2. Texture

It is the arrangement and frequency of tonal variation in an image (Figure 5.38). Texture is created by an aggregation of unit feature that may be too small to be discerned individually on the image, such as tree leaves. It determines the overall smoothness or coarseness of image features as visualised on the image. If the scale of the image is reduced, texture of any given object or area becomes progressively finer and ultimately disappears. Texture is combination of shape, size, pattern, shadow and tone.

Various features with similar reflectances can be distinguished based on their texture differences, such as the smooth texture of green grass as contrast with the rough texture of crowns of green trees. Smooth textures would have very little tonal variation, e.g., fields, asphalt, or grasslands, whereas the grey levels change abruptly in a small area, e.g., forest canopy, where rough texture is present. Sand has rough texture as compared to clay. Texture is also one of the most important elements for distinguishing features from Radar images.

3. Pattern

It relates to the spatial arrangement of visibly discernible objects (Figure 5.38). Typically, a repetition of similar tones and textures in a particular fashion will produce a distinct pattern which makes them distinct from each other. Many natural and man-made objects exhibit peculiar pattern, such as triangular, rectangular, square, pentagon, hexagon, circular or any other shape. Gardens with evenly spaced trees, and urban streets with regularly spaced houses are examples of pattern.

Table 5.7 Elements of image interpretation (Richards and Jia, 2013)

	What is it?	Example
Tone	Refers to the relative brightness or colour of objects in an image	Can distinguish between two crop fields due to different texture and shape of plants
Shape	Refers to the general form, structure, or outline of individual objects	Manmade objects tend to have straight edges, whereas natural objects have more irregular shapes
Size	**Size** of objects in an image is a function of scale	It is easy on an image to distinguish between a football stadium and a house
Pattern	Refers to the spatial arrangement of visibly discernible objects	In a housing estate, although individual houses cannot be made out, the recognisable pattern its produced
Texture	Refers to the arrangement and frequency of tonal variation in particular areas of an image	Different textures will appear differently on an image, for example, a sandy clearing within a forest will show up on the image
Shadow	May provide an idea of the profile and relative height of a target or targets which may make identification easier.	Shadow allows for identification of topographical landforms
Association	**Association** takes into account the relationship between other recognizable objects or features in proximity to the target of interest	Boats cannot be made out on their own within an image, however when in a marina, their proximity allows for identification

4. Shape

Shape can be a very distinctive clue for interpretation of various objects. It refers to the general form, configuration, or outline of individual objects (Figure 5.38). Shapes of some objects can be easily identified from stereo-photographs. Some objects are so distinctive that their images may be identified solely from their shapes. Shadow characteristics is also helpful to reveal the shape of the object. Straight edge shapes typically represent agricultural fields, while natural features, such as forest boundary, lakes, are generally more irregular in shape.

5. Size

Size of objects in an image is a function of scale of image. It is important to map the size of an object relative to other objects as well as its absolute size (Figure 5.38). The size of a feature will change on different scale images. A building may look like a point feature on a small scale image. For example, zones of land use, large buildings such as factories or warehouses would indicate commercial property, whereas small buildings would indicate residential use.

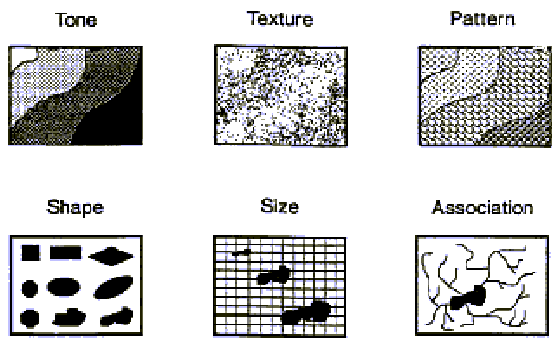


Figure 5.38 Pictorial representation of various visual interpretation elements (Garg, 2019)

6. Shadow

Shadow is helpful in interpretation of objects from images in two respects. Firstly, the shape or outline of a shadow normally provides the height of objects, and secondly the objects on the ground can be identified with respect to their shape of the shadow. However, the areas under shadow would hide the information and create difficulty in interpretation. For example, the shadows casted by various cultural features (bridges, silos, towers, etc.) can aid in their identification on air-photos/images. The shadows resulting from variations in terrain elevations can aid in assessing the natural topographic variations or geological landforms, particularly in Radar imagery.

7. Site/Association

Site refers to the topographic or geographic location of objects. It is an important clue in the identification of certain types of features (Figure 5.38). For example, certain tree species would occur on well-drained upland sites, whereas other tree species would occur on poorly drained lowland sites.

Association refers to the presence of certain features in relation to others. It takes into account the relationship between the recognizable objects/features in proximity to the object to be identified. For example, the snow cover is related to higher elevation zones, commercial properties are associated with major transportation routes.

5.17 Digital Image Interpretation Methods

This section deals with the interpretation of optical digital remote sensing images. Digital remote sensing images consist of pixels, and have a square grid structure, where one grid represents a *pixel* (i.e., the smallest element in a digital image). Each pixel is associated with a DN in a particular wavelength. Identification and separation of objects/features with respect to their DN values is called *digital image processing*. Pixels with similar DN values are grouped into various classes. In digital image processing requires a digital image and a software along with the skilled manpower for the interpretation, analysis and mapping. A comparison between visual and digital methods of interpretation is given in Table 5.8.

