REMOTE SENSING AND GIS APPLICATIONS IN AGRICULTURE

ANIL RAI
Indian Agricultural Statistics Research Institute
Library Avenue, New Delhi-110 012
anilrai@iasri.res.in

1. Introduction

The number of satellite missions dedicated to remote sensing (or Earth Observation (EO)) has increased significantly over the past decade and will further increase over the coming decade and beyond. Data from these missions offer the potential for contributing to the security of human existence on Earth in different ways. Although there have been many demonstrations of the value of EO satellites to development issues such as food production, resource management and environment characterization, it is recognized that there is a role for transferring that knowledge to developing countries, and it is hoped that this publication will stimulate, encourage and assist countries inexperienced in the use of satellite EO data to acquire the necessary expertise to exploit and realize its full potential.

The potential for social and economic benefits offered by satellite EO arise from its unique capabilities. These include the ability to provide near real-time monitoring of extensive areas of the Earth's surface at relatively low cost, as well as the capability to focus on particular land and sea surface features of interest to provide detailed, localized information. In some applications, satellite EO can offer an alternative source for data, which could be acquired by terrestrial or airborne surveying, but in a more timely and less expensive manner. In others, the availability of satellite EO data can provide a unique solution, for example where other techniques would be impractical.

The raw data from satellite EO often requires complex processing both to correct for atmospheric and geometric distortions and to derive information from the data and imagery. Additional data, such as that from in-situ sources, is sometimes used to supplement the EO data in this processing. Such data is particularly useful for calibrating or validating models.

Developments in computer technology over the past few years have resulted in the availability, at relatively low cost, of compact, high performance computers, which are well suited to the demands of satellite EO data processing. Together with the emergence of a range of commercial Geographical Information System (GIS) packages and other software tools for the manipulation of spatially referenced datasets, this has facilitated the emergence of a range of new applications of satellite EO data which have been developed or have entered operational service over the past decade.

2. Remote Sensing

The transport of information from an object to a receiver (observer) by means of radiation transmitted through the atmosphere. The interaction between the radiation and the object of interest conveys information required on the nature of the object (*eg*., reflection coefficient, emittance, roughness).

Examples

- The reflection of sunlight from vegetation will give information on the reflection coefficient of the object and its spectral variation, and thus on the nature of the object (green trees, *etc*.).
- Microwave radiation transmitted from a radar system and scattered from a rain cloud in the back direction to a receiver will give information on the raindrop size and intensity.

2.1 Passive and Active Sensing

The first example above is an example of passive remote sensing, where the reflected radiation observed originates from a natural source- the sun. The second example is an example of active remote sensing, where the scattered radiation originates from a specially designed active radar system.

2.2 Electromagnetic Radiation

Radiation can be observed either as a wave motion, or as single discrete packets of energy, photons. The two descriptions are not really contradictory. The energy is emitted as photons, but its statistical distribution over time is described by a wave.

The energy E of a photon is given by:

$$E = hv \qquad \ldots(1)$$

$$= \frac{hc}{\lambda} \qquad \ldots(2)$$

where c, v and λ are the velocity, frequency and wavelength of the radiation respectively, and h is Plank's constant.

Normally, one is dealing with a large number of photons arriving in a short time, and the radiation can be treated physically as a wave motion. However, in the visible and ultraviolet regions, very weak sources are typified by the detection of single photons. The wave theory of radiation has been developed extensively. It impacts on remote sensing in the way that radiation is reflected at a surface and transmitted, absorbed and scattered in a medium.

2.3 The Electromagnetic Spectrum

Electromagnetic radiation covers a very large range of wavelengths. In Remote Sensing we are concerned with radiation from the ultraviolet (UV), which has wavelengths of from 0.3 to 0.4 um (10^-6 m) to radar wavelengths in the region of 10 cm (10^-l m) (see Fig. 1 below). Thus the phenomena observed in the various wavelength regions differ considerably.

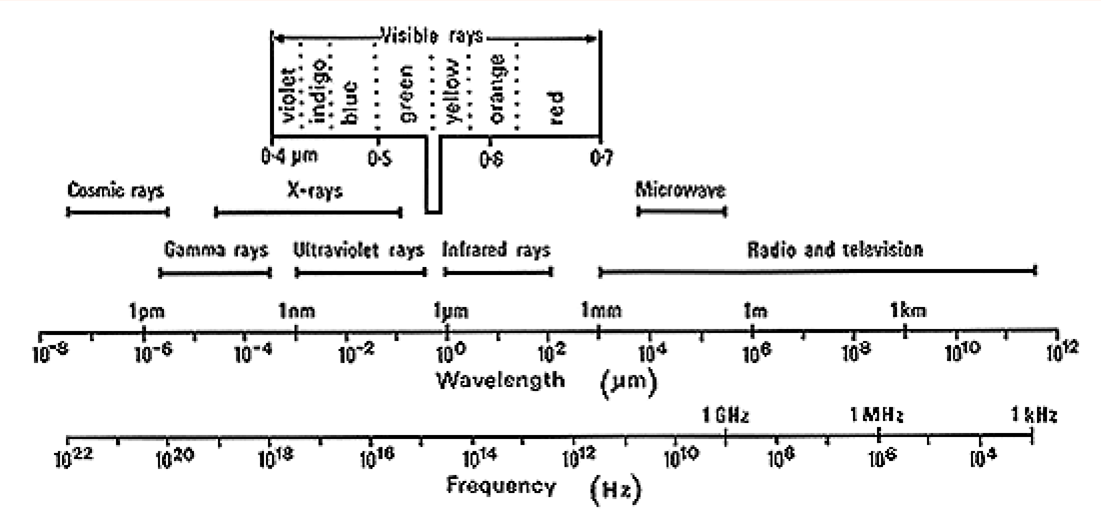

Fig. 1. Range of electromagnetic wavelengths and the transmission through the atmosphere

2.4 Reflection from Vegetation

Healthy, growing vegetation appears green because there is selective absorption in chlorophyll bands outside the green wavelengths. The absorption is only moderate so that the green light is reflected and scattered at the cellular boundaries to appear green both in reflection and transmission. Because of multiple reflections emergent natural light is non-polarized. At wavelengths beyond about 0.65 us the reflection becomes strong, indicating strong absorption in the leaf, as shown in Fig. 2.

