Lecture notes for NSCP-funded short course - Remote sensing training for resource officers and farmers

Western Australian Department of Agriculture

Follow this and additional works at: http://researchlibrary.agric.wa.gov.au/lr_scsc

Part of the Agriculture Commons, Biodiversity Commons, Databases and Information Systems Commons, and the Other Computer Sciences Commons

Recommended Citation

Western Australian Department of Agriculture. (1992), Lecture notes for NSCP-funded short course - Remote sensing training for resource officers and farmers. Department of Agriculture and Food, Western Australia, Perth. Report.

This report is brought to you for free and open access by Research Library. It has been accepted for inclusion in Soil conservation survey collection by an authorized administrator of Research Library. For more information, please contact jennifer.heathcote@agric.wa.gov.au, sandra.papenfus@agric.wa.gov.au.
IMPORTANT DISCLAIMER

This document has been obtained from DAFWA's research library website (researchlibrary.agric.wa.gov.au) which hosts DAFWA's archival research publications. Although reasonable care was taken to make the information in the document accurate at the time it was first published, DAFWA does not make any representations or warranties about its accuracy, reliability, currency, completeness or suitability for any particular purpose. It may be out of date, inaccurate or misleading or conflict with current laws, polices or practices. DAFWA has not reviewed or revised the information before making the document available from its research library website. Before using the information, you should carefully evaluate its accuracy, currency, completeness and relevance for your purposes. We recommend you also search for more recent information on DAFWA's research library website, DAFWA's main website (https://www.agric.wa.gov.au) and other appropriate websites and sources.

Information in, or referred to in, documents on DAFWA's research library website is not tailored to the circumstances of individual farms, people or businesses, and does not constitute legal, business, scientific, agricultural or farm management advice. We recommend before making any significant decisions, you obtain advice from appropriate professionals who have taken into account your individual circumstances and objectives.

The Chief Executive Officer of the Department of Agriculture and Food and the State of Western Australia and their employees and agents (collectively and individually referred to below as DAFWA) accept no liability whatsoever, by reason of negligence or otherwise, arising from any use or release of information in, or referred to in, this document, or any error, inaccuracy or omission in the information.
LECTURE NOTES FOR NSCP-FUNDED SHORT COURSE

REMOTE SENSING TRAINING

FOR RESOURCE OFFICERS

AND FARMERS

1992

WA DEPARTMENT OF AGRICULTURE
1. BASIC REMOTE SENSING PRINCIPLES

1.1 Introduction

The aim of this course is to provide an introduction to remote sensing, and knowledge and practice in image processing techniques for the extraction of information from remotely sensed data.

Aerial photography is the original form of earth remote sensing, and is still a widely used and important data source for many purposes. This course is directed principally at the analysis of digital image data recorded by satellite or airborne platforms. Remote sensing is considered to be the methodology of collecting and interpreting target information over a broad range of the electromagnetic spectrum.

To make effective use of remotely sensed data, it is necessary to understand the nature of the data, and of the display systems.

![Schematic diagram of data collection, assembly and analysis](image)

Figure 1.1  Schematic diagram of data collection, assembly and analysis
Digital data are available from a range of sensing systems. While end users may have little control over the design of these systems, they can make decisions about choosing appropriate data for particular applications. It is therefore useful to consider the practice of remote sensing as beginning with data acquisition, followed by data analysis to extract and produce appropriate information products (Figure 1.2).

![Diagram of Remote Sensing System](image)

From Lillesand & Kiefer (1999)

Figure 1.2  Schematic of Remote Sensing Systems

The general concept of applied remote sensing is the use of imagery for inventory of the earth’s resources, ultimately to produce a map of their spatial distribution and variation with time, and/or statistics describing the nature, quantity and distribution of the resource. Examples of this type of activity include:

- forest cover mapping
- topographic maps
- soil maps
- statistics relating to crops (types, distribution, yield).

1.2  Physical Basis of Remote Sensing

All of the remote sensing systems record measurements of electromagnetic radiation reflected or emitted by the target area. These data can be useful because physical differences in the composition or condition of target materials affect the spectral response. A special case of this is the visual perception of colour and texture using energy in the visible wavelengths and chemical detectors in the human eye.

The energy source for the ‘passive sensors’ in earth resource satellites is the sun (or the target itself in the thermal channels). Radar systems are ‘active’ sensors, where the energy pulses are generated by the sensing system.

It is relevant to consider the energy source, the sensor system and the atmospheric effects, as these will all affect the ability to detect differences in the ‘spectral signatures’ of target materials.

In remote sensing, we are primarily concerned with energy transfer by means of radiation. Energy that is radiated behaves in accordance with basic wave theory. That is, an electromagnetic wave is
equally and respectively spaced in time, moves with the velocity of light and has two force fields that are at right angles to each other, one electric and the other magnetic (Figure 1.3).

![Figure 1.3 Electromagnetic Wave](image)

Three measurements are used to describe electromagnetic waves: wavelength ($\lambda$), in micrometers ($\mu$m); frequency ($f$), in hertz (Hz); and velocity ($c$), in metres (m). Within a given medium, velocity is constant at the speed of light. (Hertz is the unit of frequency for 1 cycle/second.)

The relationship between velocity, frequency and wavelength of electromagnetic radiation is:

$$\text{Velocity} = \text{frequency} \times \text{wavelength}.$$  

Velocity is the speed at which wave crests advance past a point. Wavelength is the distance from one wave crest to an adjacent wave crest. Frequency is the number of waves that pass a given point in a specified period of time.

The wave velocity of electromagnetic radiation (the speed of light) changes when passing through media of different density, for example, from air to glass. The wavelength also alters in response to this velocity change, never the frequency.

As the basis of remote sensing is electromagnetic energy transfer by radiation, other fundamentals need to be considered.

(i) Max Planck in 1890 proposed the quantum theory of electromagnetic radiation, in essence that energy is transferred in units called photons, and is inversely proportional to wavelength. As an example, X-rays and ultraviolet rays are small in wavelength but high in energy, and are in fact harmful to humans. Infrared and microwaves are high in wavelength but low in energy and, like radio waves, are generally harmless to humans. From a practical remote sensing viewpoint, it means that the long wavelength/low energy bands are much more difficult to detect or sense.

(ii) Wien's Law states that the wavelength of peak emittance is inversely related to the temperature of the source. As an example, consider a piece of steel slowly heated. As the steel becomes progressively hotter, it begins to glow and change colour from pink to dull red, orange, yellow and finally white. As the source gets hotter, so the wavelength gets smaller.
1.4

(iii) Stefan’s Law, Kirchof’s Law

All matter above 0°C continuously emits electromagnetic radiation, during both day and night. Objects actually fluoresce, but because this radiation is outside the band of visible light, it cannot be sensed by the human eye. The electromagnetic energy radiated from a source is the internal or kinetic energy generated by the collision of particles within that source.

An object that will absorb all incident energy upon it is called a blackbody. Since by definition a blackbody absorbs incident energy, no energy is reflected. A blackbody is also a perfect radiator *i.e* it re-emits all its absorbed energy in a wavelength distribution pattern dependent only on the kinetic temperature. A blackbody is a physical abstraction, for no material absorbs all radiation, nor emits or radiates all absorbed energy.

The blackbody concept is the datum to which emissivity of materials is compared. The emissivity of a blackbody is 1. For all real materials, it will be less than 1. In reality, all objects are "grey" bodies.

Stefan’s Law states that the total blackbody radiation emitted is proportional to the fourth power of the absolute temperature (*e.g* the sun will emit more energy than a light globe).

Kirchof’s Law states that the emittance of an actual body is the product of the blackbody radiation and the emissivity of the real radiator.

In summary, hot objects like the sun (6000°C) will radiate short wavelengths of electromagnetic radiation high in energy which are simple to sense, while cool objects like the earth (300°C) radiate long wavelengths low in energy which are difficult to sense.

