PRINCIPLES OF REMOTE SENSING

ITC EDUCATIONAL TEXTBOOK SERIES 2

Principles of Remote Sensing

An introductory textbook

Editors

Lucas L. F. Janssen   Gerrit C. Huurneman

Authors

Wim H. Bakker   Lucas L. F. Janssen   Colin V. Reeves
Ben G. H. Gorte   Christine Pohl   Michael J. C. Weir
John A. Horn   Anupma Prakash   Tsehaie Woldai

Cover illustration of the printed book:
Paul Klee (1879–1940), *Chosen Site* (1927)
Pen-drawing and water-colour on paper. Original size: 57.8×40.5 cm.
Private collection, Munich

Cover page design: Wim Feringa

Published by:
The International Institute for Aerospace Survey and Earth Sciences (ITC),
Hengelosestraat 99,
P.O. Box 6,
7500 AA Enschede, The Netherlands

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Principles of Remote Sensing
Lucas L. F. Janssen and Gerrit C. Huurneman (eds.)
(ITC Educational Textbook Series; 2)
Second edition
ISBN 90–6164–199–3 ITC, Enschede, The Netherlands
ISSN 1567–5777 ITC Educational Textbook Series

Contents

List of Figures

List of Tables

Preface

Principles of Remote Sensing is the basic textbook on remote sensing for all students enrolled in the 2000–2001 educational programmes at ITC. As well as being a basic textbook for the institute's regular MSc and PM courses, *Principles of Remote Sensing* will be used in various short courses and possibly also by ITC's sister institutes. The first edition is an extensively revised version of an earlier text produced for the 1999–2000 programme. *Principles of Remote Sensing* and the companion volume, *Principles of Geographic Information Systems* [6], are published in the ITC Educational Textbook series. We need to go back to the 1960s to find a similar official ITC textbook on subjects related to Remote Sensing: the ITC Textbooks on Photogrammetry and Photo-interpretation, published in English and French [11, 12].

You may wonder why ITC has now produced its own introductory textbook while there are already many books on the subject available on the market. *Principles of Remote Sensing* is different in various aspects. First of all, it has been developed for the specific ITC student population, thereby taking into account their entry level and knowledge of English language. The textbook relates to the typical ITC application disciplines and among others provides an introduction into techniques which acquire sub-surface characteristics. As the textbook is used in the start of the programmes, it tries to stimulate conceptual and abstract thinking by providing and explaining some fundamental, however simple, equations (in general, no more than one equation per chapter). *Principles of Remote Sensing* aims to provide a balanced approach towards traditional *photogrammetric* and *remote sensing* subjects: three sensors (aerial camera, multispectral scanner and radar) are dealt with more or less the same detail. Finally, compared to other introductory textbooks which often focus on the technique, *Principles of Remote Sensing* also introduces processes. In this sense, it provides a frame to refer to when more detailed subjects are dealt with later in the programme.

How to use the material

Principles of Remote Sensing has been produced both as a hard-copy textbook and as an electronic document. In this way, the student is offered the optimal combination to study the subject, and, to use the book as a general reference. Each chapter gives a summary and provides questions for self test. The book comprises a glossary, an index and a bibliography. The electronic document (PDF format) enables fast navigation and quick referencing.

Acknowledgements

This textbook is the result of a process to define and develop material for a core curriculum. This process started in 1998 and was carried out by a working group comprising of Rolf de By, Michael Weir, Cees van Westen, myself, chaired by Ineke ten Dam and supported by Erica Weijer. This group put many efforts in the definition and realization of the earlier version of the two core textbooks. Ineke was also supervising the process leading to this result. My fellow working group members are greatly acknowledged for their support.

This textbook could not have materialized without the efforts of the (co–) authors of the chapters: Wim Bakker, Ben Gorte, John Horn, Christine Pohl, Colin Reeves, Michael Weir and Tsehaie Woldai. Many other colleagues contributed one way or another to either the earlier version or this version of *Principles of Remote Sensing*: Paul Hofstee, Gerrit Huurneman, Yousif Hussin, David Rossiter, Rob Soeters, Ernst Schetselaar, Andrew Skidmore, Dhruba Shrestha and Zoltán Vekerdy.

The design and implementation of the textbook layout, of both the hard-copy and electronic document, is the work of Rolf de By. Using the LaTeX typesetting system, Rolf realized a well structured and visually attractive document to study. Many of the illustrations in the book have been provided by the authors, supported by Job Duim and Gerard Reinink. Final editing of the illustrations was done by Wim Feringa who also designed the cover.

Michael Weir has done a tremendous job in checking the complete textbook on English spelling and grammar. We know that our students will profit from this.

The work on this textbook was greatly stimulated through close collaboration

with the editor of *Principles of Geographic Information Systems*, Rolf de By.

Lucas L. F. Janssen, Enschede, September 2000

Preface to the second edition

After Lucas Janssen left ITC last year, moving into another job, I was asked to be the editor for the next version(s) of this book. I got impressed about the work Lucas, in cooperation with many others, achieved.

The LaTeX frame, developed by Rolf de By is of a clear construction. It enabled me to enter new items and to modify text, equations and other objects without being familiar with this typesetting system. However, without the permanent support of Rolf I could never have finished this version.

Following many constructive suggestions received from users and reviewers, for which I am very grateful, a number of changes were realized. Especially I thank Wim Bakker for his careful reading of all chapters and his suggestions to improve the book and Wim Feringa for his contribution to the creation and modification of a number of illustrations.

The major changes made in this version are:

- A chapter on Radiometric aspects is included, treating the main distortions caused by sensors and/or the atmosphere. The text and the corresponding figures were produced by my colleague Anupma Prakash and she is acknowledged for her contribution.

- A number of new sensor systems is added to Chapter 5 on Multispectral scanners.

- On demand of the ITC working group "Core Modules," the Chapters 6 and 7 are excluded from the ITC core modules, and also a number of (sub)sections. They are indicated by a special icon in the margin, depicting a 'dangerous bend' road sign.

Gerrit C. Huurneman, Enschede, August 2001

Chapter 1

Introduction to remote sensing

1.1 Spatial data acquisition

All ITC students, one way or another, deal with georeferenced data. They might be involved in the collection of data, processing of the data, analysis of the data or actually using the data for decision making. In the end, data are acquired to yield information for management purposes: water management, land management, resources management, *et cetera*. By *data* we mean representations that can be operated upon by a computer; by *information* we mean data that has been interpreted by human beings (Principles of GIS, Chapter 1). This textbook focusses on the methods used to collect georeferenced, or geo-spatial, data. The need for spatial data is best illustrated by some examples.

- An agronomist is interested in forecasting the overall agricultural production of a large area. This requires data on the area planted with different crops and data on biomass production to estimate the yield.

- An urban planner needs to identify areas in which dwellings have been built illegally. The different types of houses and their configuration needs to be determined. The information should be in a format that enables integration with other socio-economic information.

- An engineer needs to determine the optimal configuration for siting of relay stations for a telecommunication company. The optimal configuration primarily depends on the form of the terrain and on the location of obstacles such as buildings.

- A mining engineer is asked to explore an area and to provide a map of the surface mineralogy. In addition, s/he should start to give a first estimation of the effect of water pumping on the neighbouring agricultural region.

- A climatologist would like to understand the causes of the El Niño phenomenon. For this, s/he would need data on many parameters including sea currents, sea surface temperature, sea level, meteorological parameters, energy interactions between the land and water surface, *et cetera*.

Note that all of the above examples deal with spatial phenomena; in fact, with *spatio-temporal phenomena* since time is an important dimension too. To satisfy the information requirements of the above mentioned examples a wide variety of methods will be used: conducting interviews, land surveying, laboratory measurements of samples, interpretation of satellite images, measurements by *in situ* sensors, using aerial photographs, running numerical models, *et cetera*. For our purposes, it makes sense to distinguish between ground-based and remote sensing methods.

Ground-based and Remote Sensing Methods

In principle, there are two main categories of spatial data acquisition:

- *ground-based methods* such as making field observations, taking *in situ* measurements and performing land surveying. Using ground-based methods, you operate in the real world environment (Figure 1.1).

Figure 1.1: The principle of a ground-based method: measurements and observations are performed in the real world.

- *remote sensing methods*, which are based on the use of image data acquired by a sensor such as aerial cameras, scanners or a radar. Taking a remote sensing approach means that information is derived from the image data, which form a (limited) representation of the real world (Figure 1.2).

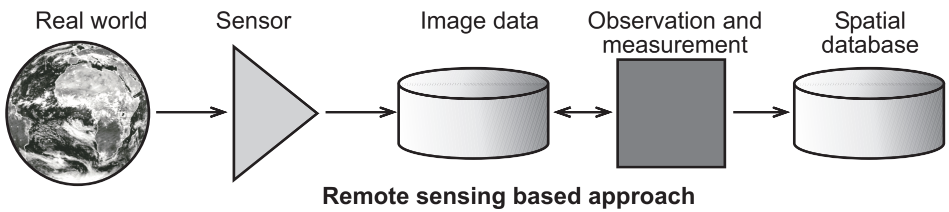


Figure 1.2: The principle of a remote sensing based method: measurement and analysis are performed on image data.

This textbook, *Principles of Remote Sensing*, introduces an overview and some first concepts of the remote sensing process. First some definitions of remote sensing will be given. In Section 1.2, some of the aspects and considerations when taking a remote sensing approach are discussed.

Remote Sensing Definitions

Three definitions of remote sensing are given below:

- Remote sensing is the science of acquiring, processing and interpreting images that record the interaction between electromagnetic energy and matter [24].
- Remote sensing is the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation [16].
- Remote sensing is the instrumentation, techniques and methods to observe the Earth's surface at a distance and to interpret the images or numerical values obtained in order to acquire meaningful information of particular objects on Earth [2].

Common to the three definitions is that data on characteristics of the Earth's surface are acquired by a device that is not in contact with the objects being measured. The result is usually stored as image data (in this book, aerial photographs are also considered as 'image data'). The characteristics measured by a sensor are the electromagnetic energy reflected or emitted by the Earth's surface. This energy relates to some specific parts of the electromagnetic spectrum: usually visible light, but it may also be infrared light or radio waves. There is a wide range of remote sensing sensors. Sensors, linked to a certain platform, are classified according to their distance from the Earth's surface: airborne and spaceborne sensors. Together, they contribute to *aerospace surveying*, which is the combined use of remote sensing and ground-based methods to collect information.

Another term to introduce here is *Earth Observation* (EO), which usually refers to spaceborne remote sensing.

Before the image data can yield the required information about the objects or phenomena of interest, they need to be processed. The analysis and *information extraction* or *information production* is part of the overall remote sensing process.

In this textbook, *remote sensing* refers to all aerospace remote sensing techniques, including aerial photography. Sometimes, because of historical reasons, aerial photos are referred to as a separate aspect of aerospace survey.

1.2 Application of remote sensing

The textbook *Principles of GIS* [6] introduced the example of studying the El Niño effect to illustrate aspects of database design and functionality of spatial information systems. The example departs from a database table, which stores a number of parameters derived from buoys, that is *in situ*, measurements. These measurements can be analysed as they are (per buoy, over time). Most often, however, spatial interpolation techniques will be used to generate maps that enable analysis of spatial and temporal patterns.

Now, let us consider the analysis of Sea Surface Temperature (SST) patterns and discuss some particular aspects related to taking a remote sensing approach to the El Niño case.

Remote sensing provides image data

The image data acquired by RS relate to electromagnetic properties of the Earth. Under many conditions the data can be related to real world parameters or features. For example, measurements in the thermal infrared wavelength can be used directly to calculate the surface temperature. This requires some processing in which, among others, the effect of the atmosphere is corrected for. SST calculation is a relatively straightforward procedure and standard SST products are publicly available.

