TN-102

DIGITAL IMAGE PROCESSING AND PATTERN RECOGNITION

NATIONAL INSTITUTE OF HYDROLOGY
ROORKEE - 247 667
1992-93

PREFACE

In December 1985 the government of the Netherlands and India agreed on an Indo-Dutch Training Programme on Water Management (WAMATRA). Phase I was concluded in 1986-87 and phase II of WAMATRA was started in August 1990. National Institute of Hydrology was associated along with other institutions in this project in few areas. One of the items was "Remote Sensing Application for Hydrological Studies". International Institute for Aerospace Survey and Earth Sciences, (ITC), The Netherlands; have to impart training to the scientists of National Institute of Hydrology, in the above mentioned area.

This technical report is based on training of Shri Anil Kumar, Scientist "B" on "Digital Image Processing and Pattern Recognition". It is a regular course of four month duration, starting generally in middle of February in each year. As per information available at ITC, the course is updated each year in view of rapid development in image processing, pattern recognition, artificial intelligence and data base systems applicable to integrated (RS+GIS) remote sensing data processing. The main objective of this course was to impart the training about the processing of remote sensing data and to extract information from these data. The training mainly consist of lectures, practicals, interaction, field work and case studies.

This report will be useful for the people who need an insight into the actual working of remote sensing data processing for direct involvement in their work environment, i.e. research scientist and teaching staff at university level.

ROORKEE

June 1991

Satish Chandra
(SATISH CHANDRA)

ACKNOWLEDGEMENT

My sincere gratitute is due to Prof Dr A M J Meijerink and the IPL staff for their assistance and availability throughout the entire course. Gratitude as well to Dr. G. C. Mishra Sc 'F' and Co-ordinator of the project for his kind help on several issues and thanks to my fellow students in IGP3 and DIS5.

Finally, my sincere gratitute to the Director, National Institute of Hydrology, Roorkee for nominating my name for this training to enhence the knowledge in this emerging field of Remote Sensing technology.

CONTENTS

1. INTRODUCTION TO REMOTE SENSING AND DIGITAL COMPUTERS

1.1 INTRODUCTION

Remote sensing of the environment comprises the measurement and recording of reflected and emitted radiance by the earth surface and atmosphere. These measurements are normally made by the instruments carried by satellite or aircraft, and are used to infer the nature and characteristics of the land and sea surface, or of the atmosphere at the time of observation. The successful application of remote sensing techniques to particular problems, whether they may be geographical, geological, or hydrological requires knowledge and skills drawn from several areas of science. An understanding of the way in which the remotely sensed data are acquired by a sensor mounted on a spacecraft needs a basic knowledge of the physics involved. The use of remotely sensed data which are inherently digital, demands some degree of mathematical and statistical skill plus some familiarity with digital computers and their operation. The high level of competence is essential in the field in which the remotely sensed data are to be used, if full use of the information contained in those data is to be made.

The purpose of processing of the remotely sensed data is to extract information from these data. The processing can be distinguished on two bases;

1. Digital Image processing is directed towards image correction, image enhancement, feature extraction for information, presentation and further visual interpretation. Finally output are in the form of images.
2. Pattern recognition is a large collection of classification and discrimination techniques implemented with analytic, statistical or heuristic decision rule.

The input data for pattern recognition is not necessarily remote sensing data only and in practice an increasing proportion of the data comes from data bases. But output are maps.

The remotely sensed data are processed digitally using statistical, logical and mathematical operation. The remote sensing images and results of image processing can be stored in GIS and may be integrated with other types of data.

With the advancement in remote sensing technique large numbers of relatively cheap personal computers and work stations are becoming available for image processing in developing as well as in developed countries. However many users and operators of these systems only receive a push button training, which does not develop concepts such as computer assisted image analysis and the appropriate use of geoinformation system. This report will familiarize the research community about the various advancement made in Image Processing techniques.

1.2 DIGITAL IMAGE PROCESSING SYSTEM

The processing of remotely sensed images is normally done by computers. It is not necessary that all the users of digital image processing systems will be involved in programming but all the users should be familiar with the way in which digital data are stored and manipulation are being done within a computer. It is well known that remotely sensed data are supplied in the form of pixel values on the scale 0-63, 0-127, 0-255 or 0-1023. Data in the form of remotely sensed images are so voluminous that they could be impossible to use if they had to be typed at a visual display terminal by a human operator. These data are transmitted to earth from a satellite in digital form and then stored on magnetic tape. These tapes are converted to a form suitable for computes. Thus, they are called computer compatible tapes or CCTs. The CCT is read by computer from a peripheral device called magnetic tape unit. Because the memory of the computer is small relative to the amount of data stored on one CCT, the data from the CCT must be placed in an external memory unit, the most common form of which is the magnetic disc. Small flexible discs (floppy) are widely used on personal computers and microprocessors systems. Depending on the computer system, each such disc can store data from 100 kb to 1 mb of data, which is not enough even for one band of a Landsat MSS image (7.6 mb).

The disc store of a computer system is on line, that is its contents are readily accessible to the CPU. Magnetic tapes are not so accessible, they have to be loaded on the magnetic tape drive and wound forwards at a typical rate of 45 or 75 inches per second until the required data are read. It would be convenient to have all the images on disc, but a simple calculation will show that it is not possible with present technology. The largest fixed disc generally can store 500 Mb of data. A Landsat MSS image from earth is composed of 2286 scan lines of 3600 pixels in each of four bands, and since each pixel value require one byte of storage, a grand total of 32 Mb of storage is required to hold single MSS image. It might appear that a 500 Mb disc would hold 15 Landsat MSS images, and in reality the disc space available is less than 500 Mb. Except for images or partial images, that are in use at a given time, the majority of Landsat images will therefore be stored on magnetic tape and transferred to disc whenever needed.

In the above paragraph the characteristic of a general purpose computer system relevant to remote sensing were described. The image display system is a peripheral device which is dedicated to a particular purpose, that of storing and manipulating digital data and converting them in analog form for display on a television monitor. In remotely sensed images, the radiance values for individual pixel are converted to a number in the range 0-255, or 0-63 depending upon the radiometric resolution of the sensor. This process is called analog to digital or A to D conversion. The image display system carries out the converse operation, that of taking a pixel value and converting it to analog form, in this case a voltage. The separate signal voltage for red, green and blue are proportional to the intensities of the red, green and blue components of the multispectral image at the point on the image corresponding to the pixel on the monitor screen. If the red, green and blue components are equal at each position than black and white image well result .

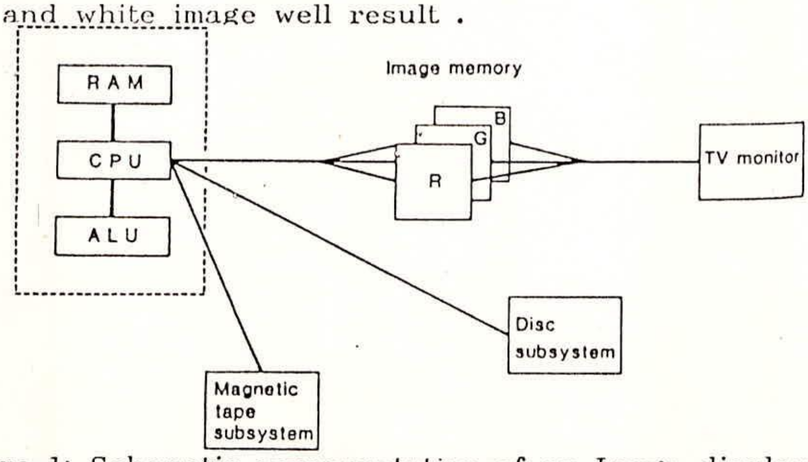


Figure 1: Schematic representation of an Image display system.

Before the digital image can be converted to a analog form for display it must first be moved into the memory of the image display system by running a program in the CPU of the host computer by reading each scan line of the image from a disc fit to the computers RAM memory, from which it is transferred to the memory of the image display subsystem. This image can be considered as a matrix, with m rows and n columns and the origin at the top left corner and each cell of the matrix is occupied by a pixel value. This report will provide the model based, object oriented image analysis and image understanding. Also within a geoinformatics contact, there is an introduction to geometric modelling, radiometric modelling etc for image prediction. Comparison with actual images provides feed back for updating the object oriented data base.

Image processing and pattern recognition are used for the abstraction of remotely sensed data into image objects with property list . Image processing is also used for geometric and radiometric correction and for visual enhancement of images.

2.0 REMOTE SENSING SATELLITE AND SENSORS

In this chapter the nature and characteristics of imaging sensors board on aircraft or spacecraft will be considered. Basically, we will restrict our discussion for the collection of image data by multispectral sensor systems. Multispectral sensor have been, and are presently, carried by American Space Shuttle.

2.1 REMOTE SENSING SENSORS

The general characteristics of imaging remote sensing instruments comprising their spatial, spectral and radiometric resolution and the number of spectral bands in which data are to be collected. Also an important property of remote sensing system, of which the instrument is a part, is the temporal resolution of the system, that is, the time that elapse between successive acquisition of same image. The satellite is of two types (i) geostationary (ii) polar orbiting type.

The INSAT, meteosat satellite is of geostationary type, whereas TIROS/NOAA, Nimbus, Landsat, IRS and SPOT satellites are of polar orbiting type. Further, the temporal resolution of geostationary type of satellite is zero whereas for a polar orbiting type it vary from days to month depending upon the altitude.

2.1.1 Spatial Resolution

The spatial resolution of an imaging system is not an easy concept to define. It can be measured in a number of different ways depending on the users purpose. In a wide ranging and comprehensive revive of the subjects, Townshend (1980) uses four separate criteria on which to base a definition of spatial resolution. These criteria are geometrical properties of the imaging system. The ability to distinguish between point targets, the ability to measure the periodicity of repetitive targets and the ability to measure the spectral properties of targets.

The most commonly used measure, based on geometric properties of imaging system is the instantaneous field of view (IFOV) of a sensor. It is defined as area on the ground that is viewed by the instrument from a given altitude at any given instant of time. The

IFOV is the most frequently cited measure of resolution, though it is not necessarily the most useful.

2.1.2 Spectral resolution

Except the microwave imaging system, the digital images collected by satellite borne sensors have been multiband or multispectral, that is individual images have been separately recorded in discrete spectral bands. The term spectral resolution refers to the width of these spectral bands. Two important point in the context of multispectral imagery are;

(i) The position in the spectrum, width and number of spectral bands will determine the degree to which individual targets can be discriminated on the multispectral image, and

(ii)The use of multispectral imagery can lead to a higher degree of discriminating power than any single band taken on its own.

2.1.3 Radiometric Resolution

Radiometric resolution or radiometric sensitivity refers to the number of digital levels used to express the data collected by the sensor. In general the greater the number of levels, the greater the detail of the information. The number of levels is commonly expressed in terms of the number of binary digits (bits) needed to store the value of maximum level. For 2 levels the number of bits is 1, while for 4, 16, 64, and 256 levels the number of bits required is 2,4,6 and 8 respectively. Thus, 6-bit data have 64 possible levels, represented by the integer value from 0 to 63 inclusive.

2.2 A BRIEF OVERVIEW OF SATELLITE REMOTE SENSING SYSTEMS

In this section a short account of the main satellite borne imaging sensor systems will be given. The sensors considered are those carried by the Landsat, SPOT, and IRS satellites.

Table 2.1 Landsat 1 to 3 Satellite and Multispectral Scanner Wave Bands

Multispectral Scanner		Landsat 1 to 3 satellite
Channel	wave bands (μm)	
4	0.5 - 0.6	Orbit: near polar,sun-synchronous
5	0.6 - 0.7	Altitude: 919 km nominal
6	0.7 - 0.8	Inclination: 99.09
7	0.8 - 1.1	Period:103 minutes
		Equatorial crossing time :0930

Revisit period: 18 days
Spatial resolution: 80m
Radiometric resolution: 6 bits (64 levels)
Swath width: 185 km

Note

1. The channel on the Landsat 4 and 5 MSS are numbered 1 to 4.
2. The radiometric resolution of Landsat 4 and 5 is 8 bit (256 level).

Table 2.2 Landsat 4 and 5 Satellite and Thematic Mapper Wave bands

Thematic Mapper		Landsat 4 and 5 satellite
Channel	wave band(μm)	
1	0.45-0.52	
2	0.52-0.60	Altitude: 705 km
3	0.63-0.69	Inclination: 98.2
4	0.75-0.90	Period:99 minutes
5	1.55-1.75	Equatorial crossing time : 0945
6	10.40-12.5	Revisit period: 16 days
7	2.08-2.35	

Swath width : 185 km
Spatial resolution : 30m (band 6: 120m)
Radiometric resolution : 8 bits (256 levels)

Table 2.3 SPOT satellite and HRV sensor detail

HRV sensor		SPOT satellite
Channel	wave bands (μm)	
1.	0.50-0.59	Orbit: near polar, sun synchronous
2.	0.61-0.68	Altitude:832 km
3.	0.79-0.89	Inclination: 98.7
		Equatorial crossing time: 1030
		Revisit period: 26 days
Panchromatic		
1.	0.51-0.73	

Spatial resolution : 20 m (multispectral)
10 m (Panchromatic)
Radiometric resolution: 8 bits (multispectral)
6 bits (Panchromatic)
Swath width: 117 km.