Table 5.8 Comparison of visual interpretation and digital interpretation methods (Garg, 2022)

S.No.	Manual Interpretation	Digital Interpretation
1	Hard copy data are required	Digital data are required
2	Requires simple and economical equipment or no equipment	Requires specialized, and often expensive equipment and software
3	Limited to analyse a single image at a time	Simultaneous interpretations of multispectral images
4	It involves subjectivity. i.e., the results may vary to some extent with different interpreters	It is an objective process which is based on the spectral signature of objects, and the images are analysed through a computer, so results are almost consistent
5	It is a qualitative method	It is a quantitative method
6	Cost-effective for small areas and for one-time interpretation	Cost-effective for large geographic areas, and for repetitive analysis.
7	No algorithm is used	Complex interpretation algorithms are required
8	Time-consuming and laborious	Speed may be an advantage
9	Difficult to change the scale of mapping	Easy to change the scale of mapping
10	Output product is in hard copy format so not compatible with other data	Compatible with other digital data

The digital images give us the flexibility to pre-process the digital pixel values in an image. The entire digital image processing consists of four basic operations: image pre-processing, image enhancement, image transformation, and image classification (Eastman, 1999). The details of these processes can be found in Garg (2022).

5.17.1 Image pre-processing

It involves the initial processing of raw image data to apply corrections for geometric distortions, calibrate the data radiometrically, and remove the noise present in the data, if any.

(A) Geometric Corrections:

It has two basic steps, as explained below-

(i) Georeferencing

Raw remote sensing data is without any geographic coordinates, and has distortions, mainly caused by the sensor geometry. Therefore, these can't be used as such for any quantitative measurements on them. Georeferencing is the conversion of image coordinates to ground coordinates by removing the distortions caused by the sensor geometry. Georeferencing is important to deal with various images, create mosaicking and compare various scenes (e.g., change assessment). It is a process of locating an entity/object in real world coordinates, also called *geo-rectification* or *geo-registration*.

The direction of satellite motion in the orbit and on-board sensors while taking images is not exactly north-south or west-east, respectively. In addition, there is a rotation of the Earth about its own axis while taking the images, so images are not perfect square but they have somewhat skewed shape. Georeferencing re-orients the image to a coordinate system representing the Earth, and making its geometry same as the Earth. Georeferenced images can be viewed, compared, and analysed with other geographic data.

To do georeferencing, the exact locations of several known points, called *Ground Control Points (GCPs)*, are required. These GCPs are normally selected as prominent objects whose geographical locations can be accurately determined either from the topographic maps or GPS survey. A minimum of four control points are required for georeferencing, however, additional control points would help increasing the accuracy of georeferencing. These GCPs are also identified on the image to be georeferenced. With these two sets of coordinates, polynomial is fitted amongst the GCPs, and *rms* error is minimized to ±1 pixel size. After georeferencing, each point on the image has real-world coordinates associated. The accuracy of the georeferencing would depend on the number, accuracy, and distribution of the control points and the choice of transformation polynomial. Normally, 2nd or 3rd order polynomial is used.

(ii) Resampling

After the georeferencing process, we may find that the pixels have been oriented differently than the way they were present in the original image coordinate system. Resampling is the process of interpolation the new DN values of the displaced pixels (new pixel location) in the new coordinate system. Three methods of resampling are commonly used, as given below (Figure 5.39).

(a) Nearest Neighbour: In this method, the attribute value of the original pixel nearest to a pixel in the output image is assigned to the corresponding cell.

(b) Bilinear Interpolation: It assigns the value to a pixel in the output image by taking weighted average of the surrounding four pixels in the original grid nearest to it.

(c) Cubic Convolution: It assigns the value to a pixel in the output image by taking weighted average of the surrounding sixteen pixels in the original grid nearest to it.

Among the three methods, nearest neighbour is a preferred method as it doesn't alter the values of the original grid cells assigned to the resampled grid cells but it produces a blocky image. The cubic convolution on the other side does change the values but is more accurate. It generates a smoother image.

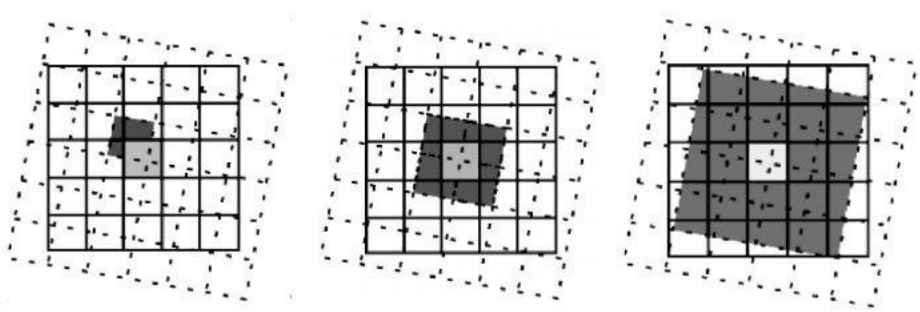


Figure 5.39 Resampling methods (Nady, 2020)

(B) Atmospheric Correction

Atmospheric correction is done to modify the DN values to remove noise, i.e., contributions to the DN due to intervening atmosphere. Lower wavelengths of visible spectrum are more subject to haze, which falsely increases the DN values. The simplest approach for its correction is known as *dark object subtraction method*, which assumes that if there was no haze present then the dark pixel will have zero DN value, e.g., deep water in near infrared will have complete absorption, and therefore those pixels will have zero DN values. But in reality, we won't find any dark pixel with zero values. For such image, the lowest DN value present in the image is determined, and is subtracted from all the DN values of the image, so that the dark (water) pixels are now zero.