Remote sensing requires ground data.

Although remote sensing data can be interpreted and processed without other information, the best results are obtained by linking remote sensing measurements to ground (or surface) measurements and observations. Highly accurate SST maps can be derived from satellite image data when combined with *in situ* temperature measurements (by buoy). This idea of complementarity between remote sensing and ground-based surveying is also part of ITC's name: *aerospace surveying*. Figure 1.3 illustrates this concept.

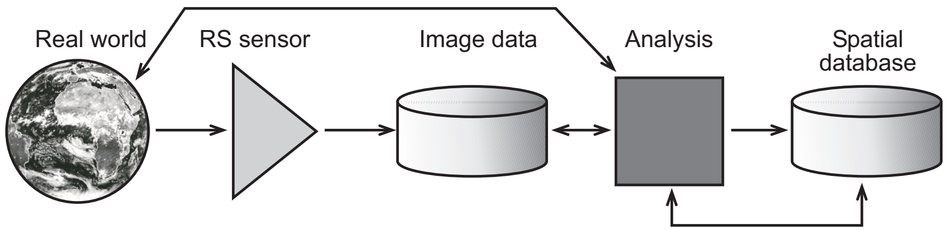


Figure 1.3: In most situations, remote sensing based data acquisition is complemented by ground-based measurements and observations.

Remote sensing provides area covering data

The spatial pattern of SST can best be studied using map-like products. These can be obtained by interpolation from *in situ* measurements by buoys. However, it is unlikely that the number and distribution of buoys is adequate to enable detection/identification of relevant surface features. The distance between the buoys network for El Niño is in the order of 2000 km (in x-direction) and 300 km (in y-direction). Using data of meteorological satellites, SST can be assessed at a level of detail up to 1 km (Figure 1.4). RS derived SST data can, for example, be used to determine an optimal density and location of buoys.

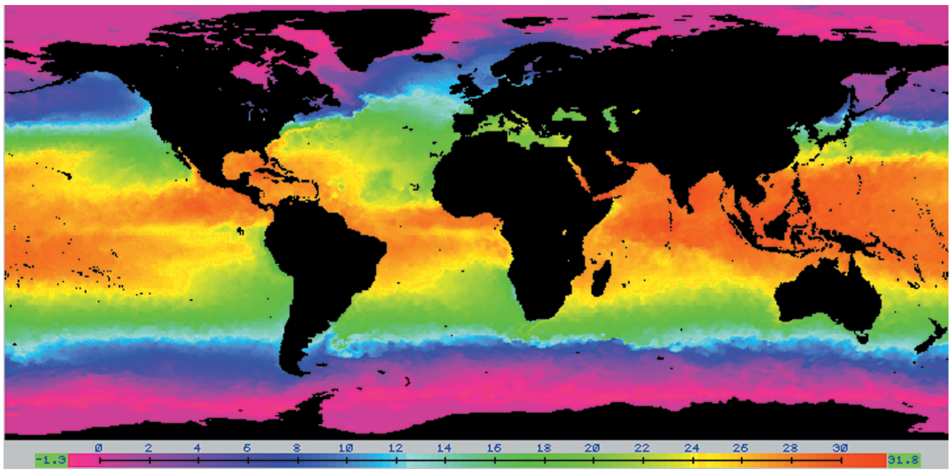


Figure 1.4: Sea surface temperature as determined from NOAA-AVHRR data. Courtesy of NOAA

Remote sensing provides surface information

In principle, remote sensing provides information about the upper few millimetres of the Earth's surface. Some techniques, specifically in the microwave domain, relate to greater depths. The fact that measurements only refer to the surface is a limitation of remote sensing. Additional 'models' or assumptions are required to estimate subsurface characteristics. In the case of SST, the temperature derived from RS only tells something about the temperature of the actual surface of the ocean. No information can be derived about subsurface currents (which is possible when using buoys).

Apart from SST, remote sensing can be used for assessing many other surface characteristics. In the context of El Niño the following parameters may be assessed from RS data: biomass (Figure 1.5), sea level (Figure 1.6), precipitation and surface wind (Figure 1.7).

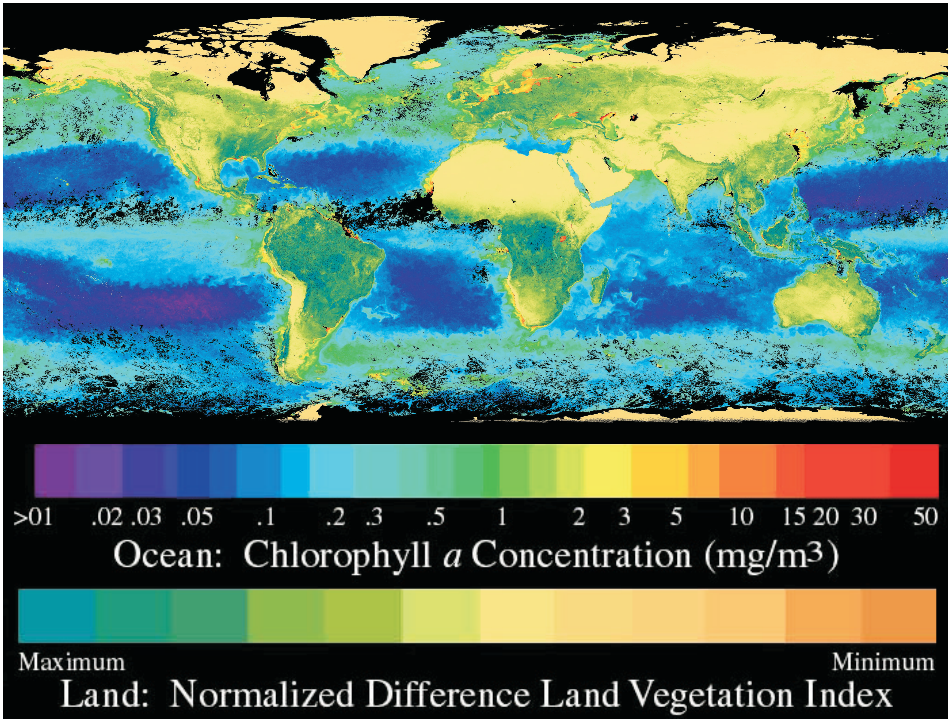


Figure 1.5: Ocean (and land) biomass as determined from Orbview-3 data. Courtesy of Orbview

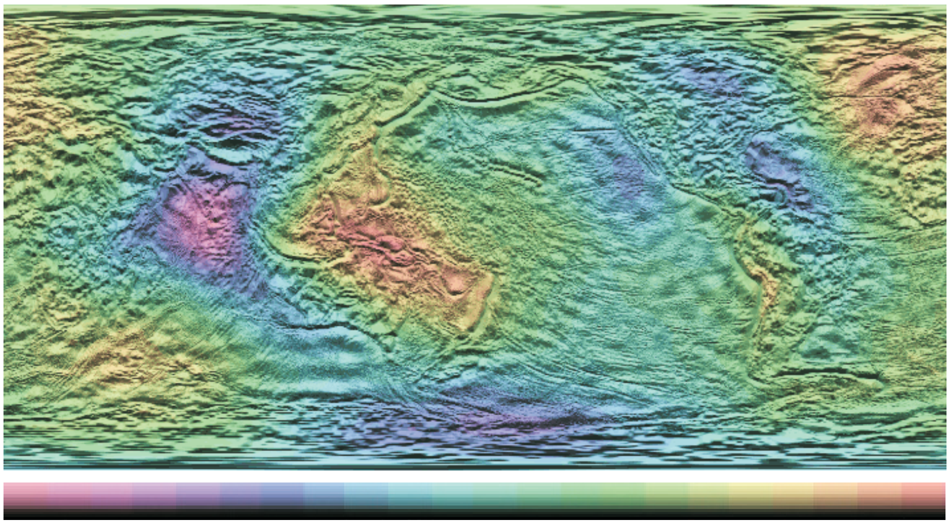


Figure 1.6: Sea surface height relative to the ocean geoid as determined from spaceborne radar and laser systems. Courtesy of University of Texas

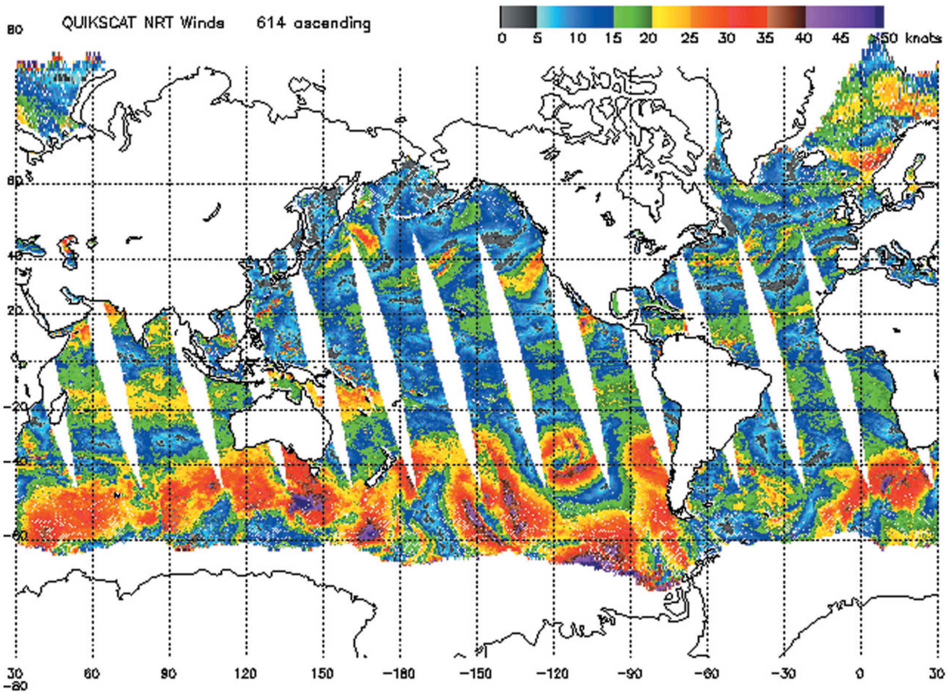


Figure 1.7: Ocean surface wind as determined from scatterometer measurements by QuickScat. Courtesy of NOAA

Remote sensing is the only way to do it

The Pacific Ocean is known for its unfavourable weather and ocean conditions. It is quite difficult to install and maintain a network of measuring buoys in this region. A RS approach is specifically suited for areas that are difficult to access. A related topic is that of acquiring global or continental data sets. RS allows data to be acquired globally using the same (or similar) sensor. This enables methods for monitoring and change detection.

Remote sensing provides multipurpose image data

In this example, SST maps are used to study El Niño. The same data could be used years later to find, for example, a relation between SST and the algae blooms around Pacific islands. SST maps are not only of interest to researchers but also to large fishing companies that want to guide their vessels to promising fishing grounds. A data set, thus, can be of use for more than one organization and can also prove useful for historical studies.

Remote sensing is cost-effective

The validity of this statement is sometimes hard to assess, especially when dealing with spaceborne remote sensing. Consider an international scientific project that studies the El Niño phenomenon. Installation and maintenance of buoys costs a lot of money. Meteorological satellites have already been paid for and the data can be considered 'free'. Remote sensing would thus be a cheap technique. However, as RS can only be used to measure surface characteristics, buoys still need to be placed to determine subsurface characteristics.

Although the above statements are related to a scientific problem of a global phenomenon, they are generic to other application contexts. Consider the above statements related to the monitoring of (illegal) urban growth around African cities using aerial photography.

