SPATIAL STATISTICS FOR REMOTELY SENSED IMAGES

Notes by Professor Dr. Qazi Mazhar Ali
Aligarh Muslim University, Aligarh

DSM- 3004

MOHAMMAD WASIQ

DSM - 3004

Spatial Statistics Remotely Sensed Images

Unit I

Spatial Data : An overview, some measurements on spatial data, Image, Image formats, types of images, basic image manipulation, reading, saving and displaying an image, image correction: radiometric and geometric, image registration.

Unit II

Image Enhancement : Point wise intensity transformation, histogram equalization and matching, linear smoothing, non-linear smoothing, low-pass filtering, high-pass filtering.

Unit III

Image Transformations : Spectral indices, Principal component transformation, Fourier transformation, wavelet transformation, image restoration, Image segmentation, Edge/Region based segmentation, spectral clustering for image segmentation, edge detection

Unit IV

Supervised and Unsupervised Image Classification : Maximum likelihood classifier, Distance based classification, parallelepiped classifier, kNN classifier, linear discrimination, support vector classifiers, neural network classifiers, tree based classifiers, clustering based unsupervised classification, k-means clustering, fuzzy k means clustering, agglomerative hierarchical clustering, Gaussian mixture clustering, Kohonen self-organizing map, hybrid supervised/ unsupervised classification, classification accuracy assessment.

Multispectral Satellite Image

A multispectral, optical/infrared image such as that shown in Figure 1.1 may be represented as a three-dimensional array of gray-scale values or pixel intensities

$$g_k(i, j),\ 1 \le i \le c,\ 1 \le j \le r,\ 1 \le k \le N,$$

where c is the number of pixel columns (also called samples) and r is the number of pixel rows (or lines). The index k denotes the spectral band, of which there are N in all. For data at an early processing stage a pixel may be stored as a digital number (DN), often in a single byte so that $0 \le g_k \le 255$.

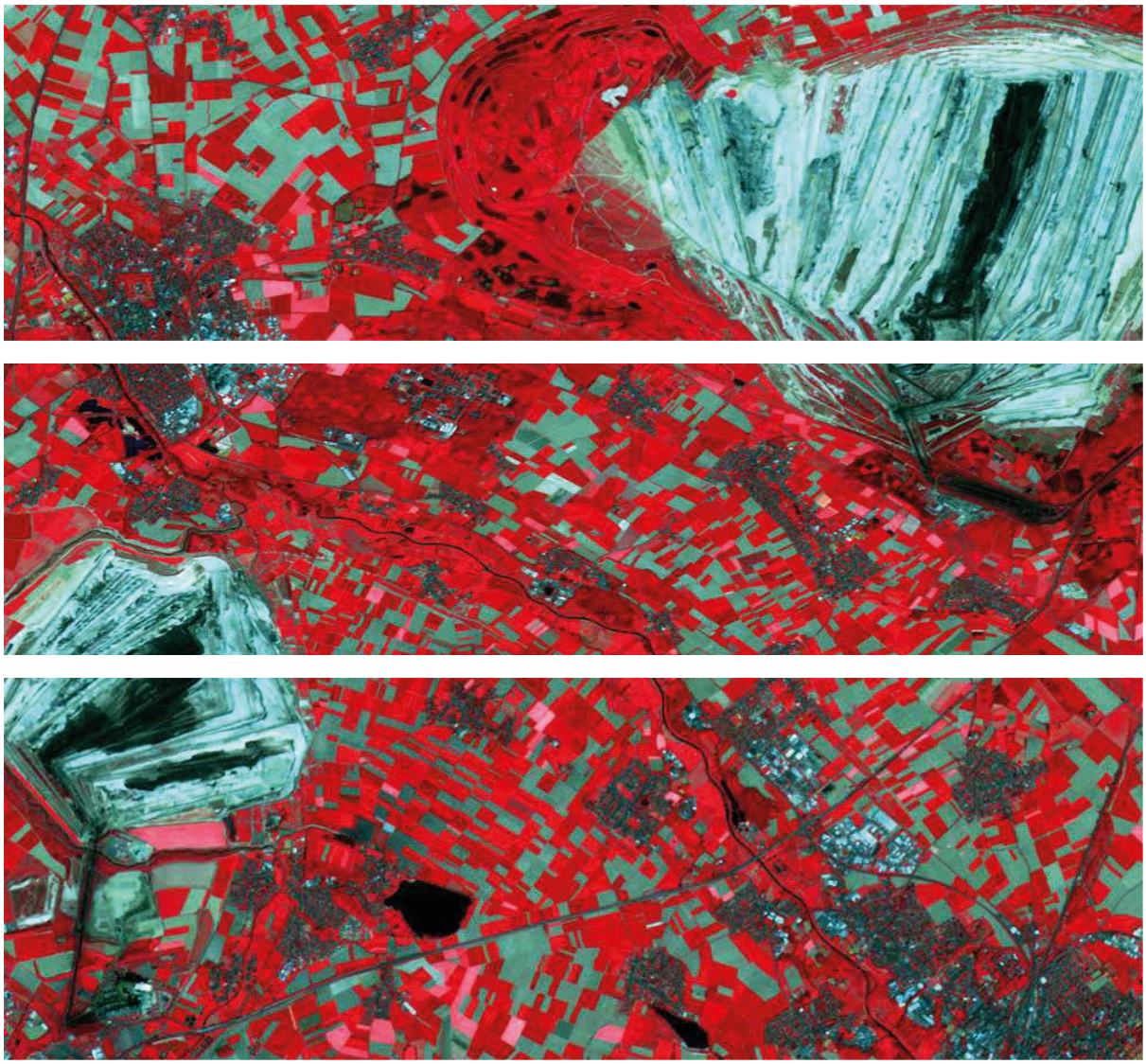


FIGURE 1.1

ASTER color composite image (1000 × 1000 pixels) of VNIR bands 1 (blue), 2 (green), and 3N (red) over the town of Julich in Germany, acquired on May 1, 2007. The bright areas are open cast coal mines.

How an image is acquired?

The gray-scale values in the various bands encode measurements of the radiance L_λ(x, y) in wavelength interval _λ due to sunlight reflected from some point (x, y) on the Earth's surface, or due to thermal emission from that surface, and focused by the instrument's optical system along the array of sensors. Ignoring all absorption and scattering effects of the intervening atmosphere, the at-sensor radiance available for measurement from reflected sunlight from a horizontal, Lambertian surface, i.e., a surface which scatters reflected radiation uniformly in all directions, is given by

$$\boldsymbol{L_\lambda(x,y) = E_\lambda . cos\theta z . R_\lambda(x,y) \, /\pi}$$

The units are [W/(m2 · sr · μm)], E_λ is the average spectral solar irradiance in the spectral band _λ, θz is the solar zenith angle, R_λ(x, y) is the surface reflectance at coordinates (x, y), a number between 0 and 1, and π accounts for the upper hemisphere of solid angle. The conversion between DN and at-sensor radiance is determined by the sensor calibration as measured (and maintained) by the satellite image provider. For example, for ASTER VNIR and SWIR L1A data,

$$\boldsymbol{L_\lambda(x,y) = A \cdot DN/G} \textbf{ +D}$$

The quantities A (linear coefficient), G (gain), and D (offset) are tabulated for each of the detectors in the arrays and included with each acquisition.

Atmospheric scattering and absorption models may be used to deduce atsurface radiance, surface temperature and emissivity or surface reflectance from the observed radiance at the sensor. Reflectance and emissivity are directly related to the physical properties of the surface being imaged.

How image is stored?

Various conventions are used for storing the image array gk(i, j) in computer memory or other storage media. In band interleaved by pixel (BIP) format, for example, a two-channel, 3 × 3 pixel image would be stored as

g1(1, 1) g2(1, 1) g1(2, 1) g2(2, 1) g1(3, 1) g2(3, 1)

g1(1, 2) g2(1, 2) g1(2, 2) g2(2, 2) g1(3, 2) g2(3, 2)

g1(1, 3) g2(1, 3) g1(2, 3) g2(2, 3) g1(3, 3) g2(3, 3)

whereas in band interleaved by line (BIL) it would be stored as

g1(1, 1) g1(2, 1) g1(3, 1) g2(1, 1) g2(2, 1) g2(3, 1)

g1(1, 2) g1(2, 2) g1(3, 2) g2(1, 2) g2(2, 2) g2(3, 2)

g1(1, 3) g1(2, 3) g1(3, 3) g2(1, 3) g2(2, 3) g2(3, 3)

and in band sequential (BSQ) format as

g1(1, 1) g1(2, 1) g1(3, 1)

g1(1, 2) g1(2, 2) g1(3, 2)

g1(1, 3) g1(2, 3) g1(3, 3)

g2(1, 1) g2(2, 1) g2(3, 1)

g2(1, 2) g2(2, 2) g2(3, 2)

g2(1, 3) g2(2, 3) g2(3, 3)

In the computer language Python, augmented with the numerical package NumPy, so-called row major indexing is used for arrays and the elements in an array are numbered from zero. This means that, if a gray-scale image g is assigned to a Python array variable g, then the intensity value g(i, j) is addressed as g[j-1,i-1]. An N-band multispectral image is stored in BIP format as an r×c×N array in NumPy, in BIL format as an r×N ×c and in BSQ format as an N × r × c array. So, for example, in BIP format the value gk(i, j) is stored at g[j-1,i-1,k-1].

How many bits do we need to store an image?

The number of bits, *b*, we need to store an image of size $N \times N$, with 2^m grey levels, is

$$b = N \times N \times m$$

So, for a typical 512 × 512 image with 256 grey levels ($m = 8$) we need 2,097,152 bits or 262,144 8-bit bytes. That is why we often try to reduce *m* and *N*, without significant loss of image quality.

Vector vs. Raster

First things first: What is the difference between vector and raster?

Raster Image Files

Raster images are constructed by a series of pixels, or individual blocks, to form an image. JPEG, GIF, and PNG are all raster image extensions. Every photo we find online or in print is a raster image. Pixels have a defined proportion based on their resolution (high or low), and when the pixels are stretched to fill space they were not originally intended to fit, they become distorted, resulting in blurry or unclear images.

In order to retain pixel quality, we cannot resize raster images without compromising their resolution. As a result, it is important to remember to save raster files at the exact dimensions needed for the application.

Vector Image Files

Vector images are far more flexible. They are constructed using proportional formulas rather than pixels. EPS, AI and PDF are perfect for creating graphics that require frequent resizing. Our logo and brand graphics should have been created as a vector, and we should always have

a master file on hand. The real beauty of vectors lies in their ability to be sized as small as a postage stamp, or large enough to fit on an 18-wheeler!

High Resolution vs. Low Resolution

Have we heard our designer talk about DPI or PPI? DPI stands for "dots per inch" and PPI translates to "pixels per inch." These units of measure are essential for determining if the density of pixels in an image is appropriate for the application we are using.

The biggest thing to note when determining what DPI or PPI we require is if we are using an image for print or web. Websites display images at 72dpi, which is low resolution; however images at this resolution look really crisp on the web. This is not the case for print. Best practices for printing an image will require it to be no less than 300dpi.

Don't try to trick the system. A lot of magic can happen in Photoshop, but creating pixels out of thin air isn't one of them. Pulling an image off of the web and trying to get it to fit the dimensions of our print project just won't work. we will end up with a pixelated image that appears stretched and distorted.

10 Types of image formats :

1. JPEG (or JPG) - Joint Photographic Experts Group
2. PNG - Portable Network Graphics
3. GIF - Graphics Interchange Format
4. TIFF - Tagged Image File
5. PSD - Photoshop Document
6. PDF - Portable Document Format
7. EPS - Encapsulated Postscript
8. AI - Adobe Illustrator Document
9. INDD - Adobe Indesign Document
10. RAW - Raw Image Formats

1. JPEG (or JPG) - Joint Photographic Experts Group

JPEGs might be the most common file type we run across on the web, and more than likely the kind of image that is in our company's MS Word version of its letterhead. JPEGs are known for their "lossy" compression, meaning that the quality of the image decreases as the file size decreases.

We can use JPEGs for projects on the web, in Microsoft Office documents, or for projects that require printing at a high resolution. Paying attention to the resolution and file size with JPEGs is essential in order to produce a nice-looking project.

JPG vs JPEG

There is no difference between the .jpg and .jpeg filename extensions. Regardless of how we name our file, it is still the same format and will behave the same way.

The only reason that the two extensions exist for the same format is because .jpeg was shortened to .jpg to accommodate the three-character limit in early versions of Windows. While there is no such requirement today, .jpg remains the standard and default on many image software programs.

PNG - Portable Network Graphics

PNGs are amazing for interactive documents such as web pages but are not suitable for print. While PNGs are "lossless," meaning we can edit them and not lose quality, they are still low resolution.

The reason PNGs are used in most web projects is that we can save our image with more colors on a transparent background. This makes for a much sharper, web-quality image.

3. GIF - Graphics Interchange Format

GIFs are most common in their animated form, which are all the rage on Tumblr pages and in banner ads. It seems like every day we see pop culture GIF references from Giphy in the comments of social media posts. In their more basic form, GIFs are formed from up to 256 colors in the RGB colorspace. Due to the limited number of colors, the file size is drastically reduced.This is a common file type for web projects where an image needs to load very quickly, as opposed to one that needs to retain a higher level of quality.

4. TIFF - Tagged Image File

A TIF is a large raster file that doesn't lose quality. This file type is known for using "lossless compression," meaning the original image data is maintained regardless of how often we might copy, re-save, or compress the original file.

Despite TIFF images' ability to recover their quality after manipulation, we should avoid using this file type on the web. Since it can take forever to load, it'll severely impact website performance. TIFF files are also commonly used when saving photographs for print.

5. PSD - Photoshop Document

PSDs are files that are created and saved in Adobe Photoshop, the most popular graphics editing software ever. This type of file contains "layers" that make modifying the image much easier to handle. This is also the program that generates the raster file types mentioned above.

The largest disadvantage to PSDs is that Photoshop works with raster images as opposed to vector images.

6. PDF - Portable Document Format

PDFs were invented by Adobe with the goal of capturing and reviewing rich information from any application, on any computer, with anyone, anywhere. They've been pretty successful so far.

If a designer saves our vector logo in PDF format, we can view it without any design editing software (as long as we have downloaded the free Acrobat Reader software), and they have the ability to use this file to make further manipulations. This is by far the best universal tool for sharing graphics.

7. EPS - Encapsulated Postscript

EPS is a file in vector format that has been designed to produce high-resolution graphics for print. Almost any kind of design software can create an EPS. The EPS extension is more of a universal file type (much like the PDF) that can be used to open vector-based artwork in any design editor, not just the more common Adobe products. This safeguards file transfers to designers that are not yet utilizing Adobe products, but may be using Corel Draw or Quark.

8. AI - Adobe Illustrator Document

AI is, by far, the image format most preferred by designers and the most reliable type of file format for using images in all types of projects from web to print, etc.