5.17.2 Image enhancement

Image enhancement is mainly carried out to improve the quality of images. Often, the images do not have the optimum contrast so objects are not clearly identified visually. Enhancement is made to make the visual appearance of image better. It is the modification of images to improve the interpretability through human vision. The main purpose of image enhancement is to increase the contrast in a low contrast image. It does not add any information to the original image but it enhances the visual appearances of already captured features. Enhancement of an image can be implemented by using various methods. They improve the image quality so that the enhanced image is better than the original image for a specific application. Before image enhancement is done it is necessary to understand the image characteristics through its *histogram*.

(A) Image Histogram

A histogram is a graphical representation of the DN values (i.e., 0-255) in an image that are displayed along x-axis, while the frequency of occurrence of these values is plotted on y-axis. Image histogram is a way to portray the information present on an image. In raw imagery, the useful data often occupies only a small portion of the available range of DN values (256 levels in an 8 bit image). In an 8-bit image, in a histogram, the x-axis will contain 256 DN values and the y-axis will display how many of each intensity value is present.

A digital image can be represented by three effective ways (Figure 5.40): (i) Pictorially, in the form of image (ii) Numerically, in the form of DN values arranged in the matrix, and (iii) Graphically, through its histogram. A single peak bell-shaped histogram is considered as the best shape of a histogram of image data. It conveys about the homogeneity and well distribution of grey levels in the image. Two peaks (bi-model) histogram indicates that there could be two pre-dominant classes (e.g., water and vegetation) present on the image.

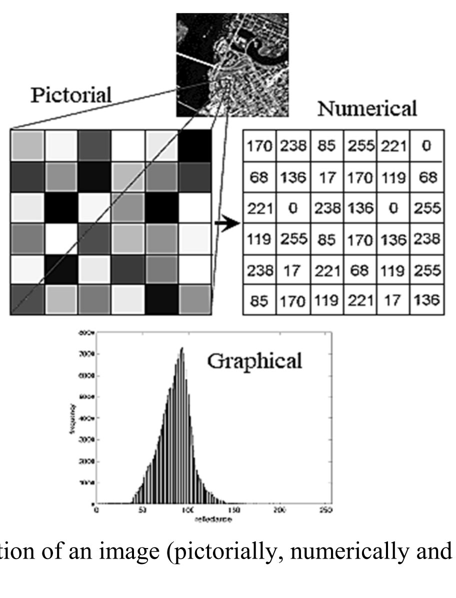


Figure 5.40 Representation of an image (pictorially, numerically and graphically) (Garg, 2019)

In reality, the shape of the histogram of an image is quite different that the ideal shape of histogram, as shown in Figure 5.41. The DN values are normally skewed either towards the lower values or towards the higher values, therefore, they indicate the presence or absence of features with higher or lower reflectances. Histograms alone can provide a lot of information about images to an interpreter even without looking at the images, such as likelihood of presence or absence of type of features, distribution etc. These help evaluate images statistically, e.g., normal, skewed, bimodal distribution, etc. The histograms are then used in individual image enhancement, image segmentation and image classification. Histograms also help matching of images across time or space.

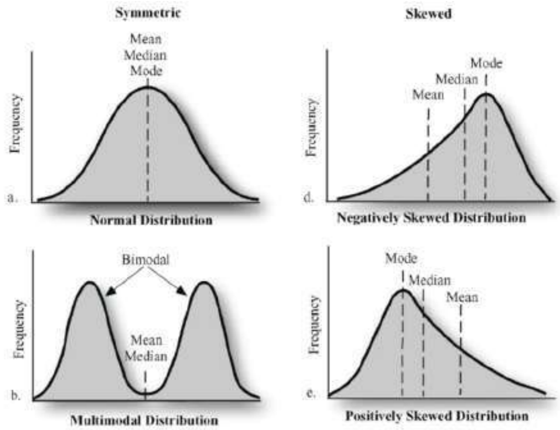


Figure 5.41 Various shapes of histogram of an image (Jenson, 1986)

Various enhancement techniques interpolate the range of DN values in an image, which can be represented graphically by its histogram.

(B) Contrast Enhancement

The contrast is defined as the maximum difference in color or intensity between two objects in an image. If the contrast is poor (low), it is impossible to distinguish between various objects and they are perceived as the same object. Contrast enhancement involves changing the original DN values so that more available range is utilised, thereby increasing the contrast between the objects and their background. It basically improves the interpretability for human viewing, and provides enhanced input to be used for image processing.