Probably, you will conclude that you do not have all the knowledge required to comment on these statements in the urban context. You may even find it difficult to identify which remote sensing technique solves your own particular information requirements. After studying this textbook *Principles of Remote Sensing* you may expect to be better equipped to consider these issues.

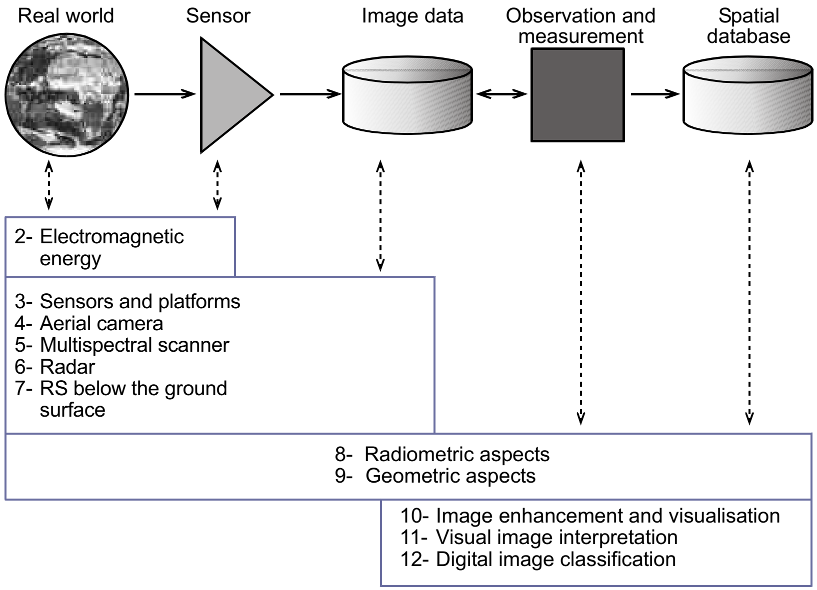


Figure 1.8: Relationship between the chapters of this textbook and the remote sensing process

1.3 Structure of this textbook

Figure 1.8 summarizes the content of Chapters 2 to 12 in relation to the overall remote sensing process.

First of all the underlying principle of remote sensing is explained: the measurement of electromagnetic energy that is reflected or emitted by the objects and materials on the Earth's surface. The main source for energy is the Sun (Chapter 2).

In the subsequent chapters (Chapters 3–7), the platform-sensor concept is explained and followed by an overview of the different types of sensors. The aerial camera, the multispectral scanner and the radar system are each introduced in a dedicated chapter. This part concludes with a summary of some specific remote sensing techniques that yield information about subsurface characteristics.

Chapters 8 and 9 introduce the radiometric and geometric aspects of remote sensing. These are important because the image data need to be integrated with other data in a GIS environment. Image data can be visualized in different ways. Chapter 10 introduces the main concepts required to understand and analyse the main classes of remote sensing pictures.

The final two chapters (11 and 12) deal with information extraction methods: visual image interpretation by a human operator, digital image classification based upon computer algorithms, and, calculation of (physical) parameters.

Summary

A large part of human activities and interest has a geographic component. For planning, monitoring and decision making, there is a need for georeferenced (geo-spatial) data.

Ground-based methods and remote sensing based methods for spatial data acquisition are distinguished. Remote sensing methods rely on the measurement of electromagnetic energy from a distance (aerospace).

A remote sensing approach is usually complemented by ground-based methods and the use of numerical models. For the appropriate choice of relevant remote sensing data acquisition you have to define the information requirements of your application.

Questions

The following questions can help to study Chapter 1.

1. To what extent are Geographic Information Systems (GIS) applied by your organization (company)?
2. Which ground-based and which remote sensing methods are used by your organization (or company) to collect georeferenced data?
3. Remote sensing data and derived data products are available on the internet. Locate three web-based catalogues or archives that comprise remote sensing image data.

4. Following the El Niño example, outline how the statements about RS relate to a particular case of your organization (e.g., land resources, geology, *et cetera*).

These are typical exam questions:

1. Explain, or give an example, how ground-based and remote sensing methods may complement each other.

2. List three possible limitations of remote sensing data.

Chapter 2

Electromagnetic energy and remote sensing

2.1 Introduction

Remote sensing relies on the measurement of electromagnetic (EM) energy. EM energy can take several different forms. The most important source of EM energy at the Earth's surface is the Sun, which provides us, for example, with (visible) light, heat (that we can feel) and UV-light, which can be harmful to our skin.

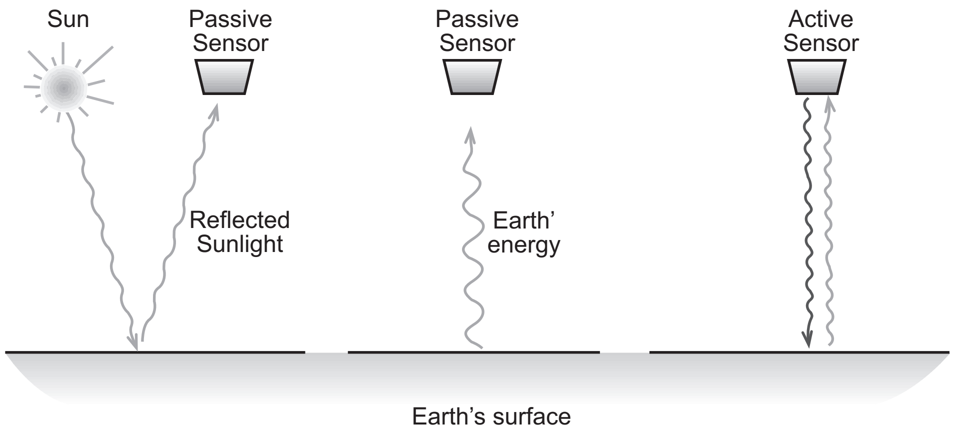


Figure 2.1: A remote sensing sensor measures reflected or emitted energy. An active sensor has its own source of energy.

Many sensors used in remote sensing measure reflected sunlight. Some sensors, however, detect energy emitted by the Earth itself or provide their own energy (Figure 2.1). A basic understanding of EM energy, its characteristics and its interactions is required to understand the principle of the remote sensor. This knowledge is also needed to interpret remote sensing data correctly. For these reasons, this chapter introduces the basic physics of remote sensing.

In Section 2.2, EM energy, its source and the different parts of the electromagnetic spectrum are explained. In between the remote sensor and the Earth's surface is the atmosphere that influences the energy that travels from the Earth's surface to the sensor. The main interactions between EM waves and the atmosphere are described in Section 2.3. Section 2.4 introduces the interactions that take place at the Earth's surface.

2.2 Electromagnetic energy

2.2.1 Waves and photons

Electromagnetic (EM) energy can be modelled in two ways: by waves or by energy bearing particles called photons. In the wave model, electromagnetic energy is considered to propagate through space in the form of sine waves. These waves are characterized by two fields, electrical (E) and magnetic (M) fields, which are perpendicular to each other. For this reason, the term *electromagnetic* energy is used. The vibration of both fields is perpendicular to the direction of travel of the wave (Figure 2.2). Both fields propagate through space at the speed of light c, which is 299,790,000 m/s and can be rounded off to $3 \cdot 10^8$ m/s.

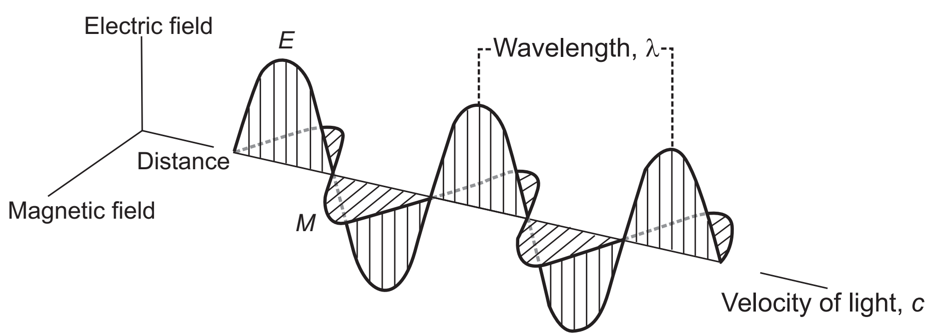


Figure 2.2: Electric (E) and magnetic (M) vectors of an electromagnetic wave.

One characteristic of electromagnetic waves is particularly important for understanding remote sensing. This is the wavelength λ that is defined as the distance between successive wave crests (Figure 2.2). Wavelength is measured in metres (m) or some factor of metres such as nanometres (nm, 10^{-9} m) or micrometres (μm, 10^{-6} m). (For an explanation of units and prefixes refer to Appendix 1).

The frequency, v, is the number of cycles of a wave passing a fixed point over a specific period of time. Frequency is normally measured in hertz (Hz), which is equivalent to one cycle per second. Since the speed of light is constant, wavelength and frequency are inversely related to each other:

$$c = \lambda \times v. \tag{2.1}$$

In this equation, c is the speed of light ($3 \cdot 10^8$ m/s), λ is the wavelength (m), and v is the frequency (cycles per second, Hz).

The shorter the wavelength, the higher the frequency. Conversely, the longer the wavelength, the lower the frequency (Figure 2.3).

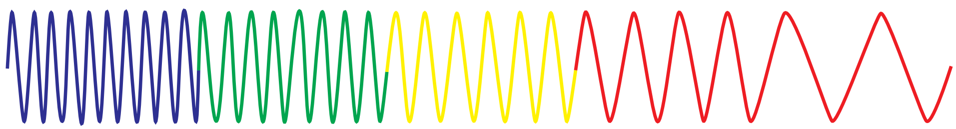


Figure 2.3: Relationship between wavelength, frequency and energy

Most characteristics of EM energy can be described using the 'wave' model as described above. For some purposes, however, EM energy is more conveniently modelled by the particle theory, in which EM energy is composed of discrete units called 'photons'. This approach is taken when quantifying the amount of energy measured by multispectral sensor (Section 5.2.1). The amount of energy held by a photon of a specific wavelength is then given by

$$Q = h \times v = h \times \frac{c}{\lambda}, \tag{2.2}$$

where Q is the energy of a photon (J), h is Planck's constant ($6.6262 \cdot 10^{-34}$ J s), and v the frequency (Hz). From Equation 2.2 it follows that the longer the wavelength, the lower its energy content. Gamma rays (around 10^{-9} m) are the most energetic, and radio waves (> 1 m) the least energetic. An important consequence for remote sensing is that it is more difficult to measure the energy emitted in longer wavelengths than in shorter wavelengths.

2.2.2 Sources of EM energy

All matter with a temperature above absolute zero (0 K, where $n\ ^\circ\mathrm{C} = n + 273\ K$) radiates EM energy due to molecular agitation. Agitation is the movement of the molecules. This means that the Sun, and also the Earth, radiate energy in the form of waves. Matter that is capable of absorbing and re-emitting *all* EM energy is known as a blackbody. For blackbodies both the emissivity, ϵ, and the absorptance, α, are equal to (the maximum value of) 1.