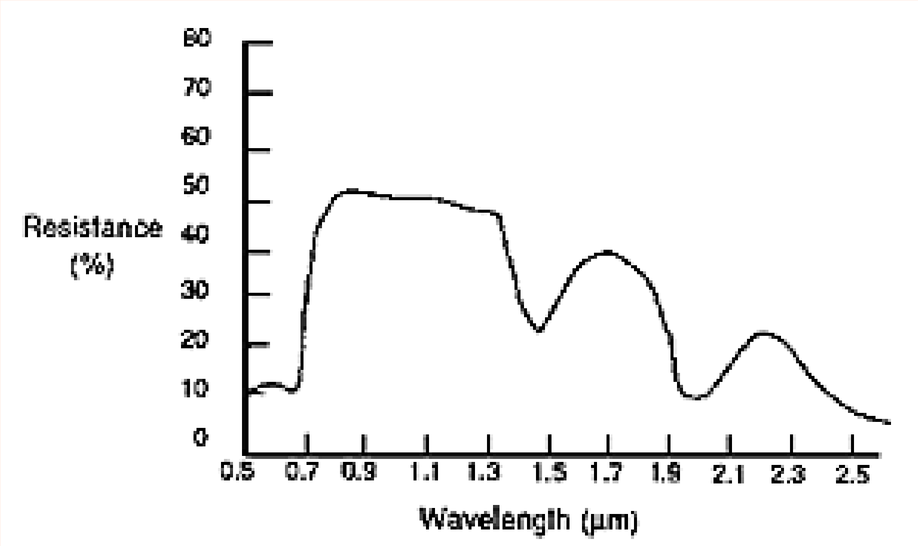

Fig.2. Wavelength dependence of reflectance of a soybean leaf

2.5 Data Resolutions

Remotely sensed data provide a synoptic or regional view of the Earth's surface as well as the opportunity to identify particular features of interest. Analysis techniques frequently relate particular data values in an image to certain ground features, or to parameters, which identify those features. However, the data acquisition methods of remote sensing implicitly involve at least one level of indirection. For example, a particular study may aim to determine vegetative cover and condition. Since such parameters are not directly measurable using remote sensing, they must be related to a property of vegetation which can be 'measured' remotely, namely reflectance. A further limitation which must be considered is that the data we collect using remote sensing only sample the potential range of measurements in the selected 'measurement space'.

Resolution refers to the intensity or rate of sampling, and extent refers to the overall coverage of a data set. Extent can be seen as relating to the largest feature, or range of features, which can be observed, while resolution relates to the smallest. For a feature to be distinguishable in the data, the resolution and extent of the measurement dimensions of the data set need to be appropriate to the measurable properties of the feature. For a feature to be separable from other features, these measurements must also be able to discriminate between the differences in reflectance from the features.

Spectral: As indicated in the preceding sections, different materials respond in different, and often distinctive, ways to EM radiation. This means that a specific spectral response curve, or spectral signature, can be determined for each material type. Basic categories of matter (such as specific minerals) can be identified on the basis of their spectral signatures alone, but may require that the spectra be sufficiently detailed in terms of wavelength intervals and covers a wide spectral range. Composite categories of matter (such as soil which contains several different minerals) however, may not be uniquely identifiable on the basis of spectral data alone.

Spatial: Spatial resolution defines the level of spatial detail depicted in an image. This may be described as a measure of the smallness of objects on the ground that may be distinguished as separate entities in the image, with the smallest object necessarily being larger than a single pixel. In this sense, spatial resolution is directly related to image pixel size. In terms of photographic data, an image pixel may be compared to grain size while spatial resolution is more closely related to photographic scale. In practical terms, the 'detectability' of an object in an image involves consideration of spectral contrast as well as spatial resolution. Feature shape is also relevant to visual discrimination in an image with long thin features such as roads showing up more readily than smaller symmetric ones. Pixel size is usually a function of the platform and sensor, while the detectability may change from place to place and time to time.

Radiometric: Radiometric resolution in remotely sensed data is defined as the amount of energy required to increase a pixel value by one quantisation level or 'count'. The radiometric extent is the dynamic range or the maximum number of quantisation levels that may be recorded by a particular sensing system. Most remotely sensed imagery is recorded with quantisation levels in the range 0&shyp;255, that is, the minimum 'detectable' radiation level is recorded as 0 while the 'maximum' radiation is recorded as 255. This range is also referred to as 8 bit resolution since all values in the range may be represented by 8 bits (binary digits) in a computer. Radiometric resolution in digital imagery is comparable to the number of tones in a photographic image &shyp; both measures being related to image contrast.

Temporal: The temporal resolution of remotely sensed data refers to the repeat cycle or interval between acquisitions of successive imagery. This cycle is fixed for spacecraft platforms by their orbital characteristics, but is quite flexible for aircraft platforms.

Satellites offer repetitive coverage at reduced cost but the rigid overpass times can frequently coincide with cloud cover or poor weather. This can be a significant problem when field work needs to coincide with image acquisition. While aircraft data are necessarily more expensive than satellite imagery, these data offer the advantage of user-defined flight timing, which can be modified if necessary to suit local weather conditions. The off-nadir viewing capability of the SPOT& shyp; HRV provides some flexibility to the usual repeat cycle of satellite imagery by imaging areas outside of the nadir orbital path. This feature allows daily coverage of selected regions for short periods and has obvious value for monitoring dynamic events such as flood or fire.

3. Digital Image Processing

The roots of remote sensing reach back into ground and aerial photography. But modern remote sensing really took off as two major technologies evolved more or less simultaneously: 1) the development of sophisticated electro-optical sensors that operate from air and space platforms and 2) the digitizing of data that were then in the right formats for processing and analysis by versatile computer-based programs. Today, analysts of remote sensing data spend much of their time at computer stations, but nevertheless still also use actual imagery (in photo form) that has been computer-processed.

Now it can be seen that the individual bands and color composites that have introduced in the previous lectures and it is interesting to investigate the power of computer-based processing procedures in highlighting and extracting information about scene content, that is, the recognition, appearance, and identification of materials, objects, features, and classes (these general terms all refer to the specific spatial and spectral entities in a scene).

Processing procedures fall into three broad categories: Image Restoration (Preprocessing); Image Enhancement; and Classification and Information Extraction. Apart from preprocessing, the techniques of contrast stretching, density slicing, and spatial filtering are discussed. Under Information Extraction, ratioing and principal components analysis have elements of enhancement but lead to images that can be interpreted directly for recognition and identification of classes and features. Also included in the third category but not discussed is Change Detection and Pattern recognition.

The data in satellite remote sensing is in the form of Digital Number or DN. It is said that the radiances, such as reflectance and emittances, which vary through a continuous range of values are digitized onboard the spacecraft after initially being measured by the sensor(s) in use. Ground instrument data can also be digitized at the time of collection. Or, imagery obtained by conventional photography is capable of digitization. A DN is simply one of a set of numbers based on powers of 2, such as 2^6 or 64. The range of radiances, which instrument-wise, can be, for example, recorded as varying voltages if the sensor signal is one which is, say, the conversion of photons counted at a specific wavelength or wavelength intervals. The lower and upper limits of the sensor's response capability form the end members of the DN range selected. The voltages are divided into equal whole number units based on the digitizing range selected. Thus, a IRS band can have its voltage values - the maximum and minimum that can be measured - subdivided into 2^8 or 256 equal units. These are arbitrarily set at 0 for the lowest value, so the range is then 0 to 255.

Preprocessing

Preprocessing is an important and diverse set of image preparation programs that act to offset problems with the band data and recalculate DN values that minimize these problems. Among the programs that optimize these values are atmospheric correction (affecting the DNs of surface materials because of radiance from the atmosphere itself, involving attenuation and scattering); sun illumination geometry; surface-induced geometric distortions; spacecraft velocity and attitude variations (roll, pitch, and yaw); effects of Earth rotation, elevation, curvature (including skew effects), abnormalities of instrument performance (irregularities of detector response and scan mode such as variations in mirror oscillations); loss of specific scan lines (requires destriping), and others. Once performed on the raw data, these adjustments require appropriate radiometric and geometric corrections.