There are many different techniques and methods of enhancing the contrast, and details can be found in Garg (2022). The *linear contrast enhancement* is the most popular technique used for image enhancement. It involves identifying the minimum and maximum DN values in the image, and applying a linear transformation to stretch the present range to occupy the full range (e.g., 0-255 in an 8-bit image). In Figure 5.42, for example, minimum DN value (occupied by actual data) in the histogram is 84 and the maximum DN value is 153, so these 70 grey levels (153- 84 =70) occupy less than one-third of the full 256 levels available. A linear stretch will uniformly expand this small range to cover almost the full range of values from 0 to 255. It would enhance the contrast in the image with light toned areas appearing lighter and dark areas appearing darker, and thus making the identification of objects/features much accurate and effective.

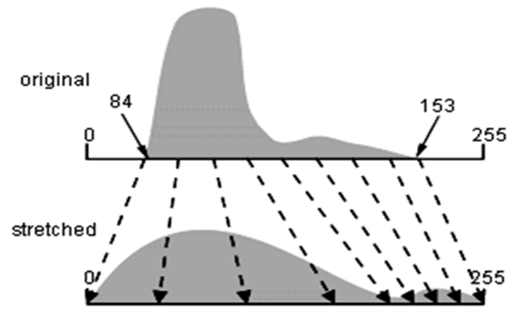


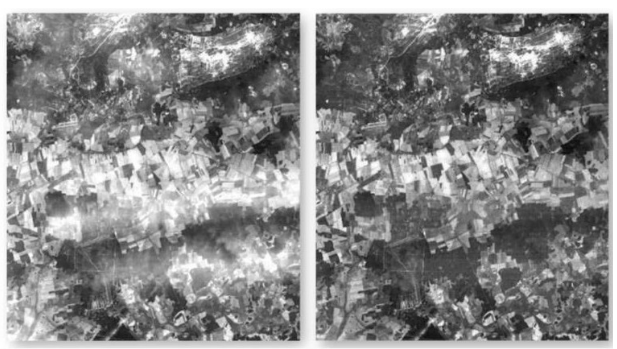
Figure 5.42 Original histogram and after linear contrast enhancement (Garg, 2022)

The linear contrast enhancement is mathematically represented as-

$$DN_{out} = ((DN_{in}-DN_{min})/(DN_{max} - DN_{min}))\text{* No. of grey levels} \quad (5.8)$$

where,
DN_{out} represents the output values in the image
DN_{in} represents the DN value at that pixel location from input image.
DN_{min} and DN_{max} are the minimum and maximum DN value, respectively, in the input image.
No. of grey levels are the total number of intensity values that can be assigned to a pixel. For example, in 8 bit images, the maximum number of grey levels is 255.

Figure 5.43 shows the results of linear contrast enhancement on Landsat ETM+ images. The main purpose of image enhancement is to increase the contrast in a low contrast image. It does not add any information to the original image but it enhances the visual appearances of already captured features.



a. Before Enhancement b. After Enhancement

Figure 5.43 (Left) original satellite image, and (Right) after linear contrast enhancement (https://www.gisoutlook.com/2019/08/digital-image-processing-image.html)

(C) Image Transformations

The image transformation is the creation of new image by using some mathematical function on the original images. The image transformation will normally yield synthetic images which are very useful for specific applications, as they enhance certain features of interest. Some examples of transformations include; simple arithmetic operations, Vegetation Indices (VI), Normalised Difference Vegetation Index (NDVI), Principal Component Analysis (PCA) and Tasselled Cap Transformations (TCT). The VI and NDVI images have been frequently used world-wide for the study of forest cover, vegetation and crop classifications. The details are given in Garg (2022).

The VI is obtained as the ratio of the near-infrared (NIR) band to the Red band-

$$VI = NIR\ band / Red\ band \tag{5.9}$$

If both the Red and NIR bands (or the VIS and NIR) have similar reflectance, the value of ratio is 1 or close to 1. Ratio values for bare soils generally are near 1; as the amount of green vegetation increases, the ratio increases. The value of ratio can increase far beyond 1 up to 30.

Most popular and commonly used vegetation index is the NDVI. The NDVI is a measure of the vegetative cover on the land surface, but it is also used to identify the water and soil. Vegetation differs from other land features because it tends to absorb strongly red wavelengths of sunlight and reflect in the near-infrared wavelengths. It is a measure of the difference in reflectance between these wavelength ranges. The NDVI is computed as-

$$NDVI = (NIR - Red) / (NIR + Red) \tag{5.10}$$

The NDVI values range between -1 and 1; value 0.5 indicates dense vegetation and values < 0 indicate no vegetation. Since vegetation has high NIR reflectance but low red reflectance, vegetated areas will have higher values as compared to non-vegetated areas. The NDVI has been used world-wide to monitor vegetation condition, vegetation health, cover and phenology over large areas, and therefore can provide early warning on droughts and famines.

5.17.3 Digital image classification

The digital classification of optical images and microwave images required different approaches, and software. In this section, optical image classification and associated algorithms have been discussed. Digital image classification is a software-based classification technique used for information extraction from optical images based on their DN values. It requires identification of spectral signature of various objects in an image, and subsequently tagging the pixels with similar spectral signatures. Each pixel (or groups of pixels) of an image may be assigned to a land use or land cover class. So, the image classification will categorise each pixel into at least one of the classes. The statistical decision rules are used which allow grouping of the pixels in different classes.

In optical remote sensing, there are broadly two classification techniques; *supervised* and *unsupervised classification*. Both the approaches of classification have their own strengths and weaknesses associated with the classification process and results of the analysis. These are briefly explained below.