Table 2.4 IRS 1A Satellite and LISS I & LISS II Sensor detail

LISS I & LISS sensor		IRS 1A
Channel	Wave bands (μm)	
1	0.45-0.52	Orbit: near polar,sun-synchronous
2	0.52-0.59	Altitude: 904 km nominal
3	0.62-0.68	Inclination: 99.049^{0}
4	0.68-0.68	Period:103 minutes
		Equatorial crossing time :1025 hrs

Revisit period: 22 days
Spatial resolution: LISS I 72.50 m
: LISS II 36.25 m
Radiometric resolution: 8 bits (256 levels)
Swath width: 148 Km

Among the many satellite sensor systems, that have been or are operating for earth observation, there are mainly four of practical importance, namely the Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM), SPOT High Resolution Visible (HRV) and IRS 1A with LISS I and LISS II cameras.

All these sensor are operating from near polar in circular orbits. They cover entire surface of the earth except the pole regions and the data are available on computer compatible tapes (CCT).

3. PREPROCESSING OF REMOTELY SENSED DATA

The image data is recorded by sensors board on satellite and these data may contain errors in geometry and in the measured brightness value of the pixel. The later are referred to as radiometric errors and can result from the instrument used to record the data and from the effect of atmosphere. The error in image geometry can arise in many way, the relative motion of a satellite, its scanners and the earth. The correction to these errors is termed as preprocessing of the data because quite logically it is carried out prior to use of the data.

Image geometry errors can arise in many ways. The relative motions of a satellite, its scanners and the earth, for example lead to a skew Image. Non-idealities in sensor themselves, the curvature of the earth and variation in position and altitude of the remote sensing platform can lead to a varying degree of severe error.

When an image is to be utilized it is necessary to make correction in brightness and geometry if the accuracy of interpretation either manually or by machine is not to be prejudiced.

It is the purpose of this chapter to discuss the nature of the radiometric and geometric errors commonly encountered in remotely sensed images and to develop computational procedures for their compensation.

3.1 SOURCES OF GEOMETRIC ERRORS

There are many sources for geometric distortion of image data than radiometric distortion and their effects are more severe. Some of the features responsible for geometric error are listed below;

(i) The rotation of the earth during image acquisition

(ii) The finite scan rate of some sensors.

(iii)The wide field of view of some sensors

(iv) The curvature of the earth

(v) Sensor non-idealities

(vi) Variations in platform altitude, attitude and velocity

(vii)Panoramic effects related to imaging geometry

Further, radiometric errors can be grouped into two categories (i) Mechanical/Electrical (ii) Atmospherical

Mechanical or Electronical errors are caused by deficiencies in the satellite or at the ground receiving station. These errors can be of dropped lines or striping. But atmospherical errors are caused by the interaction of photons with molecules and dust particles in the earth atmosphere. These errors can be of haze or skylight type.

Radiometric correction are transformations on the data with the aim to remove only those errors that are geometry independent. Further, radiometric errors are relatively constant so that they can be modelled using deterministic model because sources of errors are known.

It is the purpose of this section to discuss the nature of distortions and the measure to compensate them.

3.1.1 Earth Rotation Effects

Sensors with line scanners such as Landsat (MSS and TM), and the NOAA (AVHRR) take a finite time to acquire the Image data of a frame . The same is true for push broom scanners such as the SPOT (HRV). During the frame acquisition time the earth rotates from west to east, so that a point imaged at the end of frame would have been moved to the west when recording started. Therefore, if the lines of image data recorded were arranged for display in the manner of Figure 3.1(a), the later lines would be displaced to the east in terms of the terrain they represent. Instead to give the pixels their correct positions relative to the ground, it is necessary to shift the bottom of the image to the west by the movement of the ground during image acquisition with all intervening lines displaced proportionately as depicted in Figure 3.1(b).

Further the amount by which the image has to be skewed to the west at the end of the frame depends upon the relative velocities of the satellite, earth and the length of the image frame recorded.

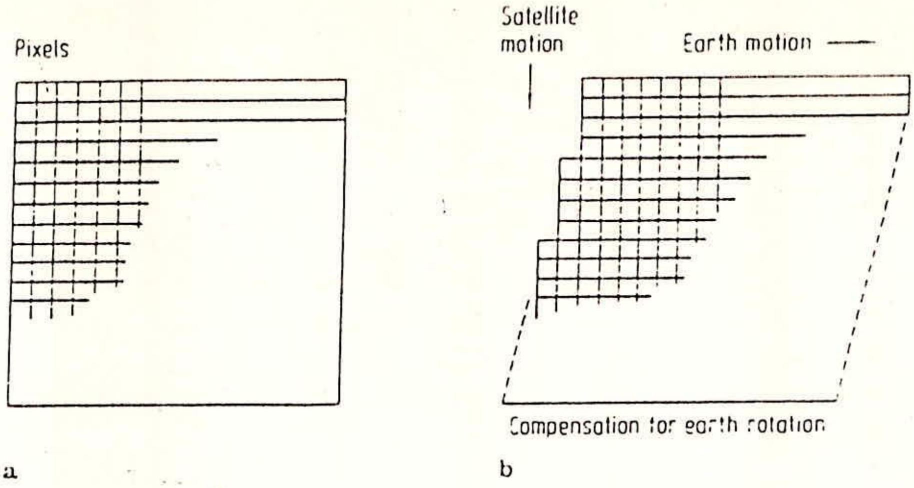


Figure 3.1: The effect of earth rotation on satellite Imagery

3.1.2 Panoramic Distortion

For scanners used on spacecraft and aircraft platforms the angular IFOV is constant. As a result the effective pixel size on the ground is larger at the extremities of the scan than at nadir

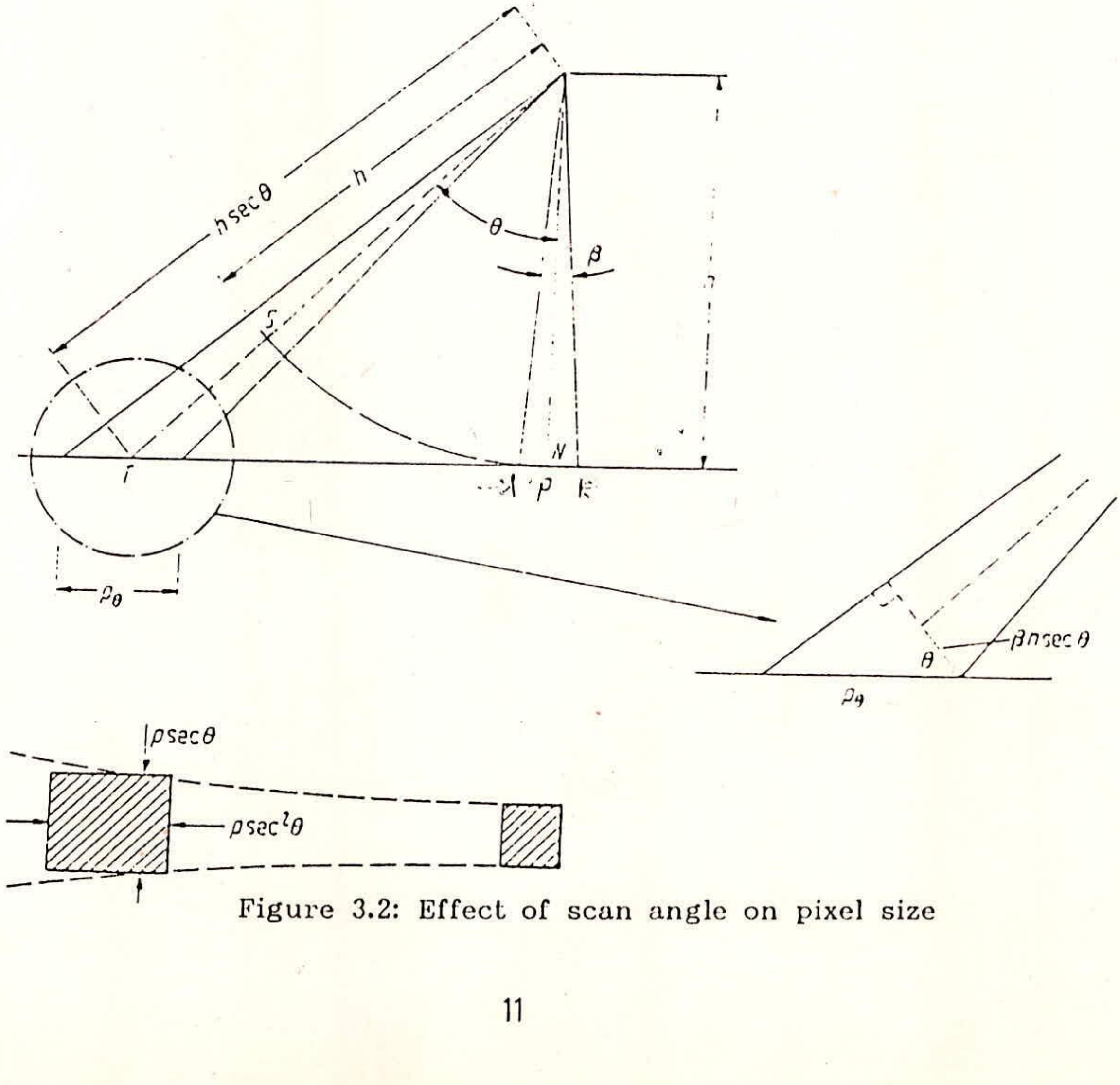


Figure 3.2: Effect of scan angle on pixel size

as illustrated in Figure 3.2. In particular, if the IFOV is β and the pixel dimension at nadir is p then, its dimension in the scan direction at a scan angle is θ is;

$$p_o = \beta h \sec^2\theta = p \sec^2\theta, \quad \text{where h is altitude.}$$

Its dimension across the scan line is p $\sec\theta$. For small values of θ, these effects are negligible.

3.1.3 Earth Curvature

Because of low altitude in the case of aircraft scanning system, the image data are not affected by earth curvature. Further, because of narrowness of swath in case Landsat and SPOT, these space system are not affected by earth curvature. However wide swath width space borne imaging system i.e. NOAA are affected. At the edges of the swath the area of the earth's surface viewed at a given angular IFOV is larger than if the curvature of the earth is ignored. The increase in pixel size can be computed by reference to the geometry of Figure 3.3. The pixel dimension in the across track direction normal to the direction of sensor is $\beta[h+r_e(1-\cos\emptyset)]\sec\theta$.The effective pixel size on the inclined earth surface is;

$$p_{-} = \beta\ \{h+r\ (1-\cos\emptyset)\}\sec\theta\ \sec(\theta+\emptyset)$$

$\beta[h+r_e(1-\cos\phi)]\sec\theta.$

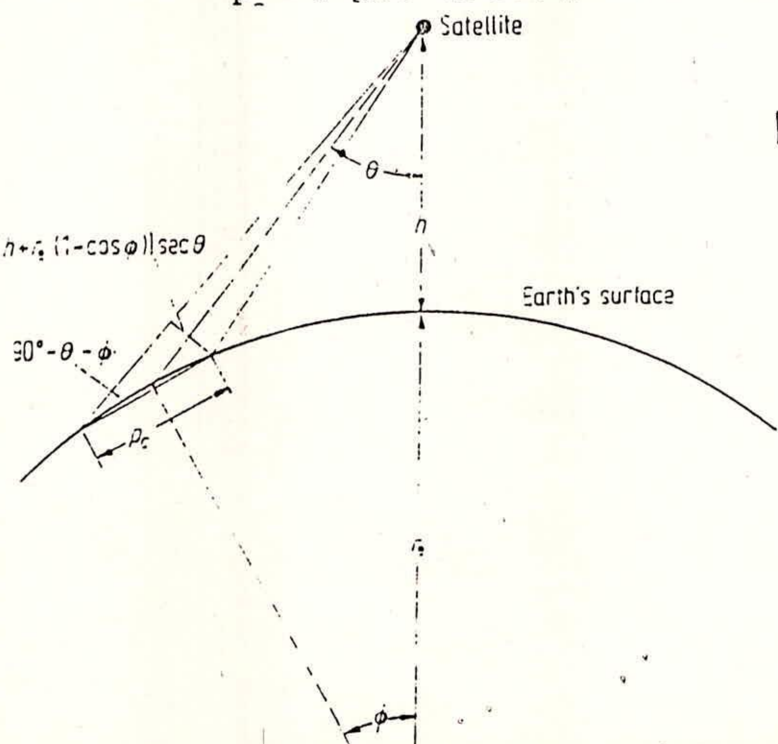


Figure 3.3: Effect of earth curvature on the size of a pixel

where βh is the pixel size at nadir and $\emptyset$ is the angle subtended at the centre of earth. It can be seen that the earth curvature introduces a significant additional compressive distortion in the image acquired by satellite such as NOAA when an image is constructed on a uniform grid. The effect of earth curvature in the along track direction is negligible.

3.1.4 Sensor Scan Nonlinearities

Line scanners that make use of rotating mirrors such as in the NOAA (AVHRR) and aircraft scanners have a scan rate across the swath, that speed is constant to the extent that the scan motor speed is constant. Systems that use an oscillating mirror, such as the Landsat MSS incur some nonlinearity in scanning near the swath edges owing to the need for the mirror to slow down the change directions.

3.1.5 Scan Time Skew

Mechanical line scanners such as the Landsat MSS and TM require a finite time to scan across the swath. During this time the satellite is moving forward leading to a skew image in the along track direction.