Resampling: Resampling is one approach commonly used to produce better estimates of the DN values for individual pixels. After the various geometric corrections and translations have been applied, the net effect is that the resulting redistribution of pixels involves their spatial displacements to new, more accurate relative positions. However, the radiometric values of the displaced pixels no longer represent the real world values that would be obtained if this new pixel array could be re-sensed by the scanner (this situation is alleviated somewhat if the sensor is a Charge-Coupled Device [CCD. The particular mixture of surface objects or materials in the original pixel has changed somewhat (depending on pixel size, number of classes and their proportions falling within the pixel, extent of continuation of these features in neighboring pixels [a pond may fall within one or just a few pixels; a forest can spread over many contiguous pixels]). In simple words, the corrections have led to a pixel that at the time of sampling covered ground A being shifted to a position that have A values but should if properly located represent ground B.

An estimate of the new brightness value (as a DN) that is closer to the B condition is made by some mathematical re-sampling technique. Three sampling algorithms are commonly used:

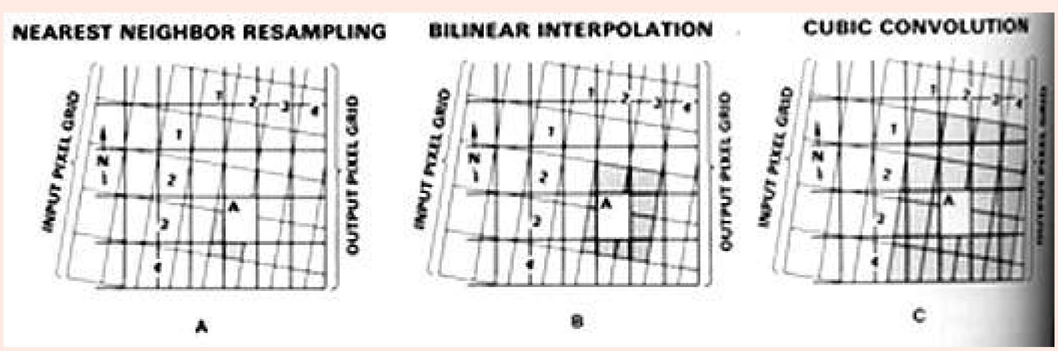

Fig. 3. Sampling algorithms

In the Nearest Neighbour technique, the transformed pixel takes the value of the closest pixel in the pre-shifted array. In the Bilinear Interpolation approach, the average of the DNs for the 4 pixels surrounding the transformed output pixel is used. The Cubic

Convolution technique averages the 16 closest input pixels; this usually leads to the sharpest image.

False Colour Composite: The first example of a colour composite, made by combining (either photographically or with a computer-processing program) any three bands of images with some choice of color filters, usually blue, green, and red. The customary false color composite made by projecting a green band image through a blue filter, a red band through green, and the photographic infrared image through a red filter.

True Colour View: By projecting IRS Bands 1, 2, and 3 through blue, green, and red filters respectively, a quasi-true color image of a scene can be generated.

In practice, we use various color mapping algorithms to facilitate visual interpretation of an image, while analytical treatment usually works with the original DN (digital number) values of the pixels. The original DN values contain all of the information in the scene and though their range of values may make it necessary to re-map them to create a good display, it doesn't add information. In fact, although visual interpretation is easier with the remapped image, re-mapping loses and distorts information thus, for analytical work, we use the original DN values or DN values translated to calibrated radiances.

With this mapping, we see a pleasing and satisfying image because it depicts the world in the general color ranges with which we are naturally familiar. We can imagine how this scene would appear if we were flying over it at a high altitude.

Other Colour Combinations: Other combinations of bands and color filters (or computer assignments) produce not only colorful new renditions but in some instances bring out or call attention to individual scene features that, although usually present in more subtle expressions in the more conventional combinations, now are easier to spot and interpret.

Contrast Stretching and Density Slicing

Almost without exception, the image will be significantly improved if one or more of the functions called Enhancement are applied. Most common of these is contrast stretching. This systematically expands the range of DN values to the full limits determined by byte size in the digital data. For IRS this is determined by the eight-bit mode or 0 to 255 DNs. Examples of types of stretches and the resulting images are shown. Density slicing is also examined. We move now to two of the most common image processing routines for improving scene quality. These routines fall into the descriptive category of Image Enhancement or Transformation. We used the first image enhancer, *contrast stretching*, to enhance their pictorial quality. Different stretching options are described next, followed by a brief look at density slicing. We will then evaluate the other routine, filtering, shortly. The contrast stretching, which involves altering the distribution and range of DN values, is usually the first and commonly a vital step applied to image enhancement. Both casual viewers and experts normally conclude from direct observation that modifying the range of light and dark tones (gray levels) in a photo or a computer display is often the single most informative and revealing operation performed on the scene. When carried out in a photo darkroom during negative and printing, the process involves shifting the gamma (slope) or film transfer function of the plot of density versus exposure (H-D curve). This is done by changing one or more variables in the photographic process, such as, the type of recording film, paper contrast, developer conditions, etc. Frequently the result is a sharper, more pleasing picture, but certain information may be lost through trade-offs, because gray levels are "overdriven" into states that are too light or too dark.

Contrast stretching by computer processing of digital data (DNs) is a common operation; although we need some user skill in selecting specific techniques and parameters (range limits). The reassignment of DN values is based on the particular stretch algorithm chosen (see below). Values are accessed through a Look-Up Table (LUT).

The fundamental concepts that underlie how and why contrast stretching is carried out are summarized in Fig. 4:

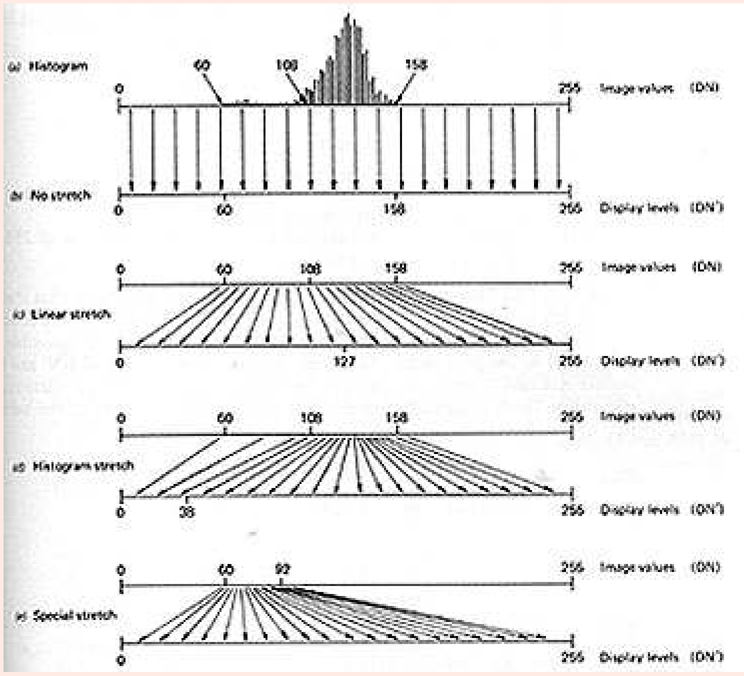

Fig. 4. Contrast stretching
From Lillesand & Kiefer, Remote Sensing and Image Interpretation, 4th Ed., 1999

In the top plot (a), the DN values range from 60 to 158 (out of the limit available of 0 to 255). But below 108 there are few pixels, so the effective range is 108-158. When displayed without any expansion (stretch), as shown in plot b, the range of gray levels is mostly confined to 40 DN values, and the resulting image is of low contrast - rather flat.