Adobe Illustrator is the industry standard for creating artwork from scratch and therefore more than likely the program in which our logo was originally rendered. Illustrator produces vector artwork, the easiest type of file to manipulate. It can also create all of the aforementioned file types. Pretty cool stuff! It is by far the best tool in any designer's arsenal.

9. INDD - Adobe InDesign Document

INDDs (InDesign Document) are files that are created and saved in Adobe InDesign. InDesign is commonly used to create larger publications, such as newspapers, magazines and eBooks.

Files from both Adobe Photoshop and Illustrator can be combined in InDesign to produce content rich designs that feature advanced typography, embedded graphics, page content, formatting information and other sophisticated layout-related options.

10. RAW - Raw Image Formats

A RAW image is the least-processed image type on this list -- it's often the first format a picture inherits when it's created. When we snap a photo with our camera, it's saved immediately in a raw file format. Only when we upload our media to a new device and edit it using image software is it saved using one of the image extensions explained above.

RAW images are valuable because they capture every element of a photo without processing and losing small visual details. Eventually, however, we'll want to package them into a raster or vector file type so they can be transferred and resized for various purposes.

As we can see from the icons above, there are multiple raw image files in which we can create images -- many of them native to certain cameras (and there are still dozens more formats not shown above). Here's a brief description of those four raw files above:

- **CR2 :** This image extension stands for Canon RAW 2, and was created by Canon for photos taken using its own digital cameras. They're actually based on the TIFF file type, making them inherently high in quality.
- **CRW :** This image extension was also created by Canon, preceding the existence of the CR2.
- **NEF :** This image extension stands for Nikon Electric Format, and is a RAW file type created by (we guessed it) Nikon Cameras. These image files actually allow for extensive editing without changing file types, provided the editing takes place using a Nikon device or Nikon Photoshop plugin.
- **PEF :** This image extension stands for Pentax Electronic Format, a RAW image file type created by Pentax Digital Cameras.

Pixel: In Computer graphics a pixel, dots, or picture element is a physical point in a picture. A pixel is simply the smallest addressable element of a picture represented on a screen.

A majority of pictures that we see on our computer screen are raster images. The selfie that we click with our mobile phone is another example of a raster image. An image is made up using a collection of pixels referred to as a bitmap.

Bitmap: In computer graphics, a bitmap is a mapping from some domain (for example, a range of integers) to bits, that is, values which are zero or one. It is also called a bit array or bitmap index.The more general term pixmap refers to a map of pixels, where each one may store more than two colors, thus using more than one bit per pixel. Often bitmap is used for this as well. In some contexts, the term bitmap implies one bit per pixel, while pixmap is used for images with multiple bits per pixel.

Raster Graphics

Raster images use bit maps to store information. This means a large file needs a large bitmap. The larger the image, the more disk space the image file will take up. As an example, a 640 x 480 image requires information to be stored for 307,200 pixels, while a 3072 x 2048 image

(from a 6.3 Megapixel digital camera) needs to store information for a whopping 6,291,456 pixels. We use algorithms which compress images to help reduce these file sizes. Image formats like jpeg and gif are common compressed image formats. Scaling down these images is easy but enlarging a bitmap makes it pixelated or simply blurred. Hence for images which need to scaled to different sizes, we use vector graphics.

File extensions: .BMP, .TIF, .GIF, .JPG

Vector Graphics

Making use of sequential commands or mathematical statements or programs which place lines or shapes in a 2-D or 3-D environment is referred to as Vector Graphics. Vector graphics are best for printing since it is composed of a series of mathematical curves. As a result vector graphics print crisply even when they are enlarged. In physics: A vector is something which has a magnitude and direction. In vector graphics, the file is created and saved as a sequence of vector statements. Rather than having a bit in the file for each bit of line drawing we use commands which describe series of points to be connected.
AS a result a much smaller file is obtained.

File extensions : .SVG, .EPS, .PDF, .AI, .DXF

Conversions:

1. **Vector to Raster :** Printers and display devices are raster devices. As a result we need to convert vector images to raster format before they can be used i.e displayed or printed. The required resolution plays a vital role in determining the size of raster file generated. Here it is important to note that the size of vector image to be converted always remains the same. It is convenient to convert a vector file to a range of bitmap/raster file formats but going down opposite path is harder.(because at times we need to edit the image while converting from raster to vector)
2. **Raster to Vector :** Image tracing in computing can be referred to vectorization and it's simply the conversion of raster images to vector images. An interesting application of vectorization is to update images and recover work. Vectorization can be used to retrieve information that we have lost. Paint in Microsoft Windows produces a bitmap output file. It is easy to notice jagged lines in Paint. In this kind of a conversion the image size reduces drastically. As a result an exact conversion is not possible in this scenario. Due to various approximations and editing that is done in the process of conversion the converted images are not of good quality.

Differences between Vector and Raster graphics

The main difference between vector and raster graphics is that raster graphics are composed of pixels, while vector graphics are composed of paths. A raster graphic, such as a gif or jpeg, is an array of pixels of various colors, which together form an image.

Raster Graphics	Vector Graphics
They are composed of pixels.	They are composed of paths.
In Raster Graphics, refresh process is independent of the complexity of the image.	Vector displays flicker when the number of primitives in the image become too large.
Graphic primitives are specified in terms of end points and must be scan converted into corresponding pixels.	Scan conversion is not required.
Raster graphics can draw mathematical curves, polygons and boundaries of curved primitives only by pixel approximation.	Vector graphics draw continuous and smooth lines.
Raster graphics cost less.	Vector graphics cost more as compared to raster graphics.
They occupy more space which depends on image quality.	They occupy less space.
File extensions: .BMP, .TIF, .GIF, .JPG	File Extensions: .SVG, .EPS, .PDF, .AI, .DXF

Raster and Vector Images in GIS :

A geodatabase is a database that is in some way referenced to locations on the earth. Coupled with this data is usually data known as attribute data. Attribute data generally defined as additional information, which can then be tied to spatial data.

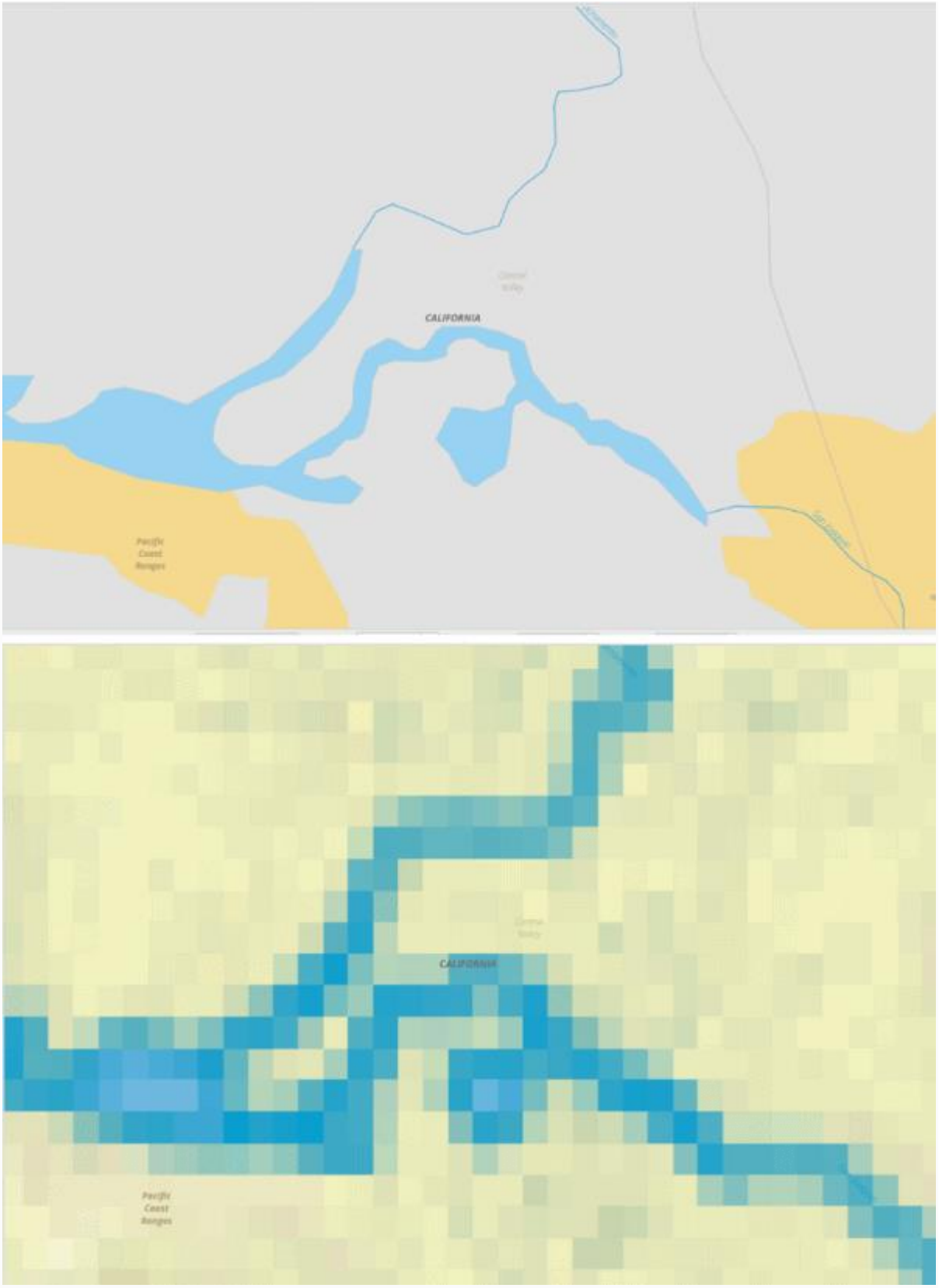
What types of GIS data are there ?

GIS data can be separated into two categories: spatially referenced data which is represented by vector and raster forms (including imagery) and attribute tables which is represented in tabular format.

Within the spatial referenced data group, the GIS data can be further classified into two different types: **vector and raster**. Most GIS software applications mainly focus on the usage and manipulation of vector geodatabases with added components to work with raster-based geodatabases.

What is the difference between raster and vector GIS data ?

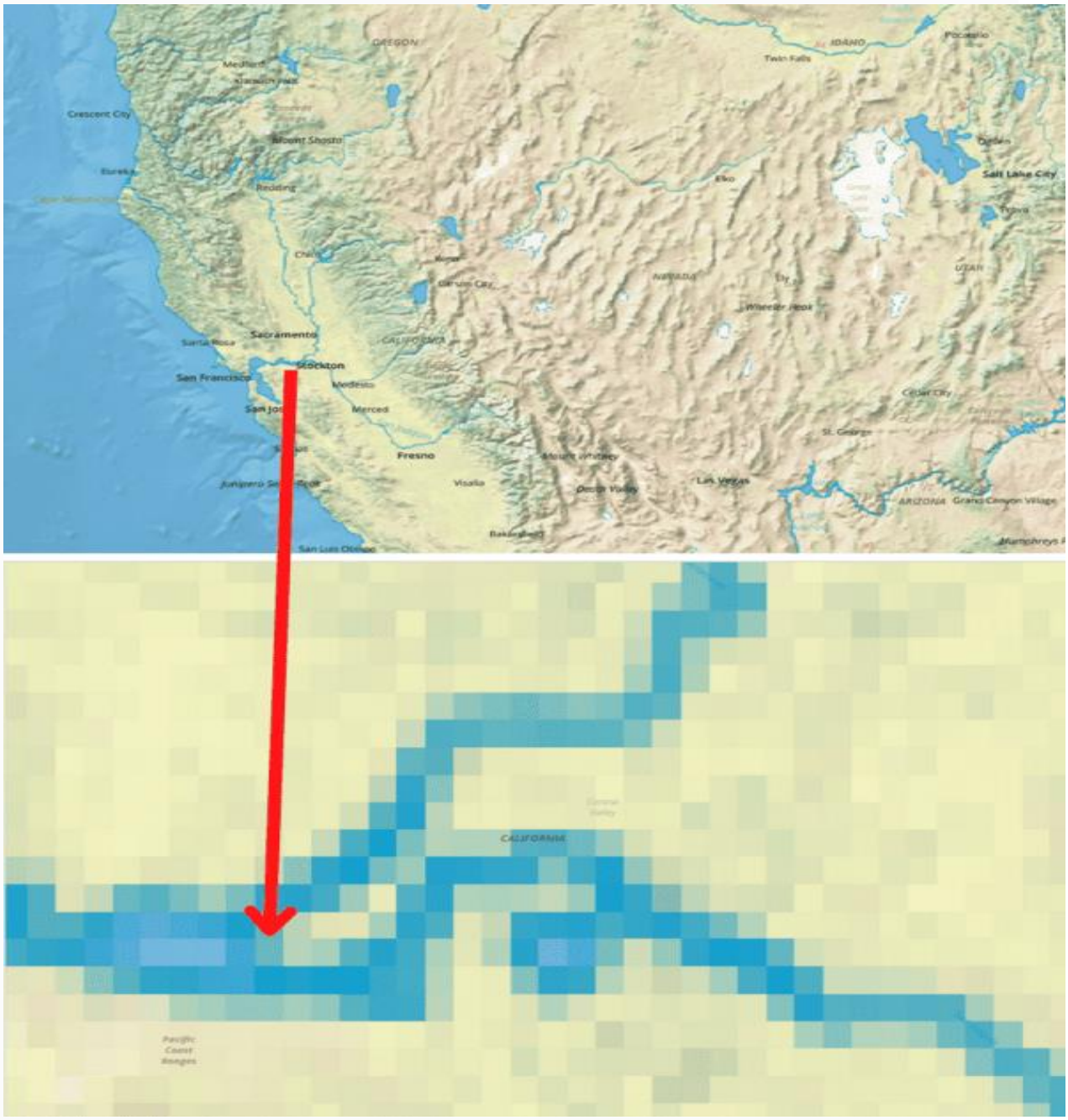
As mentioned, geospatial data can be represented in one of two main types: vector or raster.



A comparison of how rivers and lakes are shown as polygons and lines in vector data (top) and as cells in raster data (bottom). Images: Caitlin Dempsey.

Vector data is the most common type of GIS data. Most data loaded into a GIS software program tends to be in vector data. Vector data represents geographic data symbolized as points, lines, or polygons.

Raster data represents geographic data as a matrix of cells that each contains an attribute value. While the area of different polygon shapes in a data set can differ, each cell in a raster data set is the same cell. The size of the area in the real world that each cell represents is known as the spatial resolution. Raster data is most commonly found in remotely sensed data, shaded relief and topographic data, satellite imagery, and aerial imagery.



When we zoom in on raster data like this shaded relief layer, we can see each cell is the same size.

Images : Caitlin Dempsey.

Vector data

Vector data is split into three types: point, line (or arc), and polygon data.