3.1.6 Variation in Platform Altitude, Velocity and Attitude

Variation in the elevation or altitude of a remote sensing platform lead to a scale change at constant angular IFOV and field of view. The effect is illustrated in Figure 3.4 for an increase in altitude with travel at a rate that is slow compared with a frame acquisition time. Similarly, if the platform forward velocity changes, a scale change occurs in the along track direction. This is depicted in Figure 3.4(b) again for a change that occurs slowly. For a satellite platform, orbit velocity variations can result from orbit eccentricity and the non sphericity of the earth. Platform altitude changes can be resolved into yaw, pitch and roll during forward travel. These lead to image rotation along track and across track displacement as noted in Figure 3.4 c–e. For the Landsat system, these corrections are applied before the data is distributed.

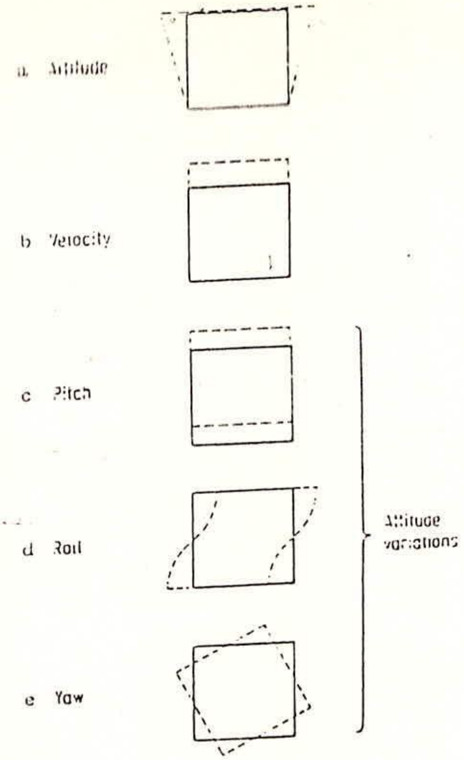


Figure 3.4 Effect of platform position and attitude on an Image

3.1.7 Aspect Ratio Distribution

The aspect ratio of an image is ratio of its vertical scale to the horizontal scale. Further, the aspect ratio of an image can be distorted by mechanism that lead to overlapping IFOV's. The most notable example of this occurs with the Landsat multispectral scanner. In this case samples are taken across a scan line too quickly compared with the IFOV. This leads to pixels having 56m centres but sampled with an IFOV of 79 m. Consequently, the effective pixel size is 79m x 56m and thus is not square. As a result if the pixels recorded by the multispectral scanner are displayed on the square grid, the image will be too wide for its height when related to the corresponding region of the ground

A similar distortion can occur with aircraft scanners, if the velocity of the aircraft is not matched to the scanning rate of sensor. Either under scanning or over scanning can occur and this lead to distortion in the along track scale of the image.

3.2 CORRECTION OF GEOMETRIC DISTORTION

Basically two techniques that can be used to correct the various types of geometric distortion present in digital image data. One is to model the nature and magnitude of the sources of distortion and use these models to establish correction formulae. This technique is effective when the types of distortion are well characterized such as that caused by earth rotation. The second approach depends upon establishing mathematical relationships between the addresses of pixels in an image and the corresponding coordinated of those points on the ground (via map). This procedure will be treated first since it is most commonly used and as a technique, it is independent of the platform used for data acquisition.

3.2.1 Use of Mapping Polynomials for Image Correction

An assumption that is made in this technique is that there is available a map of the region corresponding to the image, that is correct geometrically. Two cartesian coordinate systems well be defined, one describes the location of points in the map (x,y) and the other defines the location of pixels in the image (u,v). Now suppose these two coordinate systems can be related via a pair of mapping function f and g; so that

$$u = f(x,y) \quad (3.1\ a)$$

$$v = g(x,y) \quad (3.1\ b)$$

If these functions are known then we could locate a point in the image knowing its position on the map. With this ability we could built up a geometrically correct version of the image in the following manner.

Firstly, we define a grid over the map to act as the pixel centres in the corrected image. This grid is parallel to or indeed could in fact be the map coordinates grid described by latitude and longitude. For simplicity we will refer to this grid as display grid and by definition this is geometrically correct. We then move over the display grid pixel centre by pixel centre and use the mapping function above to find out the corresponding pixel in the image for each display grid position. These pixels are then placed on the display grid. At the conclusion of the process we have a geometrically correct image built up on the display grid utilizing the original image as a source of pixels.

While the process is a straight forward, there are some practical difficulties that must be addressed. Firstly, we do not know the explicit form of mapping function of equation 3.1. Secondly, even if we did, they may not point exactly to a pixel in the image corresponding to a display grid location. Therefore some form of interpolation may be required.

3.2.1.1 Mapping Polynomials and Ground Control Points

Because explicit form of mapping function in equation 3.1 are not known. They are generally chosen as simple polynomials of first, second or third degree. For example, in the case of second degree (or order)

$$u=a_0 + a_1x +a_2y +a_3x\ y + a_4x^2+ a_5y^2 \quad (3.2\ a)$$

$$v=b_0 + b_1x +b_2y +b_3x\ y + b_4x^2+ b_5y^2 \quad (3.2\ b)$$

These polynomials may be used if the coefficient a_i and b_i were known. However these coefficients are unknown and values can be computed by identifying sets of features on the map that can also be identified on the image. These features often referred to as ground control points (G,C,P). These are well defined and spatially small and could be road intersections, airport runway intersection, bends in rivers, prominent coast line features etc. Enough GCP may be chosen so that the required polynomial coefficient may be estimated.

3.2.1.2 Resampling

Having determined the mapping polynomials explicitly by use of the ground control points, the next step is to find points in the image corresponding to each location in the pixel grid previously defined over the map.

3.2.1.3 Interpolation

As is to be expected centres from the map registered pixel grid will not usually project to exact pixel centre locations in the image or shown in Figure 3.5 and some decision has to be made about what pixel brightness value should be chosen for placement on the new grid. Three techniques can be used for this purpose.

Nearest neighbourhood resampling simply chooses the actual pixel that has its centre nearest the point located in the image. This pixel is then transferred to the corresponding display grid location.

Bilinear interpolation uses three linear interpolations over the four pixels that surround the point found in the image corresponding to a given display grid position. Two linear interpolations are performed along the scan line and final one across the scan line to find out the brightness of the required pixel.

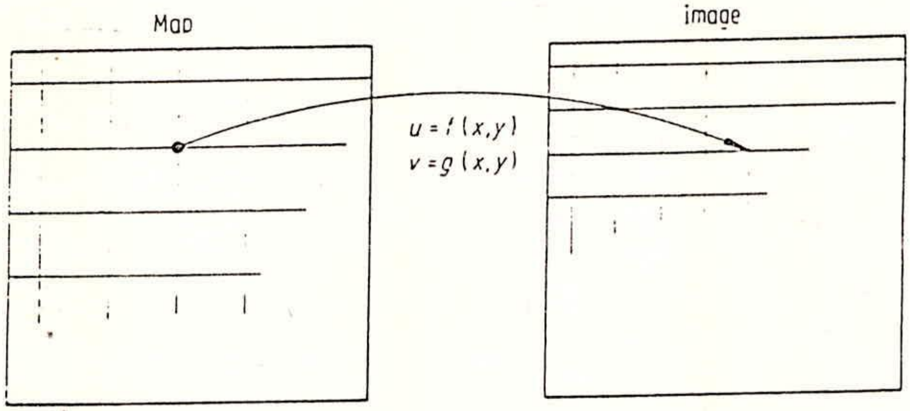


Figure 3.5: Use of mapping polynomials to locate points in the Image

Cubic convolution interpolation uses the surrounding sixteen pixels. Cubic polynomials are fixed along the four lines of four pixels surrounding the point in the image. A fifth cubic polynomials is then fitted through these to synthesize a brightness value for the corresponding location in the display grid.

3.2.2 Mathematical Modelling

If a particular distortion in image geometry can be represented mathematically then the mapping function in equation 3.1 can be specified explicitly. This obviated the need to choose arbitrary polynomials as in equation 3.2 and to use the control points to determine the polynomial coefficients. However, rather than commence with expressions that relate image coordinate (u,v) to map coordinates (x,y) it is probably simpler conceptually to start the other way around i.e. to model what the true (map) positions of pixel should be given in an image. This expression can then be inverted if required to allow the image to be resampled on to the map grid.

4.0 RADIOMETRIC CORRECTIONS

Radiometric errors are the type of systematic errors that are geometry independent and relatively constant. They can be modelled using deterministic models, because the sources/causes of the errors are known;

Radiometric errors can be grouped by their sources

- mechanical/electrical
- Atmospherical

4.1 MECHANICAL/ELECTRICAL ERRORS

Mechanical or electrical errors are caused by deficiencies in the satellite or at the ground receiving stations. Following errors may be caused due to these deficiencies;

Dropped lines: causes by temporary failure of a part of the satellite system or ground receiving station.

Striping: Variation of the offset and gain of each sensor. A systematic pattern will be visible in the image.

The corrections for these errors are often referred to as sensor corrections.

4.1.1 Dropped Line Replacement

Data correction for temporary loss of signal over part of a scan line is a cosmatic operation. One can not reconstruct the data, unless the data is redundant. In general the correction is a replacement by data from surrounding elements. For this Various techniques can be applied:

4.1.1.1 Replacement of a partly dropped line by data from the line above or below

10	13	20	21	23	20	19	18	20
10	0	0	0	0	0	0	0	19
9	12	19	21	22	19	18	17	15

10	13	20	21	23	20	19	18	20
10	0	0	0	0	0	0	0	19
9	12	19	21	22	19	18	17	15

10	13	20	21	23	20	19	18	20
10	13	20	21	23	20	19	18	20
9	12	19	21	22	19	18	17	15

10	13	20	21	23	20	19	18	20
9	12	19	21	22	19	18	17	15
9	12	19	21	22	19	18	17	15

Figure 4.1 Replacement by data from the line above or below

4.1.1.2 Replacement of a partly dropped line by the average of the line above and below

10	14	13	13	15	17	12	13	12
10	0	0	0	0	0	0	0	10
12	16	15	14	15	13	12	13	10

10	14	13	13	15	17	12	13	12
10	15	14	13	15	15	13	11	10
12	16	15	14	15	13	12	13	10

Figure 4.2 Replacement by the average of the line above and below.

These replacement techniques are based on the assumption that neighboring image elements are correlated.

4.1.2 Striping Correction

Striping means that there is no total loss of the signal but there are differences between the output from the different sensors to the same input. Originally the sensors were calibrated to give the same output for the same input, But shortly after launching differences in the input output function arise.

Sometimes, one can use calibration data for correcting the striping, else the parametric or non-parametric destriping technique can be used.

4.2 ATMOSPHERICAL ERRORS

Atmospherical errors are caused by the interaction (scattering or absorption) of photons with the dust particles in the earth atmosphere. They can have an additive effect (e.g. haze) or an multiplicative effect (e.g. skylight).

The illumination model is;

$D = (\Theta * T + \Theta * T) * R + H$

where D is detected by sensor, Θ is influence sun angle

T is the actual radiance from the sun

$\Theta * T$ is skylight

R is the reflectance property from the earth surface, and

H is haze

$D = (\Theta + \Theta) * R * T + H$

Θ: sunangle dependent attenuation, band independent

Θ: sunangle dependent, band dependent attenuation

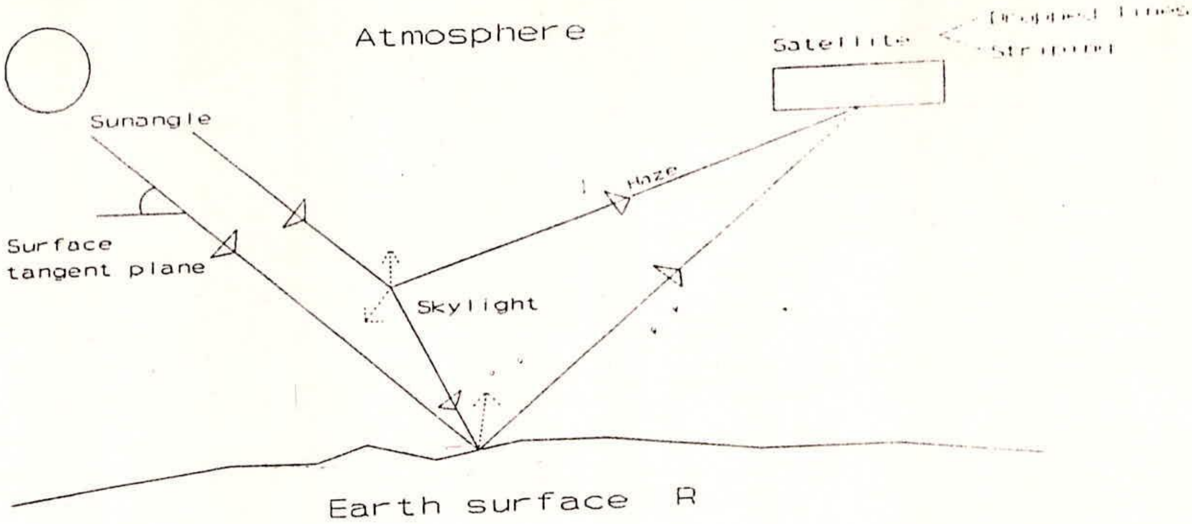


Figure 4.3 Graphical representation of the illumination model

4.2.1 Haze Correction

Haze correction is infact done by subtracting the estimated additive effect. The effect is stronger for high energy photons.

