UNIT III IMAGE INTERPRETATION AND ANALYSIS

Types of Data Products – types of image interpretation – basic elements of image interpretation - visual interpretation keys – Digital Image Processing – Pre-processing – image enhancement techniques – multispectral image classification – Supervised and unsupervised.

Types of data products

Different types of data products for LISS-III,PAN and WiFS sensors have been categorized into three groups, viz. Standard products, special products and stereo products.

Standard products:

Standard products are corrected for the following errors to the extent possible through prior knowledge and orientation parameters which are updated using orbit/attitude model.

(i) Scene-related errors: It includes earth rotation, earth shape, earth curvature and map projection.

(ii) Sensor related errors: It includes mainly detector response non-uniformity, detector array alignment related and sensor tilt.

(iii) Platform-related errors: These errors mainly include spacecraft altitude, spacecraft attitude, and sensor alignment.

Special products:

Data products with any one of the following characteristics are referred to as special products: Special radiometric data manipulation, High geometric accuracy products, registration between multi-sensor data sets, mosaicing of multiple data sets to cover the district, state, country etc.

Stereo products:

Stereo products are defined to take advantage of the stereo data acquisition capability by tilting the PAN camera.

These products are characterized by: Possibility to get a 3-D view of the area under consideration, capture of 3-D topographic data, removal of image distortions due to height variations.

Image Interpretation

Image interpretation is defined as the extraction of qualitative and quantitative information in the form of a map, about the shape, location, structure, function, quality, condition, relationship of and between objects, etc. by using human knowledge or experience. As a narrow definition, " **photo-interpretation** " is sometimes used as a synonym of image interpretation.

TYPES OF Image interpretation

The features that our brains use when we interpret an image can be grouped into six main types, summarised below and in
Figure 8-1:

1.**Tone**: variations in relative brightness or colour.

2. **Texture**: areas of an image with varying degrees of 'smoothness' or 'roughness'.

3.**Pattern**: the arrangement of different tones and textures; may indicate certain types of geology or land use.

4.**Shape**: distinct patterns may be due to natural landforms or human shaping of the land.

5. **Size**: recognition of familiar objects allows size estimation of other features; size is an important aspect of association: for instance, a 20 km-wide circular surface depression is unlikely to be a sinkhole, but might be a volcanic caldera.

4. **Shape**: distinct patterns may be due to natural landforms or human shaping of the land.

5. **Size**: recognition of familiar objects allows size estimation of other features; size is an important aspect of association: for instance, a 20 km-wide circular surface depression is unlikely to be a sinkhole, but might be a volcanic caldera.

6.**Association**: the context of features in an image, e.g. a drainage pattern.

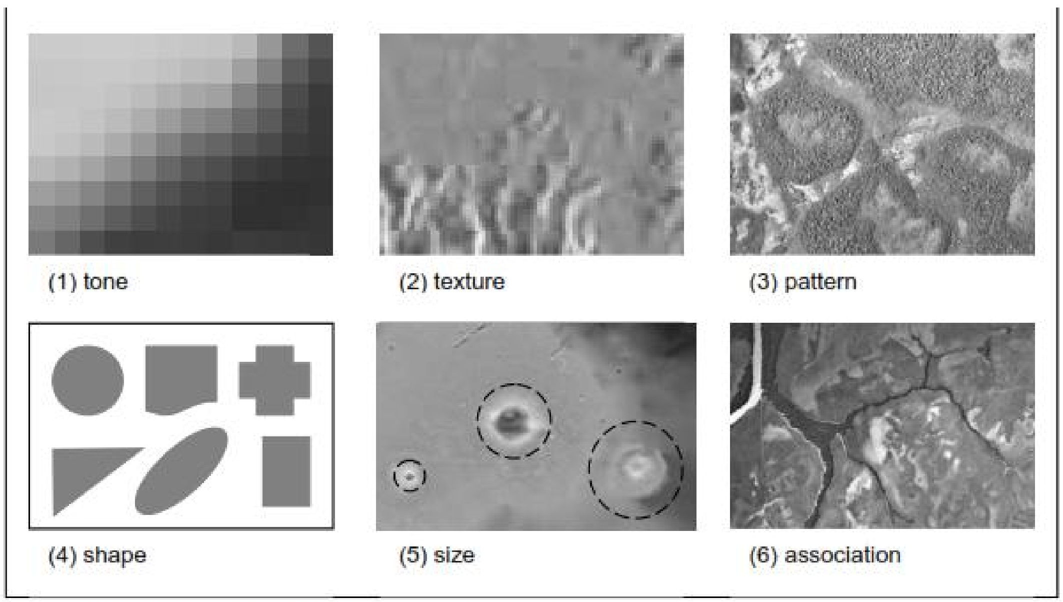


Figure 8-1 Features used in image interpretation.

Image interpretation in satellite remote sensing can be made using a single scene of a satellite image, while usually a pair of stereoscopic aerial photographs are used in photo-interpretation to provide stereoscopic vision using, for example, a mirror stereoscope. Such a single photo-interpretation is discriminated from stereo photo-interpretation (see 7.3).

Image reading is an elemental form of image interpretation. It corresponds to simple identification of objects using such elements as shape, size, pattern, tone, texture, color, shadow and other associated relationships. Image reading is usually implemented with interpretation keys with respect to each object, as explained in 7.4 and 7.5.

Image measurement is the extraction of physical quantities, such as length, location, height, density, temperature and so on, by using reference data or calibration data deductively or inductively.

Image analysis is the understanding of the relationship between interpreted information and the actual status or phenomenon, and to evaluate the situation.

Extracted information will be finally represented in a map form called an interpretation map or a thematic map.

Generally the accuracy of image interpretation is not adequate without some ground investigation. Ground investigations are necessary, first when the keys are established and then when the preliminary map is checked.