Point Data

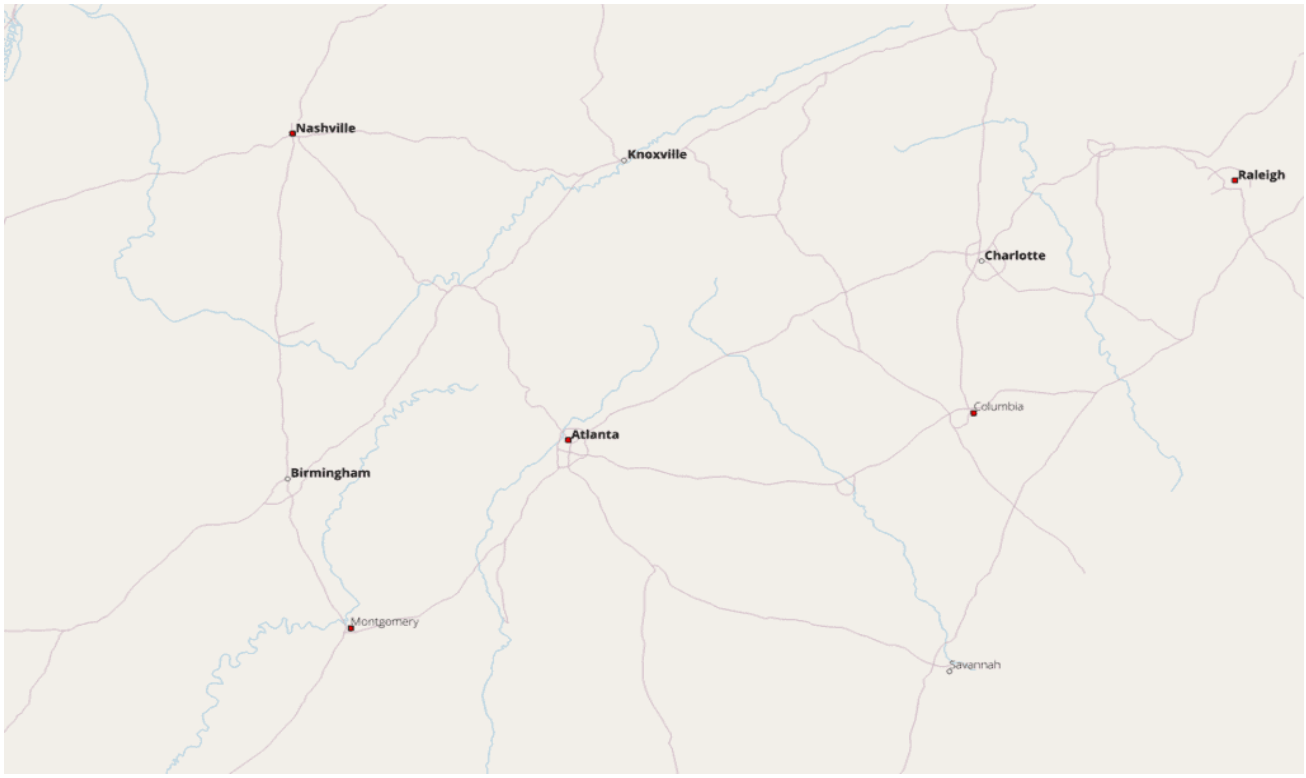
Point data is most commonly used to represent nonadjacent features and to represent discrete data points. Points have zero dimensions, therefore we can measure neither length or area with this dataset. Examples would be schools, points of interest, bridge and culvert locations. Point features are also used to represent abstract points. For instance, point locations could represent city locations or place names.

In GIS, point data can be used to show the geographic location of cities. Map: Caitlin Dempsey using Natural Earth Data.

AD

Line Data

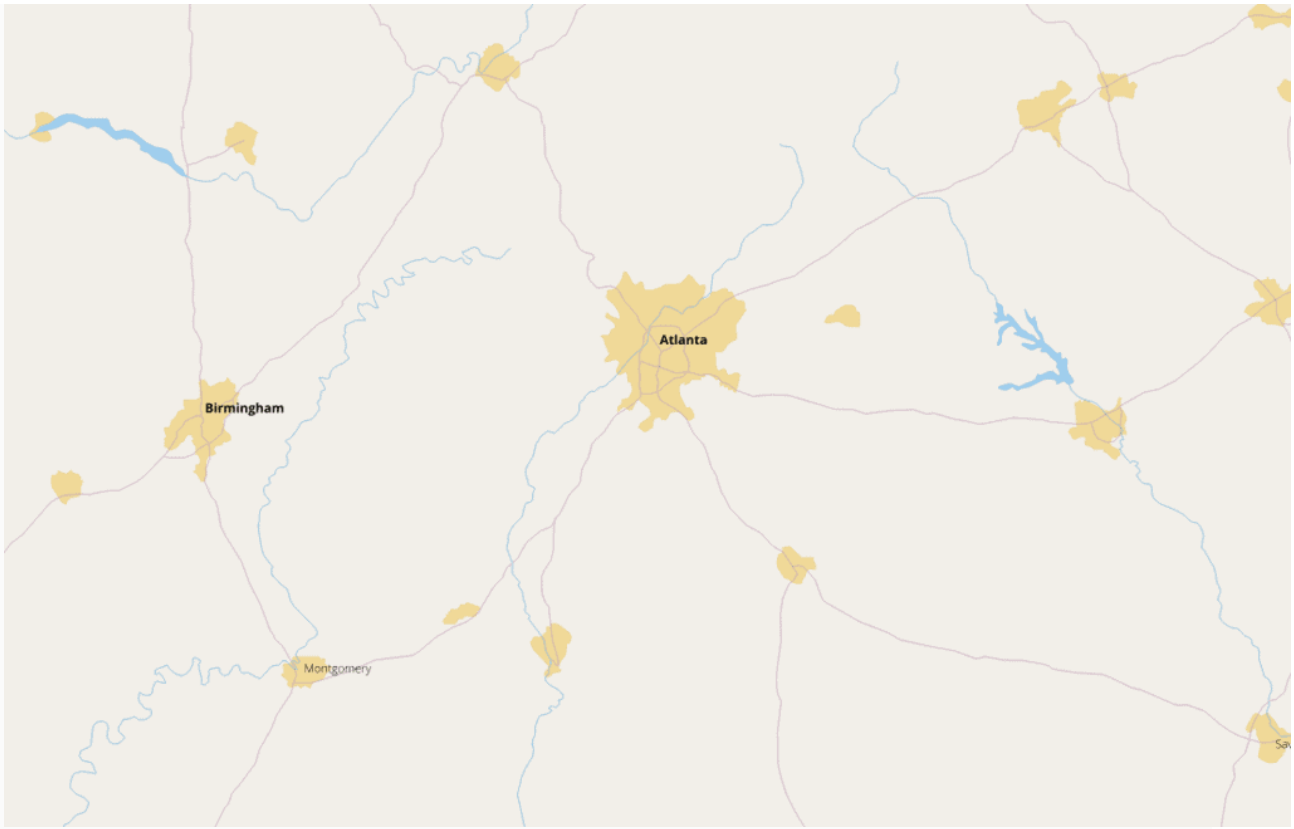
Line (or arc) data is used to represent linear features. Common examples would be rivers, trails, and streets. Line features only have one dimension and therefore can only be used to measure length. Line features have a starting and ending point. Common examples would be road centerlines and hydrology. Symbology most commonly used to distinguish arc features from one another are line types (solid lines versus dashed lines) and combinations using colors and line thicknesses. In the example below roads are distinguished from the stream network by designating the roads as a solid black line and the hydrology a dashed blue line.



In this map, roads and waterways are shown as line data. Map using Natural Earth Data.

Polygon Data

Polygons are used to represent areas such as the boundary of a city (on a large scale map), lake, or forest. Polygon features are two dimensional and therefore can be used to measure the area and perimeter of a geographic feature.
Polygon features are most commonly distinguished using either a thematic mapping symbology (color schemes), patterns, or in the case of numeric gradation, a color gradation scheme could be used.



With maps presented at a larger scale, city locations are represented as a polygon to show the extent of each city. Map made with Natural Earth Data.

AD

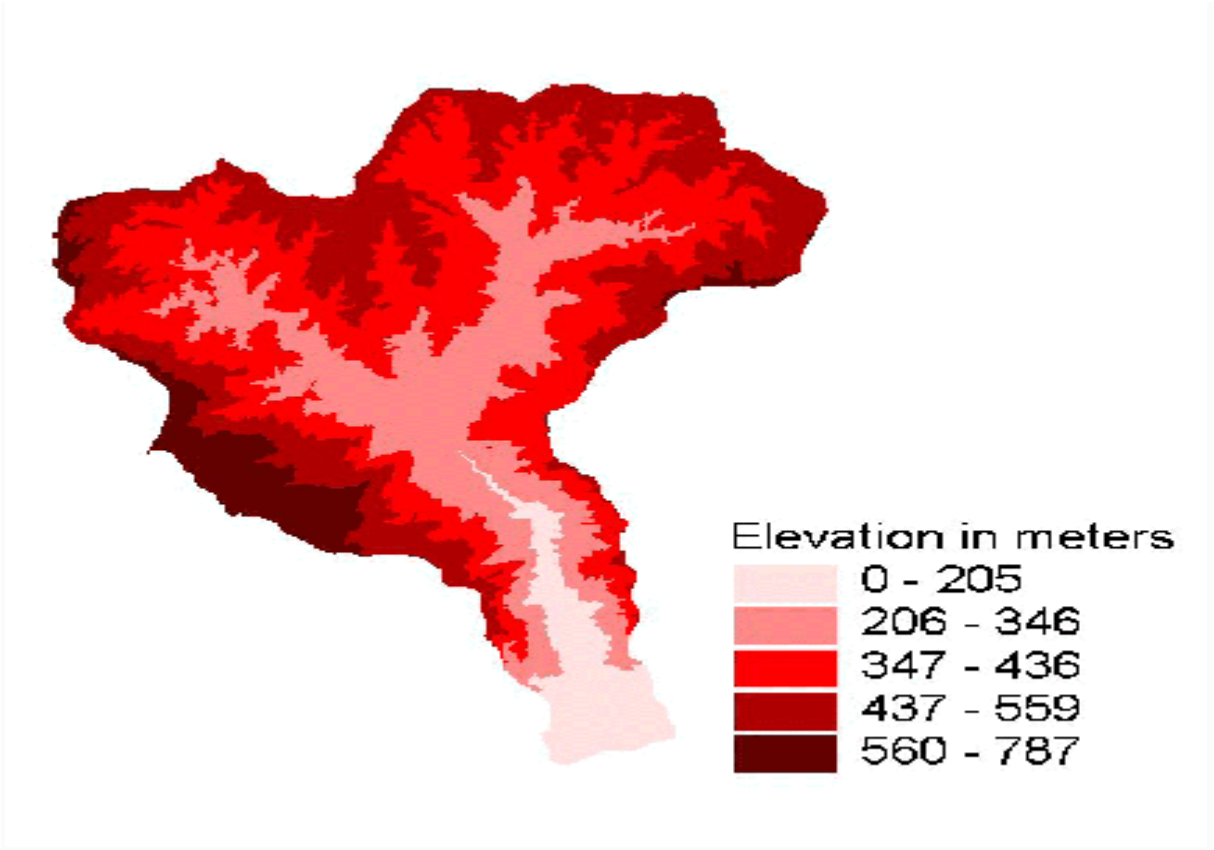
Both line and point feature data represent polygon data at a much smaller scale. They help reduce clutter by simplifying data locations.

As the features are zoomed in to, the point location of a school is more realistically represented by a series of building footprints showing the physical location of the campus.
Line features of a street centerline file only represent the physical location of the street. If a higher degree of spatial resolution is needed, a street curbwidth file would be used to show the width of the road as well as any features such as medians and right-of-ways (or sidewalks).

Raster Data

Raster data (also known as grid data) represents the fourth type of feature: surfaces. Raster data is cell-based and this data category also includes aerial and satellite imagery.
Continuous and Discrete Raster Data

There are two types of raster data: continuous and discrete. An example of discrete raster data is population density. Continuous data examples are temperature and elevation measurements. There are also three types of raster datasets: thematic data, spectral data, and pictures (imagery).



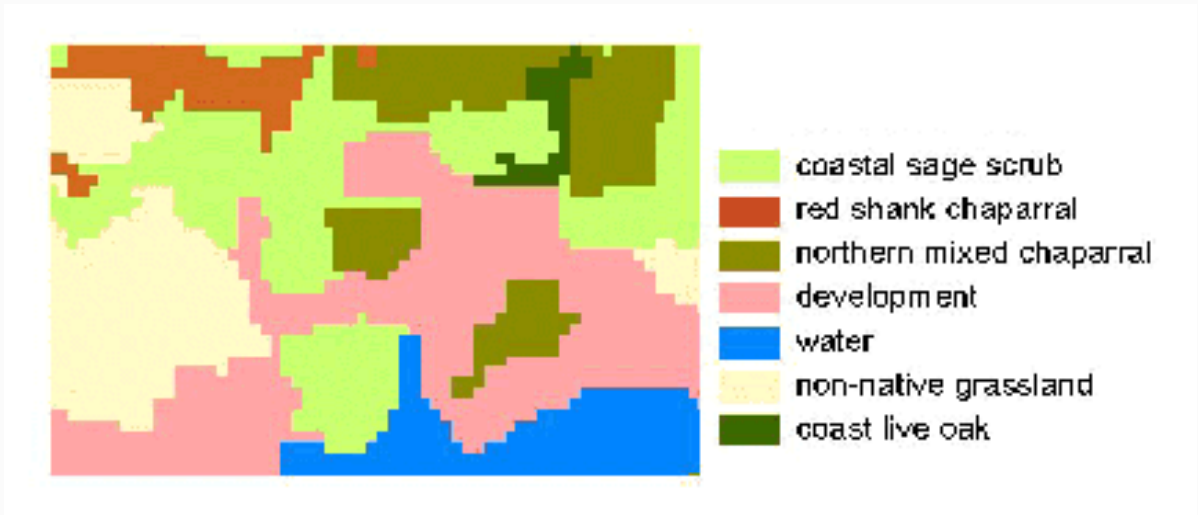
Digital Elevation Model (DEM) showing elevation.

AD

This example of a thematic raster dataset is called a Digital Elevation Model (DEM). Each cell presents a 30m pixel size with an elevation value assigned to that cell. The area shown is the Topanga Watershed in California and gives the viewer and understand of the topography of the region.

This image shows a portion of Topanga, California taken from a USGS DOQ.
AD

Each cell contains one value representing the dominate value of that cell. Raster datasets are intrinsic to most spatial analysis. Spatial hydrology modeling such as extracting watersheds and flow lines also uses a raster-based system. Spectral data presents aerial or satellite imagery which is then often used to derive vegetation geologic information by classifying the spectral signatures of each type of feature.