For all bands i,

OUTPUT'(i) = OUTPUT(i) - Haze (i)

Several techniques are used to estimate Haze;

Find samples in the subwindow which should have values equal to 0 because of their spectral properties i.e; total absorption. For example clear water bodies should have value 0 in the near infrared band. The actual value of the elements is the added Haze (i).

Use single band histograms. The additive effect can be estimated by calculating the minimum value in the histogram -1 (or -2).

Haze(i) = [Min(i) -1]

One thing must be keep in mind that the techniques to estimate Haze (i) are in fact based on the availability of 'black' elements in the scene.

The haze correction is applied to the total scene (global) where as it can be a local additive effect.

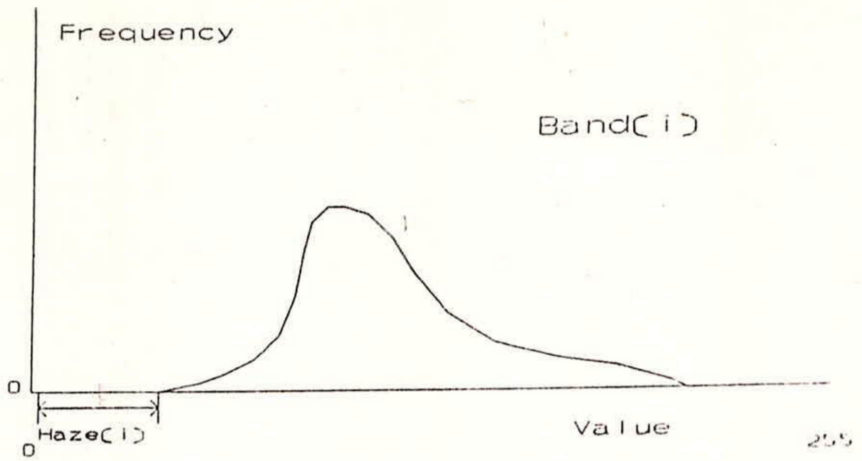


Figure 4.4 Example of the additive effect of haze on detected value

4.2.2 Skylight Correction

Skylight is in principle mainly a multiplicative effect because it influences the incoming radiance on the surface. It is caused by scattering of photon in the atmosphere on their way to the earth surface.

Correction is important for multitemporal comparison of images and for leaf area index classifications using normalised data.

For skylight correction one of the bands, from one of the dates should be used as reference. In general the near infra red band is least influenced by atmospheric effect and therefore will be the best reference. The actual effect of skylight depends on the band. Therefore the multiplicative skylight depends mainly on the relation between the band of interest and the reference band.

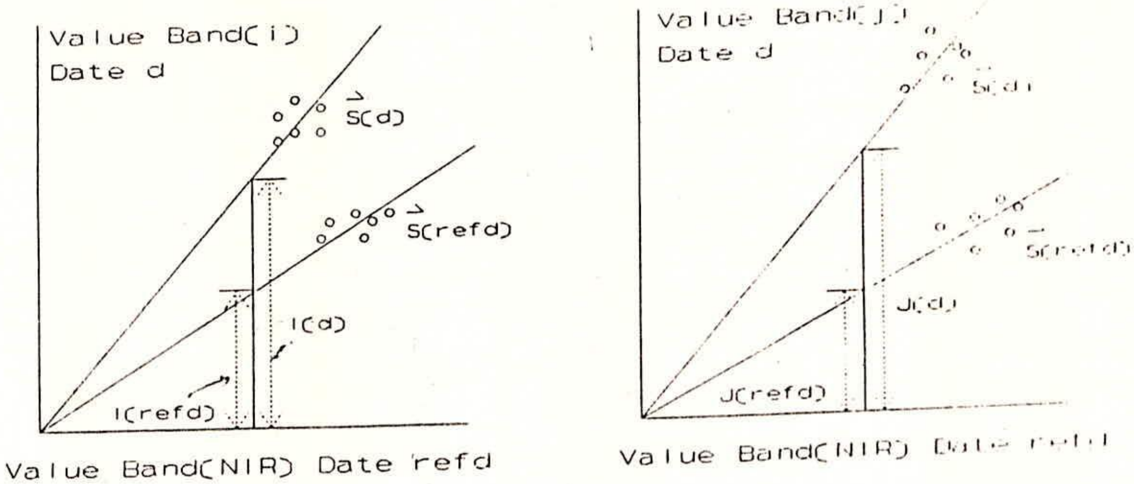
For reference band (r) and slave band (i)

OUTPUT'(i) = Θ (i,r)* OUTPUT(i)

and Θ is the skylight factor

For skylight correction one needs samples from reference ($\bar{s}$) from both dates. Estimation of Θ(i,r) requires samples of reference materials with known reflection coefficient per normalised band. Θ(4,7) can be found by ranking the angles between the two regression lines, simplified by ratioing the vertical perpendicular.

As reference materials one should use features, whose reflection parameter do not change with time. For example man made object i.e. concrete roads and parking places or bare soils.



$$I(i,NIR) = \frac{I(refd)}{I(d)} \qquad J(j,NIR) = \frac{J(refd)}{J(d)}$$

Figure 4.5 Graphical representation of skylight correction

They can be extracted from equal date ratios. Using these reference means correcting the skylight at t_1 such that the skylight conditioning are the same as at t_o.

4.2.3 Sunangle Correction

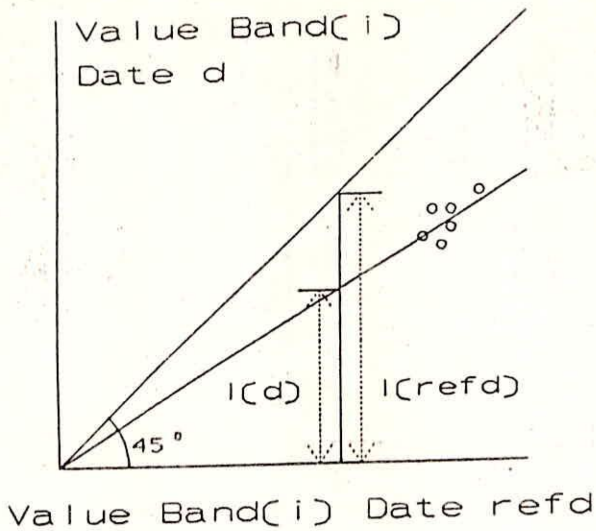
Sunangle differences have also a multiplicative effect. A lower sunangle causes more photons to be absorbed or scattered due to a longer route through the atmosphere. Correction for the variation of the sunangle is essential in change detection of un-normalised multitemporal data.

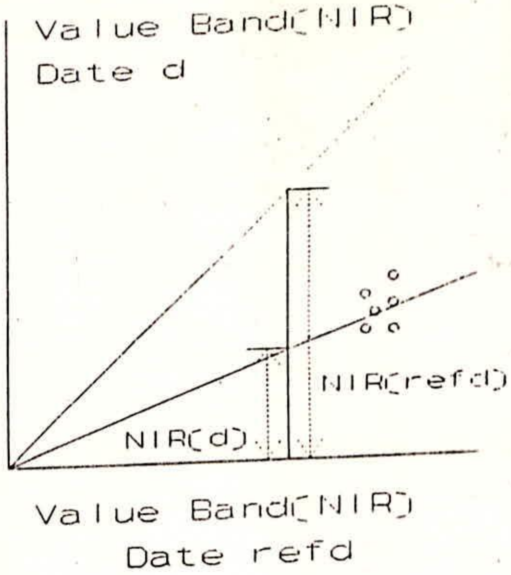
For reference date t_o Band (i)

$OUTPUT'(i,t_1) = \Theta(i) * OUTPUT\ (i,t_o)$

where Θ is the sunangle correction.

In first approximation $\Theta(i)$ is the same for all i, unless a red sunset or skylight plays an important role.





$$I(i) = \frac{I(refd)}{I(d)} \qquad J(i) = \frac{J(refd)}{j(d)}$$

Figure 4.6 Graphical representation of sunangle correction

One needs one set of stable samples on both dates. In first approximation these samples should have a regression line at 45^{o}. Estimation of $\Theta(i)$ can be accomplished by plotting reference samples in two dimensional features space with on the axes the same bands at different dates. The reference is now β=1. Again the correction factor is a ratio between the angles of the reference and the regression line from the samples, e.g. ratio of the vertical perpendicular.

5.0 COLOUR CODING TECHINIQUE AND SPECTRAL FEATURE EXTRACTION

Colour is perception by the human eye and interpretation by the human mind of incoming electro magnetic energy packages. Photon with a different energy level are perceived by humans as different colours. The sensitivity of the human eye for electro magnetic spectrum is limited to a narrow band in the total spectrum. This part of the band is known as visible light spectrum. Photon detectors can be build to detect even unvisible photons (near infra red, infra red, gamma etc.).

The special equipment onboard satellites make it possible to detect and count photons of different energies in narrow electro magnetic bands. Each sensor for each colour has a specific relation between incoming and counted photons. This relation can be represented as;

$$n_{s,c} = \lambda_{s,c} * N_c + \alpha_{s,c}$$

where λ is gain (multiplication factor ≤ 1)

α is offset (dark current, output at zero input)

s,c are sensor and colour indicator

5.1 In science several colour systems are used, created and named, but they can be subdivided into two classes;

The subtractive colour system

The additive colour system

5.1.1.The Subtractive Colour System.

The subtractive colour system is based on absorption (subtraction) of incoming photons, like in printing. One starts with an object which reflects all visible photons, e.g. white paper. By printing in blue on this paper one subtract in fact yellow (the complementary colour of blue) from the incoming radiance, only the blue photons are reflected. All other photons are absorbed. The text from this lecture note is written in black, all visible photons are absorbed.

5.1.2 The Additive Colour System

The additive colour system can be compared with a colour display. If the screen is not bombed with electrons it will be black, no photons will be emitted. If the red phosphor is activated by electrons the screen will emit red photons.

Activating also the green phosphor at the same location will cause the screen to emit red photons and green photons. This is perceived as yellow. Because of the small size of the phosphor dots, the output of the two fuses together when viewed form a distance. The additive colour system is thus based on additive, emitted radiance, and looks intuitively with the tri stimulus theory of colour vision. Because CRT display's are common equipment in digital image processing our main concern will have the additive colour system. Two additive systems are the Red-Green-Blue (RGB) and the Intensity-Hue-saturation (IHS) system

5.1.2.1 The RGB colour system

With the colours red, green and blue one can create many visible colours. We look at this colour system as a three dimensional orthogonal system. Orthogonal because all three colours are independent from each other (Figure 5.1).

The distance from the origin in either one of the axes is called the intensity of that specific colour. If the maximum intensity for red, green and blue are same, the total system is defined by a colour cube. On the diagonal from black to white one can find different tones of grey.

CRT displays uses the RGB colour system for displaying colour images.

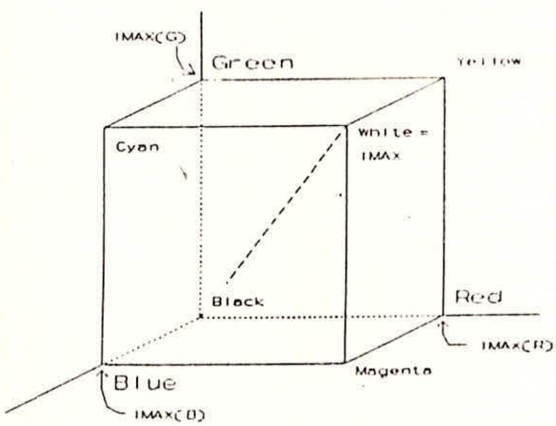
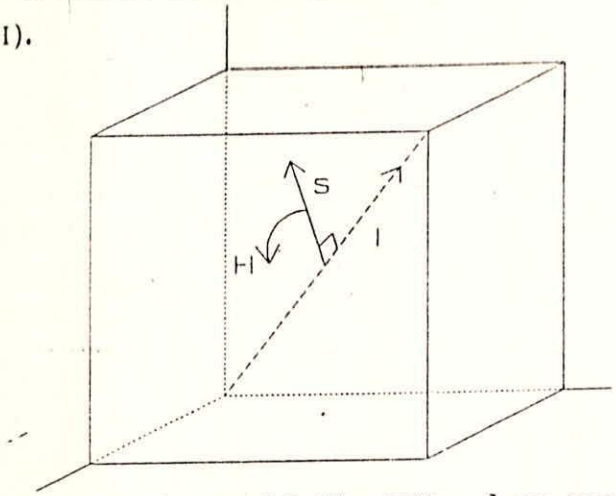


Figure 5.1 The orthogonal colour system

5.1.2.2 The IHS Colour System

The Intensity-Hue-Saturation (IHS) system is somewhat difficult to understand and manipulate because it uses two orthogonal axes and an angle to describe the three dimensions in colour space (Figure 5.2). The two axes are the intensity axis and saturation axis. The angle lies in the plane of saturation, which is orthogonal to intensity and it refers to the Hue. Sometimes they are referred to as brightness (I or B) colour purity (S) and colour (H).



I=Intensity
S=Saturation
H=Hue

Figure 5.2 The IHS colour system

5.2 COLOUR CODING TECHNIQUES

The first step in image processing has an input radiance value for a certain spectral band and as output grey tones (black and white images). This is an enhancement compared to a listing of actual values. One can not create new information via colour coding, the enhancement technique has always been latent. The following chapter will explain some colour coding techniques for image enhancement.