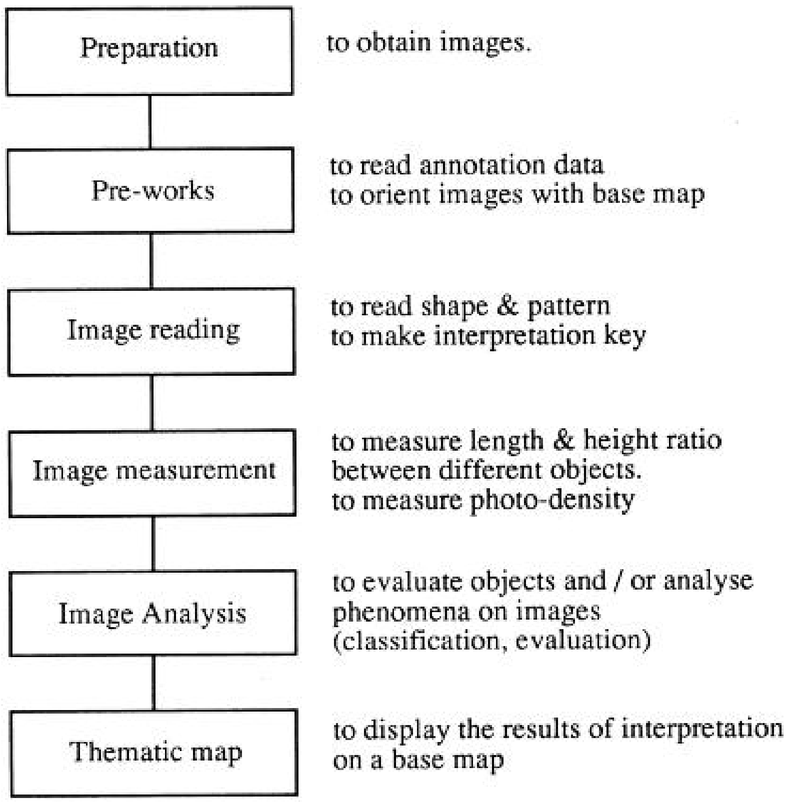


Figure 7.2.1 The image interpretation processing

7.4 Interpretation Elements

The following eight elements are mostly used in image interpretation; size, shape, shadow, tone, color, texture, pattern and associated relationship or context.

(1) **Size**:

A proper photo-scale should be selected depending on the purpose of the interpretation. Approximate size of an object can be measured by multiplying the length on the image by the inverse of the photo-scale.

(2) **Shape:**

The specific shape of an object as it is viewed from above will be imaged on a vertical photograph. Therefore the shape looking from a vertical view should be known. For example, the crown of a conifer tree looks like a circle, while that of a deciduous tree has an irregular shape. Airports, harbors, factories and so on, can also be identified by their shape.

(3) **Shadow**:

Shadow is usually a visual obstacle for image interpretation. However, shadow can also give height information about towers, tall buildings etc., as well as shape information from the non-vertical perspective-such as the shape of a bridge.

(4) **Tone**:

The continuous gray scale varying from white to black is called tone. In panchromatic photographs, any object will reflect its unique tone according to the reflectance. For example dry sand reflects white, while wet sand reflects black.In black and white near infrared infrared photographs, water is black and healthy vegetation white to light gray.

(5) **Color:**

Color is more convenient for the identification of object details. For example, vegetation types and species can be more easily interpreted by less experienced interpreters using color information. Sometimes color infrared photographs or false color images will give more specific information, depending on the emulsion of the film or the filter used and the object being imaged.

(6) **Texture**:

Texture is a group of repeated small patterns. For example homogeneous grassland exhibits a smooth texture, coniferous forests usually show a coarse texture. However this will depend on the scale of the photograph or image.

(7) **Pattern**:

Pattern is a regular usually repeated shape with respect to an object. For example, rows of houses or apartments, regularly spaced rice fields, interchanges of highways, orchards etc., can provide information from their unique patterns.

(8) **Associated relationships or context**:

A specific combination of elements, geographic characteristics, configuration of the surroundings or the context of an object can provide the user with specific information for image interpretation.

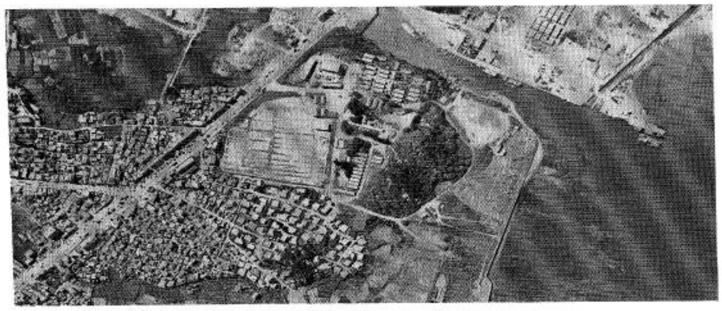


Figure 7.4.1 A sample of aerial photograph at Naha-city in Okinawa pref. (scale about 1/13,000)

[Size] Small fishery boats &large working ships
[Shape] square apartment houses & irregular low old houses
[Shadow] high buildings along the main street
[Tone] dark forest & light cultivated field, lighter coral reef & darker deep sea

Figure 7.4.2 A sample of aerial photograph at a part of Ibaragi Pref. (scale about 1/10,000)

[Texture] coarse forest area & fine young re-forest area
[Pattern] shaped housing vegetation, linearly road, meandering river and quadrangulation cultivated fields

—153—

7.5 Interpretation Keys

The criteria for identification of an object with interpretation elements is called an **interpretation key**. The image interpretation depends on the interpretation keys which an experienced interpreter has established from prior knowledge and the study of the current images. Generally, standardized keys must be established to eliminate the differences between different interpreters.

The eight interpretation elements (size, shape, shadow, tone, color, texture, pattern and associated relationship), as well as the time the photograph is taken, season, film type and photo-scale should be carefully considered when developing interpretation keys. Keys usually include both a written and image component.

Table 7.5.1 shows an example of interpretation keys for forestry mapping which have been standardized by the Japan Association for Forestry.

The keys are specified with respect to the crown's shape, rim shape of the crown, tone, shadow, projected tree shape, pattern, texture and other factors.

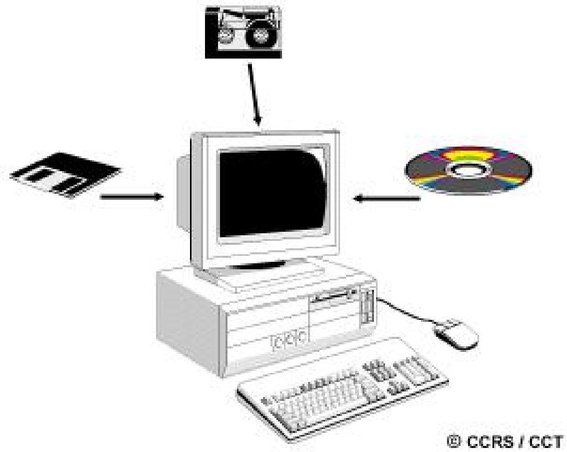
Table 7.5.2 shows an example of an interpretation key for land cover mapping with Landsat MSS images in the case of single band and false color images.

Table 7.5.1 Interpretation keys for forestry

species	crown shape	edge of crown	tone	pattern	texture
ceder	conical with sharp spear	circular and sharp	dark	spotted grain	hard and coarse
cypress	conical with round crown	circular but not sharp	dark but lighter than ceder	spotted	lard and fine
pine	cylindrical with shapeless crown	circular but unclear	light and unclear	irregularly spotted	soft but coarse
larch	conical with unclear crown	circular with unclear edge	lighter than cypress	spotted	soft and fine
fir/spruce	conical with wider crown	circular with zigzag edge	dark and clear	irregular	coarse
deciduous	irregular shapes	unclear	lighter	irregular	coarse

(by country of Japan Association of Forestry

Digital Image Processing



In today's world of advanced technology where most remote sensing data are recorded in digital format, virtually all image interpretation and analysis involves some element of digital processing. Digital image processing may involve numerous procedures including formatting and correcting of the data, digital enhancement to facilitate better visual interpretation, or even automated classification of targets and features entirely by computer. In order to process remote sensing imagery digitally, the data must be recorded and available in a digital form suitable for storage on a computer tape or disk. Obviously, the other requirement for digital image processing is a computer system, sometimes referred to as an **image analysis system**, with the appropriate hardware and software to process the data. Several commercially available software systems have been developed specifically for remote sensing image processing and analysis.