In plot c, a linear stretch involves moving the 60 value to 0 and the 158 DN to 255; all intermediate values are moved (stretched) proportionately. This is the standard linear stretch. But no accounting of the pixel frequency distribution, shown in the histogram, is made in this stretch, so that much of the gray level variation is applied to the scarce to absent pixels with low and high DNs, with the resulting image often not having the best contrast rendition. In d, pixel frequency is considered in assigning stretch values. The 108-158 DN range is given a broad stretch to 38 to 255 while the values from DN 107 to 60 are spread differently - this is the histogram-equalization stretch. In the bottom example, e, some specific range, such as the infrequent values between 60 and 92, is independently stretched to bring out contrast gray levels in those image areas that were not specially enhanced in the other stretch types.

Commonly, the distribution of DNs (gray levels) can be uni-modal and may be Gaussian (distributed normally with a zero mean), although skewing is usual. Multi-modal distributions (most frequently, bimodal but also poly-modal) result if a scene contains two or more dominant classes with distinctly different (often narrow) ranges of reflectance. Upper and lower limits of brightness values typically lie within only a part (30 to 60%) of the total available range. The (few) values falling outside 1 or 2 standard deviations may usually be discarded (histogram trimming) without serious loss of prime data. This trimming allows the new, narrower limits to undergo expansion to the full scale (0-255 for IRS data).

Linear expansion of DN's into the full scale (0-255) is a common option. Other stretching functions are available for special purposes. These are mostly nonlinear functions that affect the precise distribution of densities (on film) or gray levels (in monitor image) in different ways, so that some experimentation may be required to optimize results. Commonly used special stretches include:1) Piecewise Linear, 2) Linear with Saturation 3) Logarithmic, 4) Exponential 5) Ramp Cumulative Distribution Function, 6) Probability Distribution Function, and 7) Sinusoidal Linear with Saturation.

Spatial Filtering

Just as contrast stretching strives to broaden the image expression of differences in spectral reflectance by manipulating DN values, so spatial filtering is concerned with expanding contrasts locally in the spatial domain. Thus, if in the real world there are boundaries between features on either side of which reflectance (or emissions) are quite different (notable as sharp or abrupt changes in DN value), these boundaries can be emphasized by any one of several computer algorithms (or analog optical filters). The resulting images often are quite distinctive in appearance. Linear features, in particular, such as geologic faults can be made to stand out. The type of filter used, high- or low-pass, depends on the spatial frequency distribution of DN values and on what the user wishes to accentuate.

Another processing procedure falling into the enhancement category that often divulges valuable information of a different nature is *spatial filtering*. Although less commonly performed, this technique explores the distribution of pixels of varying brightness over an image and, especially detects and sharpens boundary discontinuities. These changes in scene illumination, which are typically gradual rather than abrupt, produce a relation that we express quantitatively as "spatial frequencies". The spatial frequency is defined as the number of cycles of change in image DN values per unit distance (e.g., 10 cycles/mm) along a particular direction in the image. An image with only one spatial frequency consists of equally spaced stripes (raster lines). For instance, a blank TV screen with the set turned on has horizontal stripes. This situation corresponds to zero frequency in the horizontal direction and a high spatial frequency in the vertical.

In general, images of practical interest consist of several dominant spatial frequencies. Fine detail in an image involves a larger number of changes per unit distance than the gross image features. The mathematical technique for separating an image into its various spatial frequency components is called Fourier analysis. After an image is separated into its components (done as a "Fourier Transform"), it is possible to emphasize certain groups (or "bands") of frequencies relative to others and recombine the spatial frequencies into an enhanced image. Algorithms for this purpose are called "filters" because they suppress (de-emphasize) certain frequencies and pass (emphasize) others. Filters that pass high frequencies and, hence, emphasize fine detail and edges, are called high pass filters. Low pass filters, which suppress high frequencies, are useful in smoothing an image, and may reduce or eliminate "salt and pepper" noise.

Convolution filtering is a common mathematical method of implementing spatial filters. In this, each pixel value is replaced by the average over a square area centered on that pixel. Square sizes typically are 3 x 3, 5 x 5, or 9 x 9 pixels but other values are acceptable. As applied in low pass filtering, this tends to reduce deviations from local averages and thus smoothes the image. The difference between the input image and the low pass image is the high pass-filtered output. Generally, spatially filtered images must be contrast stretched to use the full range of image display. Nevertheless, filtered images tend to appear flat.

Principal Components Analysis

There is a tendency for multiband data sets/images to be somewhat redundant wherever bands are adjacent to each other in the (multi-)spectral range. Thus, such bands are said to be correlated (relatively small variations in DNs for some features). A statistically based program, called Principal Components Analysis, decorrelates the data by transforming DN distributions around sets of new multi-spaced axes. The underlying basis of PCA is described in this section. Color composites made from images representing individual components often show information not evident in other enhancement products. Canonical Analysis and Decorrelation Stretching are also mentioned.

We are now ready to overview the last two types of image enhancement discussed in this article. Both are also suited to Information Extraction and Interpretation, but are treated separately from Classification (considered later in the Section). We will embark first on a quick run-through of images produced by Principal Components Analysis or PCA. PCA is a decorrelation procedure, which reorganizes by statistical means the DN values from as many of the spectral bands as we choose to include in the analysis. In producing these values, we used all bands and requested that all seven components be generated (the number of components is fixed by the number of bands, because they must be equal).

Next look at each of these components, keeping in mind that many of the tonal patterns in individual components do not seem to spatially match specific features or classes identified in the IRS bands, and represent linear combinations of the original values instead. We make only limited comments on the nature of those patterns that lend themselves to some interpretation.

Ratioing

Ratioing is an enhancement process in which the DN value of one band is divided by that of any other band in the sensor array. If both values are similar, the resulting quotient is a number close to 1. If the numerator number is low and denominator high, the quotient approaches zero. If this is reversed (high numerator; low denominator) the number is well above 1. These new numbers can be stretched or expanded to produce images with considerable contrast variation in a black and white rendition. Certain features or materials can produce distinctive gray tones in certain ratios. Three band ratio images can be combined as color composites, which highlight certain features in distinctive colors. Ratio images also reduce or eliminate the effects of shadowing.

Another image manipulation technique is ratioing. For each pixel, we divide the DN value of any one band by the value of another band. This quotient yields a new set of numbers that may range from zero (0/1) to 255 (255/1) but the majority are fractional (decimal) values between 0 and typically 2 - 3 (e.g., 82/51 = 1.6078...; 114/177 = 0.6440...). We can rescale these to provide a gray-tone image, in which we can reach 16 or 256 levels, depending on the computer display limits. One effect of ratioing is to eliminate dark shadows, because these have values near zero in all bands, which tends to produce a "truer" picture of hilly topography in the sense that the shaded areas are now expressed in tones similar to the sunlight sides.