The amount of energy radiated by an object depends on its absolute temperature, its emissivity and is a function of the wavelength. In physics, this principle is defined by Stefan-Boltzmann's Law. A blackbody radiates a continuum of

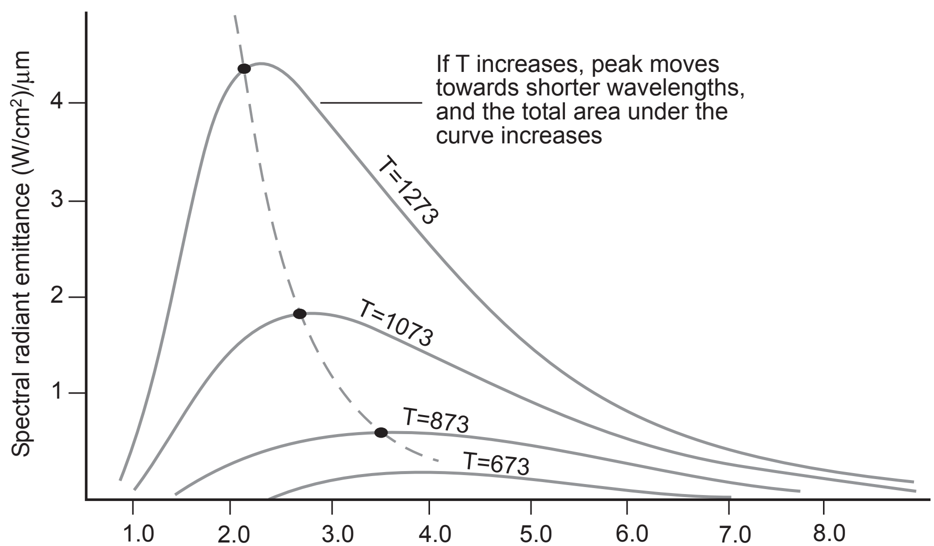


Figure 2.4: Blackbody radiation curves based on Stefan-Boltzmann's law (with temperatures in K)

wavelengths. The radiation emitted by a blackbody at different temperatures is shown in Figure 2.4. Note the units in this figure: the x-axis indicates the wavelength and the y-axis indicates the amount of energy per unit area. The area below the curve, therefore, represents the total amount of energy emitted at a specific temperature. From Figure 2.4 it can be concluded that a higher temperature corresponds to a greater contribution of shorter wavelengths. The peak radiation at 400 °C is around 4 μm while the peak radiation at 1000 °C is at 2.5 μm. The emitting ability of a real material compared to that of the blackbody is referred to as the material's emissivity. In reality, blackbodies are hardly found in nature; most natural objects have emissivities less than one. This means that only part, usually between 80–98%, of the received energy is re-emitted. Consequently, part of the energy is absorbed. This physical property is relevant in, for example, the modelling of global warming processes.

2.2.3 Electromagnetic spectrum

All matter with a certain temperature radiates electromagnetic waves of various wavelengths. The total range of wavelengths is commonly referred to as the electromagnetic spectrum (Figure 2.5). It extends from gamma rays to radio waves.

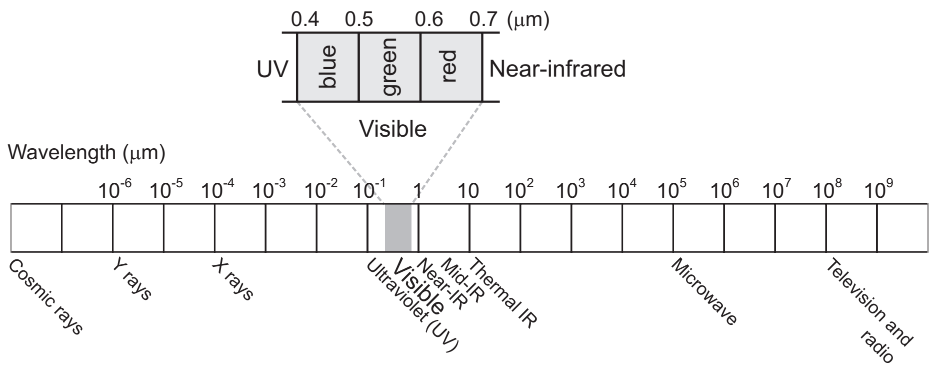


Figure 2.5: The electromagnetic spectrum, after [16]

Remote sensing operates in several regions of the electromagnetic spectrum. The *optical part of the EM spectrum* refers to that part of the EM spectrum in which optical laws can be applied. These relate to phenomena, such as reflectance and refraction, that can be used to focus the radiation. The optical range extends from X-rays (0.02 μm) through the visible part of the EM spectrum up to and including far-infrared (1000 μm). The ultraviolet (UV) portion of the spectrum has the shortest wavelengths that are of practical use for remote sensing. This radiation is beyond the violet portion of the visible wavelengths. Some of the Earth's surface materials, primary rocks and minerals, emit or fluoresce visible light when illuminated with UV radiation. The microwave range covers wavelengths from 1 mm to 1 m.

The visible region of the spectrum (Figure 2.5) is commonly called 'light'. It occupies a relatively small portion in the EM spectrum. It is important to note that this is the only portion of the spectrum that we can associate with the concept of colour. Blue, green and red are known as the primary colours or wavelengths of the visible spectrum. Section 10.2 gives more information on 'light' and perception of 'colour'.

The longer wavelengths used for remote sensing are in the thermal infrared and microwave regions. Thermal infrared gives information about surface temperature. Surface temperature can be related, for example, to the mineral composition of rocks or the condition of vegetation. Microwaves can provide information on surface roughness and the properties of the surface such as water content.

2.2.4 Active and passive remote sensing

In remote sensing, the sensor measures energy. *Passive* and *active* techniques are distinguished. Passive remote sensing techniques employ natural sources of energy, such as the Sun. Active remote sensing techniques, for example radar and laser, have their own source of energy. Active sensors emit a controlled beam of energy to the surface and measure the amount of energy reflected back to the sensor (Figure 2.1).

Passive sensor systems based on reflection of the Sun's energy can only work during daylight. Passive sensor systems that measure the longer wavelengths related to the Earth's temperature do not depend on the Sun as a source of illumination and can be operated at any time. Passive sensor systems need to deal with the varying illumination conditions of the Sun, which are greatly influenced by atmospheric conditions. The main advantage of active sensor systems is that they can be operated day and night and have a controlled illuminating signal.

2.3 Energy interaction in the atmosphere

The most important source of energy is the Sun. Before the Sun's energy reaches the Earth's surface, three fundamental interactions in the atmosphere are possible: absorption, transmission and scattering. The energy transmitted is then reflected or absorbed by the surface material. (Figure 2.6).

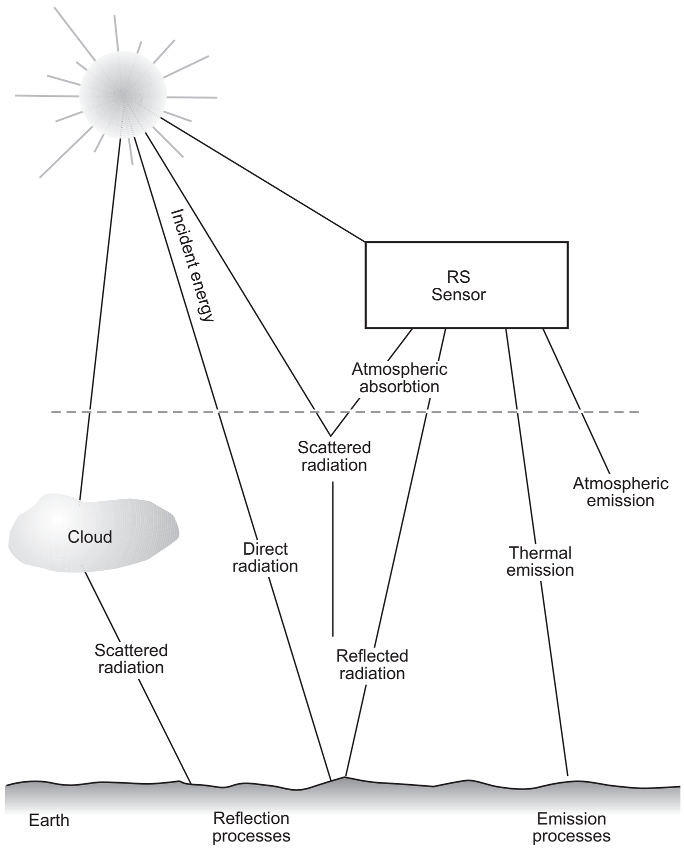


Figure 2.6: Energy interactions in the atmosphere and on the land

2.3.1 Absorption and transmission

Electromagnetic energy travelling through the atmosphere is partly absorbed by various molecules. The most efficient absorbers of solar radiation in the atmosphere are ozone (O_3), water vapour (H_2O) and carbon dioxide (CO_2).

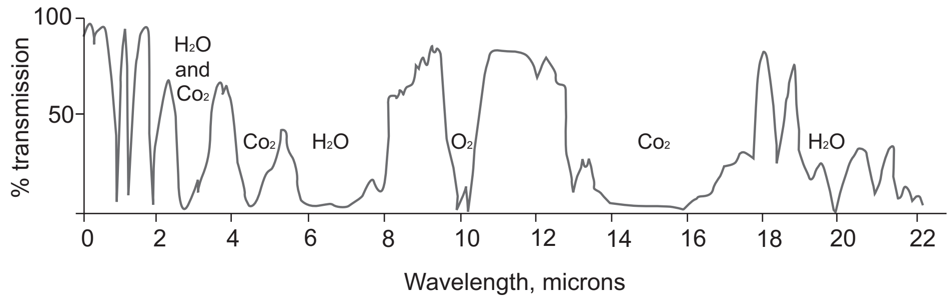


Figure 2.7: Atmospheric transmission expressed as percentage

Figure 2.7 gives a schematic representation of the atmospheric transmission in the 0–22 μm wavelength region. From this figure it may be seen that about half of the spectrum between 0–22 μm is useless for remote sensing of the Earth's surface, simply because none of the corresponding energy can penetrate the atmosphere. Only the wavelength regions outside the main absorption bands of the atmospheric gases can be used for remote sensing. These regions are referred to as the *atmospheric transmission windows* and include:

- A window in the visible and reflected infrared region, between 0.4–2 μm. This is the window where the (optical) remote sensors operate.

- Three windows in the thermal infrared region, namely two narrow windows around 3 and 5 μm, and a third, relatively broad, window extending from approximately 8 to 14 μm.

Because of the presence of atmospheric moisture, strong absorption bands are found at longer wavelengths. There is hardly any transmission of energy in the region from 22 μm to 1 mm. The more or less transparent region beyond 1 mm is the microwave region.

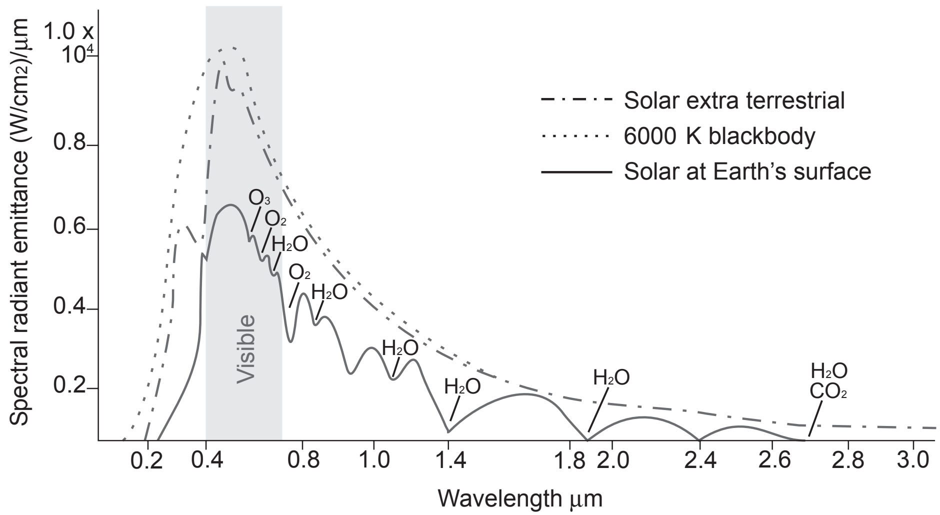


Figure 2.8: The electromagnetic spectrum of the Sun as observed with and without the influence of the Earth's atmosphere.