For discussion purposes, most of the common image processing functions available in image analysis systems can be categorized into the following four categories:

Preprocessing

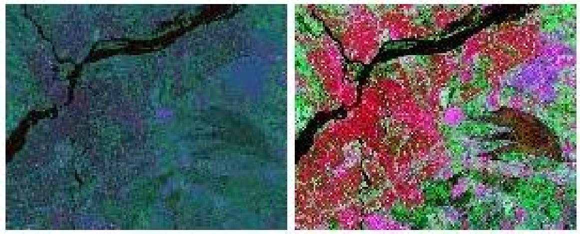
Image Enhancement

Image Transformation

Image Classification and Analysis

Preprocessing functions involve those operations that are normally required prior to the main data analysis and extraction of information, and are generally grouped **as radiometric or geometric corrections**. Radiometric corrections include correcting the data for sensor irregularities and unwanted sensor or atmospheric noise, and converting the data so they accurately represent the reflected or emitted radiation measured by the sensor. Geometric corrections include correcting for geometric distortions due to sensor-Earth geometry variations, and conversion of the data to real world coordinates (e.g. latitude and longitude) on the Earth's surface.

Digital Image Processing



The objective of the second group of image processing functions grouped under the term of **image enhancement**, is solely to **improve the appearance of the imagery** to assist in visual interpretation and analysis. Examples of enhancement functions include contrast stretching to increase the tonal distinction between various features in a scene, and **spatial filtering** to enhance (or suppress) specific spatial patterns in an image.

Image transformations are operations similar in concept to those for image enhancement. However, unlike image enhancement operations which are normally applied only to a single channel of data at a time, image transformations usually involve combined processing of data from multiple spectral bands. Arithmetic operations (i.e. subtraction, addition, multiplication, division) are performed to combine and transform the original bands into "new" images which better display or highlight certain features in the scene. We will look at some of these operations including various methods of **spectral or band** ratioing, and a procedure called **principal components analysis** which is used to more efficiently represent the information in multichannel imagery.

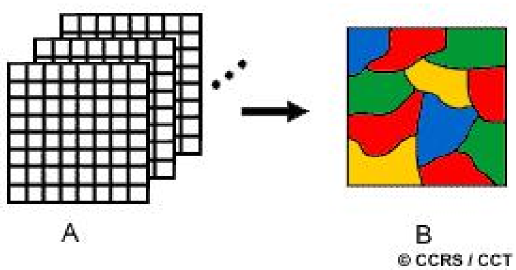


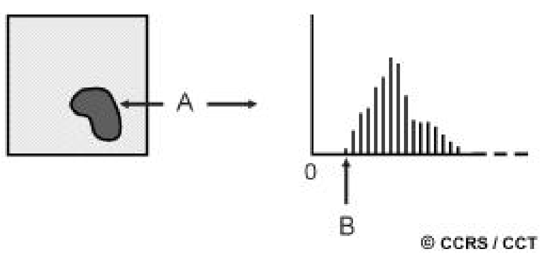
Image classification and analysis operations are used to digitally identify and classify pixels in the data. **Classification** is usually performed on multi-channel data sets (A) and this process assigns each pixel in an image to a particular class or theme (B) based on statistical characteristics of the pixel brightness values. There are a variety of approaches taken to perform digital classification. We will briefly describe the two generic approaches which are

used most often, namely
supervised
and
unsupervised
classification.
In the following sections we will describe each of these four categories of digital image processing functions in more detail.

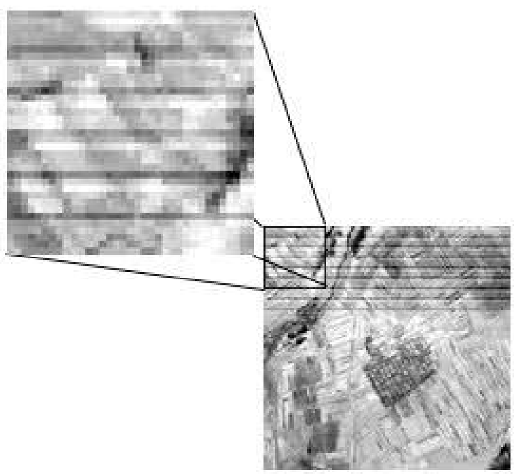
Pre-processing

Pre-processing operations, sometimes referred to as image restoration and rectification, are intended to correct for sensor- and platform-specific radiometric and geometric distortions of data. Radiometric corrections may be necessary due to variations in scene illumination and viewing geometry, atmospheric conditions, and sensor noise and response. Each of these will vary depending on the specific sensor and platform used to acquire the data and the conditions during data acquisition. Also, it may be desirable to convert and/or calibrate the data to known (absolute) radiation or reflectance units to facilitate comparison between data.

Variations in illumination and viewing geometry between images (for optical sensors) can be corrected by modeling the geometric relationship and distance between the area of the Earth's surface imaged, the sun, and the sensor. This is often required so as to be able to more readily compare images collected by different sensors at different dates or times, or to **mosaic multiple images from a single sensor** while maintaining uniform illumination conditions from scene to scene.



As we learned in Chapter 1, scattering of radiation occurs as it passes through and interacts with the atmosphere. This scattering may reduce, or attenuate, some of the energy illuminating the surface. In addition, the atmosphere will further attenuate the signal propagating from the target to the sensor. Various methods of atmospheric correction can be applied ranging from detailed modeling of the atmospheric conditions during data acquisition, to simple calculations based solely on the image data. An example of the latter method is to **examine the observed brightness values** (digital numbers), in an area of shadow or for a very dark object (such as a large clear lake - A) and determine the minimum value (B). The correction is applied by subtracting the minimum observed value, determined for each specific band, from all pixel values in each respective band. Since scattering is wavelength dependent (Chapter 1), the minimum values will vary from band to band. This method is based on the assumption that the reflectance from these features, if the atmosphere is clear, should be very small, if not zero. If we observe values much greater than zero, then they are considered to have resulted from atmospheric scattering.