Three pairs of ratio images can be co-registered (aligned) and projected as color composites. In individual ratio images and in these composites, certain ground features tend to be highlighted, based on unusual or anomalous ratio values.

Classification

This section deals with the process of classifying multispectral images into patterns of varying grey or assigned colors that represent either clusters of statistically different sets of multiband data (radiances expressed by their DN values), some of which can be correlated with separable classes/features/materials (Unsupervised Classification), or numerical discriminators composed of these sets of data that have been grouped and specified by associating each with a particular class, etc. whose identity is known independently and which has representative areas (training sites) within the image where that class is located (Supervised Classification). The principles involved in classification are mentioned briefly in this section. The approach to unsupervised classification is also described with examples and it is pointed out that many of the areas classified in the image by their cluster values may or may not relate to real classes (misclassification is a common problem).

There are two of the common methods for identifying and classifying features in images: Unsupervised and Supervised Classification. Closely related to Classification is the approach called Pattern Recognition.

Before starting, it is well to review several basic principles, with the aid of Fig. 5:

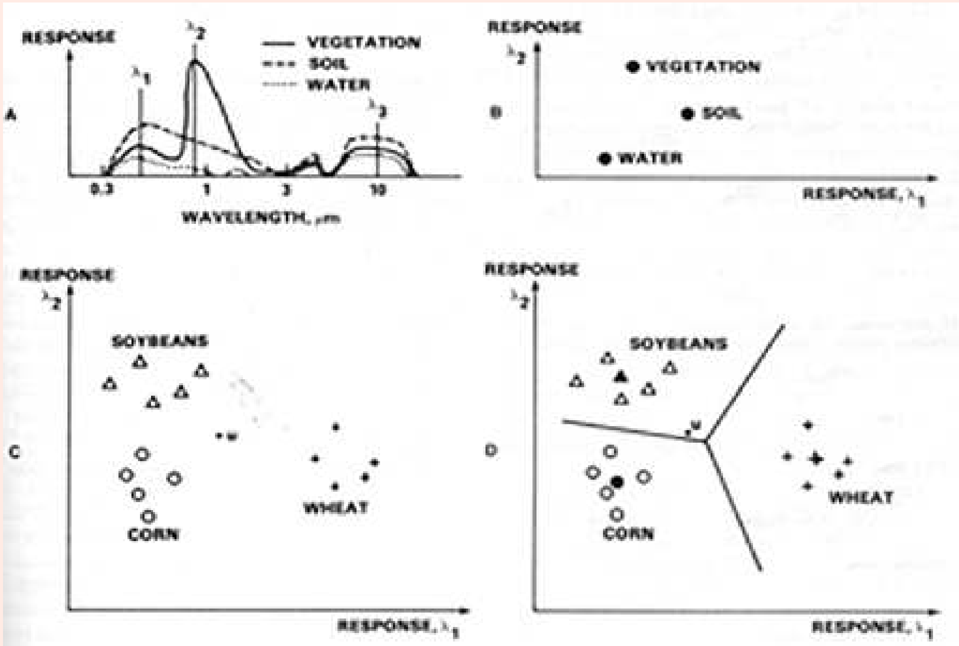

Fig. 5. Basic principles of classification

In the upper left are plotted spectral signatures for three general classes: Vegetation; Soil; Water. The relative spectral responses (reflectance in this spectral interval), in terms of some unit, *eg*., reflected energy in appropriate units or percent (as a ratio of reflected to incident radiation, times 100), have been sampled at three wavelengths. (The response values are normally converted [either at the time of acquisition on the ground or aircraft or spacecraft] to a digital format, the DNs or Digital Numbers cited before, commonly subdivided into units from 0 to 255 [2^8]).

For this specific signature set, the values at any two of these wavelengths are plotted on the upper right. It is evident that there is considerable separation of the resulting value points in this two-dimensional diagram. In reality, when each class is considered in terms of geographic distribution and/or specific individual types (such as soybeans versus wheat in the Vegetation category), as well as other factors, there will be usually notable variation in one or both chosen wavelengths being sampled. The result is a spread of points in the two-dimensional diagram (known as a scatter diagram), as seen in the lower left. For any two classes this scattering of value points may or may not overlap. In the case shown, which treats three types of vegetation (crops), they don't. The collection of plotted values (points) associated with each class is known as a *cluster*. It is possible, using statistics that calculate means, standard deviations, and certain probability functions, to draw boundaries between clusters, such that arbitrarily every point plotted in the spectral response space on each side of a boundary will automatically belong to the class or type within that space. This is shown in the lower right diagram, along with a single point "w" which is an unknown object or pixel (at some specific location) whose identity is being sought. In this example, w plots just in the soybean space.

Thus, the principle of classification (by computer image-processing) boils down to this: Any individual pixel or spatially grouped sets of pixels representing some feature, class, or material is characterized by a (generally small) range of DNs for each band monitored by the remote sensor. The DN values (determined by the radiance averaged over each spectral interval) are considered to be clustered sets of data in 2-, 3-, and higher dimensional plotting space. These are analyzed statistically to determine their degree of uniqueness in this spectral response space and some mathematical function(s) is/are chosen to discriminate the resulting clusters.

Two methods of classification are commonly used: Unsupervised and Supervised. The logic or steps involved can be grasped from Fig.6:

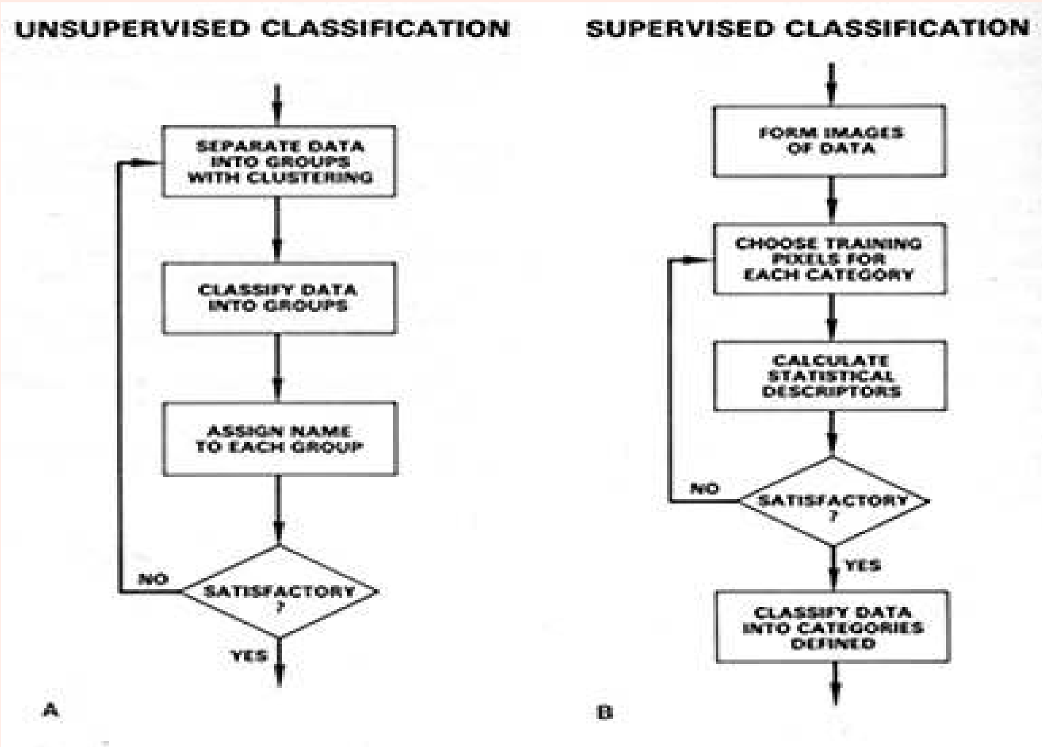

Fig. 6. Methods of classification

In unsupervised classification any individual pixel is compared to each discrete cluster to see which one it is closest to. A map of all pixels in the image, classified, as to which cluster each pixel is most likely to belong, is produced (in black and white or more commonly in colors assigned to each cluster. This then must be interpreted by the user as to what the color patterns may mean in terms of classes, etc. that are actually present in the real world scene; this requires some knowledge of the scene's feature/class/material content from general experience or personal familiarity with the area imaged. In supervised classification the interpreter knows beforehand what classes, etc. are present and where each is in one or more locations within the scene. These are located on the image, areas containing examples of the class are circumscribed (making them training sites), and the statistical analysis is performed on the multiband data for each such class. Instead of clusters then, one has class groupings with appropriate discriminant functions that distinguish each (it is possible that more than one class will have similar spectral values but unlikely when more than 3 bands are used because different classes/materials seldom have similar responses over a wide range of wavelengths). All pixels in the image lying outside training sites are then compared with the class discriminants, with each being assigned to the class it is closest to - this makes a map of established classes (with a few pixels usually remaining unknown) which can be reasonably accurate (but some classes present may not have been set up; or some pixels are misclassified.

Unsupervised Classification: In an unsupervised classification, the objective is to group multiband spectral response patterns into clusters that are statistically separable. Thus, a small range of digital numbers (DNs) for, say 3 bands, can establish one cluster that is set apart from a specified range combination for another cluster (and so forth). Separation will depend on the parameters we choose to differentiate. We can visualize this process with the aid of Fig. 7, taken from Sabins, "Remote Sensing: Principles and Interpretation." 2nd Edition, for four classes: A = Agriculture; D= Desert; M = Mountains; W = Water.

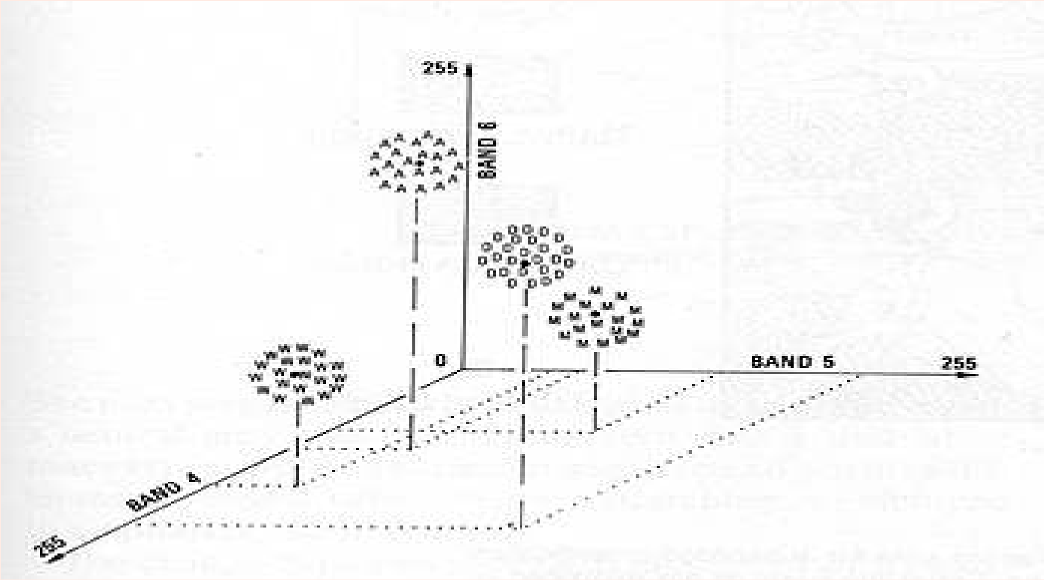

Fig. 7. Unsupervised classification

From F.F. Sabins, Jr., "Remote Sensing: Principles and Interpretation." 2nd Ed., © 1987. Reproduced by permission of W.H. Freeman & Co., New York City.

We can modify these clusters, so that their total number can vary arbitrarily. When we do the separations on a computer, each pixel in an image is assigned to one of the clusters as being most similar to it in DN combination value. Generally, in an area within an image, multiple pixels in the same cluster correspond to some (initially unknown) ground feature or class so that patterns of gray levels result in a new image depicting the spatial distribution of the clusters. These levels can then be assigned colors to produce a cluster map. The trick then becomes one of trying to relate the different clusters to meaningful ground categories. We do this by either being adequately familiar with the major classes expected in the scene, or, where feasible, by visiting the scene (ground truthing) and visually correlating map patterns to their ground counterparts. Since the classes are not selected beforehand, this latter method is called Unsupervised Classification.

The most of the image-processing program employs a simplified approach to Unsupervised Classification. Input data consist of the DN values of the registered pixels for the 3 bands used to make any of the color composites. Algorithms calculate the cluster values from these bands. It automatically determines the maximum number of clusters by the parameters selected in the processing. This process typically has the effect of producing so many clusters that the resulting classified image becomes too cluttered and, thus, more difficult to interpret in terms of assigned classes. To improve the interpretability, we first tested a simplified output and thereafter limited the number of classes displayed to 15 (reduced from 28 in the final cluster tabulation).

Supervised Classification: The principles behind Supervised Classification are considered in more detail. The fact that the pixel DNs for a specified number of bands are selected from areas in the scene that are a priori of known identity, i.e., can be named as classes of real features, materials, etc. allows establishment of training sites that become the basis of setting up the statistical parameters used to classify pixels outside these sites.

Supervised classification is much more accurate for mapping classes, but depends heavily on the cognition and skills of the image specialist. The strategy is simple: the specialist must recognize conventional classes (real and familiar) or meaningful (but somewhat artificial) classes in a scene from prior knowledge, such as, personal experience with the region, by experience with thematic maps, or by on-site visits. This familiarity allows the specialist to choose and set up discrete classes (thus supervising the selection) and the, assign them category names. The specialists also locate training sites on the image to identify the classes. Training Sites are areas representing each known land cover category that appear fairly homogeneous on the image (as determined by similarity in tone or color within shapes delineating the category). Specialists locate and circumscribe them with polygonal boundaries drawn (using the computer mouse) on the image display. For each class thus outlined, mean values and variances of the DNs for each band used to classify them are calculated from all the pixels enclosed in the site. More than one polygon can be established for any class. When DNs are plotted as a function of the band sequence (increasing with wavelength), the result is a spectral signature or spectral response curve for that class. In reality the spectral signature is for all of the materials within the site that interact with the incoming radiation. Classification now proceeds by statistical processing in which every pixel is compared with the various signatures and assigned to the class whose signature comes closest. A few pixels in a scene do not match and remain unclassified, because these may belong to a class not recognized or defined).