Raster data showing vegetation classification. The vegetation data was derived from NDVI classification of a satellite image.
AD

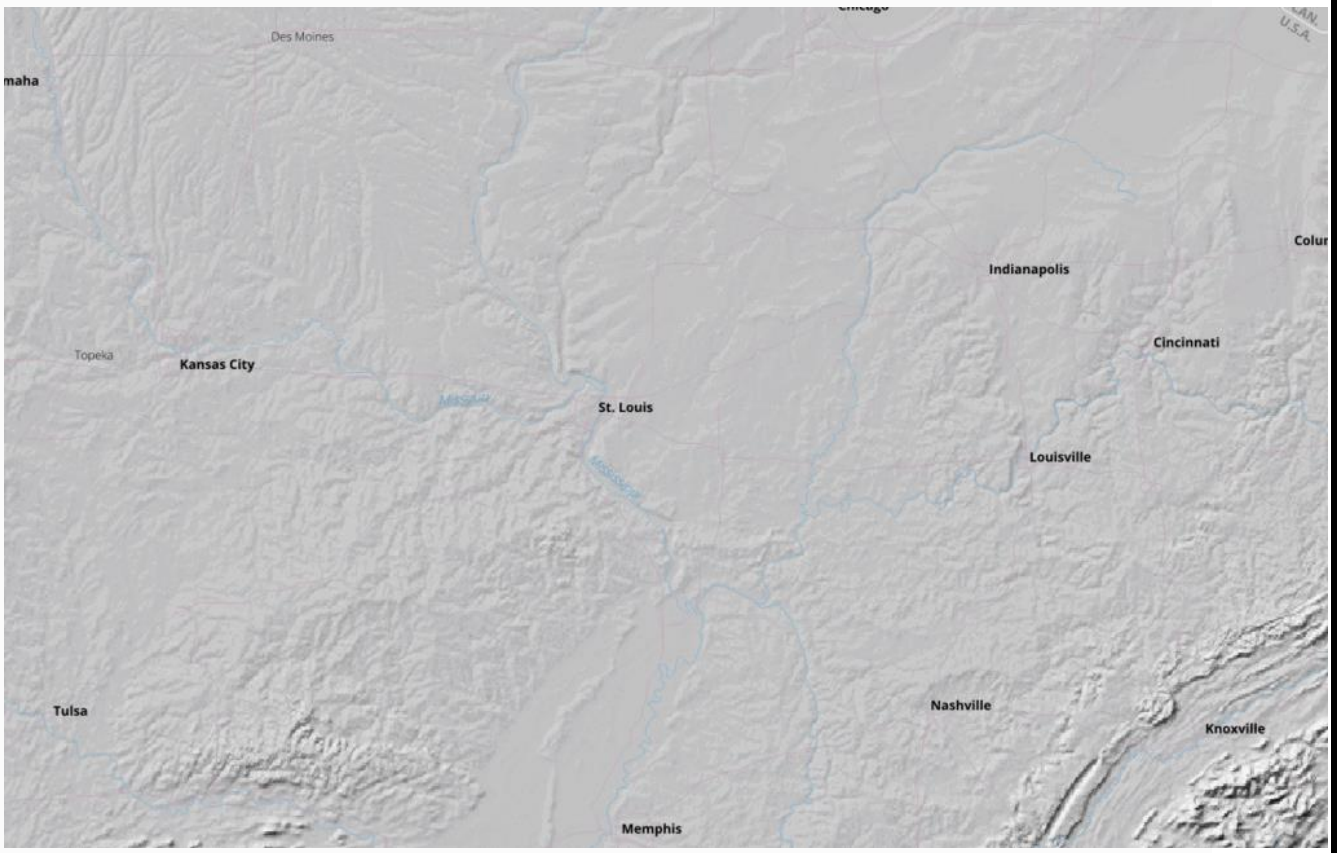
What results from the effect of converting spatial data location information into a cell based raster format is called stairstepping. The name derives from the image of exactly that, the

square cells along the borders of different value types look like a staircase viewed from the side.

Unlike vector data, raster data is formed by each cell receiving the value of the feature that dominates the cell. The stairstepping look comes from the transition of the cells from one value to another. In the image above the dark green cell represents chamise vegetation. This means that the dominate feature in that cell area was chamise vegetation. Other features such as developed land, water or other vegetation types may be present on the ground in that area. As the feature in the cell becomes more dominantly urban, the cell is attributed the value for developed land, hence the pink shading.

AD

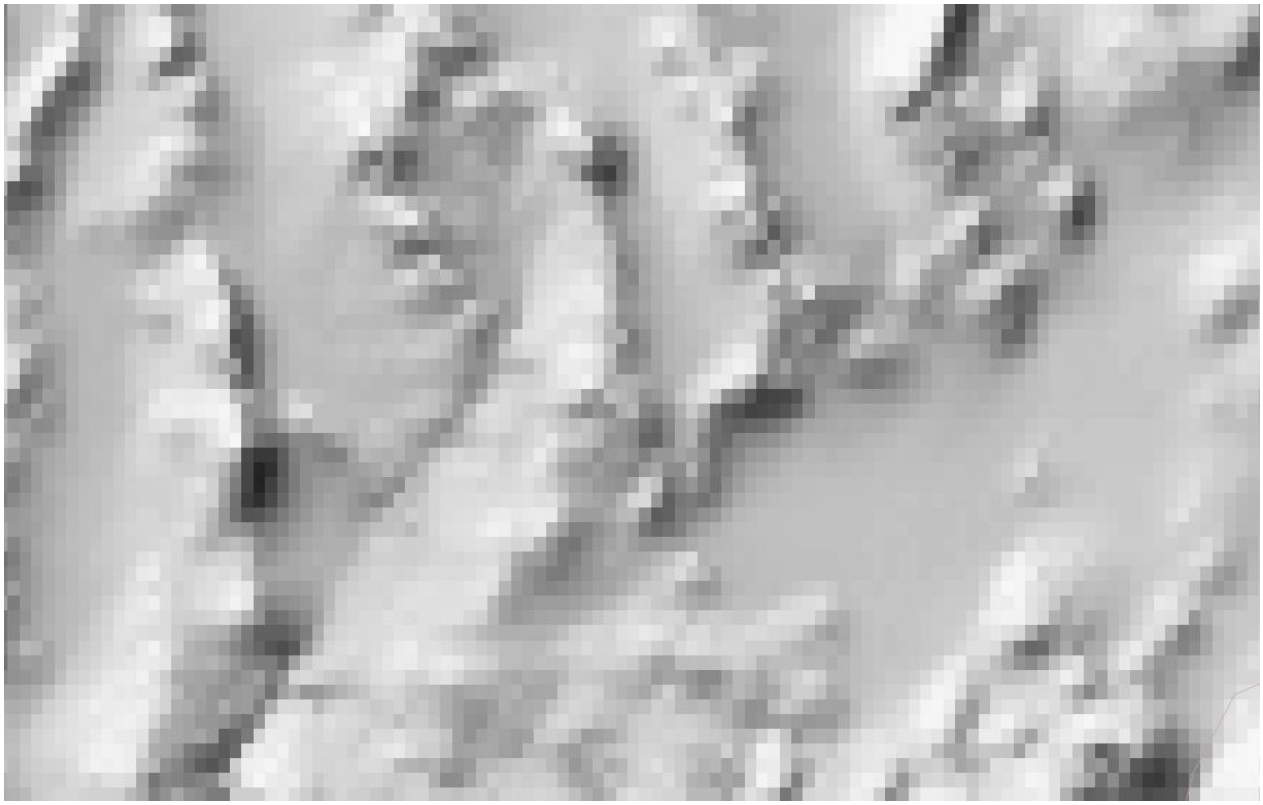
Data analysis such as extracting slope and aspect from Digital Elevation Models occurs with raster datasets.



This 10 meters shaded relief was developed from downsampled SRTM Plus elevation data. Source: Natural Earth Data.

AD

Zoom in close on a raster dataset and we will be able to see the individual cells.



Zoom in closely to a raster dataset and we will see the individual cells. Image: Grayscale shaded relief of land areas derived from downsampled SRTM Plus elevation data. Source: Natural Earth Data.

Raster Images

Aerial and satellite imagery is one type of raster data. Raster image file types include BMP, TIFF, GIF, and JPEG.

AD

Raster images accompanied by a second file known as a world file. The world file has the same name as the raster image file but has a different extension. The world file is a text file that contains the map projection information needed to properly georeference the raster image.

Scene :

The data are normally stored as raster data (referred to as “images”). Each separate image (for a place and time) is referred to as a s “scene”. As there are measurements in multiple wavelengths, a single “satellite image” has multiple observations for each pixel, that are stored in separate raster layers. In remote sensing jargon, these layers (variables) are referred to as “bands” (shorthand for “bandwidth”), and grid cells are referred to as “pixel”.

Spatial subset

Spatial subsetting can be used to limit analysis to a geographic subset of the image. Spatial subsets can be obtained by cropping of an image.

What is the purpose of image processing?

Image processing has multiple purposes.

- To improve the quality of an image in a subjective way, usually by increasing its contrast. This is called **image enhancement**.
- To use as few bits as possible to represent the image, with minimum deterioration in its quality. This is called **image compression**.
- To improve an image in an objective way, for example by reducing its blurring. This is called **image restoration**.
- To make explicit certain characteristics of the image which can be used to identify the contents of the image. This is called **feature extraction**.

How do we do image processing?

We perform image processing by using image transformations. Image transformations are performed using **operators**. An operator takes as input an image and produces another image. An example of operators are **linear operators**.

Consider O to be an operator which takes images into images. If f is an image, $O(f)$ is the result of applying O to f. O is linear if

$$O[af + bg] = aO[f] + bO[g]$$

for all images f and g and all scalars a and b.

Image Distortions

Any kind of errors present in remote sensing images are known as image distortions. Any remote sensing images acquired from either spaceborne or airborne platforms are susceptible to a variety of distortions. These distortions occur due to data recording procedure, shape and rotation of the Earth and environmental conditions prevailing at the time of data acquisition. Distortions occurring in remote sensing images can be categorised into two types:

- radiometric distortions and
- geometric distortions.

Sensors record the intensity of electromagnetic radiation (EMR) as digital numbers (DNs). These DNs are specific to the sensor and conditions under which they were measured. However, there are variations in the pixel intensities (digital numbers) irrespective of the object or scene being scanned. The recorded values get distorted due to one or more of the following factors:

- sensor ageing
- random malfunctioning of the sensor elements
- atmospheric interference at the time of image acquisition and
- topographic effects.

The above factors affect radiometry (variation in the pixel intensities) of the images and resultant distortions are known as radiometric distortions. As we know, an image is composed of finite number of pixels. The positions of these pixels are initially referenced by their row and column numbers. However, if we want to use images, we should be able to relate these

pixels to their positions on the Earth surface. Further, the distance, area, direction and shape properties vary across an image, thus these errors are known as geometric errors/distortions. This distortion is inherent in images because we attempt to represent three-dimensional Earth surface as a two-dimensional image. Geometric errors originate during the process of data collection and vary in type and magnitude. There are several factors causing geometric distortions such as:

- Earth's rotation
- Earth's curvature
- satellite platform instability and
- instrument error.

Image Corrections

As we now know that raw remote sensing images always contain significant amount of distortions, therefore, they cannot be used directly for further image analysis. The image correction involves image operations which normally precedes manipulation and analysis of image data to extract specific information. The primary aim of image correction operations is to correct distorted image data to create a more accurate representation of the original scene. Image corrections are also known as a preprocessing of remotely sensed images. It is a preparatory phase that improves quality of images and serves as a basis for further image analysis. Depending upon the kinds of errors which are present in images, the image correction functions are comprised of radiometric and geometric corrections. Radiometric correction attempts to improve the accuracy of measurements made by remote sensors pertaining to the spectral reflectance or emittance or back-scatter from the objects on the Earth surface. Geometric correction is the process of correcting geometric distortions and assigning the properties of a map to an image. Both of these corrections are made prior to actual use of remote sensing data in resource management, environmental monitoring and change detection applications by application scientists.

Image Registration

In image processing, for example, when combining the information content of image, we are interested in the relationship between two or more images. The analysis of this relationship usually be easily controlled or dealt with once a *correspondence* is set up between the images. The task of setting up this correspondence is called image registration.

Image registration is basically of two types:

1. **Image to image registration**: Here two or more images are aligned to integrate or fuse corresponding pixels that represent the same objects
2. **Image to map registration**: Here input image is warped to match the map information of a base image, while retaining the original spatial resolution.

Applications of Image registration

Image registration is the process of calculating spatial transforms which align a set of images to a common observational frame of reference, often one of the images in the set. Registration is a key step in any image analysis or understanding task where *different sources of data must be combined*. It is a critical component of computational photography, remote sensing,

automatic manufacturing processes, and medical image processing. More recently it has been used to create navigable models of a scene from a database of photographs, to remove unwanted objects from overlapping images, etc.

During registration process two situations become evident. First, it is impossible to determine which of the coordinate system points of one image matches with the corresponding points of another image, this is known as matching problem, and is also the most time consuming stage during the execution of algorithm. Second, there is need of transformation in the three dimensional information of one of the images regarding its coordinate system and its relationship to the image that was chosen as its reference. In simple words, image registration process is an automatic or manual procedure which tries to find corresponding points between two images and align them spatially to minimize a desired error, that is, a consistent distance measure between two images. There is a rapid development in image acquisition devices during last few decades and also growing amount and diversity of images that are obtained has invoked the research on automatic image registration.

The four basic steps of image registration are as follows:

- **Feature Detection:** The Salient and distinctive objects (Closed boundary regions, Edges, Contours, Line intersections, Corners, etc) in both reference image and sensed image are detected automatically or manually by a domain expert.
- **Feature Matching:** The correspondence between the features in the reference and sensed image is established. Matching method is based on image content or on the symbolic description of control point-set.
- **Transform Model Estimation:** The parameters and type of the so-called mapping functions, aligning the sensed image with the reference image, are estimated.
- **Image Resampling and Transformation:** By means of the mapping functions the sensed image is transformed

Some methods of image registration:

1. Pixel Based Method:
2. Feature Based Methods or Point Mapping Method
3. Registration Based on High Level Features (Contour Based Image Registration)
4. Multimodal Image Registration Using Mutual Information
5. Image Registration in Frequency Domain
6. Image Registration using Genetic Algorithm
7. Image Registration using Embedded Maps

Linear Contrast Stretch

Linear Contrast Stretch In its basic form the linear contrast-stretching technique involves the translation of the image PVs (Pixel values) from the observed range Vmin to Vmax to the full range of the display device (generally 0–255, which assumes an 8-bit display memory). V is a PV observed in the image under study, with Vmin being the lowest PV in the image and Vmax the highest. The PVs are scaled so that Vmin maps to a value of 0 and Vmax maps to a value of 255. Intermediate values retain their relative positions, so that the observed PV in the middle of the range from Vmin to Vmax maps to 127. Notice that we cannot map the middle of the range of the observed PVs to 127.5 (which is exactly half way between 0 and 255) because the display system can store only the discrete levels 0, 1, 2,...,255.