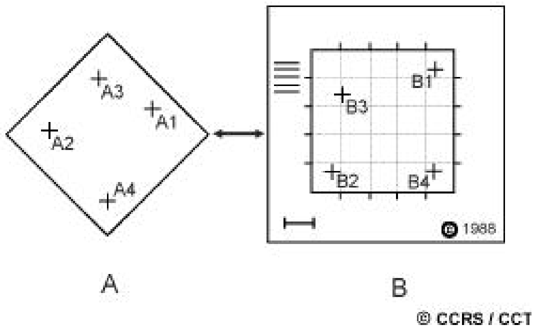


Noise in an image may be due to irregularities or errors that occur in the sensor response and/or data recording and transmission. Common forms of noise include systematic **striping** or banding and **dropped lines**. Both of these effects should be corrected before further enhancement or classification is performed. Striping was common in early Landsat MSS data due to variations and drift in the response over time of the six MSS detectors. The "drift" was different for each of the six detectors, causing the same

brightness to be represented differently by each detector. The overall appearance was thus a 'striped' effect. The corrective process made a relative correction among the six sensors to bring their apparent values in line with each other. Dropped lines occur when there are systems errors which result in missing or defective data along a scan line. Dropped lines are normally 'corrected' by replacing the line with the pixel values in the line above or below, or with the average of the two.
For many quantitative applications of remote sensing data, it is necessary to convert the digital numbers to measurements in units which represent the actual reflectance or emittance from the surface. This is done based on detailed knowledge of the sensor response and the way in which the analog signal (i.e. the reflected or emitted radiation) is converted to a digital number, called
analog-to-digital
(A-to-D) conversion. By solving this relationship in the reverse direction, the absolute radiance can be calculated for each pixel, so that comparisons can be accurately made over time and between different sensors.

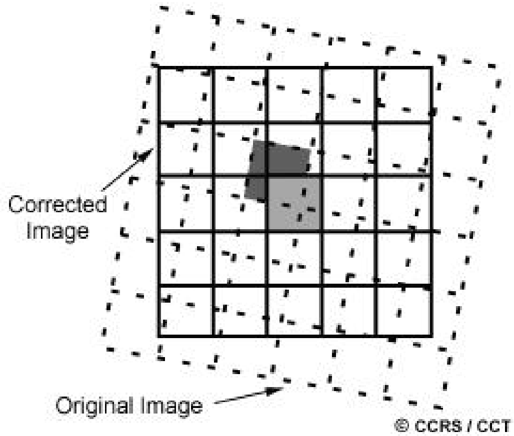
In section 2.10 in Chapter 2, we learned that all remote sensing imagery are inherently subject to geometric distortions. These distortions may be due to several factors, including: the perspective of the sensor optics; the motion of the scanning system; the motion of the platform; the platform altitude, attitude, and velocity; the terrain relief; and, the curvature and rotation of the Earth. Geometric corrections are intended to compensate for these distortions so that the geometric representation of the imagery will be as close as possible to the real world. Many of these variations are
systematic
, or
predictable
in nature and can be accounted for by accurate modeling of the sensor and platform motion and the geometric relationship of the platform with the Earth. Other
unsystematic
, or
random

, errors cannot be modeled and corrected in this way. Therefore, **geometric registration** of the imagery to a known ground coordinate system must be performed.



The **geometric registration process** involves identifying the image coordinates (i.e. row, column) of several clearly discernible points, called **ground control points** (or **GCPs**), in the distorted image (A - A1 to A4), and matching them to their true positions in ground coordinates (e.g. latitude, longitude). The true ground coordinates are typically measured from a map (B - B1 to B4), either in paper or digital format. This is **image-to-map registration**.

Once several well-distributed GCP pairs have been identified, the coordinate information is processed by the computer to determine the proper transformation equations to apply to the original (row and column) image coordinates to map them into their new ground coordinates. Geometric registration may also be performed by registering one (or more) images to another image, instead of to geographic coordinates. This is called image-to-image registration and is often done prior to performing various image transformation procedures, which will be discussed in section 4.6, or for multitemporal image comparison.



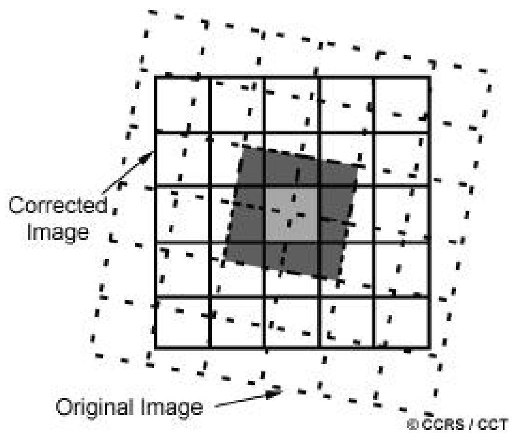
In order to actually geometrically correct the original distorted image, a procedure called **resampling** is used to determine the digital values to place in the new pixel locations of the corrected output image. The resampling process calculates the new pixel values from the original digital pixel values in the uncorrected image. There are three common methods for resampling: **nearest neighbour, bilinear interpolation**, and **cubic convolution**.

Nearest neighbour resampling uses the digital value from the pixel in the original image which is nearest to the new pixel location in the corrected image. This is the simplest method and does not alter the original values, but may result in some pixel values being duplicated while others are lost. This method also tends to result in a disjointed or blocky image appearance.

Bilinear interpolation resampling takes a weighted average of four pixels in the original image nearest to the new pixel location. The averaging process alters the original pixel values and creates entirely new digital values in the output image. This may be undesirable if further processing and analysis, such as classification based on spectral response, is to be done. If this is the case, resampling may best be done after the classification process.

Cubic convolution

resampling goes even further to calculate a distance weighted average of a block of sixteen pixels from the original image which surround the new output pixel location. As with bilinear interpolation, this method results in completely new pixel values. However, these two methods both produce images which have a much sharper appearance and avoid the blocky appearance of the nearest neighbour method.



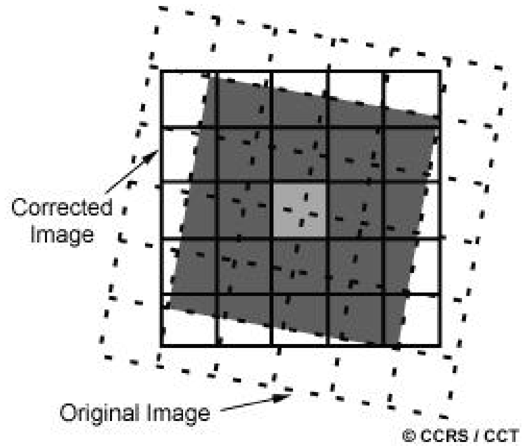


Image Enhancement
Techniques

Enhancements are used to make it easier for visual interpretation and understanding of imagery. The advantage of digital imagery is that it allows us to manipulate the digital pixel values in an image. Although radiometric corrections for illumination, atmospheric influences, and sensor characteristics may be done prior to distribution of data to the user, the image may still not be optimized for visual interpretation. Remote sensing devices, particularly those operated from satellite platforms, must be designed to cope with levels of target/background energy which are typical of all conditions likely to be encountered in routine use. With large variations in spectral response from a diverse range of targets (e.g. forest, deserts, snowfields, water, etc.) no generic radiometric correction could optimally account for and display the optimum brightness range and contrast for all targets. Thus, for each application and each image, a custom adjustment of the range and distribution of brightness values is usually necessary.