1.3 Electromagnetic Spectrum

The link between the components of the remote sensing system is electromagnetic energy. In remote sensing, we are primarily concerned with energy transfer by means of radiation. Electromagnetic radiation occurs as a continuum of wavelengths and frequencies.
Figure 1.4  The Electromagnetic Spectrum

The electromagnetic spectrum ranges from gamma rays, measured in fractions of nanometres, to larger wavelengths of the radio region measured in metres (Figure 1.4). It should be noted that the visible (to human) wavelengths (.4 μm - .7 μm) occupy only a small portion of the electromagnetic spectrum. Matter - such as moisture, dust, smoke and gases - within the atmosphere can change the character energy level of electromagnetic radiation. Scattering, absorption and refraction may occur. When selecting remote sensing systems, it is essential that sensors be chosen in regions of the electromagnetic spectrum where atmospheric transmission bands occur, in order that only minimal attenuation of electromagnetic radiation occurs.

1.4 Atmospheric Effects

The medium through which the energy propagates (usually the atmosphere) also affects the energy. Scatter and absorption of electromagnetic energy by particles such as gas molecules, aerosols and water droplets occur. The effect of these interactions depends on the application, type of sensor and the amount of atmosphere through which the energy must pass.

If the molecules and small particles in the atmosphere have diameters much less than the wavelength of the radiation, Rayleigh scatter will be present. This effect explains a blue sky - ultraviolet radiation (blue) is scattered much more than the longer visible wavelengths. A red sun in early morning or late evening results from the increased atmospheric path length of solar radiation. When aerosol particles in the atmosphere are very much larger than the radiation wavelengths, such as water droplets, the scatter is non-selective. A white cloud is the result of this kind of scatter, which scatters all wavelengths of visible light (ie blue, green, red) in equal amounts.

The extent to which the atmosphere transmits electromagnetic energy is dependent on the wavelength of the source energy (see Figure 1.5).
Figure 1.5  Atmospheric Transmission - The vertical scale gives the transmission through the atmosphere between 0% - 100%, while the wavelength is shown on the horizontal scale. Wavelengths shown as stippled are essentially not transmitted, while atmospheric windows are said to exist in areas of high transmission.

The maximum amount of reflected energy (from the sun) peaks around the 0.5 μm (green) band of visible light. The earth also radiates energy both day and night, with the maximum energy radiating at 9.7 μm wavelength in the thermal band of the IR region.
In summary:

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma Rays &lt; 0.03 μm</td>
<td>Absorbed</td>
</tr>
<tr>
<td>X-Rays .03-3 μm</td>
<td>Absorbed</td>
</tr>
<tr>
<td>Ultra Violet .03-4 μm</td>
<td>Partially absorbed</td>
</tr>
<tr>
<td>Visible .4-.7 μm</td>
<td>Transmitted - imaged with films</td>
</tr>
<tr>
<td>Infrared .7-100 μm</td>
<td>Windows separated by absorption bands</td>
</tr>
<tr>
<td>Reflected Infrared .7-3 μm</td>
<td>Contains no information on the thermal</td>
</tr>
<tr>
<td></td>
<td>properties of a target. Detectable with film</td>
</tr>
<tr>
<td></td>
<td>(photographic IR)</td>
</tr>
<tr>
<td>Thermal Infrared 3-5 μm</td>
<td>Images acquired by scanners but not by film</td>
</tr>
<tr>
<td>8-14 μm</td>
<td>Can penetrate cloud, fog and rain</td>
</tr>
<tr>
<td>Microwave .1-30 cm</td>
<td>Active form of microwave remote sensing</td>
</tr>
<tr>
<td>Radar .1-30 cm</td>
<td>No use</td>
</tr>
<tr>
<td>Radio &gt; 30 cm</td>
<td></td>
</tr>
</tbody>
</table>

1.5 Interaction Mechanisms

Incident electromagnetic radiation may interact with matter in a variety of ways.

1. It may be absorbed, and its energy converted to heat - as with a microwave oven. The microwaves cause the water molecules in the food to vibrate, producing friction, which causes the food to heat.

2. It may be transmitted or pass through the substance, which causes a change in the velocity of the incident energy. Visible light, for example, is refracted on entering glass from air. A change of wavelength will occur with a change in velocity. The effect of white light passing through a prism is an example of this.

3. It may be reflected or returned from the surface of the object, with no change to the wavelength of the incident energy, like the image from a plane mirror. Reflection is primarily determined by the colour and roughness of the surface with which the incident energy is interacting.

4. It may be scattered or deflected in all directions, again depending on the roughness of the surface.

5. It may be emitted - that is, the energy is "given off" by an object, usually at longer wavelengths, and as a function of the temperature of the object and its physical/chemical nature.

Any particular combination or mix of the five interactions can occur. These varying combinations depend on the wavelength of the incident radiation and the specific properties of the matter to which it is directed. The interactions between the incident energy and matter are recorded in a variety of ways, and are the basis for interpretation of the characteristics of the (usually on the ground) matter.
1.6 Reflectance Spectra and Spectral Signatures

The majority of remote sensing systems measure only reflected energy. That is (at any given wavelength, \( \lambda \))

\[
R_\lambda = I_\lambda - (A_\lambda + T_\lambda)
\]

*ie reflected energy = incident energy minus the absorbed and transmitted energy.*

Two important properties of the target/energy interactions are:

(i) the proportion of energy reflected, absorbed and transmitted varies with different earth surface features; and

(ii) the relative amounts of radiation reflected, absorbed and transmitted are wavelength dependent.

The *spectral reflectance* of an object at any particular wavelength is given by the ratio of the electromagnetic radiation reflected from an object to the amount of electromagnetic radiation incident upon the object.

\[
R = \frac{\text{Energy of wavelength (}\lambda\text{) reflected}}{\text{Energy of wavelength (}\lambda\text{) incident}} \times 100
\]

R is termed the reflectance, and is expressed as a percentage.

The graph of the spectral reflectance of an object over a given wavelength is termed a *spectral reflectance curve.*

These curves provide essential information on the reflectance characteristics of objects.

Figure 1.6 shows typical reflectance curves for soil and vegetation. For any soil, high reflectance is characteristic in the near-infrared region. Water shows little reflectance in the near-infrared region.

![Figure 1.6 Typical Spectral Reflectance Curves](image)
In complex environments - for example vegetation - there are variations in spectral signature due to:

(a) Illumination conditions;
(b) Site environment conditions;
(c) Plant characteristics;
(d) Atmospheric conditions; and
(e) Multichannel sensor parameters.

It is important to realise that for specific ground cover categories, signatures in complex environments are subject to many sources of variations (see, e.g., Figure 1.7).

![Reflectance Spectra of Wet and Dry Sandy Soils](image)

Figure 1.7 Reflectance Spectra of Wet and Dry Sandy Soils

Spectral curves are useful for both design of a sensing system and interpretation of remotely sensed data.

In Figure 1.8, each tree type has a range of spectral reflectance values at any wavelength.

As an example, consider why trees are green. The incident energy is the sun's visible light within the range \( \sim 0.4-0.7 \ \mu m \). The wavelengths 0.4-0.5 \( \mu m \) (blue) and 0.6-0.7 \( \mu m \) (red) are largely absorbed. What is reflected (and that which we see) is in the region 0.5-0.6 \( \mu m \) (green).

Now consider why various species of trees have different 'shades' of green. Each tree species has a unique physical composition and thus reacts differently to the incident radiation, absorbing and transmitting differently, and thus reflecting differently. The spectral sensitivity of objects to different wavelengths of electromagnetic energy results in a unique "fingerprint" of reflected radiation which is termed the spectral signature of the object.

As another example, natural and artificial grass (or turf) may look the same to the human eye. They are, however, quite different in their physical/chemical composition. As shown in Figure 1.9, their particular interaction mix is quite similar between 0.4-0.7 \( \mu m \) (visible light), but varies markedly in the infrared wavelengths.
Figure 1.8  Spectral Reflectance Curves of Some Trees

Figure 1.9  Spectral Signatures of Natural and Artificial Grass - Interpretation of interactions outside the range of wavelengths detected by the normal human eye clearly distinguishes natural and artificial grass.
Spectral reflectance curves cannot be considered as an absolute and unique representation of spectral reflectance (i.e., a spectral signature) because of the inherent variability in samples and the changes in reflectance over time. Spectral reflectance curves should be used only as a guide.