5.2.1 Linear Colour Coding

In linear colour coding there is a linear relation between input (radiance) values and output (colour) intensity. If the data is scaled between 0-255 and the output range is also from 0 to 255, the function will look as in Figure 5.3.

In general the range of input will be a subrange of 0-255 and the range of the output depend on the type of output device. Applying the above described linear function on such a data set will result in a very low contrast image, because only a very limited subrange of the output range is used (Figure 5.4). To solve this problem, one should know the MININ and MAXIN and calculate a new function

[MININ - MAXIN] ⇔ [MINOUT-MAXOUT]

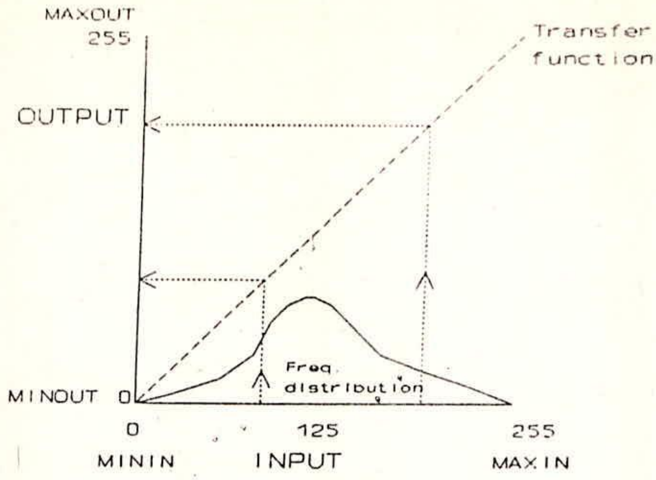


Figure 5.3 Linear colour coding of scaler data

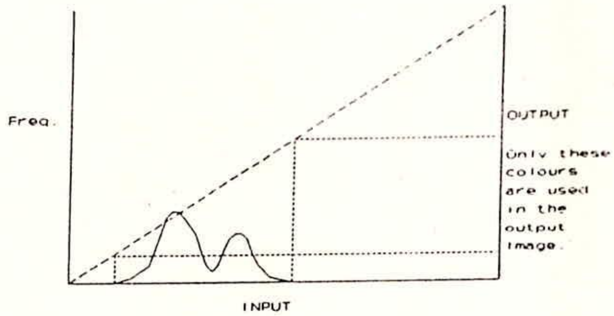


Figure 5.4: Linear transfer function over the whole input image

The MININ should be coupled to MINOUT and MAXIN to the MAXOUT to get a maximum contrast in the image. The couples between INPUT and OUTPUT can be stored in a LUT.

The procedures to take MININ and MAXIN is very sensitive for outliers. A common solutions to take the input values at 1% and 99% in a cumulative histogram. Off course one can adopt these percentage or values, depending on the circumstance (e.g. clouds) or interest (e.g. land cover or a specific subarea).

5.2.2 Piece Wise Linear Colour Coding

As in linear colour coding there is a linear relationship between input and output values, but the linear relation can vary with the input data interval. This extended linear colour coding is extremely helpful in case of a limited interest.

For example, the interesting intervals are 50-80 and 110-135, the two main peaks in the histogram. Via linear colour coding

(Figure 5.5[1]) one would loose contrast due to colours that are used for the intermediate interval (81-109=30%) of the grey tones.

Piece wise linear colour coding would change the LUT according to the Figure 5.5[2] and will result in the highest possible contrast in the two interesting intervals.

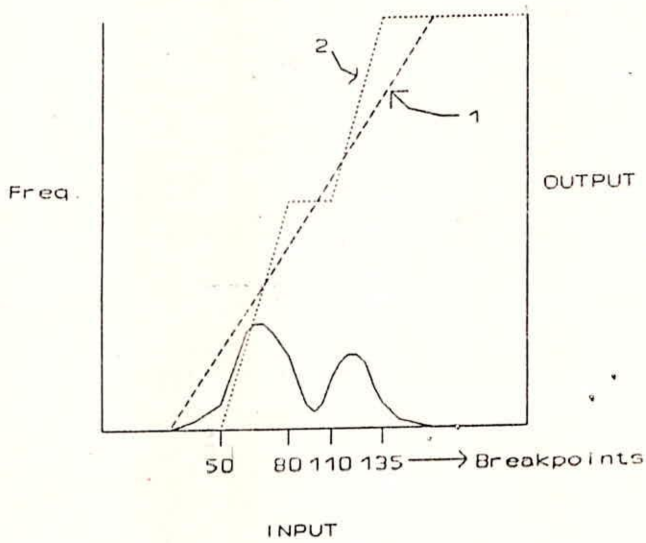


Figure 5.5: Piece wise linear transfer functions

The points at 50, 80, 110, 135 are refereed to as break points of the transfer function. Any amount of break points can be used, provided that each point input value has only one output value.

5.2.3 Equalisation of Histogram

The colour coding technique maps the cumulative histogram into a linear cumulative histogram (Figure 5.6).

Because in general histograms of remote sensing data do have a sort of gaussian distribution and lower radiance intervals. The contrast for the centre interval will be enhanced.

If the input data is changed according to the LUT after histogram equalization, the histogram of the output will change (Figure 5.7).

Application of histogram equalisation is often required when the output intensity is limited to a few levels (e.g.printers or plotters).

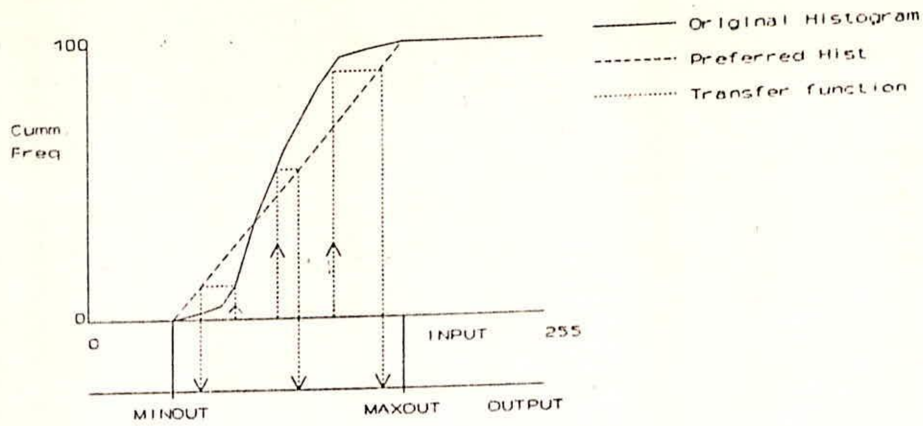


Figure 5.6 Histogram Equalisation

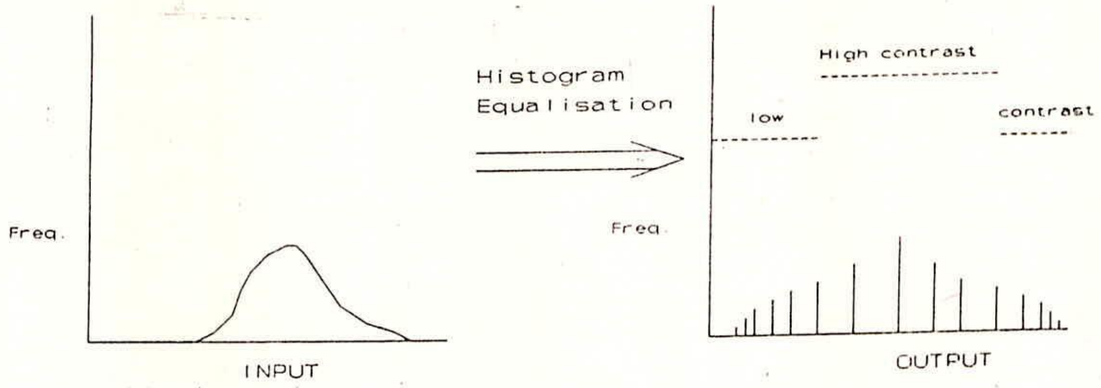


Figure 5.7 Effect of Histogram Equalisation on the Data Values

5.3 SCALER DATA INTO MORE COLOURS

5.3.1 Pseudo Colour Coding

Scalar data can also be represented in several colours, for example using the colours from the colour cube. In general one should understand that in the three dimensional colour space, you will have variations in three different ways; Intensity, Hue and saturation. Mapping scalar and ordinal data into this three dimensional space must be performed by keeping two dimensions constant and varying the other proportionally.

An obvious choice is to keep intensity constant, saturation (nearly) constant and vary the Hue (apparent colour). The ideal pseudo colour relation between Hue and pixel value is based on a circle in a plane of constant intensity

I = constant, S= Constant, H= Variable

The distance to the centre is constant. It can be proven that the relation between pixel value and R, G and B should have a sine form, but have a different phases (phase shift =120 degrees, Figure 5.9)

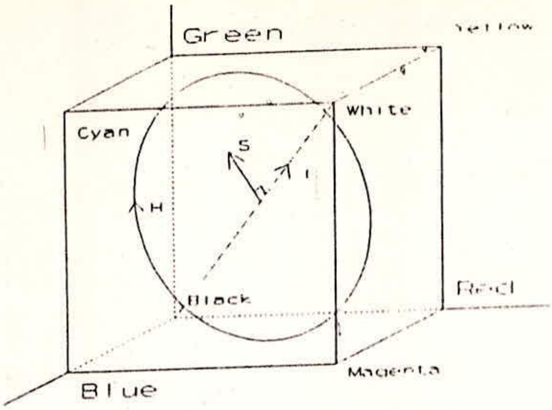


Figure 5.8 Ideal mapping in Pseudo colour.

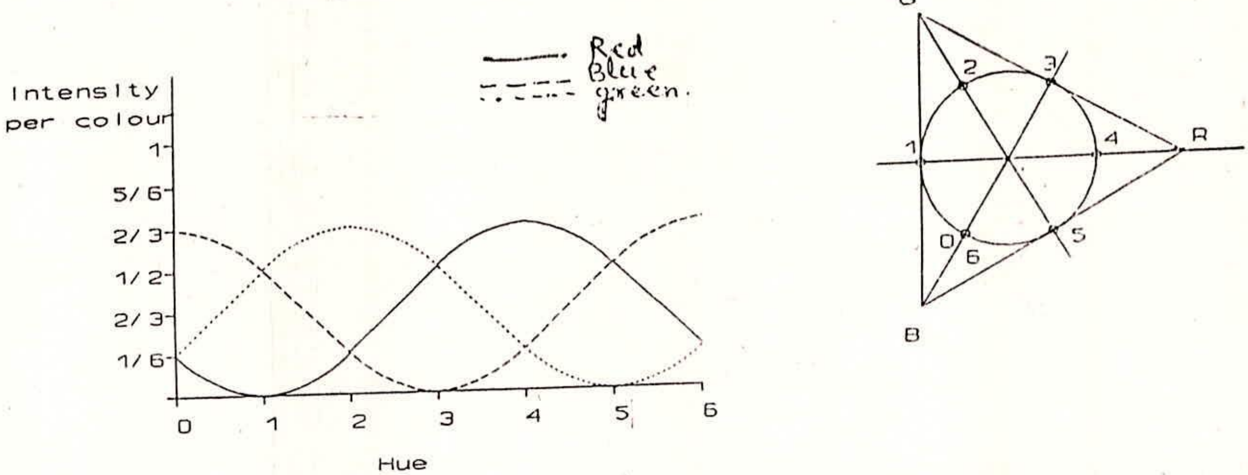


Figure 5.9 Transfer functions for ideal pseudo colour coding

A method which approaches the ideal circle is a hexagonal in a plane of constant intensity (Figure 5.10). However the saturation is not constant any more as the distance to the centre for points 1 mod(2) is smaller than for point 2 mod(2). The relation between pixel value and R, G and B are shown in Figure 5.11. It is a linear approximation of the ideal sine. These linear functions can easily be stored in a LUT.

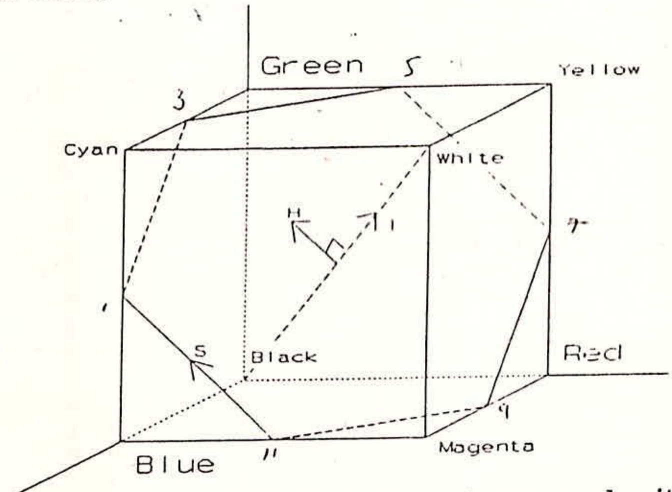


Fig.5.10: Pseudo colour coding using the hexagonal with constant I.