Some dedicated image processing systems include a hardware lookup table (LUT) that can be set so that the colour that you see at a certain pixel position on the screen is a mapping or modification of the colour in the corresponding position in the display memory. The colour code in the display memory remains the same, but the mapping function may transform its value, for example by using the linear interpolation procedure described in the preceding paragraph. The fact that the colour values in the display memory are not altered can be a major advantage if the user has adopted a trial and error approach to contrast enhancement. The mapping is accomplished by the use of a LUT that has 256 entries, labelled 0–255. In its default state, these 256 elements contain the values 0–255. A PV of, say, 56in the display memory is not sent directly to the screen, but is passed through the LUT. This is done by reading the value held in position 56in the LUT. In its default (do nothing) state, entry 56in the LUT contains the value 56, so the screen display shows an image of what is contained in the display memory. To perform a contrast stretch, we first realize that the number of separate values contained in the display memory for a given image is calculated as (Vmax −Vmin +1), which must be 256 or less for an 8-bit display. All LUT output values corresponding to input values of Vmin or less are set to zero, while LUT output values corresponding to input values of Vmax or more are set to 255. The range Vmin −Vmax is then linearly mapped onto the range 0–255, as shown in Figure 5.5. Using the LUT shown in this figure, any pixel in the image having the value 16 (the minimum PV in the image) is transformed to an output value of 0 before being sent to the digital-to-analogue converter and thence to the display screen. All input values of 191 and more are transformed to output values of 255. The range of input values between 16 and 191 is linearly interpolated onto the full dynamic range of the display device, assumed in this example to be 0–255. If a colour (RGB) image is being stretched then the process is repeated separately for each of the components (R then G then B). Figure 5.6a shows a Landsat-7 ETM+ image of the south-east corner of The Wash in eastern England. The un-stretched image is shown. The histograms of the RGB inputs (corresponding to Landsat ETM+ bands 4, 3 and 2, respectively) are shown in Figure 5.6b. The histograms are calculated from the image PVs, and are simply counts of the number of PVs having the value 0, 1,...,255.

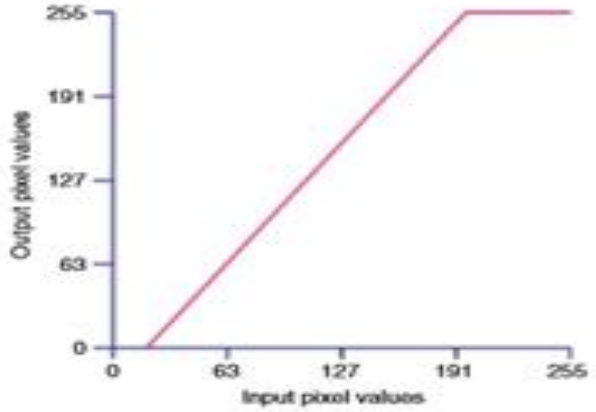


Figure 5.5 *Graphical representation of lookup table to map input pixel values 16–191 on to the full intensity range 0–255. Input values less than 16 are set to 0 on output. Input values of 191 or greater are set to 255 on output. Input values between 16 and 191 inclusive are linearly interpolated to output values 0–255.*

Histogram Equalization

The whole image histogram, rather than its extreme points, is used in the more sophisticated methods of contrast enhancement. Hence, the shape as well as the extent of the histogram is taken into consideration. The first of the two methods described here is called histogram equalization. Its underlying principle is straightforward. It is assumed that in a well-balanced image the histogram should be such that each brightness level contains an approximately equal number of PVs, so that the histogram of these displayed values is almost uniform (though not all 256 available levels are necessarily non-zero). If this operation, called histogram equalization, is performed then the entropy of the image, which is a measure of the information content of the image, will be increased. Because of the nature of remotely-sensed digital images, whose pixels can take on only the discrete values 0, 1, 2,...,255 it may be that there are 'too many' PVs in one class, even after equalization. However, it is not possible to take some of the values from that over-populated class and redistribute them to another class, for there is no way of distinguishing between one PV of 'x' and another of the same value. It is rare, therefore, for a histogram of the PVs of an image to be exactly uniformly distributed after the histogram equalization procedure has been applied.

The method itself involves, firstly, the calculation of the target number of PVs in each class of the equalized histogram. This value (call it n_t) is easily found by dividing N, the total number of pixels in the image, by 256 (the number of histogram classes, which is the number of intensity levels in the image).

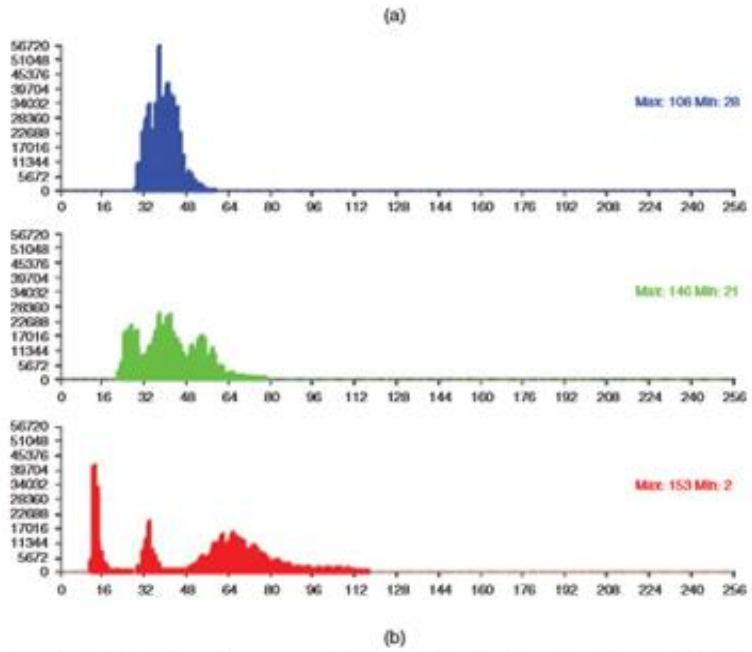


Figure 5.6 *(a) Raw Landsat-7 ETM+ false colour composite image (using bands 4, 3 and 2 as the RGB inputs) of the south-east corner of The Wash, an embayment in eastern England. The River Ouse can be seen entering The Wash. (b) Frequency histograms of the 256 colour levels used in the RGB channels. Landsat data courtesy NASA/USGS.*

Filtering Techniques

The image enhancement methods above change the way in which the information content of an image is presented to the viewer, either by altering image contrast or by coding a grey-scale image in pseudo-colour so as to emphasize or amplify some property of the image that is of interest to the user.

In symbolic terms, the information contained in an image can be represented by the following model:

image data = regional pattern + local pattern + noise = background + foreground (detail) + noise = low frequencies + high frequencies + noise

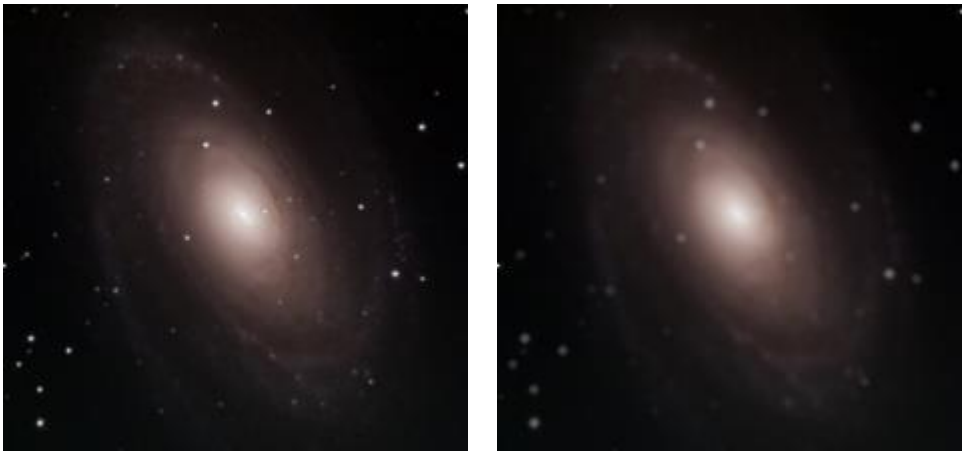
By analogy with the procedure used in chemistry laboratories to separate the components of a suspension, the techniques described in image analysis are known as filtering. A digital filter can be used to extract a particular spatial scale component from a digital image. The slowly varying background pattern in the image can be envisaged as a two-dimensional waveform with a long wavelength or low frequency; hence a filter that separates this slowly varying component from the remainder of the information present in the image is called a lowpass filter. Conversely, the more rapidly varying detail is like a two-dimensional waveform with a short wavelength or high frequency. A filter to separate out this component is called a high-pass filter. These two types of filter are considered separately. Low-frequency information allows the identification of the background pattern, and produces an output image in which the detail has been smoothed or removed from the original (input) image (hence low-pass filtering can be thought of as a form of blurring the image). High-frequency information allows us either to isolate or to amplify the local detail. If the high-frequency detail is amplified by adding back to the image some multiple of the high-frequency component extracted by the filter then the result is a sharper, deblurred image.

Spatial Domain Low-Pass (Smoothing) Filters

Before the topic of smoothing a two-dimensional image is considered, we will look at a simpler expression of the same problem, which is the smoothing of a one-dimensional pattern.

Low pass filtering (smoothing), is employed to remove high spatial frequency noise from a digital image. The low-pass filters usually employ moving window operator which affects one pixel of the image at a time, changing its value by some function of a local region (window) of pixels. The operator moves over the image to affect all the pixels in the image.

The most basic of filtering operations is called "low-pass". A low-pass filter, also called a "blurring" or "smoothing" filter, averages out rapid changes in intensity. The simplest low-pass filter just calculates the average of a pixel and all of its eight immediate neighbours. The result replaces the original value of the pixel. The process is repeated for every pixel in the image.



Before and After Low-Pass Filter

This low-pass filtered image looks a lot blurrier. But why would you want a blurrier image? Often images can be noisy – no matter how good the camera is, it always adds an amount of "snow" into the image.

Noise always changes rapidly from pixel to pixel because each pixel generates its own independent noise. The image from the telescope isn't "uncorrelated" in this fashion because real images are spread over many pixels. So the low-pass filter affects the noise more than it does the image. By suppressing the noise, gradual changes can be seen that were invisible before. Therefore a low-pass filter can sometimes be used to bring out faint details that were smothered by noise.

Filtering can be visualized by drawing a "convolution kernel". A kernel is a small grid showing how a pixel's filtered value depends on its neighbours. To perform a low-pass filter by simply averaging adjacent pixels, the following kernel is used:

+1/9	+1/9	+1/9
+1/9	+1/9	+1/9
+1/9	+1/9	+1/9

When this kernel is applied, each pixel and its eight neighbours are multiplied by 1/9 and added together. The pixel in the middle is replaced by the sum. This is repeated for each pixel in the image.

If we didn't want to filter so harshly, we could change the kernel to reduce the averaging, for example:

0	+1/8	0
+1/8	+1/2	+1/8
0	+1/8	0

The center pixel contributes half of its value to the result, and each of the four pixels above, below, left, and right of the center contribute 1/8 each. This will have a more subtle effect. By choosing different low-pass filters, we can pick the one that has enough noise smoothing, without blurring the image too much.

We could also make the kernel larger. The examples above were 3x3 pixels for a total of nine. We could use 5x5 just as easily, or even more. The only problem with using larger kernels is the number of calculations required becomes very large.

High Pass Filtering

A high pass filter is the basis for most sharpening methods. An image is sharpened when contrast is enhanced between adjoining areas with little variation in brightness or darkness. A high pass filter tends to retain the high frequency information within an image while reducing the low frequency information. The kernel of the high pass filter is designed to increase the brightness of the center pixel relative to neighbouring pixels. The kernel array usually contains a single positive value at its center, which is completely surrounded by negative values.

The above array is just an example of one possible kernel for a high pass filter. Other filters may include more weighting for the center point. A high-pass filter can be used to make an image appear sharper. These filters emphasize fine details in the image – exactly the opposite of the low-pass filter. High-pass filtering works in exactly the same way as low-pass filtering; it just uses a different convolution kernel. In the example below, notice the minus signs for the adjacent pixels. If there is no change in intensity, nothing happens. But if one pixel is brighter than its immediate neighbors, it gets boosted.

0	-1/4	0
-1/4	+2	-1/4
0	-1/4	0

Unfortunately, while low-pass filtering smooths out noise, high-pass filtering does just the opposite: it *amplifies noise*. You can get away with this if the original image is not too noisy; otherwise the noise will overwhelm the image.

High-pass filtering can also cause small, faint details to be greatly exaggerated. An over-processed image will look grainy and unnatural, and point sources will have dark donuts around them. So while high-pass filtering can often improve an image by sharpening detail, overdoing it can actually degrade the image quality significantly.

In terms of image frequencies the process of imaging or scanning involves blurring, as noted in the discussion of the point spread function (PSF), high frequencies are more heavily suppressed than are the low-frequency components of the image. It might therefore seem likely that the visual quality of an image might be improved by selectively increasing the contribution of its high-frequency components. Since the low-pass filters discussed involve some form of averaging (or spatial integration) then the use of the 'mathematical opposite' of averaging or integrating, namely the derivative function, might seem to be suited to the process of sharpening or de-blurring an image. However, a simpler way of performing an operation that is equivalent to high-pass filtering is considered before derivative-based methods as discussed below.

1. **Image Subtraction Method:** According to the model described an image can be considered to be the sum of its low and high frequency components, plus noise. The low frequency part can be isolated by the use of a low-pass filter as explained. This low-frequency image can be subtracted from the original, unfiltered, image leaving behind the high frequency component. The resulting image can be added back to the original, thus effectively doubling the high-frequency component.

2. **Derivative-Based Methods:** Other methods of high-pass filtering are based on the mathematical concept of the derivative, as noted earlier. The derivative of a continuous function at a specified point is the rate of change of that function value at that point. For example, the first derivative of position with respect to time (the rate of change of position over time) is velocity, assuming direction is constant. The greater the velocity of an object the more rapidly it changes its position with respect to time. The velocity can be measured at any time after motion commences. The velocity at time t is the first derivative of position with respect to time at time t. If the position of an object were to be graphed against time then the velocity (and hence the first derivative) at time t would be equal to the slope of the curve at the point time= t. Hence, the derivative gives a measure of the rate at which the function is increasing or decreasing at a particular point in time or, in terms of the graph, it measures the gradient of the curve.

Image Transforms

An image transform is an operation that re-expresses in a different, and possibly more meaningful, form all or part of the information content of a multispectral or greyscale image.

Spatial Transformations

The discrete Fourier and wavelet transforms convert the pixel values in a given spectral band to linear combinations of orthogonal functions of spatial frequency and distance. They may therefore be classified as spatial transformations.

Spectral Transformations

The principal components, minimum noise fraction and maximum autocorrelation factor transformations, create at each pixel location new linear combinations of the pixel intensities from all of the spectral bands and can properly be called spectral transformations.

Arithmatic Operations on Bands of Images

The term ‘transform’ is used somewhat loosely, for the arithmetic operations of addition, subtraction, multiplication and division are included, although they are not strictly transforms. These operations allow the generation of a derived image from two or more bands of a multispectral or multitemporal image. The derived image may well have properties that make it more suited to a particular purpose than the original. For example, the numerical difference between two images collected by the same sensor on different days may provide information about changes that have occurred between the two dates, while the ratio of the near-infrared (NIR) and red bands of a single-date image set is widely used as a vegetation index that correlates with difficult to measure variables such as vegetation vigour, biomass and leaf area index (LAI).