(A) Supervised Classification

Supervised classification consists of three distinct stages; *training, allocation* and *testing*, as shown in Figure 5.44. Training is the first stage where the identification of a sample of pixels of known classes is done with the help of reference data, such as field visits, existing maps and aerial photographs. The DN values of these known classes are determined to check the homogeneity. In supervised classification, an analyst uses previously acquired knowledge of an area, or a priori knowledge, to locate specific areas, or training sites, which represent homogeneous samples of known land use and/or land cover types. These classes are interactively marked on the digital image in the form of polygons. Based on statistics of these training sites, each pixel in an image is then assigned to a user-defined land use type (e.g., residential, industrial, agriculture, etc.) or land cover type (e.g., forest, grassland, snow cover, etc.).

In the second stage, the training pixels are used by the software to derive various statistics for each class, and are correspondingly assigned signatures. These samples are referred to as training areas or samples. In supervised classification, the analyst identifies the homogeneous representative samples of different land use and land cover types (information classes) of interest on the imagery. Information classes imply that the actual land cover of the area under consideration which the analyst wants to classify, like vegetation cover, agriculture, urban or water bodies. A histogram for each band of the training areas/samples can be drawn. The normal histogram with a single peak would indicate a single class but a bimodal response would indicate two class present in the training pixels. In case, some training pixels are to be dropped or new added, it can also be done here itself. Thus, the classification of image data may be improved if each of the class in training sample has one single peak in the histogram.

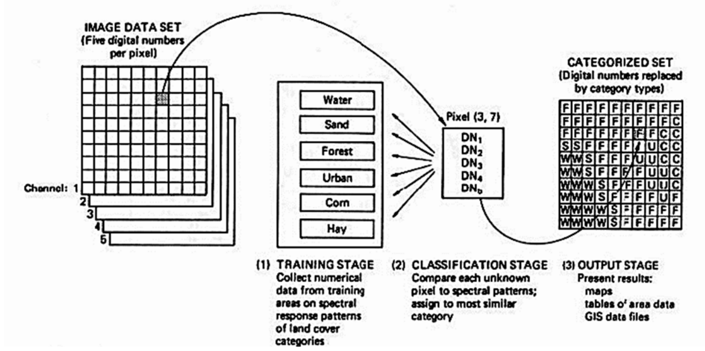


Figure 5.44 Steps in supervised classification (Alakhouri, 2014)

Training process helps in collection of statistical parameters which describe the spectral response pattern for the information classes that are being considered for image classification. Thus, a sufficient number of homogeneous training samples are required for each class to represent the tonal variation present within each class in the image. The variation in spectral reflectances of a training set per class in multispectral bands is used to derive the statistics. It is important that the training sites are well distributed throughout the image, as far possible, as they are the representative samples for the entire image. Training sets also increase the chance to incorporate the site specific variations of that class.

One of the important points in supervised classification is to avoid overlap in the training spectra for each class through the spectral plot. It is therefore important to identify the combination of bands that might be best for discrimination of classes through scatter plots between two spectral bands. The scatter plot is a graphical representation of DN values of two bands data; one on x-axis and another on y-axis. The scatter plots provide information related with the spectral separability. The training sample size is also important in classification which might vary from a minimum of 10N -100N per class, where N is the number of bands. The selection of accurate training areas is based on the analyst's familiarity and knowledge with the geographical area and actual surface cover types obtained through site surveys and present in the image. For each class, several training samples are identified from the images.

In the third stage, the remaining pixels of the image are allocated to the same class with which they show greatest similarity based on the established signature files in the second stage. Once the statistical characteristics are computed for each information class, the image is classified using any method, like *Parallelepiped classifier, Minimum distance* or *Maximum likelihood classifier*. For details of these techniques, refer Garg (2022). For the training samples, digital values (spectral signature) and their statistics in all spectral bands are used to "train" the software, and to recognize spectrally similar areas for each class. The known spectral signature for each class is then compared by the software to the matching signatures of unknown pixels and labeled the class as soon as it closely resembles the signature. This process is repeated for each unknown/unclassified pixel in the image. Figure 5.45 summarizes the supervised classification procedure in the form of a flow diagram.

It is important to note that training sites developed in one scene may or may not be replicable to entire study area due to variation in ground objects, ground conditions, illumination conditions, or atmospheric effects, which may change from one area to another. Similarly, training samples may not be usable directly across time due to lighting conditions, cloud cover response and due to growth of various vegetation types. At the end of classification, if a particular class has not been picked up properly by the software, the image is again classified by refining the training areas/samples for that class. The process allows interpreter to refine the training areas/samples several times, till a satisfactory result is obtained. It is an iterative process, and the analyst "supervises" the classification of images into a defined set of classes. That's why it is known as supervised classification method. Thus, in a supervised classification, the information classes are identified which are then used to determine the spectral classes representing them. Accuracy of supervised classification results would depend entirely on the collection of a sufficient number of training sites, and purity of samples.