Spectral reflectance may be measured by spectrophotometers in a laboratory or by using spectroradiometers and radiometers in the field. Laboratory measurements are advantageous in terms of controlled illumination, a stable environment, minimal logistic support and the ability to use instruments which may be bulky and complex. Disadvantages include a non-natural environment and artificial illumination. Field measurements ensure that measurements are made with natural illumination. Background and spectral emissivity and surface temperature can only be measured this way.

1.7 Spectral, Spatial, Temporal and Radiometric Resolution

In all applications of remotely sensed data, spectral, spatial and temporal (changes caused by time) characteristics influence the result. A combination of data from ground, aerial and satellite platforms is useful in assessing these characteristics.

Remote sensing is normally directed towards the surface of the earth and as such is designed to facilitate the managing of the earth's resources, including minerals, timber, agricultural crops, soils, water, rangelands, fish and wildlife. Whether viewed on a local, regional, national or global basis, the human demand for earth resources (both renewable and non-renewable) is increasing and the quality of the resources is deteriorating. Under such conditions, the careful management of these resources is vital. Wise management is greatly facilitated if timely, accurate inventories can periodically be made available to the resource manager.

The combination of spatial, spectral, temporal and radiometric dimensions determines the quality, quantity and type of remotely sensed data collected. Briefly, the basic elements of each are:

(i) Spatial resolution - the technical resolving power of a remote sensing system (i.e., the minimum distance between two objects at which the sensor can record them as two distinct entities);

(ii) Spectral resolution - describes the wavelength intervals and number of spectral bands where energy is recorded;

(iii) Radiometric resolution - sensitivity of a sensor to differences in reflected radiation, and the range of values it can record;

(iv) Temporal resolution - the repeat cycle time or frequency of coverage of an area.

Spatial, spectral, radiometric and temporal factors must be considered when designing/selecting a remote sensing system for a particular project. Emphasis on any one element may result in degradation of one or more of the other elements. For monitoring work the existence of archived historical imagery may be crucial.

For example, aerial photography yields superior spatial resolution compared with multispectral scanners, but multispectral scanners possess superior spectral resolution, due to their use of several sensing bands. Each band records data in a narrow range of wavelengths, providing excellent spectral resolution, but in most cases, relatively poor spatial resolution.
Selection of a remote sensing system to fulfil the requirements of a particular resource assessment or mapping task requires consideration of all four resolutions: spectral, spatial, temporal and radiometric. Depending upon the nature of the target (size, shape, aerial distribution, contrast, changing characteristics), each of these resolution components will receive a weighting according to their relative importance. In most cases, the selection of a remote sensing data source will be restricted to several readily accessible data sources, such as Landsat MSS or TM, SPOT, aerial photography and airborne MSS.

1.8 Reflectance Spectra and Digital Numbers

The reflectance spectrum of a pure material is a fundamental property which is invariant. If satellite sensors could record true reflectance values, then the physical interpretation of remotely sensed data would be much more rigorous and results could be applied across different scenes. Unfortunately, satellite sensors do not record reflectance (the proportion of reflected energy); they record, as 'digital numbers', the raw energy received at the sensor. True reflectance can only be measured at ground level by an instrument which measures simultaneously the incident and reflected energy, or the reflected energy of the target and some known standard (eg a ‘Spectralon’ or BaSO₄ plate).

The raw energy received by the satellite sensor will vary for several reasons, apart from the target material. These include

- Seasonal effects - with varying sun angles the amount of incident energy varies greatly, especially at higher latitudes.

- Atmospheric backscatter and absorption, which is wavelength dependent and variable with atmospheric conditions - pixel values are affected by scattered energy from the atmosphere and neighbouring ground areas.

- Sensor variation and 'drift' - while the sensors are calibrated at launch so that the relationship of 'at satellite' radiance and DN is known, over time the sensor quality decays.

- Angular affects for wide angle scanners such as AVHRR and airborne systems - these are a combination of atmospheric and ground reflectance properties.

The result is that the 'spectral DN curve' may bear very little relationship to the theoretical reflectance curve. However, within one date, the relative responses for different materials can be related to reflectance differences on the ground. For example, if the spectral response curve for vegetation is higher in the near-infrared region than the curve for soil, then the DN for the near-infrared band for vegetation will be higher than the DN for soil.

As an example, sample reflectance levels for clean deep water and for vegetation (healthy grass) for the TM bands are shown below. The digital counts for samples from two dates are shown for comparison.
<table>
<thead>
<tr>
<th>TM band</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflectance %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grass</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>49</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>water</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DN (0-255)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grass 12/87</td>
<td>80</td>
<td>37</td>
<td>34</td>
<td>125</td>
<td>97</td>
<td>34</td>
</tr>
<tr>
<td>grass 2/90</td>
<td>84</td>
<td>38</td>
<td>38</td>
<td>116</td>
<td>110</td>
<td>35</td>
</tr>
<tr>
<td>water 12/87</td>
<td>70</td>
<td>19</td>
<td>11</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>water 2/90</td>
<td>63</td>
<td>16</td>
<td>11</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

A further problem arises because of the initial calibration of the sensors, the dynamic range and the quantisation of response. The TM sensors are calibrated to record over the range from darker pixels (zero reflectance) to the brightest response anticipated (e.g. ice for visible bands). As a result, the dynamic range of data in an area of interest may often be small.

These issues are particularly important when examining for spectral discrimination of classes and for objective comparison over time.
2. SATELLITE CHARACTERISTICS

2.1 Landsat System

Five unmanned Landsat satellites have been launched since the first in 1972. Landsat 5 is the only Landsat satellite which is currently fully operational. Landsat 6 is proposed to be launched in 1993. Landsats 4 and 5 carry both the four-band multispectral scanning (MSS) system and the seven-band Thematic Mapper (TM) system.

The MSS system records 128 radiance levels in bands 4, 5 and 6, and 64 radiance levels in band 7, with a ground resolution cell of approximately 79 m x 79 m. The TM system records 256 radiance levels, with a ground resolution cell of approximately 30 m x 30 m, other than in band 6, which has a ground resolution of approximately 120 m x 120 m.

In summary, the characteristics of the Landsat systems are as follows:

MSS

<table>
<thead>
<tr>
<th>Wavelength ((\mu m))</th>
<th>Band</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 - 0.6 (\mu m)</td>
<td>Band 4</td>
<td>Green</td>
</tr>
<tr>
<td>0.6 - 0.7 (\mu m)</td>
<td>Band 5</td>
<td>Red</td>
</tr>
<tr>
<td>0.7 - 0.8 (\mu m)</td>
<td>Band 6</td>
<td>Near Infrared</td>
</tr>
<tr>
<td>0.8 - 1.1 (\mu m)</td>
<td>Band 7</td>
<td>Near Infrared</td>
</tr>
</tbody>
</table>

TM

<table>
<thead>
<tr>
<th>Wavelength ((\mu m))</th>
<th>Band</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.45 - 0.52 (\mu m)</td>
<td>Band 1</td>
<td>Blue/Green - good water penetration - strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td>vegetation absorbance</td>
</tr>
<tr>
<td>0.52 - 0.60 (\mu m)</td>
<td>Band 2</td>
<td>Green - strong vegetation reflectance</td>
</tr>
<tr>
<td>0.63 - 0.69 (\mu m)</td>
<td>Band 3</td>
<td>Red - strong vegetation absorbance</td>
</tr>
<tr>
<td>0.76 - 0.90 (\mu m)</td>
<td>Band 4</td>
<td>Near Infrared - high land/water contrasts - very</td>
</tr>
<tr>
<td></td>
<td></td>
<td>strong vegetation reflectance</td>
</tr>
<tr>
<td>1.55 - 1.75 (\mu m)</td>
<td>Band 5</td>
<td>Near Middle Infrared - very moisture sensitive</td>
</tr>
<tr>
<td>10.4 - 12.5 (\mu m)</td>
<td>Band 6</td>
<td>Thermal Infrared - very sensitive to soil moisture</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and vegetation</td>
</tr>
<tr>
<td>2.08 - 2.35 (\mu m)</td>
<td>Band 7</td>
<td>Middle Infrared - good geological discrimination</td>
</tr>
</tbody>
</table>