The solar spectrum as observed both with and without the influence of the Earth's atmosphere is shown in Figure 2.8. First of all, look at the radiation curve of the Sun (measured outside the influence of the Earth's atmosphere), which resembles a blackbody curve at 6000 K. Secondly, compare this curve with the radiation curve as measured at the Earth's surface. The relative dips in this curve indicate the absorption by different gases in the atmosphere.

2.3.2 Atmospheric scattering

Atmospheric scattering occurs when the particles or gaseous molecules present in the atmosphere cause the EM waves to be redirected from their original path. The amount of scattering depends on several factors including the wavelength of the radiation, the amount of particles and gases, and the distance the radiation travels through the atmosphere. For the visible wavelengths, 100% (in case of cloud cover) to 5% (in case of a clear atmosphere) of the energy received by the sensor is directly contributed by the atmosphere. Three types of scattering take place: Rayleigh scattering, Mie Scattering and Non-selective scattering.

Rayleigh scattering

Rayleigh scattering predominates where electromagnetic radiation interacts with particles that are smaller than the wavelength of the incoming light. Examples of these particles are tiny specks of dust and nitrogen (NO_2) and oxygen (O_2) molecules. The effect of Rayleigh scattering is inversely proportional to the wavelength: shorter wavelengths are scattered more than longer wavelengths (Figure 2.9).

Figure 2.9: Rayleigh scattering is caused by particles smaller than the wavelength and is maximal for small wavelengths.

In the absence of particles and scattering, the sky would appear black. In daytime, the Sun rays travel the shortest distance through the atmosphere. In that situation, Rayleigh scattering causes a clear sky to be observed as blue because this is the shortest wavelength the human eye can observe. At sunrise and sunset, however, the Sun rays travel a longer distance through the Earth's atmosphere before they reach the surface. All the shorter wavelengths are scattered after some distance and only the longer wavelengths reach the Earth's surface. As a result, the sky appears orange or red (Figure 2.10).

In the context of satellite remote sensing, Rayleigh scattering is the most important type of scattering. It causes a distortion of spectral characteristics of the reflected light when compared to measurements taken on the ground: by the Rayleigh effect the shorter wavelength are overestimated. In colour photos taken from high altitudes it accounts for the blueness of these pictures. In gen-

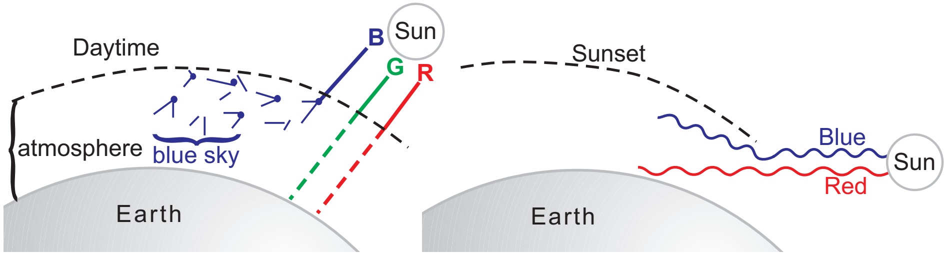


Figure 2.10: Rayleigh scattering causes us to perceive a blue sky during daytime and a red sky at sunset.

eral, the Rayleigh scattering diminishes the 'contrast' in photos, and thus has a negative effect on the possibilities for interpretation. When dealing with digital image data (as provided by scanners) the distortion of the spectral characteristics of the surface may limit the possibilities for image classification.

Mie scattering

Mie scattering occurs when the wavelength of the incoming radiation is similar in size to the atmospheric particles. The most important cause of Mie scattering are the aerosols: a mixture of gases, water vapour and dust.

Mie scattering is generally restricted to the lower atmosphere where larger particles are more abundant, and dominates under overcast cloud conditions. Mie scattering influences the entire spectral region from the near-ultraviolet up to and including the near-infrared.

Non-selective scattering

Non-selective scattering occurs when the particle size is much larger than the radiation wavelength. Typical particles responsible for this effect are water droplets and larger dust particles.

Non-selective scattering is independent of wavelength, with all wavelengths scattered about equally. The most prominent example of non-selective scattering includes the effect of clouds (clouds consist of water droplets). Since all wavelengths are scattered equally, a cloud appears white. Optical remote sensing cannot penetrate clouds. Clouds also have a secondary effect: shadowed regions on the Earth's surface (Figure 2.11).

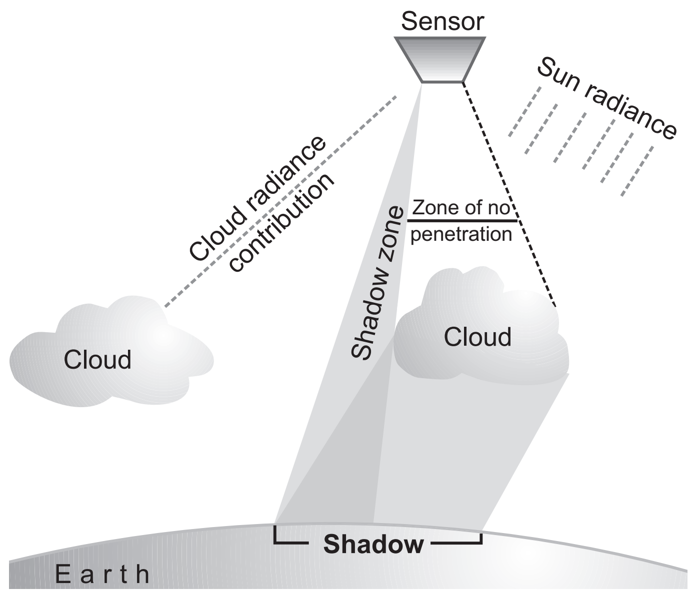


Figure 2.11: Direct and indirect effects of clouds in optical remote sensing

2.4 Energy interactions with the Earth's surface

In land and water applications of remote sensing we are most interested in the reflected radiation because this tells us something about surface characteristics. Reflection occurs when radiation 'bounces' off the target and is then redirected. Absorption occurs when radiation is absorbed by the target. Transmission occurs when radiation passes through a target. Two types of reflection, which represent the two extremes of the way in which energy is reflected from a target, are *specular reflection* and *diffuse reflection* (Figure 2.12). In the real world, usually a combination of both types is found.

- Specular reflection, or mirror-like reflection, typically occurs when a surface is smooth and all (or almost all) of the energy is directed away from the surface in a single direction. It is most likely to occur when the Sun is high in the sky. Specular reflection can be caused, for example, by a water

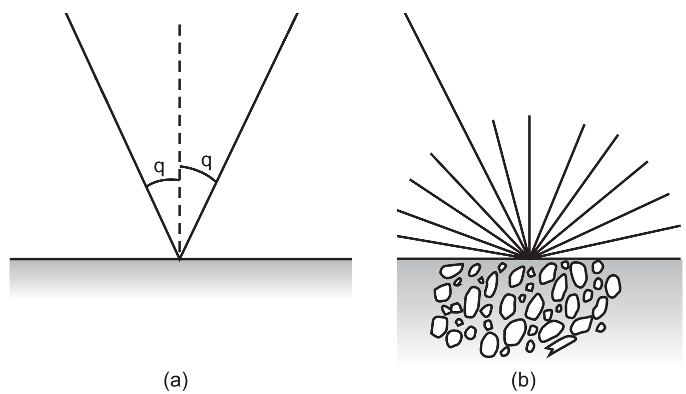


Figure 2.12: Schematic diagrams showing (a) specular and (b) diffuse reflection.

surface or a glasshouse roof. It results in a very bright spot (also called 'hot spot') in the image.

- Diffuse reflection occurs in situations where the surface is rough and the energy is reflected almost uniformly in all directions. Whether a particular target reflects specularly or diffusely, or somewhere in between, depends on the surface roughness of the feature in comparison to the wavelength of the incoming radiation.

2.4.1 Spectral reflectance curves

Consider a surface composed of a certain material. The energy reaching this surface is called *irradiance*. The energy reflected by the surface is called *radiance*. Irradiance and radiance are expressed in W/m^2.

For each material, a specific *reflectance curve* can be established. Such curves show the fraction of the incident radiation that is reflected as a function of wavelength. From such curve you can find the degree of reflection for each wavelength (e.g., at 400 nm, 401 nm, 402 nm, ...). Most remote sensing sensors are sensitive to broader wavelength bands, for example from 400–480 nm, and the curve can be used to estimate the overall reflectance in such bands. Reflectance curves are made for the optical part of the electromagnetic spectrum (up to 2.5 μm). Today, large efforts are made to store collections of typical curves in *spectral libraries*.

Reflectance measurements can be carried out in a laboratory or in the field using a field spectrometer. In the following subsections the reflectance characteristics of some common land cover types are discussed.

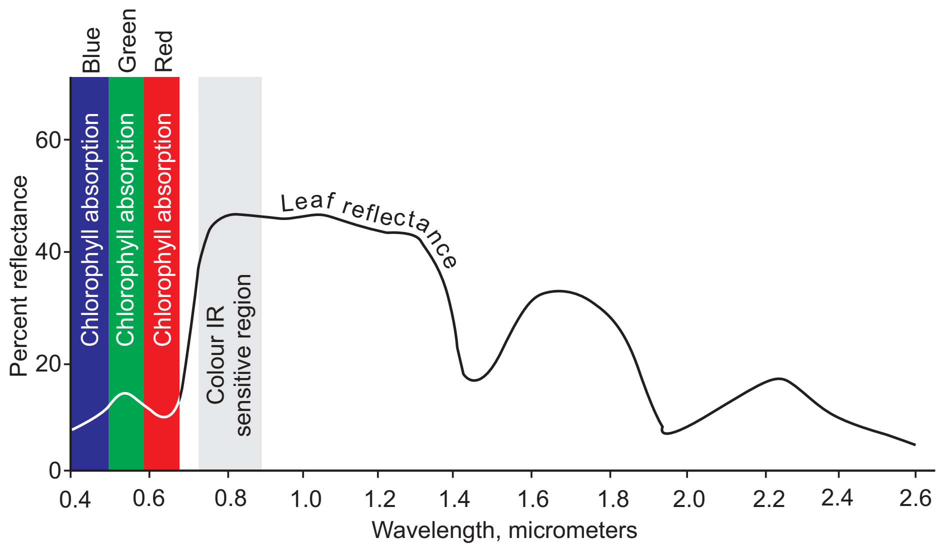


Figure 2.13: An idealized spectral reflectance curve of a healthy vegetation

Vegetation

The reflectance characteristics of vegetation depend on the properties of the leaves including the orientation and the structure of the leaf canopy. The proportion of the radiation reflected in the different parts of the spectrum depends on leaf pigmentation, leaf thickness and composition (cell structure) and on the amount of water in the leaf tissue. Figure 2.13 shows an ideal reflectance curve from healthy vegetation. In the visible portion of the spectrum, the reflection from the blue and red light is comparatively low since these portions are absorbed by the plant (mainly by chlorophyll) for photosynthesis and the vegetation reflects relatively more green light. The reflectance in the near-infrared is highest but the amount depends on the leaf development and the cell structure of the leaves. In the middle infrared, the reflectance is mainly determined by the free water in the leaf tissue; more free water results in less reflectance. They are therefore called water absorption bands. When the leaves dry out, for an example during the harvest time of the crops, the plant may change colour (for example, to yellow). At this stage there is no photosynthesis, causing reflectance in the red portion of the spectrum to be higher. Also, the leaves will dry out resulting in higher reflectance in the middle infrared whereas the reflectance in the near-infrared may decrease. As a result, optical remote sensing data provide information about the type of plant and also about its health condition.