Vegetation indices are based on a model of the distribution of data values on two or more spectral bands considered simultaneously. Two examples of these transformations are the Perpendicular Vegetation Index (PVI), which uses a two dimensional model of the relationship between vegetation and soil pixels, and the Tasselled Cap transformation, which is based on the optical (visible plus NIR) bands of a multispectral data set.

Principal Component Analysis based Transform

The widely used technique of principal components analysis (PCA) is a method of re-expressing the information content of a multispectral set of *m* images in terms of a set of *m* principal components, which have two particular properties: zero correlation between the *m* principal components, and maximum variance. The maximum variance property of principal components means that the components are extracted in order of decreasing variance. The first component is that linear combination of spectral bands that has the maximum variance of all possible linear combinations of the same spectral bands. The second principal component is that linear combination of spectral bands that has the maximum variance with respect to the remaining part of the data once the effects of the first principal component have been removed, and so on. The zero correlation property means that principal components are statistically unrelated, or orthogonal. It is usually found that much of the information in the original *m*

correlated bands is expressible in terms of the first *p* of the full set of *m* principal components, where *p* is less than *m*. This property of PCA is useful in generating a false-colour composite image. If the image set consists of more than three bands then the problem arises of selecting three bands for display in red, green and blue (RGB). Since the principal components of the image set are arranged in order of variance (which is generally assumed to correlate with information, but may also include a noise component) then the first three principal components can be used as the RGB components of a false-colour composite image. No linear combination (i.e. weighted sum) of the original bands can contain more information than is present in the first three principal components. Another use of PCA is in reducing the amount of calculation involved in automatic classification by basing the classification on *p* principal components rather than on *m* spectral bands. In addition, the *p* principal component images require less storage space than the *m*-band multispectral image from which they were derived. Hence, PCA can also be considered to be a data compression transformation.

Rather than maximize the variance of the principal components, we could choose another criterion such as maximizing the signal to noise ratio (SNR). The standard principal components procedure, despite popular belief, does not remove noise. If we can estimate the level of noise in the data then we could extract components that are arranged in order of decreasing SNR. This modification of PCA, which we call noise-adjusted PCA.

Discrete Fourier Transforms

The transforms and operations described above act on two or more image bands covering a given area. A method of examining the information content of a single-band greyscale image in terms of its frequency components is the discrete Fourier transform (DFT). It provides for the representation of image data in terms of a coordinate framework that is based upon spatial frequencies rather than upon Euclidean distance from an origin (i.e. the conventional Cartesian or *xy* coordinate system). Image data that have been expressed in frequency terms are said to have been transformed from the image or spatial domain to the frequency domain. The frequency-domain representation of an image is useful in designing filters for special purposes and in colour coding the scale components of the image.

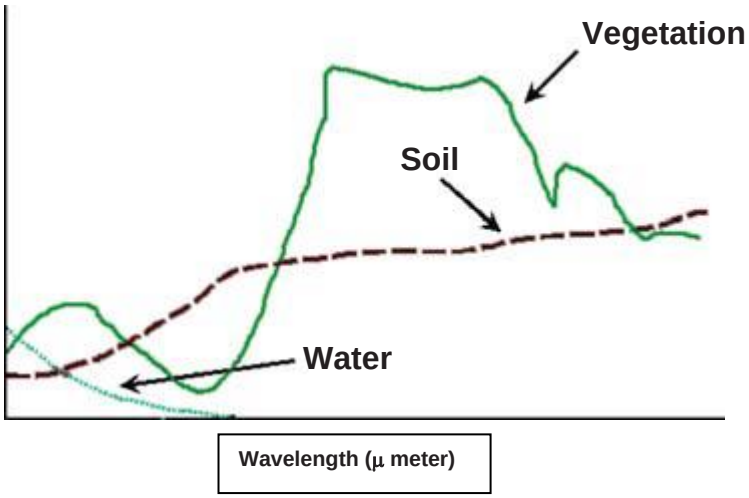
Wavelet Transform

A related transformation, called the discrete wavelet transform (DWT), represents an attempt to bridge the gap between the spatial and frequency domains, for it decomposes the input signal (which may be one-dimensional, like a spectrum collected by a field spectrometer, or two-dimensional, like an image) in terms of wavelength (1D) or space (2D) *and* scale simultaneously. One major use of wavelets is to remove noise from (i.e. ‘denoise’) one- and two-dimensional signals.

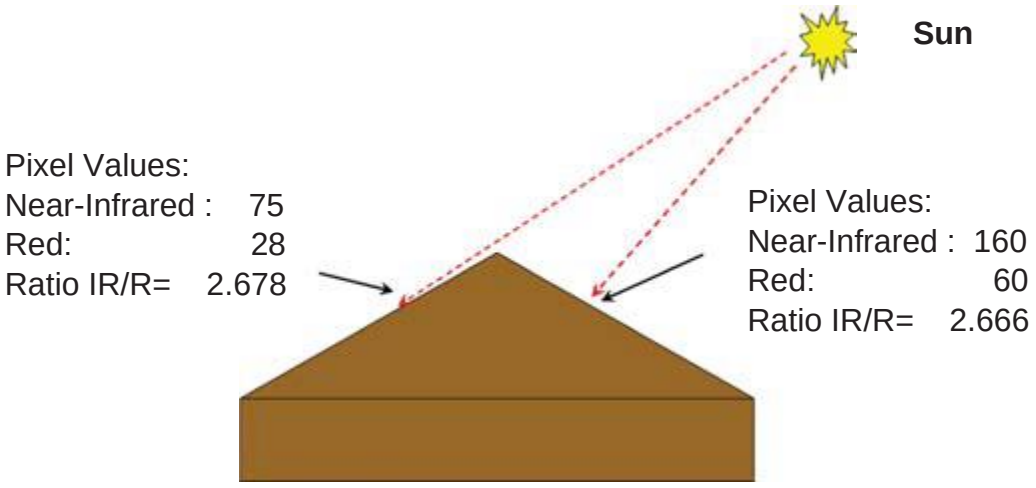
Image Division and Vegetation Indices

The process of dividing the pixel values in one image by the corresponding pixel values in a second image is known as ratioing. It is one of the most commonly used transformations applied to remotely-sensed images. There are two reasons why this is so. One is that certain aspects of the shape of spectral reflectance curves of different Earth-surface cover types can be brought out by ratioing. The second is that undesirable effects on the recorded radiances, such

as that resulting from variable illumination (and consequently changes in apparent upwelling radiance) caused by variations in topography can be reduced.



The Figure above shows the spectral reflectance curves for three cover types. The differences between the curves can be emphasized by looking at the gradient or slope between the red and the NIR bands, for example bands 3 (red) and 4 (NIR) in a Landsat ETM+ image, or bands 3 (NIR) and 2 (red) of the SPOT HRV image set. The shape of the spectral reflectance curve for water shows a decline between these two points, while that for vegetation shows a substantial increase. The spectral reflectance curve for soil increases gradually between the two bands. If a pixel value in the NIR band is divided by the equivalent value in the red band then the result will be a positive real number that exceeds 1.0 in magnitude. The same operation carried out on the curve for water gives a result that is less than 1.0, while the soil curve gives a value somewhat higher than 1.0. The greater the difference between the pixel values in the two chosen bands the greater the value of the ratio. The two images may as well be subtracted if this were the only result to be derived from the use of ratios.



The Figure above shows a hypothetical situation in which the irradiance at point *B* on the ground surface is only 50 of that at *A* due to the fact that one side of the slope is directly illuminated by the Sun. Subtraction of the values in the two bands at point *A* gives a result that is double that which would be achieved at point *B* even if both points are located on the same ground-cover type. However, the ratios of the two bands at *A* and *B* are the same because the topographic effect has been largely cancelled out in this instance. This is not always the case, as shown by the discussion below.

One of the most common spectral ratios used in studies of vegetation status is the ratio of the NIR to the equivalent red band value for each pixel location. This ratio exploits the fact that vigorous vegetation reflects strongly in the NIR and absorbs radiation in the red waveband. The result is a greyscale image that can be smoothed by a low-pass filter and density-sliced to produce an image showing variation in biomass (the amount of vegetative matter) and in leaf area index (LAI) as well as the state of health (physiological functioning) of plants.

More complex ratios involve sums of and differences between spectral bands. For example, the Normalized Difference Vegetation Index (NDVI), defined in terms of the NIR and red (R) bands as:

$$\boldsymbol{NDVI = \frac{NIR - R}{NIR + R}}$$

is preferred to the simple R : NIR ratio by many workers because the ratio value is not affected by the absolute pixel values in the NIR and R bands.

The sums and differences of bands are used in the NDVI rather than absolute values may make the NDVI more appropriate for use in studies where comparisons over time for a single area are involved, since the NDVI might be expected to be influenced to a lesser extent by variations in atmospheric conditions.

Empirically Based Image Transforms

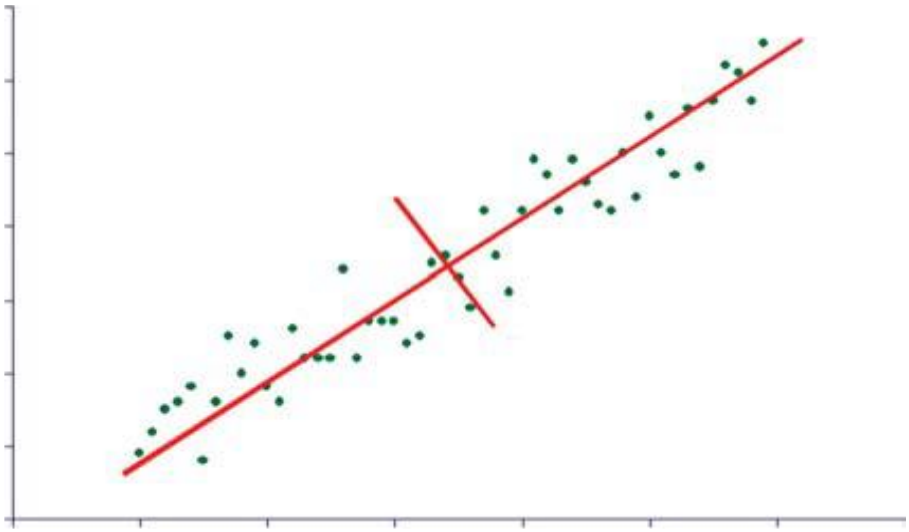
A transformation obtained, by rotating and scaling the axes e.g. of the four-dimensional space (for 4 band image of Landsat), would give a more clear view of the structure of the data. The transform is named as *Tasselled Cap* since the shape of the region of the transformed feature space that was occupied by vegetation in different stages of growth appeared like a Scottish ‘bobble hat’. Other transform; perhaps the best known is **Perpendicular Vegetation Index** (PVI) which was based on a similar idea to that of the Tasselled Cap, namely, that there is a definite axis in four-dimensional Landsat MSS space that is occupied by pixels representing soils, ranging from soils of low reflectance to those of high reflectance.

The PVI is not now widely used. Tasselled Cap or Kauth Thomas transformation is generally preferred, as it can be modified to deal with data from different sensors. However, its formulation depends on the definition of the soil line using empirical data.

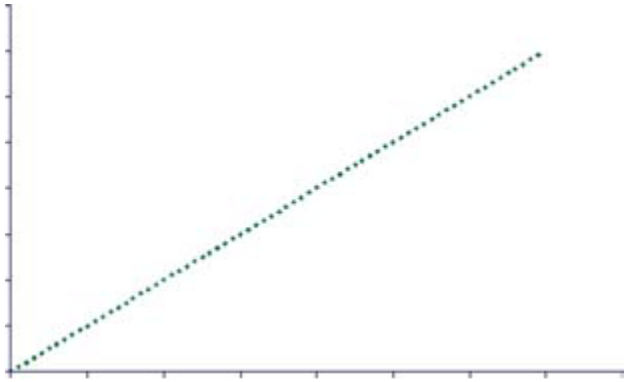
Principal Components Analysis

Standard Principal Components Analysis

Adjacent bands in a multi- or hyperspectral remotely sensed image are generally correlated. Multiband visible/NIR images of vegetated areas exhibit negative correlations between the NIR and visible red bands and positive correlations among the visible bands because the spectral reflectance characteristics of vegetation are such that as the vigour or greenness of the vegetation increases the red reflectance diminishes using the axes *AB* and *CD* rather than the conventional *x* and *y* axes might, in some cases, prove more revealing of the structures that are present within the data. Furthermore, if the variation in the direction *CD* in the following Figure contains only a small proportion of the total variability in the data then it may be ignored without too much loss of information, resulting in data compression.



This example shows that we must draw a basic distinction between the number of variables (e.g. spectral bands) in an image data set and the intrinsic dimensionality of that data set. In the following Figure the number of variables is two (*x* and *y*) but the dimensionality of the data as shown by the scatter of points is one. In the above Figure the dimensionality is again effectively one, although the number of observed variables is, in fact, two. In both examples the use of the single axis *AB* rather than the *x* and *y* axes accomplishes two aims: (i) a reduction in the size of the data set since a single coordinate on axis *AB* replaces the two coordinates on the *x* and *y* axes and (ii) the information conveyed by the set of coordinates on *AB* is greater than the information conveyed by the measurements on either the *x* or the *y* axes individually. In this context information means variance or scatter about the mean; it can also be related to the range of states or levels in the data.