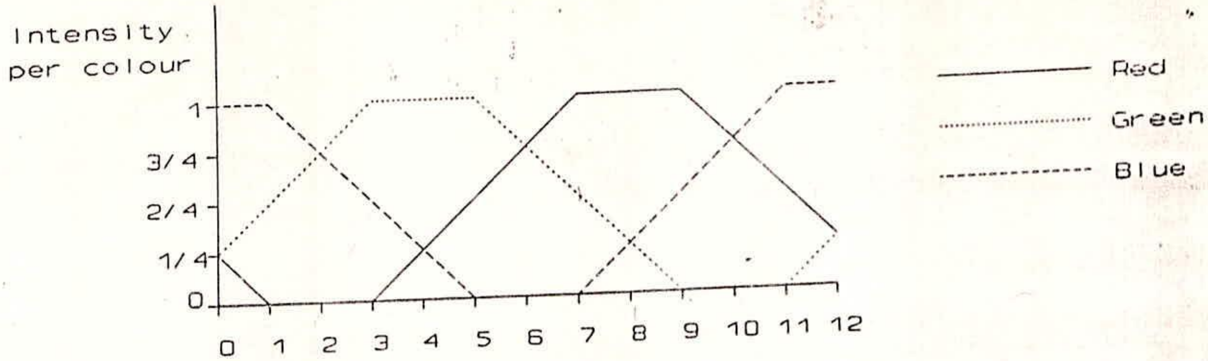
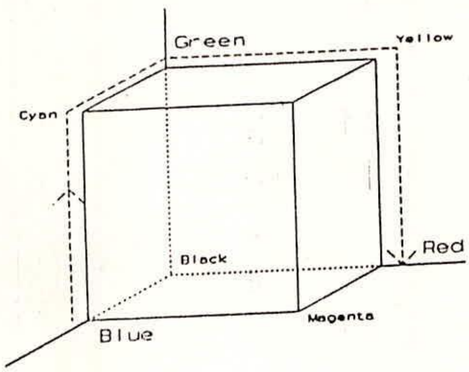


Figure 5.11: Transfer functions for pseudo colour coding via hexagonal

In fact one can easily create its own LUT, as long as the principles of constant Intensity and Saturation are preserved. Examples:Going around the 4 sides of the colour cube (Figure 5.12)



I varies (1,2)
S varies (50%, 100%)

Figure 5.12 Simple pseudo colour coding using 4 sides of the colour cube.

5.4 SPECTRAL FEATURE EXTRACTION

Feature extraction is an important step in digital image processing. In fact one should think about it before remote sensing data is ordered, e.g; what is the use of LANDSAT MSS data if features should be detected with an area of 400 m^2, what is the use of SPOT XS data if algae in the ocean should be studied. The first example is in fact referring to spatial feature extraction. Spectral feature extraction is extraction of features based on their spectral properties.

Here, I would like to discuss various techniques of feature extraction along with their application.

In general spectral feature extraction is based on transformations of the input data. To understand the effect of transformations one should carefully follow the effect on the vectors in data space. Here we will discuss two types of spectral transformation;

Commonly used spectral transformation

- Addition and subtraction
- Ratio and multiplication
- Sum normalization
- Vector normalization
- Principle components

Special spectral transformation are

- RGB to MMI and v.v.
- MMI to HSI and v.v.
- Leaf Area Index (LAI)
- Hill shading
- Spectral correlation

Spectral correlation can also be treated as a special colour coding technique using spectral features characteristics.

5.4.1 Commonly used spectral transformations

5.4.1.1 Addition and subtraction

Addition is a technique in which one adds the values of the elements from different bands to each other.

$B_{57} = (B_5 + B_7)/2$

where 5 is LANDSAT band 5 (Red)

7 is LANDSAT band 7 (NIR), and

2 is a compensation factor

To preserve on output interval of [0,255], the output should be divided. Empirically the compensation factor can be equal to the number of bands used in addition. The result is in fact an average of the input bands, which compensates the random noise component in the input data.

On the other hand, subtraction is a technique which can be used for detecting differences between image or bands. E.g. subtracting bands from different dates from each other;

$B_{d5js} = (B_{5j} - B_{5s}) + 127$

where j and s stands for January and September and 127 is a compensation factor.

Because satellite data is a representative of photon count per image element the images should not only be registered very well (geometric corrections) but also all appropriate radiometric corrections (e.g. sun angle) should be applied. Even then one should be interest in trends not in actual differences in photon count. The compensation factor is necessary because the differences can be negative.

A special application of addition and subtraction is the estimation of the leaf area Index (LAI).

5.4.1.2 Ratioing and multiplication

Ratioing or multiplication of spectral bands will not have any meaningful interpretation because we are working in a domain of photon counts. Ratioing is a trick which provides the user a wide range of possible outcomes with a range from 0 to infinity $(0,\infty)$

$$B_{r57} = (B_5/B_7)*\ 90$$

Where 5 and 7 are Landsat band 5 and Band 7 and 90 is compensation factor.

The compensation factor is needed because the ratio can be smaller than one. In the case of two band, the two ratios will map the input data on a hyperbola, so there is only one degree of freedom left.

Application consist of taking many different ratios and use them in, for example, a principle component analysis. One should always bear in mind that an improper radiometric correction e.g; haze correction, will not only influence the result of the ratio but even determine it.

5.4.1.3 Sum normalisation

Sum normalisation is used to make the input data independent of the influence of external factors such as sunangle or relief. The addition or summation of input data is a representation of this influence. The ratios between the individual input bands and this sum, results in intensity independent data. If the results are displayed they appear to be flat.

The results of normalisations always within the interval [0,1], therefore it should be multiplied with a compensation factor.

5.4.1.4 Vector normalisation

Vector normalisation is a ratio between the individual input values (scalers) and the vector length, Vector length is defined as

$$\|B\| = \sqrt{\sum B_i^2}$$

As B_i consist of photon counts in different spectral bands, Vector length has no physical meaning.

In case of two spectral bands, vector normalisation maps the input data on a sphere.

5.4.1.5 Principle Components

The principle component transformation, also referred to as the Karhunen-Loeve or KL Transformation, is a data ordering technique based on Variance. The transformation calculates a new orthogonal reference system based on the covariance matrix of a sample set (Figure 5.14). The new orthogonal reference vectors, and the eigen vectors are produced in the order of the variance they explain. If low variance PC's have the order of noise variance then these PC's can be removed from the data set. The data can be reconstructed without the noise and redundent information .

An important problem, when working with principle components is that variance by no means is equal to the information. Thus, the eigen vectors do not have any physical interpretation. Colour coding based on principle components is leading to uninterpretable images.

Another problem occurs in applications where the first PC' is always considered as the total intensity, because it contains "often" only positive factors.

One valid application is a redundancy estimation and removal, after sum normalisation. As PC's transformation is a rotation and shift of the vector base and low variance is collected in the tail

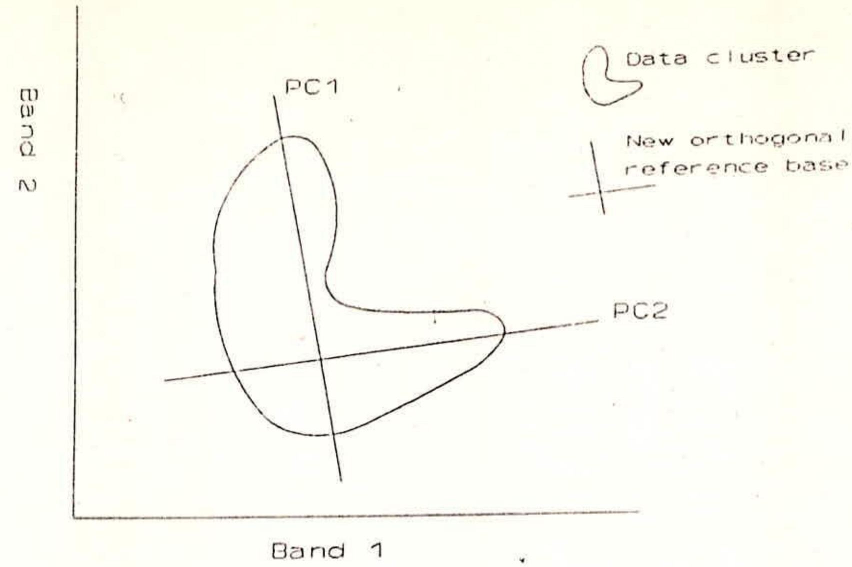


Figure 5.14: Two dimensional histogram of a general Remote sensing data set

of the PC's then it is possible to remove these low variance PC's. An even better approach is a combination of proper spectral feature extraction followed by a PC's transformation to check if any important physical features were disregarded.

5.4.2 SPECIAL SPECTRAL TRANSFORMATION

5.4.2.1 RGB to MMI transformation

In display domain the colour of each pixel is determined by the displayed intensity in each of the components of the primary colours i.e. Red, Green and Blue. Together these RGB Values can be treated as components of a colour vector in a three dimensional space.

Assuming that every colour has a range in intensity from say [0,100], then it is possible to treat the colour space as a colour cube. The colour space is a linear space in which the mixture of two additive colour is equivalent to a colour vector, a linear combination of the two mixing colour vectors.

The influence of intensity on the actual colour can be removed by sum normalisation. The effect of sum normalisation on the data is mapping the data onto a triangle spanned by maximum Red, Green and Blue. For all points within this triangle the total intensity is equal to the single colour maximum intensity i.e.100.

Although saturation and Hue can be used to describe every point within the triangle by using two orthogonal axes. An obvious choice is to have the origin of two orthogonal axes at the point where the intensity axes crosses the triangle plane. The two axes are coded M_1 and M_2. M_1 is arbitrarily chosen to go from the origin to Red and M_2 is orthogonal to M_1 thus parallel to the line from Green to Blue.

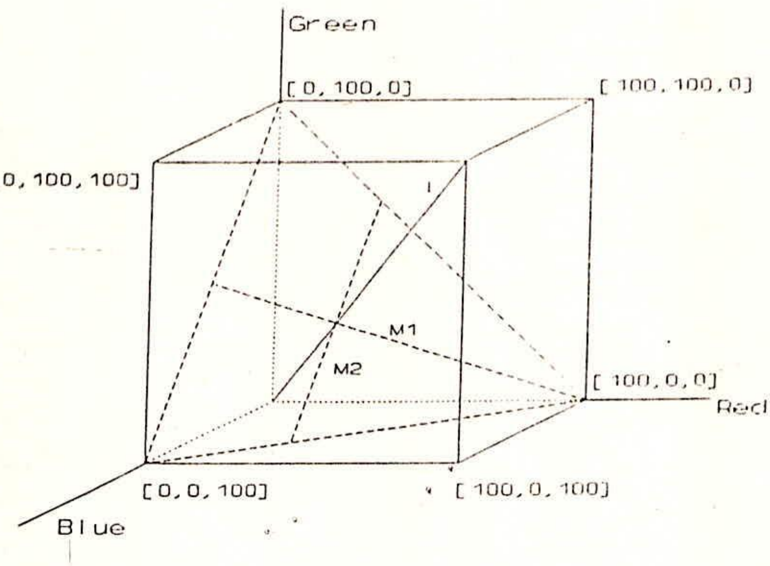


Figure 5.15 RGB colour cube with RGB colour triangle and orthogonal axes M_1 & M_2.

5.4.2.2 MMI to IHS transformations

A mathematically simple transformation is the one between MMI and HSI, because saturation is proporstional to the vector length in m_1 m_2 and Hue is related to the angle in m_1 m_2 and (Figure 5.16). So;

$$\text{Hue} = \operatorname{arctg}(M_2/M_1)$$

$$\text{hue} = \tan^{-1}(m_2/m_1)$$

$$\text{Sat} = \sqrt{(M_1^2+M_2^2)}$$

$$\text{sat} = \text{Sat}/I$$

$$\text{Int} = I$$

The inverse transformation are:

$$M_1 = \text{Sat}*\cos(\text{hue})$$

$$M_2 = \text{Sat}*\sin(\text{hue})$$

$$I = \text{Int}$$

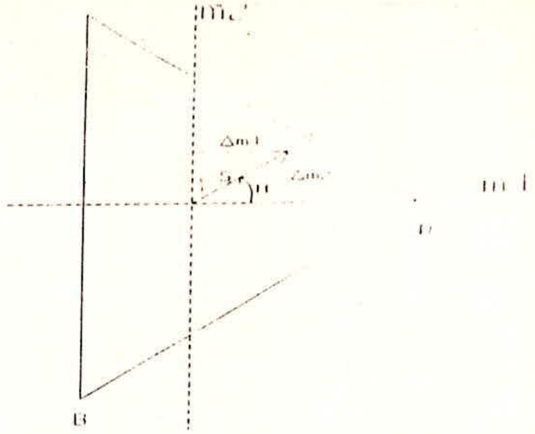


Figure 5.16: Hue and Saturation in m_1 amd m_2 space.

The MMI to IHS transformations and V.V. should not be explicitly used in digital Image processing because MMI has in fact the same physical meaning as IHS. Hue and saturation do not change when only the amount of light (I) Changes, thus I is extracted. Both IHS and MMI take into account the variation in Hue and saturation.