I. INTRODUCTION
The aim of image enhancement is to improve the interpretability or perception of information in images for
human viewers, or to provide `better' input for other automated image processing techniques.

Image enhancement techniques can be divided into two broad categories:
1. Spatial domain methods, which operate directly on pixels.
2. Frequency domain methods, which operate on the Fourier transform of an image.

Unfortunately, there is no general theory for determining what good image enhancement is? When it comes
to human perception. If it looks good, it is good! However, when image enhancement techniques are used as
pre-processing tools for other image processing techniques, then quantitative measures can determine which
techniques are most appropriate.
Apart from geometrical transformations some preliminary grey level adjustments may be indicated, to take
into account imperfections in the acquisition system. This can be done pixel by pixel, calibrating with the output
of an image with constant brightness. Frequently space-invariant grey value transformations are also done for
contrast stretching, range compression, etc. The critical distribution is the relative frequency of each grey value,
the grey value histogram.
The fingerprint images are rarely of perfect quality, due to the reasons like variations in impression condition,
skin condition, scanning devices or may be due to non-co-operative attitude of the subject. This degraded
quality of image can result in a significant number of spurious minutiae being created and genuine minutiae
being ignored. A vital step in studying the statistics of fingerprint minutiae is to reliably extract the minutiae
feature from fingerprint images. Thus it is important to employ image enhancement techniques prior to minutiae
extraction to obtain a good number of reliable estimates of minutiae locations.
The main objective of fingerprint image enhancement is to improve the ridge characteristics of the image, as
these ridges carry the information of characteristics features required for minutiae extraction. Ideally, in a welldefined

fingerprint image, the ridges and valleys should alternate and row in a locally constant direction. This
regularity facilitates the detection of ridges and consequently allows minutiae to be precisely
extracted from the thinned ridges [1]. Thus, the corruption or noise has to be reduced through image enhancement techniques to get
enhanced definition of ridges against valleys in the fingerprint images.

II. Image Enhancement

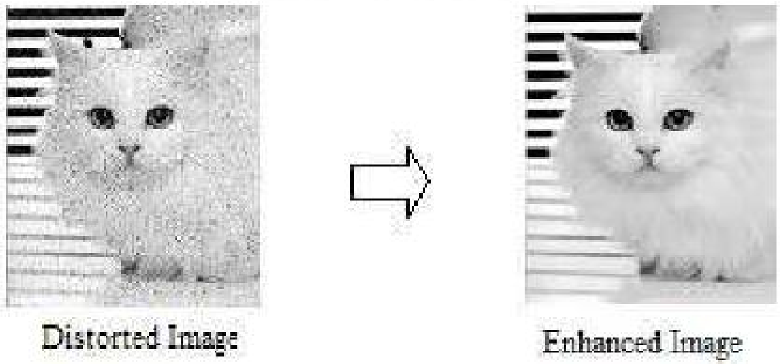


Fig.1. Showing the effect of Image Enhancement

Fig.1. Showing the effect of Image Enhancement
Image enhancement is basically improving the interpretability or perception of information in images for
human viewers and providing `better' input for other automated image processing techniques. The principal
objective of image enhancement is to modify attributes of an image to make it more suitable for a given task and
a specific observer. During this process, one or more attributes of the image are modified. The choice of
attributes and the way they are modified are specific to a given task. Moreover, observer-specific factors, such
as the human visual system and the observer's experience, will introduce a great deal of subjectivity into the
choice of image enhancement methods. There exist many techniques that can enhance a digital image without
spoiling it. The enhancement methods can broadly be divided in to the following two categories:

1. Spatial Domain Methods
2. Frequency Domain Methods

In spatial domain techniques [10] we directly deal with the image pixels. The pixel values are manipulated to
achieve desired enhancement. In frequency domain methods, the image is first transferred in to frequency
domain. It means that, the Fourier Transform of the image is computed first. All the enhancement operations are
performed on the Fourier transform of the image and then the Inverse Fourier transform is performed to get the
resultant image. These enhancement operations are performed in order to modify the image brightness, contrast
or the distribution of the grey levels. As a consequence the pixel value (intensities) of the output image will be

modified according to the transformation function applied on the input values into image g using T. (Where T is
the transformation. The values of pixels in images f and g are denoted by r and s, respectively. As said, the pixel
values r and s are related by the expression,

$$s=T(r) \qquad (1)$$

Where T is a transformation that maps a pixel value r into a pixel value s. The results of this transformation
are mapped into the grey scale range as we are dealing here only with grey scale digital images. So, the results
are mapped back into the range [0, L-1], where L=2k, k being the number of bits in the image being considered.
So, for instance, for an 8-bit image the range of pixel values will be [0, 255].

Many different, often elementary and heuristic methods [11] are used to improve images in some sense. The
problem is, of course, not well defined, as there is no objective measure for image quality. Here, we discuss a
few recipes that have shown to be useful both for the human observer and/or for machine recognition. These
methods are very problem-oriented: a method that works fine in one case may be completely inadequate for
another problem. In this paper basic image enhancement techniques have been discussed with their
mathematical understanding. This paper will provide an overview of underlying concepts, along with algorithms
commonly used for image enhancement. The paper focuses on spatial domain techniques for image
enhancement.

IMAGE ENHANCEMENT TECHNIQUES

A. Spatial domain methods
The value of a pixel with coordinates (x, y) in the enhanced image is the result of performing some operation
on the pixels in the neighborhood of (x, y) in the input image, F. Neighborhoods can be any shape, but usually
they are rectangular.
B. Histogram Equalization

Histogram equalization is a method in image processing of contrast adjustment using the image's histogram.
In histogram equalization we are trying to maximize the image contrast by applying a gray level transform which tries to flatten the resulting histogram. It turns out that the gray level transform that we are seeking is simply a scaled version of the original image's cumulative histogram. That is, the graylevel transform T is given by T[i] = (G-1)c(i), where G is the number of gray levels and c(i) is the normalized cumulative histogram of the original image.