Multispectral image data sets generally have a dimensionality that is less than the number of spectral bands. For example the fourband Landsat MSS Tasselled Cap transform produces two significant dimensions (brightness and greenness) while the six-band Landsat TM Tasselled Cap transform defines three meaningful functions (dimensions). The purpose of principal components analysis is to define the number of dimensions that are present in a data set and to fix the values of the coefficients which specify the positions of that set of axes which point in the directions of greatest variability in the data (such as axes *AB* and *CD* in the above Figure). These axes or dimensions of variability are always uncorrelated. A principal components transform of a multispectral image (or of a set of registered multitemporal images) might therefore be expected to perform the following operations:

- estimate the dimensionality of the data set and
- identify the principal axes of variability within

the data. These properties of PCA (sometimes also known as the Karhunen-Lo`eve transform) might prove to be useful if the data set is to be compressed, for example for transmission over a slow connection. Also, relationships between different groups of pixels representing different land cover types may become clearer if they are viewed in the principal axis reference system rather than in terms of the original spectral bands, especially as the variance of the data set is concentrated in relatively fewer principal components. Variance is often associated with information. The data compression property is useful if more than three spectral bands are available. A conventional RGB colour display system relates a spectral band to one of the three colour inputs (RGB). The Landsat TM provides seven bands of data, hence a decision must be made regarding which three of these seven bands are to be displayed as a colour composite image. If the basic dimensionality of the TM data is only three then most of the information in the seven bands will be expressible in terms of three principal components. The principal component images could therefore be used to generate a RGB false-colour composite with principal component number 1 shown in red, number 2 in green and number 3 in blue. Such an image contains more information than any combination of three spectral bands. The positions of the mutually perpendicular axes of maximum variability in the two-band data set shown in Figure 6.10b can be found easily by visual inspection to be the lines *AB* and *CD*. If the number of variables (spectral bands) is greater than three then a geometric solution is impracticable and an algebraic procedure must be sought. The direction of axis *AB* in Figure 6.10b is defined by the sign of the correlation between variables *x* and *y;* high positive correlation results in the scatter of points being restricted to an elliptical region of the twodimensional space defined by

the axes *x* and *y*. The line *AB* is, in fact, the major or principal axis of this ellipse and *CD* is the minor axis. In a multivariate context, the shape of the ellipsoid enclosing the scatter of data points in a *p*-dimensional space is defined by the variancecovariance matrix computed from *p* variables or spectral bands. The variance in each spectral band is proportional to the degree of scatter of the points in the direction parallel to the axis representing that variable, so that it can be deduced from Figure 6.11a that for the circular distribution the variances of variables *X* and Y (represented by GH and EF) are approximately equal. The covariance defines the shape of the ellipse enclosing the scatter of points. Figure 6.11a shows two distributions. One (green outline) has a high positive covariance while the other (blue outline) has a covariance of zero. The mean of each variable gives the location of the centre of the ellipse (or ellipsoid in a space of dimensionality higher than two). Thus, the mean vector and the variance-covariance matrix define the location and shape of the scatter of points in a *p*-dimensional space. The information contained in the variances and covariances of a set of variables is used again in the definition of the maximum likelihood classification procedure.

The relationship between the correlation matrix and the variance–covariance matrix sometimes leads to confusion. If the *p* variables making up the data set are measured on different and incompatible scales (for example three variables may be measured, respectively, in metres above sea level (elevation), in millibars (barometric pressure) and kilograms (weight)) then unit changes in the variances of these variables are not directly comparable – how many metres are equal to 1000 millibars? The importance of the variance in defining the scatter of points in a particular direction has already been stressed, so it is clear that if the variance is to be used in defining the shape of the ellipsoid enclosing the scatter of points in the *p*-dimensional space then the scales used to measure each variable must be comparable.

The Discrete Fourier Transform

The coefficients of the Tasselled Cap functions, and the eigenvectors associated with the principal components, define coordinate axes in the multidimensional data space containing the multispectral image data. These data are re-expressed in terms of a new set of coordinate axes and the resulting images have certain properties, which may be more suited to particular applications. The Fourier transform operates on a single-band (greyscale) image, not on a multispectral data set. Its purpose is to break down the spatial variation in grey levels into its spatial scale components, which are defined to be sinusoidal waves with varying amplitudes, frequencies and directions. The coordinates *k*1 and *k*2 of the tw odimensional space defined by the axes *U, V* in which these scale components are represented are given in terms of frequency (cycles per basic interval). This representation is called the frequency domain whereas the normal row/column coordinate system in which images are normally expressed is termed the spatial domain. The Fourier transform is used to convert a single-band image from its spatial domain representation to the equivalent frequency domain representation, and vice versa.

The idea underlying the Fourier transform is that the greyscale values forming a single-band image can be viewed as a three-dimensional intensity surface, with the rows and columns defining two axes (*x* and *y)* and the grey level intensity value at each pixel giving the third (*z*) dimension. A series of waveforms of increasing frequency and with different orientations is fitted to this intensity surface and the information associated with each such waveform is calculated. The Fourier transform therefore provides details of (i) the frequency of each of the scale components (waveforms) fitted to the image and (ii) the proportion of information

associated with each frequency component. Frequency is defined in terms of cycles per basic interval where the basic interval in the across-row direction is given by the number of pixels on each scan line, while the basic interval in the down-column direction is the number of scan lines. Frequency could be expressed in terms of metres by dividing the magnitude of the basic interval (in metres) by cycles per basic interval. Thus, if the basic interval is 512 pixels each 20m wide then the wavelength of the fifth harmonic component is *(*512 × 20*)/*5 or 2048 m. The first scale component, conventionally labelled zero, is simply the mean grey level value of the pixels making up the image. The remaining scale components have increasing frequencies (decreasing wavelengths) starting with 1 cycle per basic interval, then 2, 3, . . . , *n*/2 cycles per basic interval where *n* is the number of pixels or scan lines in the basic interval.

Two-Dimensional Fourier Transform

If a function is defined over a two-dimensional grid then the differences would be that (i) the scale components would be two-dimensional waveforms and (ii) each scale component would be characterized by orientation as well as by amplitude. The squared amplitudes of the waves are plotted against frequency in the horizontal and vertical directions to give a two-dimensional amplitude spectrum, which is interpreted much in the same way as the one-dimensional amplitude spectrum, the major differences being:

1. The frequency associated with the point [*k*1, *k*2] in the two-dimensional amplitude spectrum is given by:

$$\boldsymbol{k_{12} = \sqrt{k_1^2 + k_2^2}}$$

where the basic intervals given by each axis of the spatial domain image are equal, or by:

$$\boldsymbol{k_{12} = \sqrt{{k_1}/{n_1 \triangle t_1} + {k_2}/{n_2 \triangle t_2}}}$$

where the basic intervals in the two spatial dimensions of the image are unequal. In the latter case $n_1 \triangle t_1$and $n_2 \triangle t_2$ are the lengths of the two axes, *n*1 and *n*2 are the number of sampling points along each axis, and *t*1 and *t*2 are the sampling intervals. This implies that frequency is proportional to distance from the centre of the amplitude spectrum which is located at the point [0, 0] in the centre of the amplitude spectrum diagram.

2. The angle of travel of the waveform whose amplitude is located at point (*k*1, *k*2) in the amplitude spectrum is perpendicular to the line joining the point (*k*1, *k*2) to the centre (DC) point of the spectrum (0, 0). Consider an 512 × 512 pixel image made up of a set of horizontal lines spaced 16 rows apart. These lines are represented digitally by rows of 1s against a background of 0s. The amplitude spectrum shows a set of symmetric points running horizontally through the origin (centre point or DC). The points are so close that they give the appearance of a line. They represent the amplitudes of the waveforms reconstructed from the parallel, horizontal lines which could be considered to lie on the crests of a series of sinusoidal waveforms progressing down the image from top to bottom. Since the direction of travel could be topbottom or bottomtop the amplitude spectrum is symmetric and so the two points closest to the DC represent the amplitudes of the set of waves whose wavelength is equal to the spacing

between the horizontal lines. The two points further out from the DC represent a spurious waveform, which has a wavelength equal to double the distance between the horizontal lines. Such spurious waveforms represent a phenomenon termed *aliasing* which can be defined as the case in which several different frequency curves fit all or some of the available data points. It is a result of sampling a continuous signal.

++

The Fourier transform is **a representation of an image as a sum of complex exponentials of varying magnitudes, frequencies, and phases**. The Fourier transform plays a critical role in a broad range of image processing applications, including enhancement, analysis, restoration, and compression. In oteher words, the Fourier transform is a mathematical operation that converts amplitude as a function of time to amplitude as a function of frequency.

The Fourier transform can be used **to interpolate functions and to smooth signals**. For example, in the processing of pixelated images, the high spatial frequency edges of pixels can easily be removed with the aid of a two-dimensional Fourier transform.

Let $g(i, j)$, $i = 0, \ldots, c-1$, $j = 0, \ldots, r-1$, represent a gray-scale image. Its discrete

inverse Fourier transform is

$$g(i,j) = \sum_{k=0}^{c-1} \sum_{l=0}^{r-1} \hat{g}(k,l) e^{i2\pi \left(ik/c + jl/r \right)}$$

and the discrete Fourier transform is

$$\hat{g}(k,l) = \frac{1}{cr} \sum_{k=0}^{c-1} \sum_{l=0}^{r-1} g(i,j) e^{-i2\pi \left(ik/c + jl/r \right)}$$

The frequency coefficients in $\hat{g}(k,l)$ in the above equations are complex numbers.

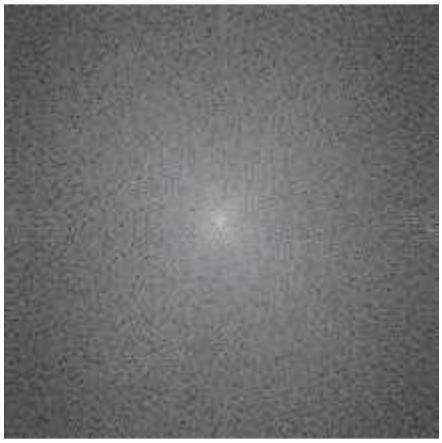


FIGURE 3.1

Logarithm of the power spectrum for the 3N band of the ASTER image of Figure 1.1.

Edge Detection

A high-pass filtered image that is added back to theoriginal image is a high-boost filter and the result is a sharpened or de-blurred image. The high-pass filtered image can be used alone, particularly in the study of the location and geographical distribution of 'edges'. An edge is a discontinuity or sharp change in the greyscale value at a particular pixel point and it may have some interpretation in terms of cultural features, such as roads or field boundaries, or in terms of geological structure or relief. We have already noted that the first difference can be computed for the horizontal, vertical and diagonal directions, and the magnitude and direction of the maximum spatial gradient can also be used. Other methods include the subtraction of a low-pass filtered image from the original or the use of the Roberts Gradient. A method not so far described is the Sobel nonlinear edge operator which is applied to a 3 × 3 window area. The value of this operator for the 3 × 3 window defined by:

$$A\ B\ C$$

$$D\ E\ F$$

$$G\ H\ I$$

is given for the pixel underlying the central window weight (E) by the function: $S = X2 + Y\,2$

where,

$$X = (C + 2F + I) - (A + 2D + G)$$

$$Y = (A + 2B + C) - (G + 2H + I)$$

This operation can also be considered in terms of two sets of filter weight matrices. X is given by the following weight matrix, which determines horizontal differences in the neighbourhood of the centre pixel:

$$\begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix}$$

while Y is given by a weight matrix which involves vertical differences:

$$\begin{matrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{matrix}$$

The Roberts and Sobel techniques produced a too-intense enhancement of local edges but did not remove the regional patterns. One of the many uses of edge-detection techniques (image

segmentation) is in the enhancement of images for the visual identification and analysis of geological lineaments, which are defined as mappable, simple or composite linear features whose parts are aligned in a rectilinear or slightly curvilinear relationship and which differ distinctly from the pattern of adjacent features and which presumably reflect a subsurface phenomenon.

The sub-surface phenomena to which the definition refers are presumed to be fault and joint patterns in the underlying rock. However, linear features produced from remotely-sensed images using the techniques described in this section should be interpreted with care.

identify lineaments.

Other applications of edge-detection techniques include the determination of the boundaries of homogeneous regions (segmentation) in an image from the Sobel filter is related to the degree of contrast between the pixels on either side of the edge in question. Landsat data courtesy NASA/USGS.

Unit- IV Supervised and Unsupervised Image Classification

Classification of Remotely sensed Images

The process of classification consists of two stages. The first is the recognition of categories of real-world objects. In the context of remote sensing of the land surface these categories could include, for example woodlands, water bodies, grassland and other land cover types, depending on the geographical scale and nature of the study.

The second stage in the classification process is the labelling of the entities (normally pixels) to be classified. In digital image classification these labels are numerical, so that a pixel that is recognized as belonging to the class 'water' may be given the label '1', 'woodland' may be labelled '2', and so on. The process of image classification requires the user to perform the following steps:

1. Determine a priori the number and nature of the categories in terms of which the land- cover is to be described.

2. Assign labels to the pixels on the basis of their properties using a decision-making procedure, usually termed a classification rule or a decision rule.

Sometimes these steps are called classification and identification (or labelling), respectively. The classification stage is normally based on a predetermined number of classes that, one hopes, can be observed on the ground at the chosen spatial scale. These are the target or information classes. Clustering, which is described next, produces classes that are more or less spectrally distinct, and these are called spectral classes. They may correspond to information classes, providing the spatial scales match each other.

In contrast to the classification procedure, the process of clustering does not require the definition of a set of categories in terms of which the land surface is to be described. Clustering is a kind of exploratory data analysis or data mining procedure, the aim of which is to determine the number (but not initially the identity) of land cover categories that can be separated in the area covered by the image, and to allocate pixels to these categories. Identification of the clusters or categories in terms of the nature of the land cover types is a separate stage that follows the clustering procedure.

These two approaches to pixel labelling are known in the remote sensing literature as supervised and unsupervised classification procedures, respectively. They can be used to segment an image into regions with similar attributes. Although land cover classification is used above as an example, similar procedures can be applied to clouds, water bodies and other objects present in the image. In all cases, however, the properties of the pixel to be classified are used to label that pixel. In the simplest case, a pixel is characterized by a vector whose elements are its grey levels in each spectral band. This vector represents the spectral properties of that pixel.

Supervised Classification:

Supervised classification is the technique most often used for the quantitative analysis of remote sensing image data. At its core is the concept of segmenting the spectral domain into regions that can be associated with the ground cover classes of interest to a particular application.

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 assign them distinct 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 appears fairly homogeneous on the image (as determined by similarity in tone or colour within shapes delineating the category).

Generally supervised classification is essentially a mapping from the measurement space of the sensor to a field of labels that represent the ground cover types of interest to the user. It depends on having enough pixels available, whose class labels are known, with which to train the classifier. In this context ‘‘training’’ refers to the estimation of the parameters that the classifier needs in order to be able to recognize and label unseen pixels. The labels represent the classes on the map that the user requires. The map is called the thematic map, meaning a map of themes. The two statistical supervised classifiers that analysts most readily use are the classifiers based on Linear Discriminant Function (LDF) and the Quadratic Discriminant Function (QDF).

Parametric Classification

The supervised parametric classification techniques assume that the observed feature matrices for each spectral class in the digital image come from a known probability distribution and make inferences about the parameters of the classes under this assumption. The training data is used to estimate the parameters or other constants required to operate the chosen classification algorithm. If the algorithm involves explicit parameters, such as the mean vector and covariance matrix for the multivariate normal distribution, then the technique is called parametric.

Maximum Likelihood Classifier

Maximum likelihood classifier (MLC) is one of the most common supervised classification techniques used with remote sensing image data, and was the first rigorous algorithm to be employed widely. MLC is derived from Bayes decision rule where likelihood of each pixel belonging to one of the pre-defined set of classes is calculated under the assumption of known probability distributions of underlying data classes, and the pixel is allocated to the class for which it has the maximum likelihood. Under the most general setting of MLC, the underlying information classes are assumed to be the form of multivariate normal model. The most commonly used parametric distribution models in designing a MLC are Gaussian distribution. Further, depending upon the associated assumptions of equal and unequal covariance matrices for component classes, MLC is formulated as a Bayes-normal-linear rule resulting in linear decision boundaries and Bayes-normal-quadratic rule resulting in quadratic decision boundaries respectively.