(B) Unsupervised classification

This method is almost opposite to the supervised classification process. The unsupervised classification is very useful technique to classify the remotely sensed image where the field data/reference data are not available. Here, the image is classified purely based on the spectral variations of the classes to identify the major classes that already existed in the image. Unsupervised classification techniques do not require training sample signatures, prior to analysis of the scene. Statistical algorithms group DN values with similar pixels into various spectral classes, and later analyst will identify or combine these spectral classes into information classes (Jensen 2005). Spectral classes are grouped, solely based on DN values in the data, and then these DNs are matched by the analyst to information classes. Several clustering algorithms are used to determine the natural (statistical) groupings in the image data. The basic assumption in unsupervised classification is that a particular land cover in a scene would form a single cluster. Thus, the algorithm classifies the pixel data based on similar properties of the data itself. Figure 5.45 shows various steps involved in the unsupervised classification procedure.

In unsupervised classification, only major land classes are separated as clusters, while for smaller classes it may be not be possible. The decision for the number of clusters may be based on histogram analysis of the reflectance values. Usually, the analyst specifies the number of groups or clusters (classes) to be identified from the scene. The analyst also specifies the parameters related to the separation distance among the clusters, number of iterations required, and the variation (like standard deviation) within each cluster, as input to software. Often, the number of peaks as seen in the histogram can also be considered as the number of clusters present in the scene. The iterative clustering process may result into some clusters that the analyst wants to subsequently combine, or clusters that should be split further, based on ancillary/reference data available for the site. Alternatively, the complete process can be re-started by changing the input parameters into the software, till a satisfactory result is obtained. Thus, unsupervised classification is faster and less dependent on human intervention.

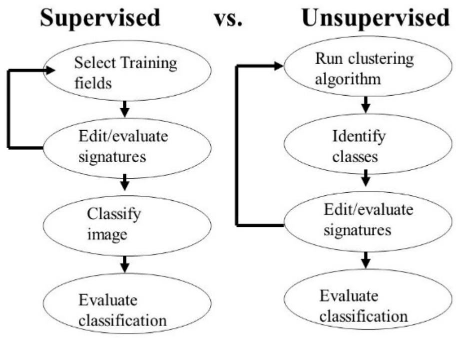


Figure 5.45 Broad steps involved in supervised classification and unsupervised classification procedures

There are two most popular clustering methods used for unsupervised classification; K-means and Iterative Self-Organizing Data Analysis Technique (ISODATA). These two methods rely purely on spectrally pixel-based statistics and require no prior knowledge of the characteristics of the themes being studied. In K-means approach, classes are determined statistically by assigning the pixels to the nearest cluster mean based on all available bands. The result of the K-means clustering could be influenced by the number of cluster centers specified, the choice of the initial cluster centre, the sampling nature, properties of the data, and clustering parameters. The ISODATA method uses the minimum spectral distance to assign a cluster to each pixel. The method requires a specified number of arbitrary cluster means or the means of existing signatures. It then processes the data repetitively, so that those means shift to the means of the clusters in the data. The input to ISODATA is number of clusters: 10 to 15 per desired land cover class, convergence threshold: percentage of pixels whose class values should not change between iterations; generally, set to 95%, and the maximum number of iterations: ideally, the convergence threshold should be reached. It should set "reasonable" parameters so that convergence is reached before iterations run out. In the iterations, pixels assigned to clusters with closest spectral mean; mean recalculated; pixels reassigned. The process continues until maximum iterations or convergence threshold reached. A graphical example is shown in Figure 5.46, where left graph shows the results of clustering by ISODATA after first iteration, and right graph shows the results after the second iteration. One may see the changes in the size of clusters. This way iteration continues till the specified threshold value is matched.

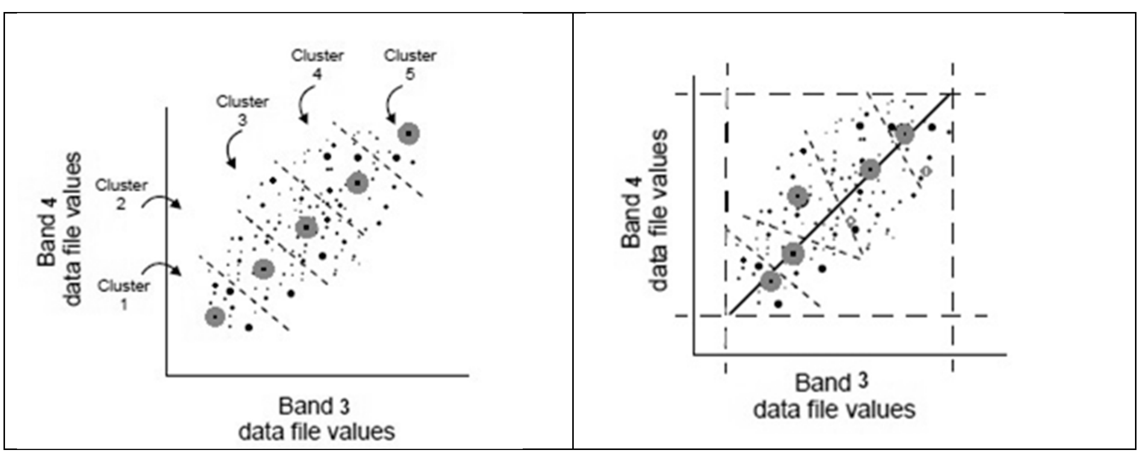


Figure 5.46 ISODATA Clustering techniques, result after (left) first iteration, and (right) after second iteration (Richards, 2013)

The supervised technique has some advantage over the unsupervised approach, as in supervised approach, information categories are distinct first, and then their spectral separability is examined while in the unsupervised approach, the software determines spectrally separable classes based, and then defines their information values (Lillesand and Keifer 1994). Unsupervised classification is easy to apply, and does not require analyst specified training samples. It automatically converts raw image data into useful information as long as there is higher classification accuracy (Langley et al., 2001). But, the disadvantage is that the classification process has to be repeated, if new classes are added. Figure 5.47 presents the results of supervised and unsupervised classification.