2.1.1 Landsat Orbit and Scanning Characteristics

The Landsat satellites are in a circular, near-polar, sun-synchronous orbit. Landsats 1-3 are at an altitude of 918 km, while Landsats 4 and 5 are at 705 km altitude. Earth coverage is available between 81° North and 81° South latitudes. All satellites are inclined at approximately 99° from the equator (Figure 2.1).
2.1.2 Detectors

The MSS total field of view is approximately $12^\circ$ ($6^\circ$ from the nadir), with six detectors per band, at 79 m x 79 m resolution (79 m x 57 m after processing). For a normal 185 km x 185 km scene, this results in approximately 3240 pixels x 2340 rows, recording between 0 and 127 (0 and 63 for band 7) video levels (Figure 2.2). The TM total field of view is approximately $15^\circ$ ($7.5^\circ$ from the nadir), with 16 detectors for the non-thermal bands and four for band 6. Resolution is 30 m x 30 m for the non-thermal bands, and 120 m x 120 m for Band 6, each recording 0 - 255 quantised video levels.
Figure 2.2  Landsat Pixel Characteristics

As the earth rotates east-west, the satellites are orbiting approximately 9° off north-south. One complete orbit takes 103 minutes for Landsats 1-3, and 99 minutes for Landsats 4-5. Each swath from the satellite covers only 185 km (E-W) approximately; successive orbits are 2760 km apart for Landsats 1-3. Earth coverage is not successive for each orbit. For Landsats 1-3, adjacent orbit paths (side by side) are sensed in the one day. For Landsats 4 and 5, the swath coverage is also 185 km (E-W) and successive orbits are 2752 km apart. Adjacent swaths for Landsats 4 and 5 are imaged 7 days apart. Complete earth coverage for Landsats 1-3 is 18 days, for Landsats 4 and 5, 16 days.

The sidelap between adjacent orbit paths is approximately 14% of the swath at the equator and 70% at the polar latitudes. The sidelap provides some, be it small, stereoscopic viewing at the equator and more so at the polar latitudes.
2.4

All Landsat sensors transmit their images to earth through a ground receiving station. Australia has such a facility at Alice Springs.

Paths and rows are used to specify the location of Landsat scenes.

2.2 SPOT Satellite

The Systeme Probatoire de Observation de la Terre (SPOT) has French, Belgium and Swedish participation. Spot 1 was deployed in February 1986, while Spot 2 was launched in January 1990. Spot 3 is under construction. SPOT has an altitude of 850 km and is in a near-polar sun-synchronous orbit with a repeat of 26 days. However, it can observe any area of the earth at least seven times during this period, and in mid-latitude regions up to eleven times.

The satellite does not use a conventional multispectral scanner, but rather a pushbroom sensor, which has a number of advantages over the conventional Landsat scanning devices. In panchromatic mode, a spatial resolution of about 10 m is achieved, and in multispectral mode, a resolution of about 20 m. The resolution has resulted in SPOT imagery being used successfully for mapping at scales of 1:100 000. SPOT imagery can be recorded at Alice Springs.

The pushbroom scanner, unlike the rotating mirror system of Landsat, has no moving parts and does not sweep side-to-side to sense reflected radiation. The sensing is accomplished using charge-coupled devices (CCD). A 6000-element subarray is used for panchromatic sensing, and three 3000-element subarrays are used for multispectral sensing.

One of the advantages of SPOT for mapping is the ability to program the scanner to point at a target with an off-nadir angle of 27° - that is, within a region of 950 ± 50 km width on the earth, centred on each satellite track. The ground coverage of each swath is 60 km at 27° off-nadir. Any given point on the equator can be observed on seven different passes, and on eleven different passes at latitude 45°. This means that scenes can be selected with suitable base height ratios (> 0.5) which will permit photogrammetric mapping.

The SPOT sensor systems are as follows:

- **Panchromatic**: 0.51 - 0.73 μm (10 m resolution)
- **MSS**: Band 1: 0.50 - 0.68 μm (20 m resolution)
  - Band 2: 0.68 - 0.79 μm
  - Band 3: 0.79 - 0.89 μm

In comparison with the Landsat output, the following observations can be made:

- SPOT scanner has no moving parts, and hence greater reliability and longer life expectancy
- One array is used for each spectral channel
- There are no inherent complex geometric distortions, as in optical mechanical scanner systems
- The pushbroom system requires 6000 detectors per channel to calibrate
2.5

- Higher geometric accuracy
- Higher radiometric accuracy
- Higher spatial resolution
- Lower spectral resolution

2.2.1 Nadir Viewing

The two HRV instruments can be pointed to cover adjacent fields (Figure 2.3). In this configuration, the total swath width is 117 km (nadir) and the two fields overlap by 3 km. Since the distance between adjacent ground tracks at the equator is approximately 108 km, complete earth coverage can be obtained with this fixed setting of instrument fields.

Figure 2.3 Coverage of SPOT
2.2.2 Off-Nadir Viewing

By appropriately selecting the orientation of the pointing mirror, which is controlled from the ground, it is possible to observe any region of interest within a 950 km-wide strip on the satellite ground track, i.e., the observed region need not necessarily be centred on the ground track.

The width of the swath actually observed varies between 60 km for nadir viewing and 80 km for extreme off-nadir viewing.

The program of observations to be made is controlled by the satellite's onboard computer. A sequence of recorded images may include both modes of instrument operation (multispectral and panchromatic mode) and changes in each instrument's viewing directions.

![Diagram of satellite views and ground track](image)

Figure 2.4 SPOT Characteristics

2.2.3 Revisit Capabilities

Were the satellite instruments only capable of nadir viewing, the revisit frequency for any given region would be 26 days. This interval is unacceptable for the observation of phenomena evolving on timescales ranging from several days to a few weeks, especially where cloud cover hinders the acquisition of useable data.
However, with SPOT, the capability for off-nadir viewing during satellite passes in the vicinity of a region of interest considerably increases the revisit possibilities.

The following revisit frequencies can be achieved. During the 26-day period separating two successive satellite passes over a given point on the Earth’s surface, and taking into account the steering capability of the instruments, a point in question could be observed on seven different passes if it were on the equator and on eleven if at a latitude of 45°C. A given region can be revisited on dates separated alternately by one, four or (occasionally) five days.

2.2.4 Stereoscopic Viewing Capabilities

Another important possibility offered by the off-nadir viewing capability of the HRV instruments is that of recording stereoscopic pairs of images of a given scene (i.e., images recorded at different viewing angles during successive satellite passes in the vicinity of the scene concerned).

Two observations can be made on successive days such that the two images correspond to pointing angles on either side of the vertical. In such cases, the ratio between the observation base (or distance between the two satellite positions) and the height (or satellite altitude) is approximately 0.75 at the equator and 0.50 at a latitude of 45°.

The main applications for stereoscopic imagery are in photogrammetry and photointerpretation (morphogeology, hydrographic studies, etc).

2.3 AVHRR

NOAA operates two polar orbiting environmental satellites with advanced very high resolution radiometer (AVHRR) cross-track scanners. Each satellite orbits the earth 14 times daily and acquires complete earth coverage every 24 hours. The swath width is 2700 km in width, with a ground resolution cell of 1.1 km by 1.1 km. Spectral bands on the AVHRR are shown in Figure 2.5.

