differences in channel 4
- maximum entropy in the spatial grey level differences in channel 4
- circular autocorrelation statistic with radius 1 in channel 1

These features are a combination of those with the highest ranking by either assessment method (F-ratio or Fisher distance) but neglecting any which are well correlated (>0.8) with higher ranking features.

It is important to note that all but one of the textural features are spatial features and that the spatial frequency features are not, in general, as useful. However, for one cloud pair (b-k) a spatial frequency feature is the best discriminator.

It has been suggested in Ref 5 that features which have low discriminating power by themselves may have much higher discriminating power when combined. The circular Moran autocorrelation statistics, although not very useful individually, are powerful in combination (Ref 6) because the four measures respond to clouds of different size.

6. REFERENCES

1. Ebert, E., 1987, A pattern recognition technique for distinguishing surfaces and cloud types in the polar regions, J.Climat.Appl.Meteor., Vol.26, No.10, pp1412-1427.
2. Liljas, E., 1984, Processed satellite imageries for operational forecasting, Report of SMHI.
3. Parikh, J., 1977, A comparative study of cloud classification techniques, Remote Sensing of Environment 6, pp67-81.
4. Cliff, A.D. and Ord, J.K., 1973, Spatial Autocorrelation, Pion, London, 178pp.
5. Weszka, J.S., Dyer, C.R., and Rosenfeld, A., 1976, A comparative study of texture measures for terrain classification, IEEE Trans. on Systems, Man and Cybernetics, pp269-285.
6. Gu, Z., 1988, Texture analysis techniques for multispectral cloud classification, M.Phil. thesis, University of Edinburgh.