Bare soil

Surface reflectance from bare soil is dependent on so many factors that it is difficult to give one typical soil reflectance curve. However, the main factors influencing the soil reflectance are soil colour, moisture content, the presence of carbonates, and iron oxide content. Figure 2.14 gives some reflectance curves for the five main types of soil occurring in the USA. Note the typical shapes of most of the curves, which show a convex shape between 500–1300 nm with dips at 1450 and 1950 nm. These dips are so-called *water absorbtion bands* and are caused by the presence of soil moisture. The iron-dominated soil has a quite different reflectance curve that can be explained by the iron absorbtion dominating at longer wavelengths.

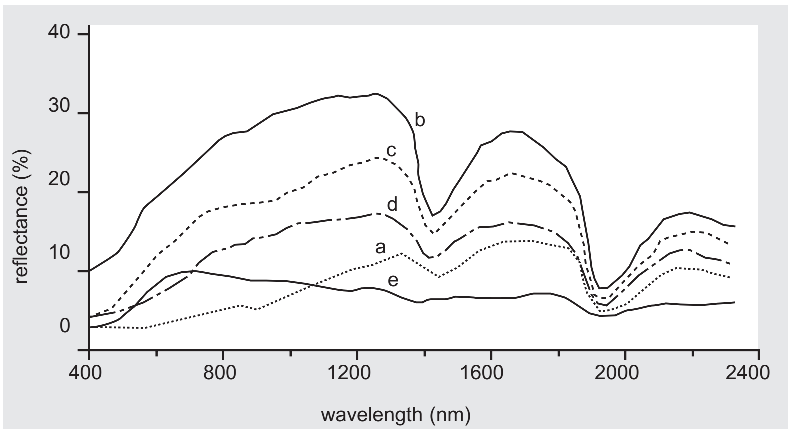


Figure 2.14: Reflectance spectra of surface samples of five mineral soils, (a) organic dominated, (b) minimally altered, (c) iron altered, (d) organic affected and (e) iron dominated (from [17])

Water

Compared to vegetation and soils, water has the low reflectance. Vegetation may reflect up to 50%, soils, up to 30–40% while water reflects at the most 10% of the incoming radiation. Water reflects EM energy in the visible up to the near-infrared. Beyond 1200 nm all energy is absorbed. Some curves of different types of water are given in Figure 2.15 The highest reflectance is given by turbid (silt loaded) water and water containing plants with a chlorophyll reflection peak at the green wavelength.

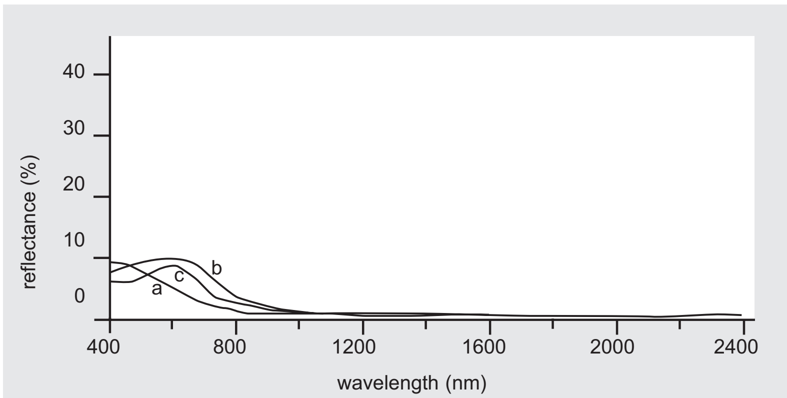


Figure 2.15: Typical effects of chlorophyll and sediments on water reflectance: (a) ocean water, (b) turbid water, (c) water with chlorophyll (from [17])

Summary

Remote sensing is based on the measurement of Electromagnetic (EM) energy. EM energy propagates through space in the form of sine waves characterized by electrical (E) and magnetic (M) fields, which are perpendicular to each other. EM can be modelled either by waves or by energy bearing particles called photons. One property of EM waves that is particularly important for understanding remote sensing is the wavelength (λ), defined as the distance between successive wave crests measured in metres (m), nanometres (nm, 10^{-9} m) or micrometres (μm, 10^{-6} m). The frequency is the number of cycles of a wave passing a fixed point in a specific period of time and is measured in hertz (Hz). Since the speed of light is constant, wavelength and frequency are inversely related to each other. The shorter the wavelength, the higher the frequency and *vice versa*.

All matter with a temperature above the absolute zero (0 K) radiates EM energy due to molecular agitation. Matter that is capable of absorbing and re-emitting all EM energy received is known as a blackbody. All matter with a certain temperature radiates electromagnetic waves of various wavelengths depending on its temperature. The total range of wavelengths is commonly referred to as the electromagnetic spectrum. It extends from gamma rays to radio waves. The amount of energy detected by a remote sensing system is a function of the interactions on the way to the object, the object itself and the interactions on the way returning to the sensor.

The interactions of the Sun's energy with physical materials, both in the atmosphere and at the Earth's surface, cause this energy to be reflected, absorbed, transmitted or scattered. Electromagnetic energy travelling through the atmosphere is partly absorbed by molecules. The most efficient absorbers of solar radiation in the atmosphere are ozone (O_3), water vapour (H_2O) and carbon

dioxide (CO_2).

Atmospheric scattering occurs when the particles or gaseous molecules present in the atmosphere interact with the electromagnetic radiation and cause it to be redirected from its original path. Three types of scattering take place: Rayleigh scattering, Mie Scattering and Non-selective scattering.

When electromagnetic energy from the Sun hits the Earth's surface, three fundamental energy interactions are possible: absorption, transmission, and reflectance. Specular reflection occurs when a surface is smooth and all of the energy is directed away from the surface in a single direction. Diffuse reflection occurs when the surface is rough and the energy is reflected almost uniformly in all directions.

Questions

The following questions can help you to study Chapter 2.

1. What are advantages/disadvantages of aerial RS compared to spaceborne RS in terms of atmospheric disturbance?

2. How important are laboratory spectra in understanding the remote sensing images?

These are typical exam questions:

1. List the two models used to describe electromagnetic energy.

2. How are wavelength and frequency related to each other (give a formula)?

3. What is the electromagnetic spectrum?

4. List and define the three types of atmospheric scattering?

5. What specific energy interactions take place when EM energy from the Sun hits the Earth's surface?

6. In your own words give a definition of an atmospheric window.

7. Indicate True or False: Only the wavelength region outside the main absorption bands of the atmospheric gases can be used for remote sensing.

8. Indicate True or False: The amount of energy detected by a remote sensing sensor is a function of how energy is partitioned between its source and the materials with which it interacts on its way to the detector.

Chapter 3

Sensors and platforms

3.1 Introduction

In Chapter 2, the underlying principle of remote sensing was explained. Depending on the surface characteristics electromagnetic energy from the Sun or active sensor is reflected or energy may be emitted by the Earth itself. This energy is measured and recorded. The resulting data can be used to derive information about surface characteristics.

The measurements of electromagnetic energy are made by sensors that are attached to a static or moving platform. Different types of sensors have been developed for different applications (Section 3.2). Aircraft and satellites are generally used to carry one or more sensors (Section 3.3). General references with respect to missions and sensors are [13, 14]. Please refer to ITC's Aerospace Sensor Database for a complete and up-to-date overview.

The sensor-platform combination determines the characteristics of the resulting image data. For example, when a particular sensor is operated from a higher altitude the total area imaged is increased while the level of detail that can be observed is reduced (Section 3.4). Based on your information needs and on time and budgetary criteria, you can determine which image data are most appropriate (Section 3.5).

3.2 Sensors

A sensor is a device that measures and records electromagnetic energy. Sensors can be divided into two groups. *Passive sensors* depend on an external source of energy, usually the Sun (although sometimes the Earth itself). The group of passive sensors cover the electromagnetic spectrum in the range from less than 1 picometre (gamma rays) to over 1 metre (micro and radio waves). The oldest and most common type of passive sensor is the (photographic) camera.

Active sensors have their own source of energy. Measurements by active sensors are more controlled because they do not depend upon the (varying) illumination conditions. Active sensors include the laser altimeter (using infrared light) and radar. Figure 3.1 gives an overview of the types of the sensors that are introduced in this section. The camera, the multispectral scanner and the imaging radar are explained in more detail in Chapters 4, 5 and 6 respectively. For more information about spaceborne remote sensing you may refer to ITC's Aerospace Sensor Database.

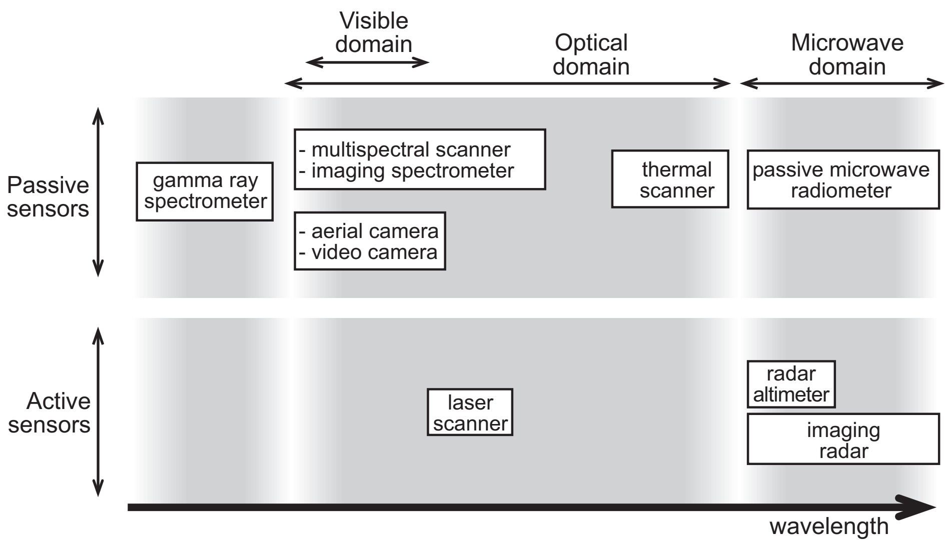


Figure 3.1: Overview of the sensors that are introduced in this chapter

3.2.1 Passive sensors

Gamma-ray spectrometer

The gamma-ray spectrometer measures the amount of gamma rays emitted by the upper soil or rock layers due to radioactive decay. The energy measured in specific wavelength bands provides information on the abundance of (radio isotopes that relate to) specific minerals. Therefore, the main application is found in mineral exploration. Gamma rays have a very short wavelength on the order of picometres (pm)). Because of large atmospheric absorption of these waves this type of energy can only be measured up to a few hundred metres above the Earth's surface. Example data acquired by this sensor are given in Figure 7.1.

Aerial camera

The camera system (lens and film) is mostly found in aircraft for aerial photography. Low orbiting satellites and NASA Space Shuttle missions also apply conventional camera techniques. The film types used in the camera enable electromagnetic energy in the range between 400 nm and 900 nm to be recorded. Aerial photographs are used in a wide range of applications. The rigid and regular geometry of aerial photographs in combination with the possibility to acquire stereo-photography has enabled the development of 'photogrammetric procedures' for obtaining precise 3D coordinates. Although aerial photos are used in many applications, principal applications include medium and large scale (topographic) mapping and cadastral mapping. Today, analogue photos are often scanned to be stored and processed in digital systems. Various examples of aerial photos are shown in Chapter 4.