<table>
<thead>
<tr>
<th>Band</th>
<th>Wavelength, µm</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.55 to 0.68</td>
<td>Red - for daytime clouds and vegetation</td>
</tr>
<tr>
<td>2</td>
<td>0.73 to 1.10</td>
<td>Reflected IR - for shorelines and vegetation</td>
</tr>
<tr>
<td>3</td>
<td>3.55 to 3.93</td>
<td>Thermal IR - for hot targets such as fires and volcanoes</td>
</tr>
<tr>
<td>4</td>
<td>10.50 to 11.50</td>
<td>Thermal IR - for sea temperature and for daytime and nighttime clouds</td>
</tr>
<tr>
<td>5</td>
<td>11.50 to 12.50</td>
<td>Thermal IR - recorded only on the NOAA 7 satellite</td>
</tr>
</tbody>
</table>

Figure 2.5 AVHRR Characteristics

AVHRR images are suitable for environmental monitoring on a macro scale (i.e., continental scale), as only a few images are needed to cover a continent such as Australia. The daily repeat coverage of AVHRR allows scientists to closely monitor a variety of earth surface features. However, the resolution of such imaging may be a limitation in many mapping projects.
3. IMAGE FILES, DISPLAY SYSTEMS AND COLOUR SPACE

To make effective use of hard-copy output and of image processing systems, it is useful to have a basic idea of the way the system handles data, and colour displays.

3.1 An Image File

An image file on the computer is a stream of coded numerical values. This may be visualised as layers of numerical grids, each layer representing a band. The important aspects of the image file are:

- the file size, which is a product of
  - number of lines (rows)
  - number of elements (pixels)
  - number of bands
- the data type and storage 'packing' (1-bit, 1-byte, 2-byte, 4-byte)

The data storage unit or 'packing' is an important concept and is a function of the dynamic range of the data. Landsat and Spot provide 8-bit (1-byte) data which can store a range of values from 0-255. One-bit data can only be 0 or 1, *i.e.* 'off' or 'on'. If a sensor has a higher dynamic range (*e.g.* NOAA), data may be stored as 2-byte integers. Other data sources may require storage of real numbers which require 4-bytes per data item.

The image processing system keeps track of the file structure by means of header information stored with the file or linked to it. The normal structure for image files is

```
Header information (fixed size)

data records

Trailer information
```

The actual way that these records are arranged is called the image file format, and varies between systems. For the user, this can become a problem when attempting to transfer data between systems.

When manipulating image data in disk-to-disk operations, the user needs to be aware of the limits of values which can be stored. For example, negative values or values greater than 255 cannot be stored in byte data files.

3.2 Image Display

When image data are displayed, the digital values in the file are retrieved and mapped to colour values on the display monitor. Systems such as ER Mapper and DISIMP support 24-bit colour display (8 bits for each of Red, Green and Blue). They also often support separate graphics overlay planes which can be turned off or on via the display controller. The display system also has a 'screen coordinate' system, with limits in size.
3.2

In RGB display, the calculated display values for each pixel are stored in a ‘frame buffer’. Separate values are mapped for each colour gun to the brightness values (typically in the range 0-255) for each screen pixel. The colour displayed is the combination of the levels of the three primary colours. The colour values are actually stored in three colour look-up tables (LUTS), which speeds up the display process. When the colour enhancement is changed, it is the LUT values which are altered - the values calculated from the file are not altered.

For single channel grey level or ‘pseudo-colour’ display, a colour look-up table (LUT) of 256 values is set up, and the image data are displayed in the appropriate colour by reference to this table. Graphics planes contain only values 0/1, but the ‘on’ values can be assigned any colour by altering a look-up table entry.

When using any image display operation, it is important to understand how the numerical image values are mapped to colours and locations on the screen.

3.3 Colour Space and Colour Mixing

Computer screens are additive display systems; colours are created by adding intensities of the Blue, Green and Red primary colours, which are controlled by ‘colour guns’. (The complementary colours of yellow, magnetic and cyan form the so-called subtractive primaries which are used in photographic processing.)

The human visual system can perceive many thousands of different shades of colours. However, it does not detect differences in brightness so well; perhaps only 50-100 levels of grey tones can be differentiated. Thus the computer RGB display may not produce ideal results for visual interpretation.

Colour space can be represented by a colour cube, with each point representing intensities on the RGB axes (Figure 3.1). White is formed when the three colours are at maximum intensity, while black results when all three colours are ‘off’. How do cyan, magenta and yellow arise?

Figure 3.1 Schematic Diagram of the Colour Cube
3.3

An alternative representation of colours is defined by Intensity, Hue and Saturation (IHS). Transformation to IHS colour space may improve representation of certain image features, but is not dealt with in this course.

Single band images are often displayed as grey-tone images. Alternatively a colour palette is applied by 'density slicing' the data and using a different colour for values in the range from low to high. There is no consensus on an ordered colour scheme for this representation.

It is important to recognise that the colours on a hard-copy image or computer screen arise by assigning bands or indices to the three colour guns, and then adjusting the intensity of the colour for each colour gun. Each colour represents an (R, G, B) combination, and this relates to a particular combination of digital counts, which in turn relates to an underlying spectral response curve.

3.4 Band Combinations

Landsat MSS data are often presented as a false colour composite, the band 7 response being assigned to the red gun, band 5 to the green, and band 4 to the blue gun.

With Landsat Thematic Mapper data having seven wavelength bands, a number of combinations can be made. The more common include:

1. Natural colour composite, where bands 1, 2 and 3 are coloured blue, green and red;
2. False colour composite, where bands 2, 3 and 4 are coloured blue, green and red; and
3. Infrared composite, where bands 7, 5 and 4 are coloured blue, green and red.

3.5 Other Data Inputs

Any digital data in a grid or raster format may be treated as image data by the display system. Scan digitisers can capture photographs, text or any image in colour or panchromatic mode. The digital capture of maps and airphotographs is particularly useful in some applications.

Vector or raster data can be output from a GIS system and incorporated with image data provided that the spatial base is consistent. Image processing systems vary in the ease with which they can handle vector data, but most systems can readily use rasterised polygon data from a GIS.
4. BASIC RADIOMETRIC ENHANCEMENTS

4.1 Image Histograms

Each image band is an ordered stream of numbers. To optimise the display of the data, it is necessary to understand the range and distribution of the values. A histogram summarises and represents the data distribution. A frequency histogram shows the number of pixels at each numerical value in the range. The minimum and maximum values are illustrated, as well as the 'shape' or distribution of the data. The percentiles of the distribution are also useful statistics for enhancement purposes. Histograms may also be represented numerically as counts or percentages of data points at each value.

Figure 4.1  Histogram of a single channel of Image Data

4.2 Image Enhancement Techniques

Modern image processing systems offer a wide range of enhancement techniques. Image enhancement methods are generally applied to each band of a multispectral image. Image enhancement techniques are closely related to the data and target characteristics, but are largely independent of sensor characteristics.

Image enhancements are applied to map the digital values (in the histogram) to the available colour range to produce a useful image.

To produce an image with the optimum contrast ratio, it is important to use the entire brightness range of the display medium.

When the brightness range of a scene does not occupy the full range of digital values (eg 0 - 255), techniques called contrast stretching can be used. The values present in the scene are altered according to a mathematical model (and processed at display time), so that they occupy the full brightness range of the display medium.
In contrast enhancement operations, certain trade-offs result from the stretch employed. One type of stretch will enhance certain parts of the data, and suppress others. The stretch to be employed must be considered in the light of the results required.

Forms of stretch include:

(i) Linear Contrast Stretch

The mapping of input digital values to output brightness values is a linear function as represented in the diagrams below. The user sets lower and upper limit values; digital numbers less than the lower limit are displayed with zero brightness (black), while pixel values above the upper limit are displayed at maximum brightness. These limits are often set by examining the image histogram. The slope of the linear stretch function represents the contrast (i.e. the change in brightness for a change in DN value).

![Linear Stretch Mapping](image)

Figure 4.2 Linear Stretch Mapping

(ii) Logarithmic, exponential, and power stretches

Again, lower and upper limits are set by the user, and one of these functions is used to map the DN to output values. As these functions are non-linear, the contrast varies across the range, and suitable values can be set to increase colour contrast of low or high values.