When we want to specify a non-flat resulting histogram, we can use the following steps:

1. Specify the desired histogram g(z)
2. Obtain the transform which would equalize the specified histogram, Tg, and its inverse Tg^{-1}
3. Get the transform which would histogram equalize the original image, s=T[i]
4. Apply the inverse transform Tg^{-1} on the equalized image, that is $z=Tg^{-1}[s]$

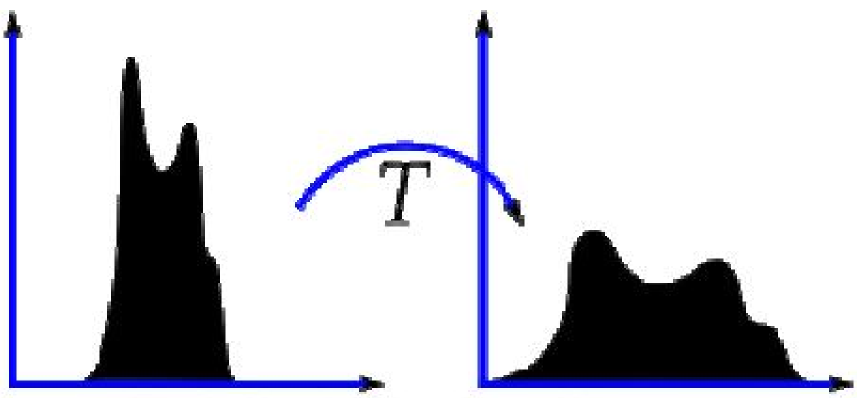


Fig. 2. Showing the effect of histogram

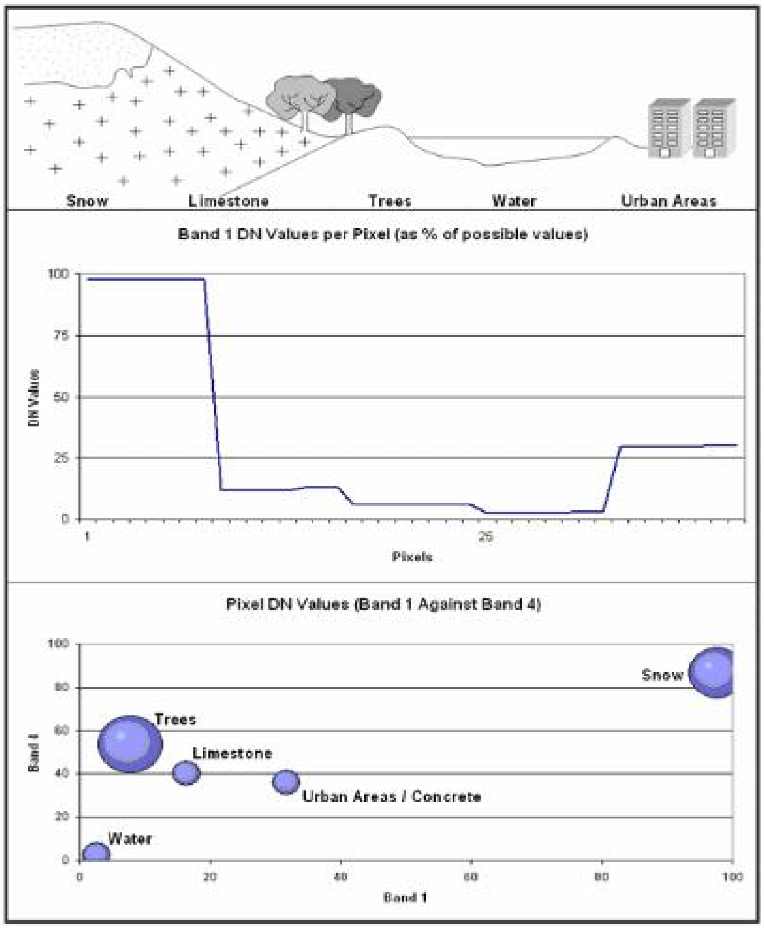
The method is useful in images with backgrounds and foregrounds that are both bright or both dark. In particular, the method can lead to better views of bone structure in x-ray images, and to better detail in photographs that are over or under-exposed. A key advantage of the method is that it is a fairly straightforward technique and an invertible operator. So in theory, if the histogram equalization function is known, then the original histogram can be recovered. The calculation is not computationally intensive. A disadvantage of the method is that it is indiscriminate. It may increase the contrast of background noise, while decreasing the usable signal.

Image classification

The subdivision of Earth surface features into different types on the basis of their spectral responses is known as classification. There are two types of classification: unsupervised and supervised, with the latter being more complicated but also potentially more accurate.

Unsupervised classification is carried out by the image processing software without any initial input, or 'training', from the user. The process can be illustrated with reference to Figure 8-12. Diagram A shows four different land cover types, each with a distinct spectral signature. The differing spectral responses can be more effectively distinguished by plotting the DN values from the bands of the original Landsat imagery against each other, as illustrated in
Figure 8-12B. This produces distinct clusters of DN values, along with a
modified image on which each cluster is colour-coded. The user then has to allocate each cluster to its corresponding land cover type, providing a legend, thereby turning the colour-coded cluster analysis image into a land cover map.

Supervised classification requires the user to select training areas containing about 100 pixels of each land cover type: these pixels are used to 'teach' a computer to recognise the spectral responses of each land cover type. A legend for the ensuing land cover map is built up as the user inputs the training areas for each land cover type. The software then uses the training areas to derive statistical summaries of each land cover type's spectral response, from which it goes on to classify all of the remaining pixels in the image. The purer the sample of pixels in the training area for each land cover type, the better the accuracy of the ensuing classification.
Classification based on a spectral reflectance from a single band and then from two bands is illustrated graphically in
Figure 8-12. Including a third band in the classification routinewill improve the discrimination between land cover types, as illustrated in the 3-D graph in
Plate 13. Using multispectral or hyperspectral imagery, classification software can utilise *n* bands in *n* dimensions, giving increasingly better separations between land cover types – in some cases allowing automated mapping based solely on the varying spectral responses along scan lines. The maths behind image classification can be complex: the reader is referred to Lillesand & Kiefer (2000), Drury (2001), or Mather (1999) for useful summaries



Multispectral image Classification Methods

General procedures in image classification

Classification is the most popularly used information extraction techniques in digital remote sensing. In image space *I*, a *classification unit* is defined as the image segment on which a classification decision is based. A classification unit could be a pixel, a group of neighbouring pixels or the whole image. **Conventional multispectral classification techniques** perform class assignments based only on the spectral signatures of a classification unit. **Contextual classification** refers to the use of spatial, temporal, and other related information, in addition to the spectral information of a classification unit in the classification of an image. Usually, it is the pixel that is used as the classification unit.

General image classification procedures include (Gong and Howarth 1990b):

(1) Design image classification scheme: they are usually information classes such as urban, agriculture, forest areas, etc. Conduct field studies and collect ground infomation and other ancillary data of the study area.

(2) Preprocessing of the image, including radiometric, atmospheric, geometric and topographic corrections, image enhancement, and initial image clustering.

(3) Select representative areas on the image and analyze the initial clustering results or generate training signatures.

(4) Image classification

Supervised mode: using training signature

unsupervised mode: image clustering and cluster grouping

(5) Post-processing: complete geometric correction & filtering and classification decorating.

(6) Accuracy assessment: compare classification results with field studies.