5.4.2.3 Leaf Area Index (LAI)

The LAI is the area of leaves when projected on the ground, under the constraint of a certain average leaf angle. This LAI is often confused with the general normalised green vegetation index (NGVI) based on the near infra red and red spectral channels from LANDSAT MSS data.

$$NGVI = (B_7 - B_5)/(B_7 + B_5)*90$$

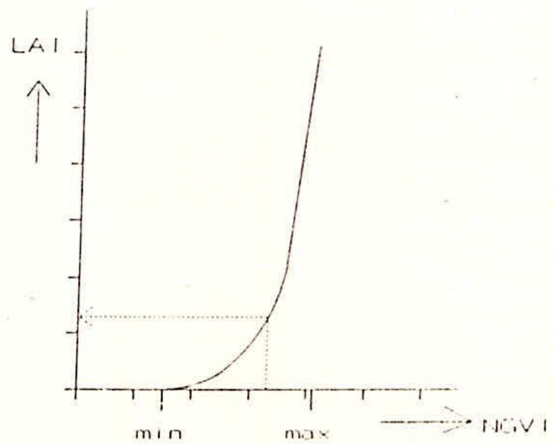


Figure 5.17 Exponential LUT for NGVI to LAI conversion

The LAI is now estimated from the NGVI. For practicals situations with leaf angles varying between o (horizontal) and 45^0 degree, we may assume a logarithmic mixture of a high LAI(6-8) vegetation canopy and bare soil in a sum normalised data space. Thus processing of the NGVI through an exponential look up table will result in a better estimation of LAI (Figure 5.17).

5.4.2.4 Hill Shading

Hill shading is a technique to visualise digital elevation data. It involves actually scene irradiance of a hilly terrain. One has already the digital elevation available but the proper way of representing has to be selected. Hill shading can be seen as a colour coding technique but if properly executed one will have a new feature to work with.

The digital elevations in general derived from point samples, profiling a contour map (iso lines). It is ordinal data and often displayed as contour lines. Coding of elevation in grey scales gives a strong effect of contouring if few steps are used. Pseudo colour coding is more pleasing and often used for cartographical purposes.

Hill shading requires first the gradient vector of the surface of which elevation gradient is to be determined. An additional synthetic file and a simple change of indices produces surface normal vectors. If the assumption of Lambertian surface reflectors is made, the inner product of a user specified sun direction vector with the surface normals will result in hill shading values.

5.4.2.5 Spectral Correlation

As correlation is the inner product of data vectors and reference vectors, spectral correlation is a method which uses spectral characteristics vector as a new vector base to decompose the data vector. For MSS remote sensing data 98% of the natural materials can be described by three reference vectors, i.e bare soil, dense vegetation and clear water. It is even possible to use the spectral correlation technique to map N dimensional data into these three reference vectors without loosing to much information. The data in this new system can easily be displayed (in a pseudo natural composit) 'on a graphical screen using RGB colours. Each reference vector can be seen as a filter. Thus three filters are sufficient to describe 98% of the natural variance. In case of traditional MSS data using band pass filters, the spectral data vectors represent sampled spectral data. It is not a continuous data set. Thus the integral needed for correlation become a sum in this sampled domain;

$B' = \Sigma B_i . R_i$ where,

Σ stands for sum $i = 1$ to N

B_i is the data vector

and R_i is the reference vector

This is interpreted as the inner product of vector B with reference vector R. The exact definition of correlation requires a normalisation with the vector norm of both vectors. Because vector norm has no physical meaning in the context of photon count, the inner product of the sum normalised data vector and the normalised reference vector will be used as a spectral correlation measure.

6.0 LOCAL OPERATORS AND FILTERS

Enhancement of an image can be done either by changing image contrast or by coding a black and white image in pseudo colour, so as to emphasize or amplify some property of the image that is of interest to the user. In each case, however, the digital values held in the image are unchanged; all that is done is to modify the look up tables in such a way that the physical appearance of the image on the monitor is changed. Method, for selectively emphasizing or suppressing information at different spatial scales contained in image data are described here. For example, we may wish to suppress the noise pattern caused by detector imbalance that is sometimes seen in Landsat MSS and TM images. On the other hand, we may wish to emphasize some feature or features of interest, such as curvilinear boundaries between areas that are relatively homogeneous in terms of their tone or colour, in order to sharpen the image and reduce blurring. In this technique the image data may be changed.

By analogy with the procedure used in the chemistry to separate components of a suspension, the technique is known as filtering. A digital filter can be used to extract a particular component from a digital image. The slowly varying background pattern can be envisaged as a waveform with a long wavelength or low frequency, hence a filter which separates this component from the remainder of the information present in an image is called a low pass filter. Conversely, the more rapidly varying detail is like a waveform with a short wavelength or high frequency, a filter to separate out this component is called a high-pass filter.

Two types of filter will be considered, one which extracts low-frequency information and one which extracts high-frequency information.

The low frequency information will allow the identification of the background pattern, and will look as if the detail has been smoothed or removed from the image. The high frequency information will allow us either to isolate or to amplify the local detail.

6.1 LOW-PASS FILTERS

Before the topic of smoothing of an image, a two dimensional image is considered. We will look at a simpler expression of the same problem, which is the smoothing of one dimensional pattern. Figure 6.1 is a plot between pixel value (Vertical axis) against position along a scan line of a image. The underlying pattern is partially obscured by the presence of local pattern and random noise. If the local variable and the random noise, were to be removed then the overall pattern would become more clear and a general description of the trends in the data would be more easily seen. The solid line in Figure 6.1 is a plot of the observed pixel values against position along the scan line, while the dotted line and the broken line represent the output from median and moving average filters, which will be described subsequently. Both filters produce smoother plots than the raw data curve, and the trend in the data can be seen more easily. Local sharp fluctuations are removed. These represent the high-frequency component of the data and may be the result of local anomalies or of noise.

6.1.1 Moving Average Filter

If the coordinate on the horizontal axis of Figure 6.1 is denoted by the index j, then the moving average filtered value at any point j is X'_j. The calculation of X'_j depends on the number of local values used, this number is always an odd positive integer. The broken line in Figure 6.1 is based on a five point moving average, defined by

$$X'_j = (X_{j-2} + X_{j-1} + X_j + X_{j+1} + X_{j+2})/5$$

Five raw data values (X) centered on point j are summed and averaged to produce one output value (X'_j).

A two dimensional moving average filter is defined in terms of its horizontal and vertical dimensions. Like the one dimensional moving average filter, these dimensions must be odd, positive and integer. However they need not be equal. A two dimensional moving average is described in terms of its size, such as 3x3. To begun with, the window is placed in the top left corner of the image to be filtered and the average value of the elements

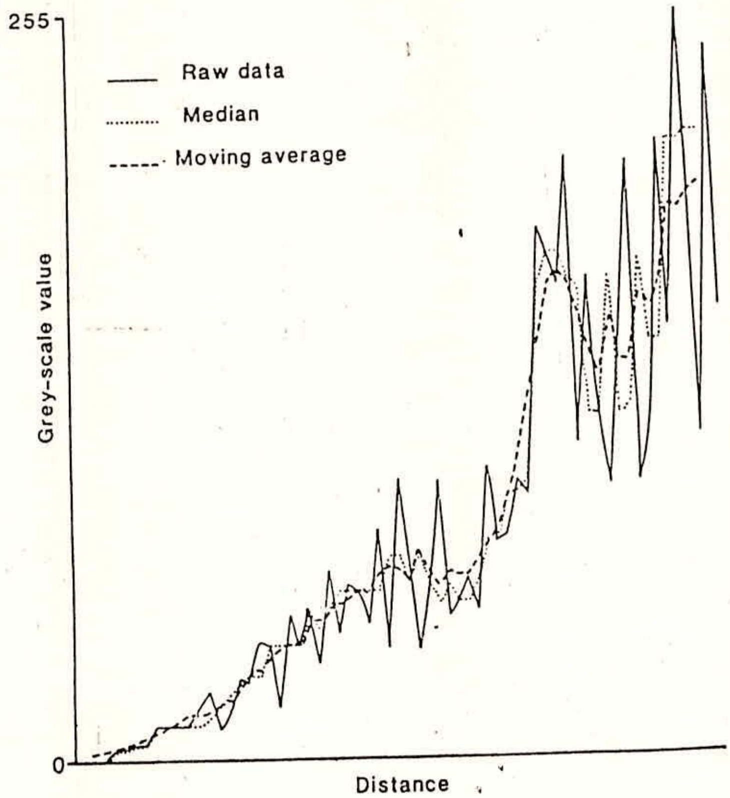


Figure 6.1: Effect of median filter and moving average filter

in the input image that are covered by window is computed. This value is placed in the output image corresponding to the central element of the window.

The unfiltered margin of the Image can be filled with zeros or the unaltered pixels from the corresponding pixels of the input image and can be placed there.

Generally, these missing rows and columns are filled with zeros in order to keep the input and output images of the same size. The effect of the moving average filter is to reduce the overall variability of the image and lower its contrast. At the same time those pixels that have larger or smaller values than their neighbourhood average are respectively reduced or increased in value so that local detail is lost. Noise components, such as the banding patterns evident in MSS images, are also reduced in magnitude by the averaging process, which can be considered as a smearing or blurring operation . Therefore, after filtering, the detail has been lost but the general pattern of variation is far clearer and easier to describe and interpret. The moving average filter described in this section is an example of a general class of filters called box filters.

6.1.2 Median Filter

An alternative type of smoothing filter utilizes the median of the neighbourhood elements rather than the mean. There are two reasons why the median is felt to be a superior choice. Firstly, the median of a set of n numbers is always equal to one of the values present in the set. Secondly, the median is much less sensitive to errors or to extreme data values as compared to average.

Therefore, isolated extreme pixel values, which might represent noise, are removed by the median filter. It also follows that the median filter preserve edges better than a moving average filter, which blurs or smooths the grey levels in the neighbourhood of the central point in the filter window. It is clear that, while both removes oscillations, the median filter removes isolated spikes more successfully and better preserved edges, defined as pixels at which the gradient or slope of greylevel value changes markedly. Synthetic aperture radar images display a noise pattern called speckle. This is seen as a random pattern of bright points over the image. The median filter has been used by a number of workers in order to eliminate this speckle without unduly blurring the sharp features of the image.

Increasing the window size for a moving average or median filter results in a greater degree of smoothing since more pixels are included in the window. From the above discussion, It can be concluded that the median filter might be preferred to the moving average filter for the reasons given earlier, it might be rejected if the computational cost was too high.

6.1.3 Adaptive Filter

Both the mean and the moving average filter apply a fixed set of weights to all areas of the image irrespective of the variability of the grey levels underlying the filter window. Several authors have considered smoothing methods in which the filter weights are calculated for each window position, the calculations being based on the mean and variance of the grey levels in the area of the image underlying the window. Such filters are termed adaptive filters. Their use is particularly important in the attenuation of the multiplicative noise effect known as speckle which affects synthetic aperture radar images.

6.2 HIGH - PASS FILTERS

The process of imaging or scanning involves blurring. High frequencies are suppressed relative to the low frequency components of the image. It would seem intuitively obvious that the visual quality of an image could be improved by selectively increasing the contribution of its high frequency components. A simpler way of performing a high-pass filtering is considered before the derivative-based methods are discussed.

6.2.1 Image Subtraction Method

An image can be considered to be the sum of its low and high frequency components. This high frequency and low frequency part of the image can be isolated by the use of a low-pass and high-pass filter as explained earlier. This low-frequency image can be subtracted from the original unfiltered image, leaving behind the high frequency component. The resulting image can be added back to the original, thus effectively doubling the high frequency component.

6.2.2 Derivative Based Methods

Other methods of high-pass filtering are based on the mathematical concept of the derivative. The derivative of a continuous function is the rate of change of the function at a point.

In terms of a continuous grey scale image, the first derivative of this function is the rate of change of greyscale value over space. Further, In those areas of Image that are homogeneous, all three derivatives will be small. Where, there is a rapid change in the greyscale values, for example at a coastline in a near-infrared image, the gradient of the image at that point will be high. These lines or edges of sharp change in grey level can be thought of as being represented by the high frequency component of the image as mentioned earlier, the local variation from the overall background pattern is due to high frequency components. The first derivative or gradient of the image, therefore identifies the high frequency portion of the image.

The second derivative of the image is known as Laplacian operator. It can be realised that the effect of imaging through the atmosphere and the use of lenses in the optical system is to diffuse the radiance emanating from a point source so that the image of a sharp point source, so appears as a circular blob. The subtraction of the Laplacian is equivalent to removing the diffused element of the signal from a given pixel. Another possible explanation is that the value recorded at any point contains a contribution from the neighbouring pixels. This is not an unreasonable hypothesis, the contribution could consist of the effect of diffuse radiance, that is, radiance from other pixels that has been scattered into the field of view of the sensor. The Laplacian operator effectively subtract this contribution.

6.3 EDGE DETECTION

A high pass filtered image that is added back to the original image is a high boost filter and the result will be sharpened or deblurred image. The high-pass filtered image can be used alone, particularly in the study of location and geographical distribution of "edges". An edge is a discontinuity or sharp change in the greyscale value at a particular pixel point and it may have some interpretation in terms of geological structure or relief. We have already noted that the first difference can be compute for the horizontal, vertical and diagonal directions and the magnitude and direction of the maximum spatial gradient can also be used. Other methods include the subtraction of a low-pass filtered image from the original or the use of a Roberts gradient.

One of the main uses of edge-detection technique has been in the enhancement of images for the identification and analysis of geological lineaments, which are defined as "mappable, simple or composite linear features whose parts are aligned in a rectilinear or slightly curvilinear relationship and which differ distinctly from a pattern of adjacent features and which presumably reflect a sub surface phenomenon". The sub surface phenomena to which the definition refers are presumed to be fault and joint patterns in the underlying rock. However, linear features produced from remotely sensed images using the techniques described above should be interpreted with care since other features such as roads, railways, canals, rivers, field boundaries, forest clearcuts for power lines, or changes in landuse also produce "edges", which could be misinterpreted as lineaments by an uncritical observer.