Many of the classes in general are almost self-evident ocean water, waves, beach, marsh, shadows. In practice, we could further sequester several such classes. For example, we might distinguish between ocean and bay waters, but their gross similarities in spectral properties would probably make separation difficult. Other classes that are likely variants of one another, such as, slopes that faced the morning sun as IRS flew over versus slopes that face away, might be warranted. Some classes are broad-based, representing two or more related surface materials that might be separable at high resolution but are inexactly expressed in the IRS image. In this category we can include trees, forests, and heavily vegetated areas (the golf course or cultivated farm fields).

Note that software does not name them during the stage when the signatures are made. Instead, it numbers them and names are assigned later. Several classes gain their data from more than one training site. Most of the software has a module that plots the signature of each class.

Minimum Distance Classification: One of the simplest supervised classifiers is the parallelopiped method. But we employ a (usually) somewhat better approach (in terms of greater accuracy) known as the Minimum Distance classifier. This sets up clusters in multidimensional space, each defining a distinct (named) class. Any pixel is then assigned to that class if it is closest to (shortest vector distance).

We initiate our exemplification of Supervised Classification by producing one using the Minimum_Distance routine. The software program acts on DNs in multidimensional band space to organize the pixels into the classes we choose. Each unknown pixel is then placed in the class *closest* to the mean vector in this band space We can elect to combine classes to have either color themes (similar colors for related classes) and/or to set apart spatially adjacent classes by using disparate colors.

Maximum Likelihood Classification: The most powerful classifier in common use is that of Maximum Likelihood. Based on statistics (mean; variance/covariance), a (Bayesian) Probability Function is calculated from the inputs for classes established from training sites. Each pixel is then judged as to the class to which it most probably belongs. This is done with the IRS data, using three reflected radiation bands. The result is a pair of quite believable classification maps whose patterns (the classes) seem to closely depict reality but keep in mind that several classes are not normal components of the actual ground scene, *eg*., shadows.

In many instances the most useful image processing output is a classified scene. This is because you are entering a partnership with the processing program to add information from the real world into the image you are viewing, in a systematic way, in which you try to associate names of real features or objects with the spectral/spatial patterns evident in individual bands, color composites, or PCI images. The most of the software are capable of producing both unsupervised and supervised classifications.

4. Geographic Information System

The Collation of data about the spatial distribution of significant properties of the earth's surface has been an important part of activities of organised societies from the ancient times to the present day, spatial data have been collected and collated by the surveyors, geographers, navigators, etc. and these were used to plan and make decisions about the future activities of the societies. As scientific study of the earth advanced, so the new material needed to be mapped. The developments in the assessment and the understanding of the natural resources- agriculture, soil science, ecology, geomorphology, land and geology that began in the nineteenth century have continued today, provided new material to be mapped. The need for spatial data and spatial analysis has not been restricted to earth scientist. Urban planners and cadastral agencies need detail information about the distribution of land and resources in town and cities. The collection and compilation of data and the publication of printed maps is a costly and time consuming business.

Consequently, the extraction of single theme from general purpose maps was prohibitively expensive as it requires redrawing map by hand. Since, most of the earth resources are highly correlated with each other the need was felt to overlay different thematic maps over each other for better understanding the various processes and activities on the earth surface, which was not possible through conventional technique, also, there was a serious difficulty to handle the tabular data or attribute data in conjunction with spatial features.

The developments in the field of computer technology have given new direction to handling and using spatial data for assessment, planning and monitoring. The concept of using the computers for making maps and analysing them was initiated with the SYMAP-Synagraphic mapping system, developed by Harvard School of Computer Graphics in the early 1970. Since then, there has been wide range of automated methods for handling maps using computers. The history of using computers for mapping and spatial analysis shows that there have been parallel developments in automated data capture, data analysis and presentation in several broadly related fields. All these efforts have been oriented towards the same sort of operation- namely to develop a powerful tools for collecting, storing, retrieving at will, transforming, integrating and displaying spatial and non-spatial data from the real world for a particular set of purpose. These set of tools constitute Geographic Information System (GIS). The GIS can be used to solve broader range of problems as comparable to any isolated system for spatial or non-spatial data alone. For example using a GIS:

- Users can interrogate geographical features displayed on computer map and retrieve associated attribute information for display or further analysis.
- Maps can be constructed by querying or analysing attribute data.
- New sets of information can be generated by performing spatial operations.
- Different items of attribute data can be associated with one another through a shared location codes.

The GIS field is characterised by a great diversity of applications and concepts developed in many areas- agriculture, statistics, computer science, graphics, mathematics, surveying, cartography, geology, geography, database technology, resource management and decision making etc. The diversification of applications leads to different concepts and methods of GIS, thus making a proper definition difficult. For the purpose of understanding, the following definition of GIS encompasses most of the concepts.

A GIS is a specific information system applied to geographical data and is mainly referred to as a system of hardware, software, and procedures designed to support the capture, management, manipulation, analysis, modelling and display of spatially-referenced data for solving complex planning and management problems.

While many other graphical packages could handle spatial data- say AUTOCAD and other statistical packages, GIS is distinct in its capability to perform spatial operations of integration, it is this characteristic of GIS that helps in distinguishing it from other graphical packages.

4.1 Data in GIS

Broadly, the basic data for any GIS application can be categorised as:

- **Spatial Data** consisting of maps which have been prepared either with the help of field surveys or with the help of interpreted remotely sensed data *(RS).* Remote sensing data is a classic source of data on natural resources for a region and provides a record of the continuum of resource status because of its repetitive coverage. Remotely sensed data in the form of satellite imageries can be used to study and monitor land features, natural resources and dynamic aspects of human activities for preparation of thematic maps.

- **Non-Spatial Data** is attributes as complimentary to the spatial data and describe what is at a point, along a line or in a polygon and as a socio-economic characteristics from census or other sources. The attributes of a soil category could be the depth of soil, texture, erosion, drainage etc. and for geological category could be the rock type, its age, major composition etc. The socio-economic characteristics could be the demographic data, occupation data for a village etc.

The GIS will have to be the workhorse of integrated database system as both spatial and non-spatial data need to be handled. The GIS package offers efficient utilities to handle both these data sets and also allows for the spatial database organisation along with non-spatial database organisation. It is also capable to transform as well as integrate these two different kinds of information.

Typologies of spatial data in GIS: The spatial data in GIS is generally described by X,Y co-ordinates and descriptive data are best organised in alphanumeric fields. The GIS features can be classified in to four categories, first three of which pertains to spatial data.

Points refer to a single place and usually considered as having no dimension or having a dimension which is negligible when compared to study area. There is a large number of examples of point data such as the distribution of plants in forest, village location, industrial locations, cities etc.

Line represents the linear features and consists of series of X, Y co-ordinate pairs with discrete beginning and ending point. Line features have length attributes, rivers, streams, road networks, *etc*.