The following diagram shows the major steps in two types of image classification:

Supervised:

$\boxed{\text{Image}} \rightarrow \boxed{\text{Supervised Training}} \rightarrow \boxed{\text{Pixel Labelling}} \rightarrow \boxed{\text{Accuracy Assessment}}$

Unsupervised

$\boxed{\text{Image}} \rightarrow \boxed{\text{Clustering \& Cluster Analysis}} \rightarrow \boxed{\text{Clustering \& Cluster Grouping}}$

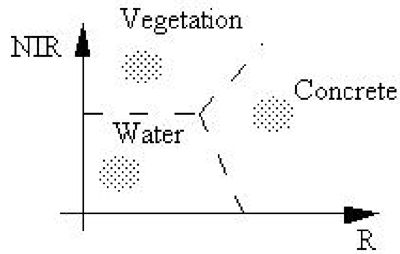
$\rightarrow \boxed{\text{Accuracy Assessment}}$

In order to illustrate the differences between the supervised and unsupervised classification, we will introduce two concepts: information class and spectral class:

Information class: a class specified by an image analyst. It refers to the information to be extracted.

Spectral class: a class which includes similar grey-level vectors in the multispectral space.

In an ideal information extraction task, we can directly associate a spectral class in the multispectral space with an information class. For example, we have in a two dimensional space three classes: water, vegetation, and concrete surface.



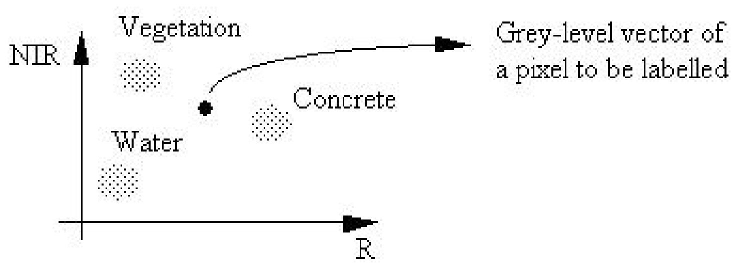
By defining boundaries among the three groups of grey-level vectors in the two-dimensional space, we can separate the three classes.

One of the differences between a supervised classification and an unsupervised one is the ways of associating each spectral class to an information class. For supervised classification, we first start with specifying an information class on the image. An algorithm is then used to summarize multispectral information from the specified areas on the image to form class signatures. This process is called supervised training. For the unsupervised case,however, an algorithm is first applied to the image and some spectral classes (also called clusters) are formed. The image analyst then try to assign a spectral class to the desirable information class.

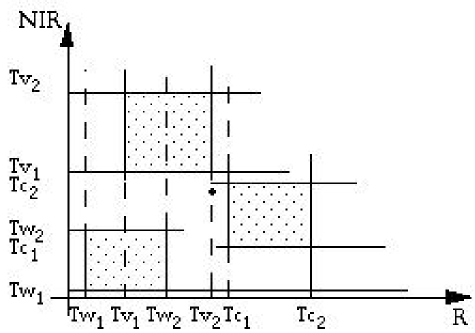
Supervised classification

Conventional Pixel-Labelling Algorithms in Supervised Classification

A pixel-labelling algorithm is used to assign a pixel to an information class. We can use the previous diagram to discuss ways of doing this.



From the above diagram, there are two obvious ways of classifying this pixel.



(1) Multidimensional thresholding

As in the above diagram, we define two threshold values along each axis for each class. A grey-level vector is classified into a class only if it falls between the thresholds of that class along each axis.

The advantage of this algorithm is its simplicity. The drawback is the difficulty of including all possible grey-level vectors into the specified class thresholds. It is also difficult to properly adjust the class thresholds.

(2) Minimum-Distance Classification

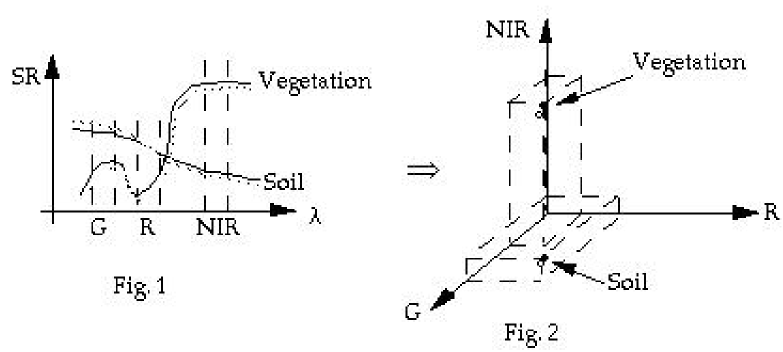


Fig. 1 shows spectral curves of two types of ground target: vegetation and soil. If we sample the spectral reflectance values for the two types of targets (bold-curves) at three spectral bands: green, red and near-infrared as shown in Fig. 1, we can plot the sampled values in the three dimensional multispectral space (Fig. 2). The sampled spectral values become two points in the multispectral space. Similar curves in Fig. 1 will be represented by closer points in Fig. 2 (two dashed curves in Fig. 1 shown as empty dots in Fig. 2. From Fig. 2, we can easily see that distance can be used as a similarity measure for classification. The closer the two points, the more likely they are in the same class.

We can use various types of distance as similarity measures to develop a classifier, i.e. minimum-distance classifier.

In a minimum-distance classifier, suppose we have nc known class centers
C = {C1, C2, ..., Cnc}, Ci, i = 1, 2, ..., nc is the grey-level vector for class i.

C_i, i = 1, 2, ..., nc is the grey-level vector for class i.

$$C_i = \begin{cases} (DN_{i1},\ DN_{i2},\ \ldots,\ DN_{inb})^T & \text{in digital number form.} \\ (r_{i1},\ r_{i2},\ \ldots, r_{inb})^T & \text{in spectral reflectance form.} \end{cases}$$

As an example, we show a special case in Fig. 3 where we have 3 classes (nc = 3) and two spectral bands (nb = 2)

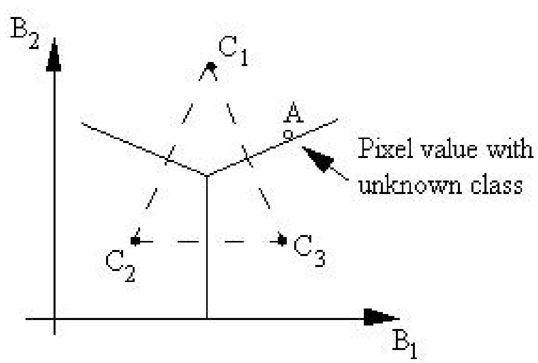


Fig. 3

If we have a pixel with a grey-level vector located in the B1-B2 space shown as A (an empty dot), we are asked to determine to which class it should belong. We can calculate the distances between A and each of the centers. A is assigned to the class whose center has the shortest distance to A.

In a general form, an arbitrary pixel with a grey-level vector g = (g1, g2, ..., gnb)T,

is classified as Ci if

d(Ci, g) = min (d(Ci1,g1), d(Ci2,g2), ..., d(Cinb,gnb))

Now, in what form should the distance d take? The most-popularly used form is the Euclidian distance

$$d_e(Ci,g) = \begin{cases} \sqrt{\sum_{j=1}^{nb} (C_{ij} - g)^2} & \text{Numerical form.} \\ [(C_i - g)^T (C_i - g)]^{1/2} & \text{matrix form.} \end{cases}$$

The second popularly used distance is Mahalanobis distance

$$d_m(C_i,g) = [(g - C_i)^T\, V^{-1} (g - C_i)]^{1/2}$$

where V-1 is the inverse of the covariance matrix of the data.