Other applications of edge-detection techniques include the determination of the boundaries of homogenous regions in synthetic aperture radar images in order to segment those images (Quegan and Wright 1984).

7.0 STATISTICAL PATTERN RECOGNITION,CLASSIFICATION AND FEATURE SELECTION

Classification of pixels making up a remotely sensed images involves associating each pixel in the image with a label describing a particular class. Classes are themes so the resulting image is a thematic map.

The decision to classify an image cell into any particular class is a statistically intelligent 'guess' with some associated probability of errors. It is essentially a decision making process, in which one is trying to minimize the errors in the classification of scene elements.

An intuitively satisfying and mathematically 'manageable' classification theory is maximum likelihood or Baye's optimal classification, other methods can be made to work well in various ways, but they often lack a supporting theory.

7.1 PARAMETRIC CLASSIFICATION

7.1.1 Baye's Theory

Suppose we measure some feature of a scene and have to decide to which of the two classes a pixel belongs. This is a one dimensional, two class classification problem in the feature domain of the image. If a large number of pixels are available that may be considered representative of each class we can calculate a relative frequency histogram of the feature for each class and consider these to be approximating to the continuous probability density functions of an infinite sample of data. These state-conditional probability density functions, $p(x/1)$ and $p(x/2)$, have unit area and describe the probability of a pixel having a feature value x given that the pixel is in class 1 or class 2 respectively.

Each probability density function (histogram) may be scaled by the a priori probability, $p(i)$, that class i occurs in the image area of interest. These scaled probability function, $p(x/i)p(i)$, represent the probability that a pixel has a feature value x and is in the class i.

In remote sensing the priori probabilities may be estimated from external sources of information about the scene such as ground surveys, existing maps or historical data.

To make a classification decision for a pixel we need to know the posteriori probabilities that the pixel belongs to each of the training classes, given that the pixel has the feature value x. This probability, p(i/x), may be calculated with Baye's rule

$$p(i/x) = p(x/i)\ p(i)/p(x) \tag{7.1}$$

Where

$$p(x) = \sum_{i=1}^{2} p(x/i)\ p(i) \tag{7.2}$$

A decision rule may now be formed with a posteriori probabilities of equation (1.3). If a pixel has feature value x, an intuitively satisfying approach is to assign the pixel to class 1 if p(1/x) is greater than p(2/x). Since p(x) is the same for both classes in equation (1.3). It can be ignored in a comparison of the two and we can write as the Baye's decision rule

a pixel belongs to class 1 if $p(x/1)p(1) > p(x/2)p(2)$

a pixel belongs to class 2 if $p(x/2)p(2) > p(x/1)p(1)$

This is the maximum -Likelyhood discrimination technique.

In the very unlikely situation that the two posteriori probabilities are exactly equal i.e.

$$p(1/x) = p(2/x) \tag{7.3}$$

or $$p(x/1)\ p(1) = p(x/2)\ p(2) \tag{7.4}$$

A decision can not be made from the class probabilities. A tie-breaking process then must be employed, such as using the classification of an adjoining previously classified pixel or randomly choosing either class 1 or class 2. It can be shown that the Baye's decision rule minimizes the average probability of error over the entire classified data set, if all the classes have normal (Gaussian) probability density functions.

The total probability of classification error is given by the area under the overlapping portions of a posteriori probability functions as shown in Figure 7.1. The total probability of error is the sum of the probabilities that an incorrect decision was made on either side of the class partition. It is easy to see that the Baye's optimal partition minimizes this error because a shift of the partition to the right or left will include a larger area from either class 2 or class 1 respectively, thus increasing the total error.

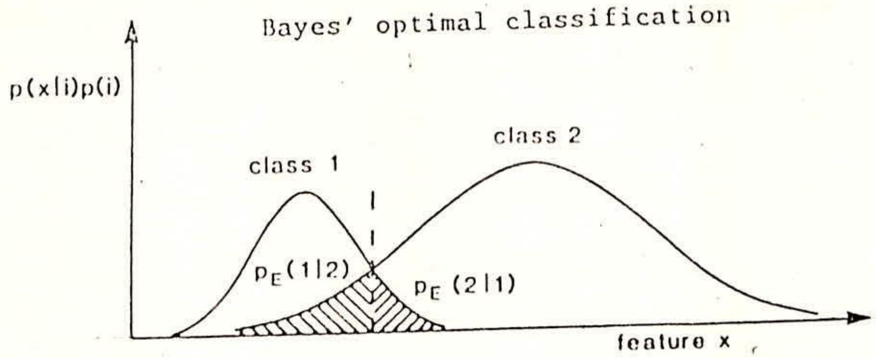


Figure 7.1: Probability of error, P_E for a maximum likelyhood classification

7.1.2 The minimum distance classifier

A commonly used algorithm for image classification is the minimum-distance classifier or linear discrimination. With this algorithm, each unknown pixel with feature vector x is classified by assigning it to the class whose mean vector M_i, is closest to x. In addition to the obvious intuitive appeal and computational simplicity of this approach, it can be shown that it is a very special case of the general maximum likelyhood classifier.

7.2 NON-PARAMETRIC CLASSIFICATION

Nearly all non-parametric classification methods are based on attempts to estimate $P(x/c_i)$ and apply in Baye's rule to construct a classifier.

7.2.1 Nearest neighbour methods

Nearest neighbour methods are in a sense based on a "reverse histogram" estimator. In the histogram method, the number of points falling into a fixed Volume is used, in the nearest neighbour (NN) method, the volume which contain a fixed number of points is used as an estimator. We may define the K-NN estimate $p(x/c_i)$, based on a sample of size m_i, as:

$$P_k(x/c_i) = \frac{k}{m} \frac{1}{A(k,x)_i}$$

Where A(k,x) is the volume which contains the K nearest sample points to X. To complete the definition we have to say what we mean by 'near', in other words we must define our matrix. We can show that $P_k(x/c_i)$ is asymptotically unbiased and consistent if;

$$\lim_{k \Rightarrow \infty} \frac{k}{m_i} = 0$$

As this condition is usually satisfied, the K-NN estimate is thus usually a good one .

7.2.2 K-NN Classification rule

By a slight modification the K-NN estimate can be turned into a classification rule that makes no reference to A(k,x) and hence is easy to use. If we have a sample of size M consisting of m_i samples from each class (i.e. $M = \sum m_i$) and find the K_{th}NN to x ignoring class membership, and if the K points contained in A(K,x) consist of K_1, K_2,..K_g points from each class, we can estimate $P_k(x/c_i)$ by;

$$p_k(x/c_i) = \frac{k_i}{m_i} \frac{1}{A(k,x)}$$

Using this estimate in Baye's rule gives;

assigning to class i, if

$$p(c_i) \frac{k_i}{m_i} \frac{1}{A(k,x)} > p(c_j) \frac{k_j}{m_j} \frac{1}{A(k,x)}$$

for all j≠i

cancelling the factors 1/A(k,x), we have;

$$\frac{P(c_i)}{m_i} k_i > \frac{p(c_j)}{m_j} k_j$$

Which is very simple rule. If the sample sizes m are proportional to $P(c_i)$, or if $P(c_i)$ is estimated by m_i/M, then the rule becomes;

$$K_i > K_j$$

Which is the most usual form of the K-NN classification rule and amounts to assigning the new class to the class that the majority of its K nearest neighbours belong to.

The problem with this rule is the selection of K and the large amount of storage required. To a certain extent the first problem is artificial in that there are good theoretical reasons for supposing that the error rate of the 1-NN classifier will be better than any other value of K. Thus in most cases there is no need to consider anything more complicated than a nearest neighbour classifier. The second problem can be solved by removing all those points in the sample which are not 'important', i.e. do not influence the classification. These are located along the classification boundaries.

7.3 CLASSIFICATION

In reality, the spectral radiance of a given surface material is not characterised by a single, deterministic curve, but by a family of curves with a range of variability caused by many natural factors. Sensor properties, such as detector noise, can cause additional variation in the measured radiance. The pixel spectral measurement, therefore, form a cluster of vectors for each class Figure 7.3. Separation of the classes is now more difficult because, although water can still be distinguished from the other two classes; soil and vegetation overlap (a common situation) and require a compromise partition. To a large extent, or ability to perform an accurate classification of a given multispectral image is determined by the extent of overlap between class signatures. As discussed earlier, one compromise that can be achieved is minimization of the average error in the classification.

In general, the separation of all classes requires more than two spectral bands. Because the clusters occur in K-dimensional space, the class partitions are surfaces in K dimensions. At this point, we introduce the general term feature to describe each dimension of this K- dimensional space. As will be seen later, spectral bands are not the only possible components of this space, other image-derived properties may be useful, for example spatial texture or spectral ratios. The word feature accommodates this broader scope.

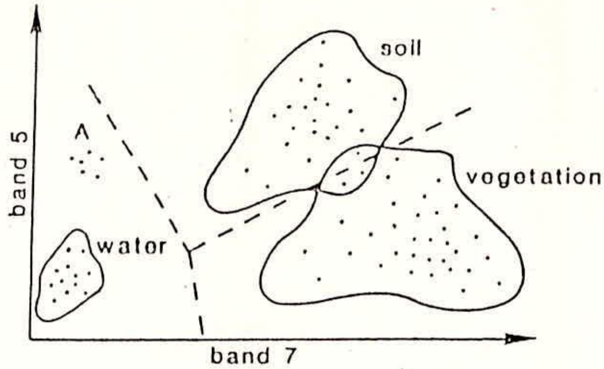


Figure 7.3: Typical two-band signatures for real data

The distribution of actual data in feature space is usually difficult to visualize if k is greater than 2. The number of classes is typically six or more, which further confuses the display of class clusters in more than two dimensions.

The first step of any classification procedure is the training of computer program to recognise the class signatures of interest. This aspect of classification is critical for the success of entire process and often occupies a majority of the analysis time. To train the computer program, we must supply a sample of pixels from which class signatures, e.g., mean vectors and covariance matrices, can be developed.

For supervised classification, the analyst uses prior knowledge derived from field surveys, photo interpretation, and other sources, about small regions of the image to be classified to identify those pixels that belong to the class of interest.

The feature signatures of these identified pixels are calculated and used to recognise pixels with similar signatures throughout the image. For unsupervised training, the analyst employs a computer algorithm that locates naturally occurring concentrations of feature vectors from a heterogeneous sample of pixels. These computer specified clusters are then assumed to represent feature class in the image and are used to calculate class signatures. The computer-derived classes remain to be identified. However, they may or may not belong to class of interest to the analyst.

Supervised and unsupervised classification are complement to each other, the former imposes the analyst knowledge of the area in the analysis to constrain the results, and the latter determines the inherent structure of the data, unconstrained by external knowledge about the area. A combination of the two techniques is often used to take advantage of the characteristics of each.

7.3.1 Supervised

For supervised classification a representative area for each desired class must be located in the image. It is important that the training area be a homogeneous sample of the respective class, but at the same time, the range of variability for the class must be included. Thus more than one training sample set per class is often used. Field surveys, aerial photographs and existing maps are used to verify the training sites.

If there is considerable variation within the class, the selection of training sites can be laborious, and it is impossible to made it certain that a comprehensive set of training samples for each class has been specified.

One important statistical aspect of selecting training data is that a sufficient number of pixels must be used to estimate the class signature properties accurately. If Baye's Maximum-likelyhood classifier is used and normal class distributions are assumed, the class mean vectors and covariance matrices must be calculated. However to obtain reliable class statistics, 10 to 100 training pixels per class per feature are typically needed. The number of training pixels required for a given signature accuracy increases with the increase in the within class variability

7.3.2 Unsupervised

In defining image areas for unsupervised training, the analyst does not need to be concerned with the homogeneity of the sites. Often, the sites are purposely choose to be heterogeneous to ensure that all possible classes and their respective within class variability are included. The pixels within the training areas are submitted to a clustering algorithm that determines the "natural" groupings of the data in the K dimensional feature space. Each cluster then is assumed to represent the probability distribution for one class. The assignment of identifying levels to each class may be done by the analyst at this point or after classification of full image. Because unsupervised trainings does not necessarily require any information about the area being classified, beyond what is in the image itself. It may be useful for delineating homogeneous areas for potential supervised training sites.

REFERENCES

1. Townshend;J. R. G. (1980), The spatial resolving power of Earth resources satellites: a review, NASA Technical Memorandum 82020. Goddard spacelight centre, Green belt, Maryland.

2. Quegan and Wright 1984; Automatic segmentation techniques for satellite borne synthetic aperture radar (SAR) images. Proceeding of the tenth Anniversary International Conference of the Remote Sensing society - review and perview, Remote Sensing Society, University of Nottingham 161-167.

3. Kostwinder H.R. and Bakx J.P.G; Radiometric Corrections, L.N. DIP5 1991.

4. Huurneman G.C.; Geometric Corrections, L.N. DIP5 1991.

5. Bakx, J.P.G.; Colour coding and spectral feature extraction, L.N. DIP5 1991.

6. Gorte B.G.H., Operators and filters, L.N. DIP 5 1991.

7. Ellis M.C.; Statistical Pattern Recognition classification and feature selection in Image processing L.N. DIP5 1991.