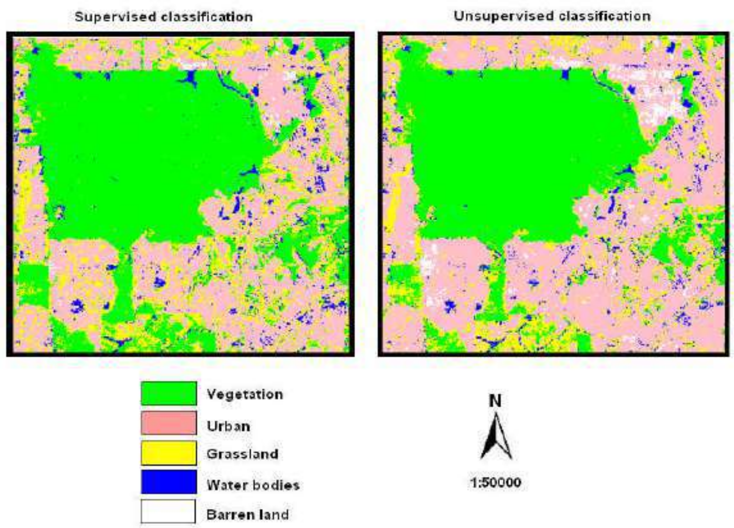


Figure 5.47 (a) Supervised, and (b) unsupervised classification of SPOT 5 image of the area (Ismail et.al., 2009)

While using high resolution images, it is still difficult to obtain satisfactory results by using supervised and unsupervised methods alone. Researchers have therefore developed advanced classification procedures to get high speed and better accuracy. These are summarised in Table 5.9.

Table 5.9 Various image classification techniques along with their salient characteristics (Garg, 2022)

Methods	**Examples**	**Characteristics**
Supervised	Maximum Likelihood, Minimum Distance, and Parallelepiped classification etc.	Analyst identifies the training sites to represent classes and each pixel is classified based on the statistical analysis
Unsupervised	ISODATA and K-means etc.	Prior ground information is not required. Pixels with similar spectral characteristics are grouped according to specific statistical criteria
Parametric	Maximum Likelihood classification and Unsupervised classification, etc.	Data are normally distributed, Prior knowledge of class density functions
Non-parametric	Nearest-neighbor classification, Fuzzy classification, Neural networks and support Vector machines, etc.	No prior assumptions are made
Non-metric	Rule-based Decision tree classification	Can operate on both real-valued data and nominal scaled data statistical analysis
Hard (parametric)	Supervised and Unsupervised classifications	Classification using discrete categories

Soft (non-parametric)	Fuzzy Set classification logic	Considers the heterogeneous nature of real-world, Each pixel is assigned a proportion of the in land cover type found within the pixel
Pre-pixel		Classification of the image pixel by pixel
Object-oriented		Image regenerated into homogenous objects, Classification preformed on each object and pixel
Hybrid approaches		Includes expert systems and artificial intelligence

5.17.4 Accuracy assessment

The increased use of remote sensing data and techniques has made analysis faster and more powerful, but the spectral and spatial complexity in the images have also created increased possibilities for errors. Thematic maps generated from remotely sensed image may contain some errors because of; (i) geometric error, (ii) atmospheric error, (iii) clusters incorrectly labeled after unsupervised classification, (iv) training sites incorrectly labeled for supervised classification, and (v) separability of classes.

Image classification is considered to be incomplete without estimating the accuracy of classification. It measures the agreement between a reference (assumed as 100% accurate) and a classified image of unknown accuracy. The accuracy assessment or error analysis is the quantitative comparison to validate the classified map with the actual reference data/image.

Accuracy assessment of land use or land cover classification obtained by remote sensing images is necessary to evaluate the quality of classification. Accuracy assessment is a critical step in the use of the results of analyses from remotely sensed data. A typical approach for validation and accuracy assessment is to use a statistically sound sampling design to select a sample of ground locations (number of pixels) in the scene. The land use or land cover classification assigned at these ground locations is actually compared with the true classification to ascertain the accuracy. An *error matrix* or *confusion matrix* is thus generated, as shown in Figure 5.48. An error matrix is a common tool to assess the accuracy of classification. Error matrix compares the pixels or polygons in a classified image against the ground reference data (Jensen 2005). A confusion matrix (or error matrix) is usually a quantitative method of characterising the image classification accuracy. It shows a correspondence between the classification results with respect to a reference image. The error matrix is created with the help of reference data which includes thematic map, aerial photos, ground surveys, GPS surveys and a high resolution satellite data used for classification.