If the Mahalanobis distance is used, we call the classifier as a **Mahalanobis Classifier**.

The simplest distance measure is the city-block distance

$$d_c(C_i,g) = \sum_{n=1}^{nb} |C_{ij} - g_j|$$

For dm and de, because taking their squares will not change the relative magnitude among distances, in the minimum distance classifiers, we usually use d_m^2 and d_e^2 as the distance measures so as to save some computations.

Class centers C and the data covariance matrix V are usually determined from training samples if a supervised classification procedure is used. They can also be obtained from clustering.

For example, there are ns pixels selected as training sample for class Ci.

$$C_i = (C_{i1}, C_{i2}, ..., C_{inb})$$

$$C_{ij} = \frac{1}{ns}\sum_{k=1}^{ns} DN_{jk}$$

where j = 1, 2, ..., nb

k = 1, 2, ..., ns

If there are a total of nt pixels selected as training samples for all the classes

$$nt = \sum_{i=1}^{nc} Ns_i$$

The average vector M = (m1, m2, ..., mns) will be obtained.

$$m_i = \frac{1}{nt}\sum_{k=1}^{nt} DN_{i,k}$$

i = 1, 2, ..., nb.

k = 1, 2, ..., nt.

The covariance matrix is then obtained through the following vector form

$$V = \frac{1}{nt = 1} \sum_{k=1}^{nb} (DN - M) \bullet (DN - M)^T$$

where $DN = (DN, DN_2, ..., DN_{nb})^T$

$M = (m_1, m_2, ..., m_{nb})^T$

(3) Maximum Likelihood Classification (MLC)

MLC is the most common classification method used for remotely sensed data. MLC is based on the Baye's rule.

Let C = (C1, C2, ..., Cnc) denote a set of classes, where nc is the total number of classes. For a given pixel with a grey-level vector x, the probability that x belongs to class ci is P(Ci|x), i = 1, 2, ..., nc. If P(Ci|x) is known for every class, we can determine into which class x should be classified. This can be done by comparing P(Ci|x)'s, i = 1, 2, ..., nc.

x => ci, if P(Ci|x) > P(Cj|x) for all j # i. (1)

However, P(Ci|x) is not known directly. Thus, we use Baye's theorem:

P(Ci|x) = p(x|Ci) ï P(Ci)/P(x)

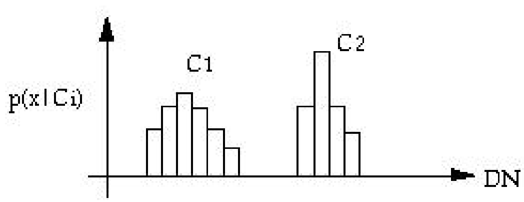
where

P(Ci) is the probability that Ci occurs in the image. It is called *a priori* probability.

P(x) is the probability of x occurring in each class ci.

$$P(x) = \sum_{i=1}^{nc} p(x|C_i) \bullet P(C_i)$$

However, P(x) is not needed for the classification purpose because if we compare P(C1|x) with P(C2|x), we can cancel P(x) from each side. Therefore, p(x|Ci) i = 1, 2, ..., nc are the conditional probabilities which have to be determined. One solution is through statistical modelling. This is done by assuming that the conditional probability distribution function (PDF) is normal (also called, Gaussian distribution). If we can find the PDF for each class and the *a priori* probability, the classification problem will be solved. For p(*x|ci) we use training samples.



For one-dimensional case, we can see from the above figure that by generating training statistics of two classes, we have their probability distributions. If we use these statistics directly, it will be difficult because it requires a large amount of computer memory. The Gaussian normal distribution model can be used to save the memory. The one-dimensional Gaussian distribution is:

$$p(x|C_i) = \frac{1}{\sqrt{2\pi} \bullet \delta_i} \cdot \exp\{-(x-\mu_i)^2/(2\delta_i^2)\}$$

where we only need two parameter for each class μ_i and δ_i , i = 1, 2, ..., nc

μ_i the mean for Ci

δ_i the standard deviation of Ci

μ_i, δ_i can be easily generated from training sample.

For higher dimensions,

$$p(x|C_i) = \frac{1}{(2\pi)^{nb/2} \cdot \sqrt{|V_i|}} \exp\{-\frac{1}{2}(x-\mu_i)^T V_i^{-1} \cdot (x-\mu_i)\}$$

where nb is the dimension (number of bands)

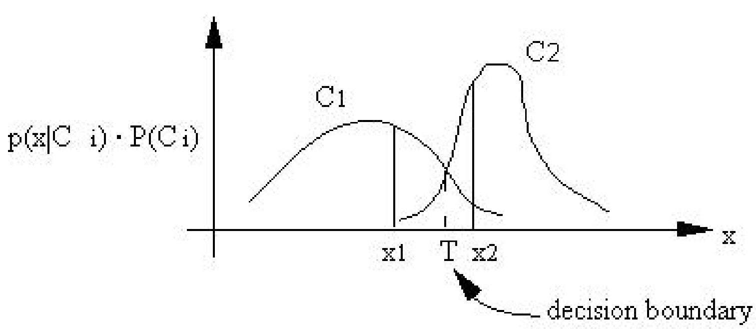
μ_i is the mean vector of ci

Vi is the covariance matrix of Ci

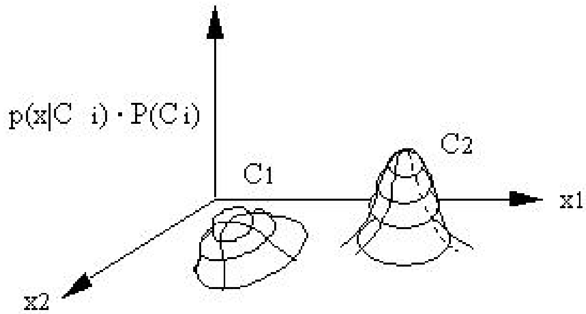
P(Ci) can also be determined with knowledge about an area. If they are not known, we can assume that each class has an equal chance of occurrence.

i.e. P(C1) = P(C2) = ... = P(Cnc)

With the knowledge of p(x|Ci) and P(Ci), we can conduct maximum likelihood classification. p(x|Ci) ï P(Ci) i = 1, 2, ..., nc can be compared instead of P(Ci|x) in (1).



The interpretation of the maximum likelihood classifier is illustrated in the above figure. An x is classified according to the maximum p(x|Ci) ï P(Ci). x1 is classified into C1, x2 is classified into C2. The class boundary is determined by the point of equal probability.



In two-dimensional space, the class boundary cannot be easily determined. Therefore we don't use boundaries in maximum likelihood classification and, instead, we compare probabilities.