Let the classes be represented by ω_i , $i = 1,...,M$ where M is the number of classes. In determining the class or category to which a pixel with measurement vector x belongs, the conditional probabilities

$$p\ (\omega_i | x),\ i = 1,...,M$$

play a central role. The vector *x* is a column vector of the brightness values for the pixel in each measurement band. It describes the pixel as a point in spectral space. The probability $p\ (\omega_i | x)$ tells us the likelihood that ω_i is the correct class for the pixel at position *x* in the spectral space. If we knew the complete set of $p\ (\omega_i | x)$ for the pixel, one for each class, and then we could label the pixel, classify it according to the decision rule.

$$x \in \omega_i \textbf{ if } p\ (\omega_i | x) > p\ (\omega_j | x) \text{ for all } j \neq i. \qquad (1)$$

This says that the pixel with measurement vector *x* is a member of class ω_i if $p\ (\omega_i | x)$ is the largest probability of the set.

Despite the simplicity of (1) the $p\ (\omega_i | x)$ are unknown. What we can find relatively easily, though, are the set of class conditional probabilities $p(x | \omega_i)$; which describe the chances of finding a pixel at position *x* in spectral space from each of the classes ω_i. Those conditional probability functions are estimated from labeled training data for each class. The desired $p\ (\omega_i | x)$ in (1) and the available $p(x | \omega_i)$ estimated from training data, are related by Bayes theorem

$$p\ (\omega_i | x) = p(x | \omega_i)\ p(\omega_i) / p(x) \qquad (2)$$

in which $p(\omega_i)$ is the probability that pixels from class ω_i appear anywhere in the image. The $p(\omega_i)$ are referred to as prior probabilities or sometimes just priors, and

$$p(x) = \sum_{i=1}^{M} p(x | \omega_i) p(\omega_i) \qquad (3)$$

The term $p(x)$ is the probability of finding a pixel with measurement vector *x* in the image, from any class which remains uniform over all observations and hence, can be ignored for classification so that the classification criterion in equation (1) reduces to

$$x \in \omega_i \text{ if } p(x | \omega_j)\, p(\omega_j) > p(x | \omega_i)\, p(\omega_i) \qquad i \neq j \qquad (4)$$

The decision rule of (4) is more acceptable than that of (1) since the $p(x | \omega_i)$ known from training data and it is conceivable that the priors $p(\omega_i)$ are also known or can be estimated.

Non-Parametric Classification Techniques

Non-parametric classification methods also referred as the distribution free methods do not impose any distributional assumptions on the observed feature matrices and hence, are considered robust for a wide variety of class distributions as long as the spectral signatures of the information classes are distinct. A variety of non-parametric classifiers are available in

image analysis literature. Among the statistical non-parametric supervised classification algorithms, parallelepiped and minimum distance classifiers are the most frequently used ones.

Parallelepiped Classifier

The parallelepiped classifier is a very simple supervised classifier that is trained by finding the upper and lower brightness values in each spectral dimension. Often that is done by inspecting histograms of the individual spectral components in the available training data, shown in Figure 1(a). Together the upper and lower bounds in each dimension define a multidimensional box or parallelepiped; Figure 1(b) displays the segmentation of the feature space into the parallelepipeds for two dimensional classification problems. Unknown pixels are labeled as coming from the class of the parallelepiped within which they lie.

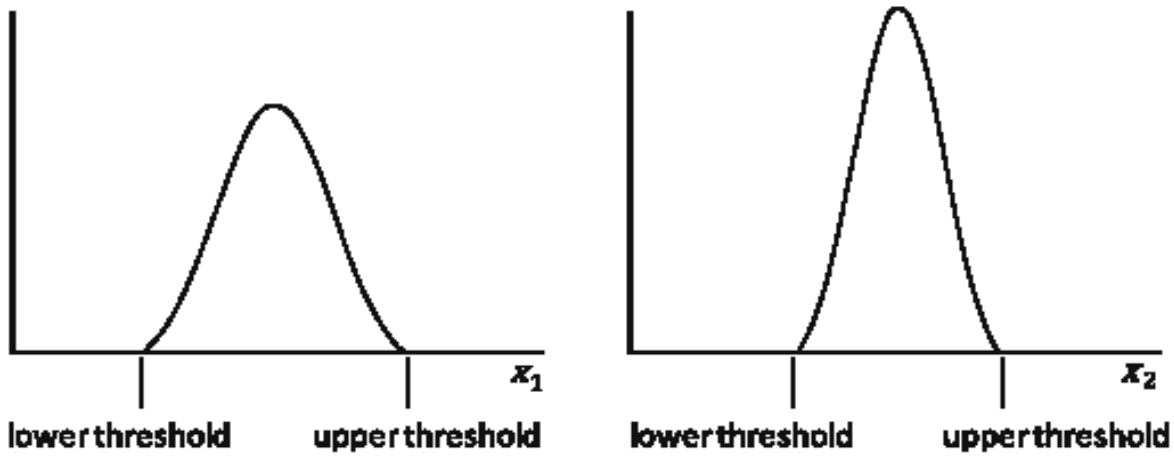


Figure 1(a): Setting the parallelepiped boundaries by inspecting class histograms in each band

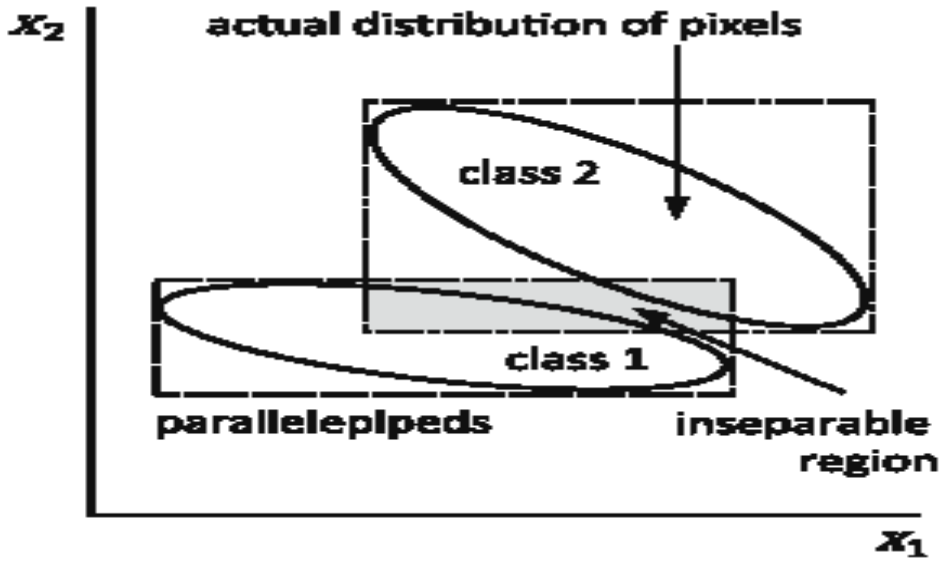


Figure 1(b): Classification of correlated data showing regions of inseparability.

While it is a very simple, and fast, classifier, it has several limitations. First, there can be considerable gaps between the parallelepipeds in spectral space; pixels in those regions cannot be classified. By contrast, the maximum likelihood and minimum distance rules will always label unknown pixels unless thresholds are applied. Secondly, for correlated data some parallelepipeds can overlap, as illustrated in Figure 1(b), because their sides are always parallel to the spectral axes.

Minimum Distance Classifier

Minimum distance classifier (MDC) is another distribution-free simple statistical classifier which somehow resembles the k-means clustering algorithm in that it also calculates distance based dissimilarity measures for decision making. As the name suggests, this classifier calculates the distance between a pixel and the centroids of the training data classes and accordingly decides to assign the pixel to the nearest class. Although, other distance measures can also be used for calculating the distances between pixel and the class centers, Mahalanobis distance and Euclidean distance measures are most generally used in practice.

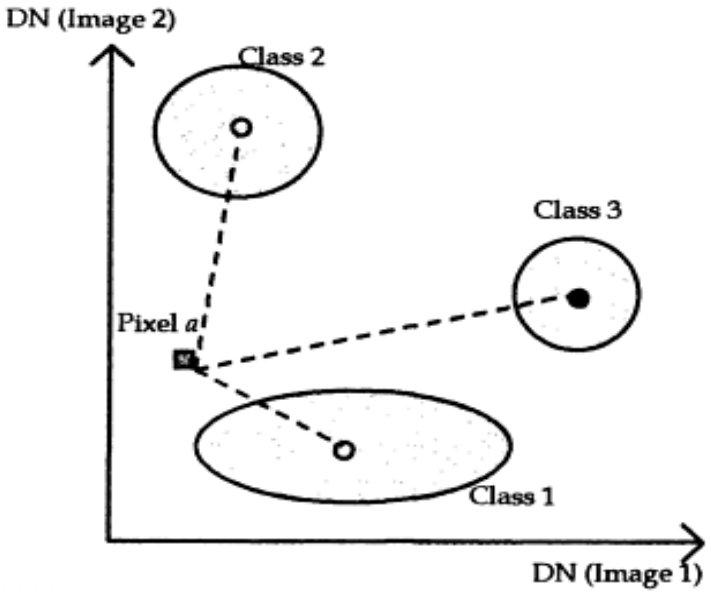


Figure 2: Minimum distance classifier using Euclidean distance.

Figure 2 shows the pixel is at minimum distance from class 1 and hence belonging to class 1 by the minimum distance classification algorithm.

Classification Accuracy Assessment

The most common tool used for the classification accuracy assessment is in terms of a confusion (or error) matrix. Since the risk is seldom deterministic, we might also need to check the reliability of the risk estimation. This is achieved using the random agreement measures. This section refers to a class of discrete multivariate techniques which take into account the chance allocations made by a classifier for assessing its performance. The most popular among them is kappa measure which is suggested by Cohen (1960) and is therefore often referred to as Cohen's kappa. The Kappa coefficient uses all of the information in the confusion matrix

in order for the chance allocation of labels to be taken into consideration. It provides a better measure of the accuracy of a classifier than the overall accuracy as it consider inter-class agreement and is calculated as:

$$\kappa = \frac{p_0 - p(k)}{1 - p(k)}$$

where, κ is the kappa coefficient, p_0 is the proportion of overall agreement and $p(k)$ is the chance agreement probability defined by Cohen as,

$$p_0 = \sum_{i=1}^{m} \frac{a_{ii}}{N}$$

$$p(k) = \sum_{i=1}^{n} \frac{(r_i \times c_i)}{N^2}$$

where, *N* is the total number of observations. The higher the value of kappa coefficient κ, the better the classification performance. In the ideal situation, when all the pixels are correctly classified, it takes value equal to 1.

Linear Discriminant Function

Linear discriminant analysis (LDA) is based on the generalization of Fisher's linear discriminant function. A method used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or more commonly, for dimensionality reduction before classification.

Let us consider two independent random samples of sizes n_1 and n_2 respectively, from each of two p – variate populations, and that a method which best distinguishes between these samples was required. The only assumption is that the dispersion matrices in these two populations were equal, otherwise the populations were completely unspecified. With this assumption, the data can be summarized by computing sample mean vectors $\bar{X}_1$ and $\bar{X}_2$ and the pooled within-sample covariance matrix S. Fisher (1938), then looked for the linear combination $w = a'\underline{X}$ of responses that gave maximum separation of the group means when measured relative to the within-group variance of the data. This linear combination he found, by maximizing $\{a'(\bar{X}_1 - \bar{X}_2)\}^2 / a'Sa$, to be $w = (\bar{X}_1 - \bar{X}_2)'S^{-1}\underline{X}$. The given group separation is maximized by this function, then a sensible allocation rule can be constructed by allocating $\underline{X}$ to π_1 if $(\bar{X}_1 - \bar{X}_2)'S^{-1}\underline{X}$ is greater than some constant k, and otherwise to π_2.

Note that this function is of exactly the same form as to allocate $\underline{X}$ to π_1 if $L(\underline{X}) > \log_e k$, otherwise to π_2 where $L(\underline{X}) = (\bar{X}_1 - \bar{X}_2)'S^{-1}\{\underline{X} - \frac{1}{2}(\bar{X}_1 + \bar{X}_2)\}$. The function L is known as the sample linear discriminant function. Since the portion

$\frac{1}{2}(\overline{X}_1 - \overline{X}_2)' S^{-1}(\overline{X}_1 + \overline{X}_2)$ of the latter is merely a sample-based constant and can be absorbed into right-hand side of inequality. The function $(\overline{X}_1 - \overline{X}_2)' S^{-1} \underline{X}$ is generally known as Fisher's linear discriminant function (LDF).

K-means Algorithm

The k-means algorithm an exploratory classification algorithm should require little, if any, user interaction. The workings of such a technique, called the k-means clustering algorithm, are now described by means of an example. Figure3 shows two well-separated groups of points in a two-dimensional feature space. The members of each group are drawn from separate bivariate-normal distributions. It is assumed that we know that there are two groups of points but that we do not know the positions of the centres of the groups in the feature space. Points '10' and '20' represent a first guess at these positions. The 'shortest distance to center' decision rule, as described earlier, is used to label each unknown point (represented by a dot in the figure) with a '1' or a '2' depending on the relative Euclidean distance of the point from the initial cluster centers, labeled '10' and '20'. Thus, the (squared) Euclidean distances to cluster centers 1 and 2 (d2 q1 and d2 q2) are computed for each point q, and q is allocated the label '1' if d2 q1 is less than d2 q2 or the label '2' if d2 q2 is less than d2 q1 . If the two squared distances are equal, then the point is arbitrarily allocated the label '1'.

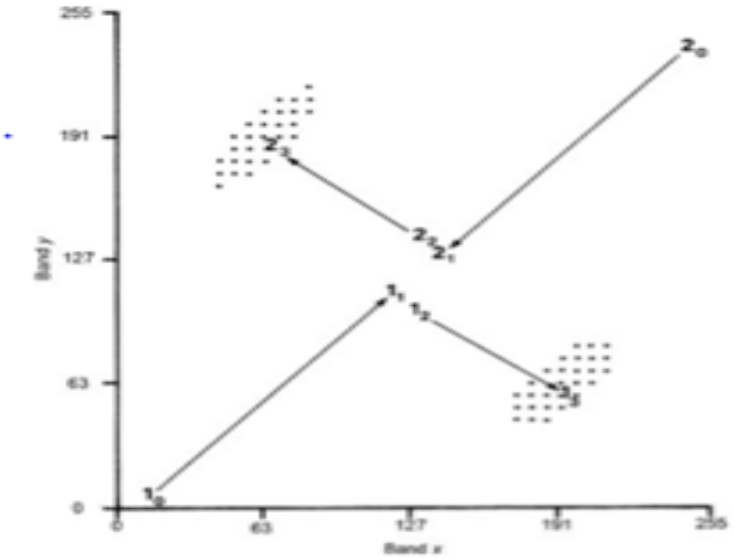


Figure 3. Illustrating the iterative calculation of centroid positions for two well-separated groups of points in a two-dimensional feature space defined by axes labelled Band x and Band y. Points 10 and 20 migrate in three moves from their initial random starting positions to the centres of the two clouds of points that represent the two classes of pixels.