Video camera

Video cameras are sometimes used to record image data. Most video sensors are only sensitive to the visible colours, although a few are able to record the near-infrared part of the spectrum (Figure 3.2). Until recently, only analogue video cameras were available. Today, digital video cameras are increasingly available, some of which are applied in remote sensing. Mostly, video images serve to provide low cost image data for qualitative purposes, for example, to provide additional visual information about an area captured with another sensor (e.g., laser scanner or radar).

Figure 3.2: Analogue false colour video image of 'De Lopikerwaard' (NL). Courtesy of Syntoptics

Multispectral scanner

The multispectral scanner is an instrument that mainly measures the reflected sunlight in the optical domain. A scanner systematically 'scans' the Earth's surface thereby measuring the energy reflected from the viewed area. This is done simultaneously for several wavelength bands, hence the name 'multispectral scanner'. A wavelength band is an interval of the electromagnetic spectrum for which the average reflected energy is measured. The reason for measuring a number of distinct wavelength bands is that each band is related to specific characteristics of the Earth's surface. For example, reflection characteristics of 'blue' light give information about the mineral composition; reflection characteristics of 'infrared light' tell something about the type and health of vegetation. The definition of the wavebands of a scanner, therefore, depends on the applications for which the sensor has been designed. An example of multispectral data for

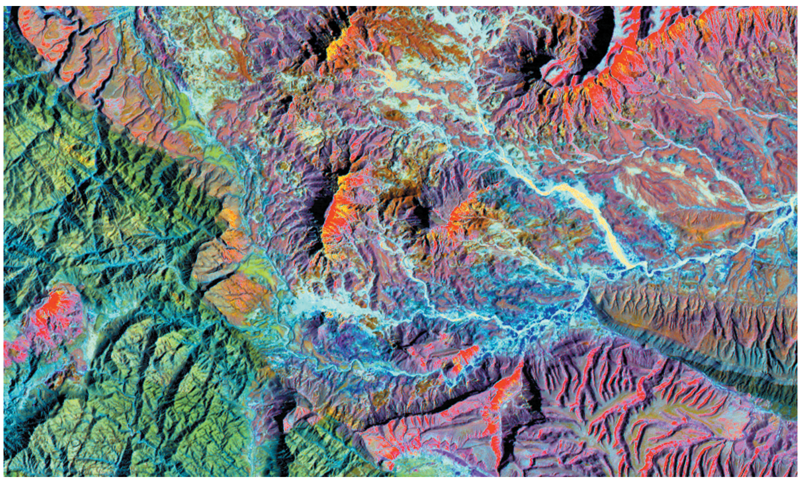


Figure 3.3: Landsat TM colour composite (RGB=457) of Yemen.

geological applications is given in Figure 3.3.

Imaging spectrometer

The principle of the imaging spectrometer is similar to that of the multispectral scanner, except that spectrometers measure only very narrow (5–10 nm) spectral bands. This results in an almost continuous reflectance curve per pixel rather than the values for relatively broad spectral bands. The spectral curves measured depend on the chemical composition of the material. Imaging spectrometer data, therefore, can be used to determine mineral composition of the surface

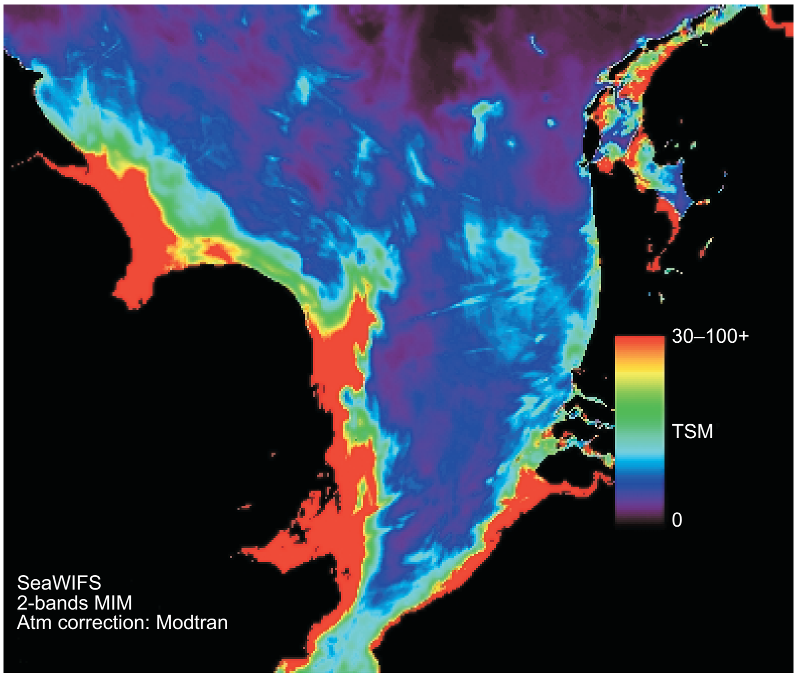


Figure 3.4: Total Suspended Matter concentration of the North Sea derived from SeaWIFS (OrbView2) data. Courtesy of CCZM, Rijkswaterstaat

or the chlorophyll content of the surface water (Figure 3.4).

Thermal scanner

Thermal scanners measure thermal data in the range of 10–14 μm. Wavelengths in this range are directly related to an objects temperature. Data on cloud, land and sea surface temperature are extremely useful for weather forecasting. For this reason, most remote sensing systems designed for meteorology include a thermal scanner. Thermal scanners can also be used to study the effects of drought ('water stress') on agricultural crops, and to monitor the temperature of cooling water discharged from thermal power plants. Another application is in the detection of coal fires (Figure 3.5).

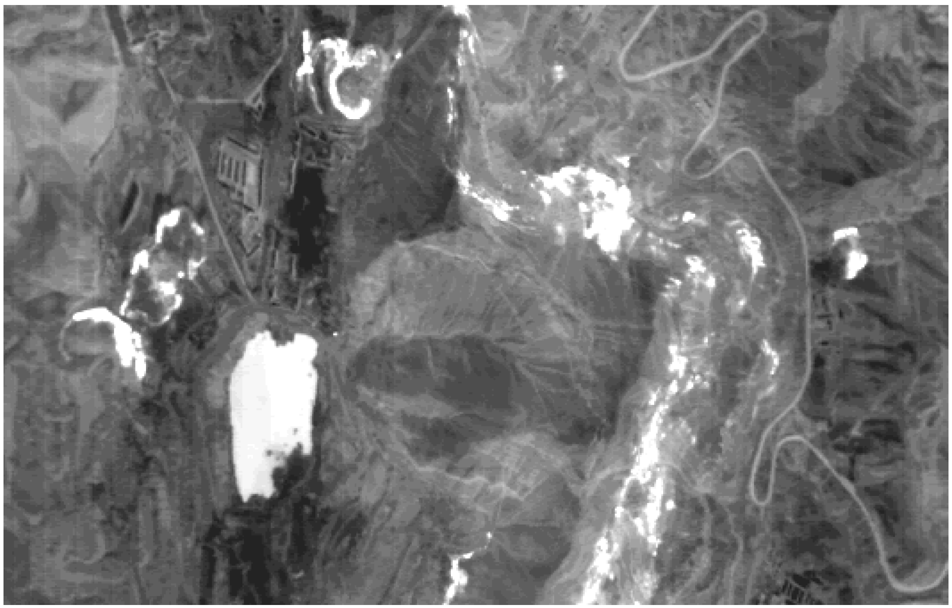


Figure 3.5: Night-time airborne thermal scanner image of a coal mining area. Dark tones represent the relatively cold surfaces, whilst light tones represent the relatively warm spots. Most of the warm spots are due to underground coal fires apart from the largest light patch which is a lake.

Radiometer

EM energy with very long wavelengths (1–100 cm) is emitted from the soil and rocks on, or just below, the Earths surface. The depth from which this energy is emitted depends on the properties, such as water content, of the specific material. Radiometers are used to detect this energy. The resulting data can be used in mineral exploration, soil mapping and soil moisture estimation (Figure 3.6).

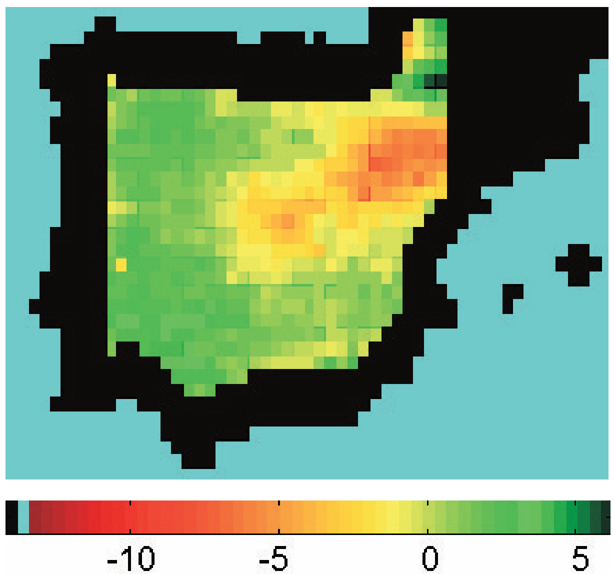


Figure 3.6: Map of the Iberian Peninsula showing the long-term trend in soil moisture change (vol.%) over a 10-year period. Map is based on Nimbus/SMMR observations. Courtesy of Free University Amsterdam / NASA

3.2.2 Active sensors

Laser scanner

Laser scanners are mounted on aircraft and use a laser beam (infrared light) to measure the distance from the aircraft to points located on the ground. This distance measurement is then combined with exact information on the aircraft's position to calculate the terrain elevation. Laser scanning is mainly used to produce detailed, high-resolution, Digital Terrain Models (DTM) for topographic mapping (Figure 3.7). Laser scanning is increasingly used for other purposes, such as the production of detailed 3D models of city buildings and for measuring tree heights in forestry.

Figure 3.7: Digital Terrain Model (5 m grid) of the St. Pietersgroeve (NL). Courtesy Survey Department, Rijkswaterstaat

Radar altimeter

Radar altimeters are used to measure the topographic profile parallel to the satellite orbit. They provide profiles (single lines of measurements) rather than 'image' data. Radar altimeters operate in the 1–6 cm domain and are able to determine height with a precision of 2–4 cm. Radar altimeters are useful for measuring relatively smooth surfaces such as oceans and for 'small scale' mapping of continental terrain models. Sample results of radar altimeter measurements are given in Figure 1.6 and Figure 7.2.

Imaging radar

Radar instruments operate in the 1–100 cm domain. As in multispectral scanning, different wavelength bands are related to particular characteristics of the Earth's surface. The radar backscatter (Figure 3.8) is influenced by the illuminating signal (microwave parameters) and the illuminated surface characteristics (orientation, roughness, di-electric constant/moisture content). Since radar is an active sensor system and the applied wavelengths are able to penetrate clouds, it has 'all-weather day-and-night' acquisition capability. The combination of two radar images of the same area can provide information about terrain heights. Combining two radar images acquired at different moments can be used to precisely assess changes in height or vertical deformations (SAR Interferometry).