(iii) Histogram Equalisation

The original pixel values are redistributed to produce a uniform density of pixels along the digital value axis. This redistribution results in the greatest contrast enhancement being applied to the most populated range of brightness values in the original image. The functional shape is actually the same as that of the cumulative histogram.

(iv) Gaussian Stretch

A Gaussian stretch is a non-linear stretch that fits the original histogram to a Gaussian distribution curve between specified lower and upper digital values. The result of this stretch is to enhance the contrast within the tails of the histogram.
4.3 Linear Stretch - an example

Consider a hypothetical sensing system with image output levels which vary from 0 - 255. Figure 4.3 illustrates a histogram of brightness levels recorded in one spectral band over a scene.

![Histogram and Default Stretch](image)

Figure 4.3 Histogram and Default Stretch

Assume that the output device can output levels between 0 and 255. The histogram scene values occur only in the range 60 to 158. If the image values are used directly, the display device would be using only a small portion of the full range of output device display colour levels. Display tones 0 - 59 and 159 - 255 would not be used. The visual information in the scene is compressed into a small range of display values, reducing the ability to differentiate radiometric detail. A more useful display would result if the range of image levels could be expanded to fill the whole output range available, *ie* if the range 60 to 158 could be expanded linearly to fill the total available output range from 0 to 255 (Figure 3.2). This expansion is referred to as a linear stretch.

![Display Values after Linear Stretch](image)

Figure 4.4 Display Values after Linear Stretch

Here, the lower level of the brightness values is mapped to the minimum value of the output display; the upper level of the brightness values is mapped to the maximum value of the output display; and the intermediate brightness values are mapped linearly to the corresponding output display values.

While display tones will now be expanded, this linear stretch still assigns as many display levels to each of the rarely occurring image values as to the frequently occurring image values. The bulk of the image data (109 - 158) is confined to half the output display levels (128 - 255). Although
better than the direct display, the naive linear stretch would still not provide the most effective display of the data.

An obvious improvement is to map only the main body of values - say from 105 to 160 - to the full range of output display levels.

4.4 Scatter Plots and 2-D Histograms

The two-dimensional scatter plot can be a useful aid for visualising the joint data distribution of two image channels. It is particularly useful for examining band relationships of samples of the data.

A two-dimensional histogram represents the density of points in each cell of the scatter plot domain. The counts can be represented as heights or by contour lines. Such a plot may also be represented using colour enhancement to represent heights. A non-linear stretch may be necessary to highlight interesting features of the data distribution.

Note that the marginal distribution of the 2-D histogram (that is the counts collapsed onto the two axes) are the usual single band histograms already discussed.

![2-D Histogram](image)

Figure 4.5 Density Representation of a 2-D Histogram
5.1 DATA PROCESSING AND DATA CALIBRATION

5.1 Digital Numbers and Reflectance

The image data value measured for a pixel is not the actual reflectance of that pixel; it is a digital response to the at-satellite radiance energy.

\[ L = rIT + L_p \]

- **L** is total at-satellite radiance energy
- **r** is the actual reflectance of the ground element
- **I** is the incident energy (irradiance)
- **T** is the atmospheric transmission for the path
- **L_p** is the atmospheric path radiance - the component due to scatter and reflection from the atmosphere.

In general, all of these coefficients are not known. 'Reflectance' can only be recovered by assuming certain values for the physical parameters. An additional complication arises because of the interference from surrounding pixels due to atmospheric scattering. This effect is relatively larger for small pixels and for shorter wavelengths. A linear relationship between input energy and output DN is assumed to hold.

The actual quantisation of DN from at-satellite radiance is known (though not perfectly) for each detector from calibration measurements taken on board the satellite. As multiple detectors are used, differing detector responses may introduce noise in the image, often seen as horizontal line striping, or in the case of SPOT, as vertical column striping.

![Diagram](image)

Figure 5.1 Schematic Detector Response Functions (from Richards, 1986)

5.2 Noise Removal

Noise removal means the altering of pixel numbers to remove spurious effects such as detector line striping and data drop-outs. Destriping algorithms usually calculate image statistics for each detector, and adjust these by either histogram matching or an adjustment to produce equal means.
and variances. In practice, their application over a whole image may produce imperfect results. Destriping procedures are best applied to relatively uniform areas using statistics from such areas.

Note that the quantisation of data to whole numbers may introduce apparent speckle or striped which is impossible to remove, especially for Landsat MSS data which have been rescaled to the 8-bit range.

For problems such as data dropout, and speckle in images, data values can be replaced by averages from surrounding pixels. In practice, this is a slow process. Automatic smoothing can remove speckle or 'high frequency' noise, but it also degrades the image information.

Fortunately, the quality of imagery now available is very high and noise removal may not be necessary. Nevertheless, it is always good practice to inspect imagery band-by-band for possible noise.

5.3 Atmospheric Correction and Calibration

To relate the digital values to physical properties, it is highly desirable to be able to calibrate the digital numbers to reflectance values. In practice, there is no general method to do this and several practical and theoretical approaches have been attempted.

If \( I, T \) and \( L_p \) are known in the relationship in Section 5.1, then reflectance can be recovered. The atmospheric path effect (\( L_p \)) may be estimated using the values for a "dark pixel", if such a pixel can be found. The irradiance can be estimated from knowledge of the sun angle at the time of the overpass. The atmospheric transmission can only be estimated using physical models of the atmosphere. Thus a perfect solution for reflectance remains an elusive problem. In addition, it is necessary to know how the detector translates the radiance (\( L \)) to digital numbers. A linear relationship may be assumed

\[
DN = b + g L
\]

where \( b \) (the intercept) is called the offset
and \( g \) (the slope) is known as the gain of the detector.

5.3 Dark Pixel Subtraction

As a first approximation, the DN values for a true dark pixel can be used to estimate the path radiance effect for each wavelength. Subtraction of these values from all pixels in the scene will theoretically remove path radiance effects so that zero DN corresponds to zero reflected energy.

A dark pixel may be either

(1) a known material which has zero reflectance for the wavelength (deep clean water for NIR and SWIR); or

(2) a pixel which receives zero illumination, *i.e.* shadowed areas in rugged terrain or (perhaps) shadow beneath dense clouds - such a pixel should be dark for all bands.

An assumption of the method is that the atmospheric path radiance is constant over the scene. This is not a reasonable assumption for wide angle scanners such as AVHRR, and atmospheric modelling is required.
5.4 Correction to Like-Values Using Scene Targets

If a scene contains large, truly invariant, targets of both bright and dark materials, then scene-to-scene calibration to like values is possible using the linear function defined by these targets. If the reflectance of the targets is known, then calibration to reflectance is possible.

In practice, such invariant targets are rare. Moreover, there are considerable difficulties in measuring ground reflectance on the scale of satellite pixels.

Clean, deep ocean water may be considered an invariant target, but suitable bright targets are harder to find, and of course the ocean may not always be present in a scene.

If such targets do exist, they provide the basis of a reasonable practical approach to scene-to-scene normalisation.

Location of "invariant" bright and dark targets in a scene pair is recommended. It is not necessary to know atmospheric or detector response functions to adopt this approach. It should be noted that the MIR bands (TM 5 & 7) present particular problems in the search for invariant targets.

The atmospheric backscatter is greatest at short wavelengths (eg TM band 1).
6. **RADIOMETRIC INDICES**

So far we have concentrated on the display of single image channels. The data values of different bands may be combined in different ways to produce new "indices" for each pixel. Spectral indices are often derived as a result of analyses, or basic physical principles related to the problem.

Since the band DN's are just numbers, they can be combined using arithmetic operations. The results will be simply another set of numbers which may be stored on disk as an image file, or manipulated in the same way as the original data.

Note here that raw calculated index values may not be suitable for storage in, say, a single-byte range (0-255). For example, the index value may be negative, or too large or a real number. The index may need to be 'rescaled', or stored as a different data type.

Commonly used indices include

- ratios of band values $b_i/b_j$
- linear combinations of bands $\Sigma c_i b_i$

  (includes simple sum and difference indices)

- some combination of the above.