To select the ground samples, five important sampling techniques are often used. These are:

1. ***Simple Random Sampling***: observations are randomly placed, and no rules are used. It is done using a completely random process.
2. ***Systematic Sampling***: observations are placed at equal intervals according to a planned strategy.
3. ***Stratified Random Sampling***: a minimum number of observations are randomly placed in each class. **Sampling** points are generated proportionate to the distribution of classes in the image
4. ***Stratified Systematic Un-aligned Sampling***: a grid is laid out which provides even distribution of randomly placed observations.
5. ***Cluster Sampling***: randomly placed “centroids” used as a base of several nearby observations. The nearby observations can be randomly selected, systematically selected, etc.

		Reference image			
		A	B	C	Total
Classified image	a	37	3	7	∑a = 47
	b	9	25	5	∑b = 39
	c	11	2	43	∑c = 56
	Total	∑A = 57	∑B = 30	∑C = 55	N = 142

Figure 5.48 Error matrix or Confusion matrix

It is not practically possible to test each and every pixel in the classification image, and therefore a representative sample of reference points in the image with known class values is used. Ground reference pixels earlier used to train the classification algorithm are not used now for the assessment of classification accuracy. Normally half of the training samples are used for classification, and remaining half for estimating the accuracy. The accuracy of a classification can be defined as; (i) Overall accuracy, (ii) Producer's accuracy, (iii) User's accuracy, and (iv) Kappa coefficient.

The overall accuracy of the classification map is determined by dividing the total correct pixels (sum of the major diagonal) by the total number of pixels in the error matrix. The total number of correct pixels in a category is divided by the total number of pixels of that category as derived from the reference data (i.e., the column total). It indicates the probability of a reference pixel being correctly classified, and is a measure of *omission error*. This statistics is called the *producer's accuracy* because the producer (the analyst) of the classification is interested in how well a certain area can be classified. Thus, the errors in classification can further be obtained individually for each class. The classification accuracy for each class indicates if there is any over-estimation or under-estimation in classification based on the input from reference data.

Columns of the table in Figure 5.48 are the reference (ground truth) classes, while rows are the classes of classified image whose accuracy is to be assessed. Various cells of the Table show number of pixels for all the possible correlations between the ground truth and the classified image. The diagonal elements of the matrix are highlighted which contains the number of correctly identified pixels. Classification overall accuracy is obtained by dividing the sum of these diagonal pixels by the total number of pixels. It is computed as:

$$\{(37 + 25 + 43) / 142\} \; 100 \approx 0.74 \qquad (5.11)$$

The overall accuracy measures the accuracy of the entire image without reference to the individual categories. It is sensitive to differences in sample size, and biased towards classes with larger samples. In addition, the accuracy of individual class needs to be assessed. The non-diagonal cells in the matrix contain classification errors, i.e., the number of pixels in reference image and the classified image don't match. There are two types of errors: under-estimation (omission errors) and over-estimation (commission errors).

If the total number of correct pixels in a class is divided by the total number of pixels that were actually classified in that category, the result is a measure of *commission error*. This measure, called the *user's accuracy* or *reliability*, is the probability that a pixel classified on the map actually represents that category on the ground. For any class, error of commission occurs when a classification procedure assigns pixels to a certain class that don't belong to it in reality. Number of pixels incorrectly allocated to a class is found in column cells of the class above and below the main diagonal. The sum of these yellow is divided by the total number of class pixels to get the commission error for class A:

$$\{(9 + 11) / 57\} \; 100 \approx 35\% \qquad (5.12)$$

The Producer's accuracy is the number of correctly identified pixels divided by the total number of pixels in the reference image. For class A, it is:

$$(37 / 57) \; 100 \approx 65\% \qquad (5.13)$$

The *Commission error* quantifies the Producer's accuracy. The sum of Error of Commission and Producers Accuracy for class A is 100%.

The probability a reference pixel is being properly classified measures the "Omission error" (reference pixels improperly classified are being omitted from the proper class). For any class, error of omission occurs when pixels that in fact belong to one class, are included into other classes. In confusion matrix, number of such pixels is found in the row cells to the left and to the right from the main diagonal. The sum of these orange cells is divided by the sum by the total number of pixels in the classified image to compute omission error:

$$\{(3 + 7) / 47\} \; 100 \approx 21\% \qquad (5.14)$$

User's accuracy is the number of the correctly identified pixels of a class, divided by the total number of pixels in that class in the classified image. For class A, it is computed as:

$$(37 / 47) \; 100 \approx 79\% \qquad (5.15)$$

The Omission error quantifies the User's accuracy. The sum of Error of Omission and User's Accuracy for class A is 100%.

The overall accuracy of classification can sometimes be misleading, as it does not reveal if error was evenly distributed between the classes or if some classes were really badly classified and some really good. Therefore, it is always better to compute the values of commission and omission errors. It is likely that the overall accuracy of an image classification might be quite high, whereas an individual class may have a low accuracy. If the accuracy of individual classes is most important to users, the classification results can't be accepted as such, even if the overall accuracy is coming out be higher. Additionally, the producer's and user's accuracy can't be separately used as an indicator of the classification accuracy, as these values do not provide the complete details.

Unit Summary

This unit discusses various remote sensing data products and their utilisation. It focusses mainly on the use of optical satellite images for carious purposes. The technical terms used in remote sensing are defined. Various comonents of remote sensing and interaction of EMR with the atmosphere are discussed. Laws governing black body are described. Resolutions are very important while dealing with the satellite images, so these have been detailed out in the unit.