Polygons are closed features defined by set of linked lines enclosing an area. Polygons are characterised by area and perimeter. Administrative boundaries, landuse categories, city boundary are some of the examples.

Attributes are either the qualitative characteristics of the spatial data or descriptive information about geographical features. Attributes are stored in form of tables where each column of the table describes one attribute and each row of the table corresponds to a feature.

Data Structure: In order to represent the spatial information along their attributes a data model, which is set of logical definitions or rules for characterising the geographical data, is adopted. The data model represents the linkage between the real world domain of geographical data and computer or GIS presentations of these features. Different type of structures has been used as far as GIS is concerned. They are Raster model, Vector model, Quadtree model etc. The first two are most popular in GIS packages available in the market.

Raster Model represents the image with help of square lattice grids. In this case system stores an image by assigning a series of values (generally an integer ranging between 0 to 255) to each cell identified by its Cartesian co-ordinates in space.

Vector Model represents the geographical feature by a set of co-ordinates vectors as xy-coordinates define points, lines and polygons. The basic premise of the vector based structuring is to define a two-dimensional space where features are represented by coordinates on the two axes.

The vector image is more pleasant to the eye and more accurate representation to the reality. This implies that a vector based GIS is a better way to produce graphical output. On the other hand vector image is more complex and requires more advanced technology in terms of hardware. Further more from analysis point of view data in vector format presents more problems if we want to apply complex spatial statistical procedures. In contrasts, due to its regularity, a raster image is more easily accessible and is the natural format for many spatial techniques. Further more it requires a less advanced technology. On the other hand raster based image produce less pleasant graphic outputs. Further a raster image is based on quantization of reality and as such it can lead to serious estimation errors when we are interested in geometric and topologic characteristics like area or perimeters. The choice between vector and raster is crucial one and depends on specific aim for which GIS is designed.

4.2 GIS Data Base Design

The GIS has two distinct utilisation capabilities, first pertaining to querying and obtaining information and the second pertaining to integrated analytical modelling. However, both these capabilities depend upon the core of the GIS database that has been organised. Generally, a proper database organisation needs to ensure the following:

- flexibility in the design to adapt to the needs of different users
- controlled and standardised approach to data input and updation
- system of validation checks to maintain the integrity and consistency of the data elements
- level of security for minimising damage to the data
- minimising redundancy in data storage

While the above is general consideration for database organisation, in GIS domain the consideration is more pertinent because of the varied types and nature of data that need to be organised and stored. The design of the database will include three major elements;

Conceptual Design basically laying down the application requirements and specifying the end utilisation of the database. The conceptual design is independent of hardware and software and could be a wish list of utilisation goal.

Logical Design is the specification of the database vis-à-vis, a particular GIS package. This design set out the logical structure of the data base elements and is determined by the GIS package.

Physical Design pertains to the hardware and software characteristics, and requires consideration of file structure, memory and disk space, access and speed, *etc.*

Each stage is inter-related to the next stage of the design and impacts the organisation in a major way. The success or failure of a project on GIS is determined by the strength of the design and a good deal of time must be allocated to the design activities.

4.3 Integrated Modelling in GIS

Integration, in a GIS context, is the synthesis of spatial and non-spatial information with in the frame work of an application. By performing the operations across the two sets of information in tandem, a far richer set of questions can be answered and a far broader range of problems can be solved than in a system that handle just attribute or spatial data alone. All problems of GIS based integration involves a conjuctive analysis of multi-parameter data. The multi-parameter data include different spatial inputs say, maps of land use, soils, slopes, terrain etc. and other no-spatial data sets. The GIS allows for the integration of these data sets so as to obtain a composite information set. However, the important aspect is the interpretation and the analysis of the integrated information sets. Toward this there are two aspects that are important for integrated model building, first is the criterion that defines the logic for the analysis of the composite information set and second is the relative importance or weightage of each of the parameter for the end objectives.

4.4 Sources of Data

The sources of spatial data have undergone deep changes in last decades. The traditional sources, like census, military archives, official budgets, air photographs and adhoc surveys, can be integrated with the widespread use of new sources like satellite images. Recently, with the rapid developments in the field of information technology specially networking have increased many fold the sources as well the amount of data which can be integrated in any GIS system linked to it. The various sources of data can be classified in to three broad categories. (a) Census and administrative records. These sources provide a complete geographical coverage of the phenomenon studied at a given level of disaggregation. (b) Sample surveys. This source of data usually do not provide a complete coverage of the phenomenon; samples generally provide estimates which are significant only at a broader level of spatial aggregation. (c) air and satellite photographs which provide a complete image of limited number of phenomena(mainly environmental) on discrete regular grid.

4.5 Errors in Data

The maps containing spatial information and stored in a GIS is, in general, contaminated by errors. The practice of storing large masses of data into computerised information system is bound to make this problem even worse. Errors in spatial database can be classified in to three types. Conceptual errors arise from the process of translating real world features in to map objects. Process errors arise when spatial information is converted in to map form. Source errors arise from discrepancy between reality and its mapped representation. A second classification is possible by distinguishing between location and attribute errors. Location errors arise from the uncertainty as to where a geographical object is. This kind of error refers to the disagreement between boundaries on the ground and on the map or between points on the ground and on the map. It also includes errors arising when points, lines and areal data are digitised for the purpose of computerised storage. In contrast, Attributes errors arise from our uncertainty about the values assign to a geographical object and it is mainly associated with the need to use sample data, which surrogate information aggregate rather than individual data as well as imperfections in the measuring device by which attribute values are recorded. The presence of errors in a GIS data can lead to serious problems especially when we perform automated GIS operations that involves convolution of two or more maps. In this resulting output map will contain a combinations of the errors contained in the various source maps and will distort the resulted output in unpredictable way.

4.6 Important GIS Packages

Some of the important GIS packages are:

- ARC/INFO is one of the first GIS packages that was available commercially and is a package used all over the world. It has been developed by Environmental System Research Institute, USA. It is available on wide range of platforms-PC's, workstations, PRIME systems. It is also available on variety of operating systems-DOS, UNIX, VMS, *etc*.
- PAMAP is a product of Graphic Limited, Canada and is an integrated group of software products designed for an open system environment. The package is modular and is designed to address the wide range of mapping and analysis requirements of the natural resource sector. It is also available on variety of platforms - Pentium, 486/386/286 PC's, UNIX, SUN, VAX system.
- MAPINFO is a popular package translated in to several languages and ported to several platforms like Windows, Macintosh, Sun, and HP workstations.
- GRASS, a public domain UNIX package with large established user base which actually contributes to the code that is incorporated in to new versions.
- ISROGIS is a state-of-art GIS package with efficient tools of integration and manipulation of spatial and non-spatial data and consists of a set of powerful module. It is available on PC platforms on MS-Windows and on UNIX and SUN platforms.
- IDRISI has been developed by Clarke University, USA, and an inexpensive PC based advanced features including good import export facility, a new digitisation module, and some image processing facilities.
- GRAM is a PC based GIS tool developed by IIT, Bombay. It can handle both vector and raster data and has functionality for raster based analysis, image analysis, *etc*.