Numerous combinations may be created from, say, a 6-band Thematic Mapper image. It is instructive to consider the spectral features which are enhanced by some simple combinations.

A difference or contrast index, *eg* TM4-TM3, measures the spectral shape or slope between the two bands. A ratio index such as TM4/TM3 measures the relative values in these two bands in a slightly different way.

A linear combination with positive coefficients is a 'brightness' index of some kind. A simple sum of bands corresponds to the area under the spectral curve.

It is important to consider whether data correction should be applied before using spectral indices. The theoretical use of ratios (below) assumes that dark pixels have a DN of zero.

### 6.1 Ratio Images

Ratio images are prepared by dividing the digital value for a pixel in one band, by the corresponding digital value in another band.

On a ratio image, the extreme (black or white) tones of the (grey) scale represent the maximum difference in spectral reflectivity between the two bands. The darkest tones are targets for which the denominator of the ratio is greater than the numerator. Conversely, the numerator is greater than the denominator for the lightest tones. For example, with Landsat MSS data, the highest reflectance for vegetation occurs in the near-infrared bands 6 and 7. Thus vegetation has a dark signature for ratios with MSS band 6 and 7 in the denominator. Vegetation has a bright signature on a 4/5 image because reflectance is higher for band 4 (green) than for band 5 (red).
One advantage of ratio images is that a material tends to have the same ratio value, regardless of variations in illumination. This removal of illumination differences also eliminates the expression of topography on ratio images, provided the effects are proportional.

Ratio images show the variations in slopes of the spectral reflectance curves between the two wavelength bands. These variations are useful for distinguishing among rock types because the main spectral differences in the visible and near-infrared regions occur in the slopes of the reflectance curves. A disadvantage of ratio images is that differences in albedo are suppressed. Thus dissimilar materials with different albedos, but similar slopes, may be inseparable on ratio images.

The simplest case of ratio images is to divide an individual band by another band. Other ratio combinations are possible including:

- an individual band divided by the average of all the other bands, resulting in a normalised ratio image; and
- dividing the difference between two bands by their sum.

6.2 Spectral Indices and Vegetation

Spectral Response of Individual Leaves

Typical ranges of reflectance, absorptance and transmittance spectra for a healthy, green leaf over shortwave wavelengths are shown in Figure 6.1. Beyond the ends of the range in the ultraviolet at the shorter wavelengths, and in the mid-infrared at the longer wavelengths (i.e. longer than 2700 nm), leaf reflectance is at a low and relatively uniform level, generally less than 5%.

![Figure 6.1 Reflectance Regions](image-url)
Many studies have been oriented toward the extraction of a meaningful, simplified data set for characterising vegetation cover. The simplest approach uses only the visible red (R) and the near-infrared (NIR) channels. Vegetation indices, derived from such data, may be expressed in several forms, including:

a) **Subtractive Vegetation Index (SVI),** where
   \[ \text{SVI} = \text{NIR} - \text{R} \]
   e.g. TM4 - TM3

b) **Ratio Vegetation Index (RVI),** where
   \[ \text{RVI} = \frac{\text{NIR}}{\text{R}} \]
   e.g. TM4/TM3

c) **Indices related to the RVI,** such as the Normalised Difference Vegetation Index, where
   \[ \text{NDVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R})} \]
   e.g. (TM4 - TM3)/(TM4 + TM3)

It is sometimes argued that these three vegetation indices are functionally equivalent, that is, that phytogeographical decisions made using any one of the indices could be made equally well using one of the other two. When looking at a scatterplot of remotely sensed near-infrared and red measurements, generally a triangular plot is observed (Figure 6.2). The apices of the triangle can be related respectively to dark bare soil (or dark dead vegetation), light bare soil, and healthy green vegetation. The values for core soils are chiefly aligned near the first bisectrix of the NIR-R plane - this is the "soil-line". RVI or NDVI may be displayed as lines with increasing slopes, or, alternatively, in terms of the distance of parallel lines to the soil line.

![Figure 6.2 Scatterplot of TM Band 4 (NIR) against TM Band 3 (R).](image)

Between about 400 and 700 nm (ie in the invisible wavelengths), leaf spectral response is complicated by the range of pigments which may be present. Usually, the most important pigments are chlorophylls, which are present in a number of spectral forms and which together comprise about 70% of all leaf pigments. However, in addition to the chlorophylls, there are also carotenoids (accessory pigments in the photosynthesis - absorbing 400 - 500 nm wavelengths), which comprise the bulk of the remaining pigments, chlorophyll and carotenoid precursors, and
various secondary plant products. Typically, there is high absorption by leaves in the 400 - 700 nm wavebands, with an absorption minimum at 500 nm, the green part of the spectrum.

Between 700 and 1300 nm (ie in the near-infrared wavelengths), there is only slight absorption of radiation by healthy, green plants, resulting in high reflection and transmission. Beyond 1300 nm (ie in the mid-infrared wavelengths), absorption increases, due to leaf water.

**Spectral Response from Plant Canopies**

Although green leaves have a particularly distinctive reflectance spectrum, it is not possible to extrapolate directly from hemispherical leaf reflectance measurements made in the laboratory to reflectance measurements of natural canopies made by satellite-borne sensors. A variety of factors besides the spectral responses of the vegetation target itself qualitatively and quantitatively alter the radiance from a canopy measured by a satellite-borne sensor.

These factors, which have been investigated both empirically and through the use of mathematical models, include the spectral response of the non-vegetation components of a scene, the spectral resolution of the sensor, the radiometric accuracy of the sensor, the instantaneous field of view of the sensor, the viewing and illumination geometry, and variations in atmospheric conditions.

**Use of Satellite Sensors**

The best spectral detail for phytogeographical applications is provided by the Thematic Mapper. Currently, this is the only satellite-borne sensor which measures reflectance in all three of the phytogeographically significant shortwave wavebands, *ie* in the visible (leaf pigment absorption), the near-infrared (leaf reflection) and the middle-infrared (leaf water absorption). Availability of data in all three of these wavebands is especially useful for purposes of assessing vegetation condition. If large-scale imagery is to be preferred for detailed mapping purposes, the largest scale data that are available currently are produced by the sensor aboard SPOT.
7. IMAGE ENHANCEMENT USING NEIGHBOURHOOD FILTERS

The preceding section discussed how pixel values could be modified by calculating multispectral indices to improve information content. In this section, neighbourhood operators which calculate a pixel value based on the values in a spatial neighbourhood around the pixel will be considered. The spatial relationships of the raw data values for the band (or index) affect the calculated pixel value.

7.1 Spatial Filters

The simplest form of spatial filter is a square (sometimes rectangular) grid of numerical weights which are used to calculate the new value for the central pixel. The square normally has an odd number of pixels on each side; the simplest case is a 3 x 3 filter window. The grid of numerical weights is called by many names, including filter, template, kernel, window and box-car filter.

![Image Grid and Template Window](image)

Figure 7.1 Image Grid and Template Window

The operation of the template calculates a new brightness value \( b_{ij} \) for the central pixel as a combination of the neighbouring values as

\[
b_{ij} = \sum \sum w(m,n) b(m,n),
\]

where the sum is over the filter window and \( w(m,n) \) are the template or filter weights. The template or filter moves across the original image and the calculation is done for each pixel. Edge pixels may either be set to a null value, or some 'boundary reflectance' method used to synthesise values for the calculations.

The weights in the filter may be positive or negative. It is necessary in some systems to rescale the new values to a suitable range. 3 x 3 filters are used to illustrate the techniques, but use of larger filters is no more complicated (it just takes longer to do the calculations!). The actual size of the filters should be chosen in relation to the features of interest and the pixel size.

7.2 Smoothing Filters

The simplest smoothing filter is
7.2

\[
\begin{bmatrix}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1 \\
\end{bmatrix}
\quad \text{or} \quad \frac{1}{9} \begin{bmatrix}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1 \\
\end{bmatrix}
\]

which calculates the average value of the $3 \times 3$ neighbourhood. Any template with non-negative weights can be seen as smoothing filter of some kind *eg*

\[
\begin{bmatrix}
0 & 1 & 0 \\
1 & 1 & 1 \\
0 & 1 & 0 \\
\end{bmatrix}
\quad \text{or} \quad \begin{bmatrix}
0 & 1 & 0 \\
1 & 2 & 1 \\
0 & 1 & 0 \\
\end{bmatrix}
\]

The effect of using different weights is best appreciated by simple examples of their use.

Smoothing filters remove or suppress high frequency noise or features in an image. This may have a pleasing effect in ‘homogeneous’ areas, but leads to a “blurring” of edge features and a degradation of information. The filter operation can be modified by a threshold to be applied only when the discrepancy of new and old values is small (assumed noise). If the difference is large, the original value is preserved. Adaptive filtering is a more complex approach to modify the operation so that edge information can be preserved in the smoothing operation.

Median smoothing filters operate slightly differently. The median is calculated instead of the mean of the window values, and replaces the central value.

**Low-Pass Filters**

Convolution filter techniques process each pixel in turn by passing a window over the image (such as a 3-by-3 box centred on each pixel being processed). The filter calculates a new value for all the pixels in an image, based on the current value of the pixel and that of its (usually) immediate eight neighbours. The low-pass filters to be described remove the high frequency data, by smoothing out isolated points with high local variance. Given the sample of data below, the smoothed value at the central points would be as shown:

<table>
<thead>
<tr>
<th>Sample Data</th>
<th>6</th>
<th>9</th>
<th>7</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

←3x3 box
The mean filter generally produces smooth homogeneous areas with sharp continuous false boundaries values, due to the averaging. This filter removes linear features and tends to change the size of the areas.

7.3 Edge Enhancement

Edge-enhancement filters are used to sharpen image products for visual interpretation, and also to detect field boundaries for agricultural and surveying purposes as well as for detecting lineaments in geology.

Edges are detected by creating an intensity value which is high at edge pixels and low in homogeneous areas. This method is used to accentuate areas of change in continuous surfaces. Natural features such as faults, joints, lineaments and vegetation discontinuities caused by drainage patterns and soil changes may also be highlighted. Filters used for edge enhancement are specifically designed to highlight linear features, whereas contrast enhancement techniques emphasise all variations within the scene. Edge enhancement may be preferable, because many contrast variations are very subtle and contrast enhancement techniques may not adequately highlight the feature.

Edge-enhancement techniques are accomplished using digital filters of two different types: directional and non-directional filters.

Non-directional filters use a (say 3 x 3) kernel, an example of which follows:

<table>
<thead>
<tr>
<th>0</th>
<th>-1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>4</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>
The central value is placed over the pixel of interest and all nine pixels are multiplied by their respective coefficients (0, -1, 4) and summed. The total is then added to the central pixel value giving a kernel of the form

```
   0   -1   0
  -1    9  -1
   0   -1   0
```

This value replaces the original value to produce the edge-enhanced image.

The centre of the filter is successively moved from pixel to pixel. The example below illustrates the effect of a filter, showing the data and plots of profiles of the sample and enhanced data. In the western portion of the lineage, the contrast has been increased from 40/35 (1.14) to 45/30 (1.5). In the eastern portion of the lineage, a north-south trending edge has been contrast altered from 45/40 (1.12) to 50/35 (1.43). In both cases, the edge is significantly enhanced.

Figure 7.3  Original and Filtered Data
7.4 Image Texture

The human visual system makes extensive use of texture in recognition of objects and image interpretation. Digital image processing does not have the same ability to do this, but many measures of 'texture' may be proposed based on the image values. Texture has been shown to be important in recognition and classification of different vegetation types.

The simplest measures of image texture are based on neighbourhood template windows - for example, the standard deviation of the band values in a $3 \times 3$ window.

The texture values highlight areas of rapid changes in the spectral data. Their use depends on the distribution of ground objects within a class. For example, with MSS spatial resolution, an open woodland may appear smooth (low texture), whereas with TM data it may appear quite variable as pixels contain different portions of vegetation and open ground. Calculation of a texture band using NIR or a 'vegetation index' may help to identify classes of similar mean spectral response but different spatial arrangement.
8. REGISTRATION AND RECTIFICATION OF IMAGES

For change detection and integration of different data sources, it is often necessary to register two or more images to a common base, or to rectify imagery to a map base.

The process of geometric correction to a defined base grid is that of finding in the distorted image the image coordinates of the ground pixel which corresponds to each base grid location. This operation is carried out for all base grid coordinate points. Writing $X,Y$ for the base coordinates, and $x,y$ for the coordinate system in the distorted image, it is necessary to derive and apply functions

$$x = f_1(X,Y)$$
$$y = f_2(X,Y)$$

for each $X,Y$ on the base grid. The digital values in the distorted image can then be extracted and written to a new 'rectified' file.

8.1 Fitting the Functions

For Landsat and Spot data, polynomial functions are adequate for $f_1$ and $f_2$. For wide angle scanners (eg AVHRR) and airborne systems, more complicated functions may be required.

Possible forms of polynomial to use are

1. linear (first order)
   $$x = a + bX + cY$$
   (3 coefficients)

2. linear + cross-term
   $$x = a + bX + cY + dXY$$
   (4 coefficients)

3. quadratic
   $$x = a + bX + cY + dXY + eX^2 + fY^2$$
   (6 coefficients)

4. cubic
   $$x = a + bX + cY + dXY + eX^2 + fY^2 + gX^3 + hX^2Y + iXY^2 + jX^3 + jY^3$$
   (10 coefficients!)

The steps in image rectification or registration are

1. Choose a base image, area and grid size
2. Locate a well-distributed set of ground control points (GCP's) in both coordinate systems
3. Fit suitable polynomial 'warp' functions $f_1$, $f_2$. (If the fit is satisfactory, then proceed, otherwise check the GCPs and repeat)
4. Resample or warp the distorted image to the base coordinate system
5. Merge the base and warped images.

In practice, the location of ground control points can be a time-consuming process.

The fitting procedure should always be run carefully to ensure that the calculated function is a 'good fit'. The fitting procedure is a least-squares regression procedure. If it is used carefully, the user should be able to check the GCP’s and choose an appropriate functional model.
For a good fit, it is necessary to have many more GCP's than the number of coefficients in the model.

The regression fit will produce equations, whether the GCP's are accurately located or not. The regression output should be used to check the prediction of the GCP's. The residual mean square error, and the individual GCP's residuals, should be inspected. It is useful to plot the residuals against the fitted values for x and y to check that the model is reasonable and to identify influential or atypical GCP coordinates.

If your image processing system will not allow these checks, it is better to run the regression procedures in some statistical package such as SAS or GLIM.

For Landsat TM or Spot data, a linear or quadratic function should be adequate. If a cubic is required to reduce the rms error, it is likely that some of the GCP's are in error.

Figure 8.1 GCPs in base and distorted grids
8.2 Resampling

The base and distorted image grids will not exactly coincide. In relocating the pixel values to the new coordinates, values cannot be retrieved which exactly correspond to the new grid area. Different resampling strategies may be employed for different needs:

1. nearest neighbour - this simply transfers the values from the closest pixel to the new coordinates and does not alter the spectral values
2. bilinear interpolation - this transfers a weighted average of the four closest pixels
3. cubic convolution - this uses a value calculated from 16 pixels in the distorted image; the result is somewhat sharper than that provided by method (2).

For many applications, it is desirable to use the original unresampled data and rectify the results at a later stage. Note that it is possible and often useful to warp map data to image coordinates in such operations.
8.4

8.3 Rectification and Registration by Triangulation

Airborne scanner images suffer from complex and unpredictable distortions due to aircraft movement and interactions of the scanner geometry and ground relief variations.

Polynomial warping is not suitable for such images, as the GCP's in one area may not relate to those elsewhere. 'Local' correction can be achieved by triangulation of a network of GCP's and linear resampling within each triangle.