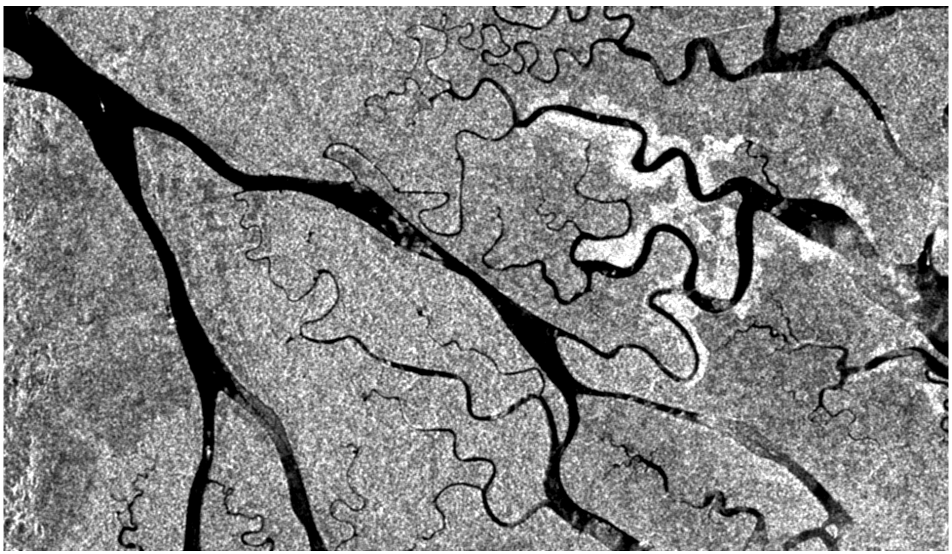


Figure 3.8: ERS SAR image of a delta in Kalimantan, Indonesia. The image allows three different forest types to be distinguished.

3.3 Platforms

In remote sensing, the sensor is mounted on a platform. In general, remote sensing sensors are attached to moving platforms such as aircraft and satellites. Static platforms are occasionally used in an experimental context. For example, by using a multispectral sensor mounted to a pole, the changing reflectance characteristics of a specific crop during the day or season can be assessed.

Airborne observations are carried out using aircraft with specific modifications to carry sensors. An aircraft that carries an aerial camera or a scanner needs a hole in the floor of the aircraft. Sometimes Ultra Light Vehicles (ULVs), balloons, Airship or kites are used for airborne remote sensing. Airborne observations are possible from 100 m up to 30–40 km height. Until recently, the navigation of an aircraft was one of the most difficult and crucial parts of airborne remote sensing. In recent years, the availability of satellite navigation technology has significantly improved the quality of flight execution.

For *spaceborne* remote sensing, satellites are used. Satellites are launched into space with rockets. Satellites for *Earth Observation* are positioned in orbits between 150–36,000 km altitude. The specific orbit depends on the objectives of the mission, e.g., continuous observation of large areas or detailed observation of smaller areas.

3.3.1 Airborne remote sensing

Airborne remote sensing is carried out using different types of aircraft depending on the operational requirements and budget available. The speed of the aircraft can vary between 140–600 km/hour and is, among others, related to the mounted sensor system. Apart from the altitude, also the aircraft orientation affects the geometric characteristics of the remote sensing data acquired. The orientation of the aircraft is influenced by wind conditions and can be corrected for to some extent by the pilot. Three different aircraft rotations relative to a reference path are possible: roll, pitch and yaw (Figure 3.9). An Inertial Measurement Unit (IMU) can be installed in the aircraft to measure these rotations. Subsequently the measurements can be used to correct the sensor data for the resulting geometric distortions.

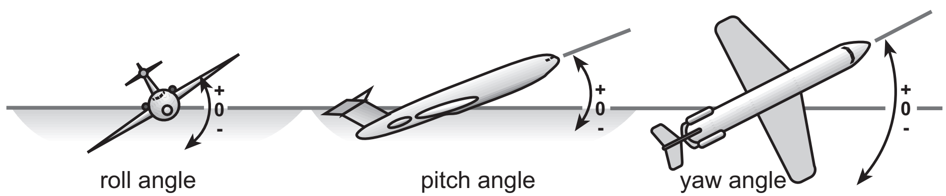


Figure 3.9: The three angles (roll, pitch and yaw) of an aircraft that influence image data acquired.

Today, most aircraft are equipped with satellite navigation technology, which yield the approximate position (RMS-error of less than 30 m). More precise positioning and navigation (up to decimetre accuracy) is possible using so-called 'differential approaches'. In this textbook we refer to satellite navigation in general, which comprises the American GPS system, the Russian Glonass system and the proposed European Galileo system.

In aerial photography the 'measurements' are 'stored' on hard-copy material: the negative film. For other sensors, e.g., a scanner, the digital data can be stored on tape or mass memory devices. Tape recorders offer the fastest way to store the vast amount of data. The recorded data are only available after the aircraft has returned to its base.

Owning, operating and maintaining survey aircraft, as well as employing a professional flight crew is an expensive undertaking. In the past, survey aircraft were owned mainly by large national survey organizations that required large amounts of photography. There is an increasing trend towards contracting specialized private aerial survey companies. Still, this requires basic understanding of the process involved. A sample contract for outsourcing aerial photography is provided by the American Society of Photogrammetry and Remote Sensing at their ASPRS web site.

3.3.2 Spaceborne remote sensing

Spaceborne remote sensing is carried out using sensors that are mounted on satellites. The monitoring capabilities of the sensor are to a large extent determined by the parameters of the satellite's orbit. Different types of orbits are required to achieve continuous monitoring (meteorology), global mapping (land cover mapping), or selective imaging (urban areas). For remote sensing purposes, the following orbit characteristics are relevant :

- *altitude*, which is the distance (in km) from the satellite to the mean surface level of the Earth. Typically, remote sensing satellites orbit either at 600–800 km (polar orbit) or at 36,000 km (geo-stationary orbit) distance from the Earth. The distance influences to a large extent which area is viewed and at which detail.

- *inclination angle*, which is the angle (in degrees) between the orbit and the equator. The inclination angle of the orbit determines, together with field of view of the sensor, which latitudes can be observed. If the inclination is 60° then the satellite flies over the Earth between the latitudes 60° South and 60° North; it cannot observe parts of the Earth at latitudes above 60°.

- *period*, which is the time (in minutes) required to complete one full orbit. A polar satellite orbits at 800 km altitude and has a period of 90 minutes. A ground speed of 28,000 km/hour is almost 8 km/s. Compare this figure with the speed of an aircraft, which is around 400 km/hour. The speed of the platform has implications for the type of images that can be acquired (time for 'exposure').

- *repeat cycle*, which is the time (in days) between two successive identical orbits. The revisit time, the time between two subsequent images of the same area, is determined by the repeat cycle together with the pointing capability of the sensor. Pointing capability refers to the possibility of the sensor-platform to 'look' sideways. Pushbroom scanners, such as those mounted on SPOT, IRS and IKONOS (Section 5.4), have this possibility.

The following orbit types are most common for remote sensing missions:

Polar, or near polar, orbit. These are orbits with inclination angle between 80 and 100 degrees and enable observation of the whole globe. The satellite is typically placed in orbit at 600–800 km altitude.

Sun-synchronous orbit. An orbit chosen in such a way that the satellite always passes overhead at the same local solar time is called sun-synchronous. Most sun-synchronous orbits cross the equator at mid-morning (around 10:30 h). At that moment the Sun angle is low and the resultant shadows reveal terrain relief. Sun-synchronous orbits allow a satellite to record images at two fixed times during one 24-hour period: one during the day and one at night. Examples of near polar sun-synchronous satellites are Landsat, SPOT and IRS.

Geostationary orbit. This refers to orbits in which the satellite is placed above the equator (inclination angle is 0°) at a distance of some 36,000 km. At this distance, the period of the satellite is equal to the period of the Earth. The result is that the satellite is at a fixed position relative to the Earth. Geostationary orbits are used for meteorological and telecommunication satellites.

Today's meteorological weather satellite systems use a combination of geostationary satellites and polar orbiters. The geo-stationary satellites offer a continuous view, while the polar orbiters offer a higher resolution (Figure 3.10).

The data of spaceborne sensors need to be sent to the ground for further analysis and processing. Some older spaceborne systems utilized film cartridges that fell back to a designated area on Earth. In the meantime, practically all Earth Observation satellites apply satellite communication technology for *downlink* of the

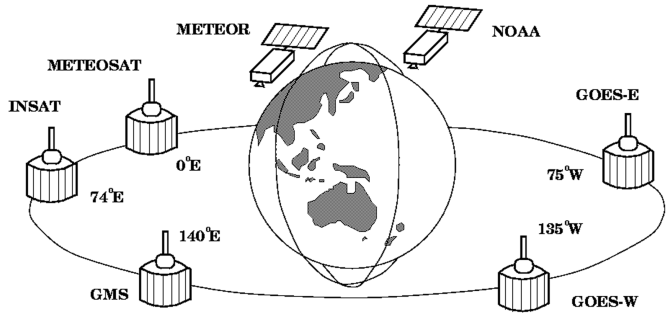


Figure 3.10: Meteorological observation system comprised of geo-stationary and polar satellites.

data. The acquired data are sent down to a receiving station or to another communication satellite that downlink the data to receiving antennæ on the ground. If the satellite is outside the range of a receiving station the data can be temporarily stored by a tape recorder in the satellite and transmitted later. One of the trends is that small receiving units (consisting of a small dish with a PC) are being developed for local reception of image data.

3.4 Image data characteristics

Remote sensing image data are more than a picture—they are measurements of EM energy. Image data are stored in a regular grid format (rows and columns). The single elements are called *pixels*, an abbreviation of 'picture elements'. For each pixel, the measurements are stored as Digital Number values or DN-values. Typically, for each wavelength band measured, a separate 'layer' is stored (Figure 3.11).

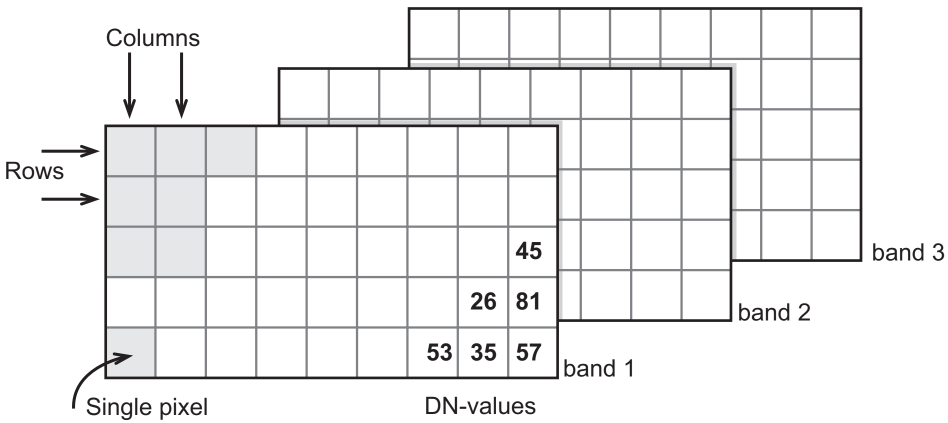


Figure 3.11: An image file comprises a number of bands. For each band the Digital Number (DN) values, corresponding to the measurements, are stored in a row-column system.

The 'quality' of image data is primarily determined by characteristics of the sensor-platform system. These sensor-platform characteristics are usually referred to as:

1. *Spectral resolution* and *radiometric resolution*, referring respectively to the part of the Electro Magnetic spectrum measured and the differences in energy that can be observed.

2. *Spatial resolution*, referring to the smallest unit-area measured, it indicates the minimum size of objects that can be detected.

3. *Revisit time*, the time between two successive image acquisitions over the same location on Earth.

The resolutions need a more complicated, sensor specific, explanation. These are explained in the respective chapters (camera, multispectral scanner, radar). The image data produced by the sensor-platform system are related to the above mentioned resolutions. Characteristics of *image data* are:

1. *Image size*: the number of rows and columns in a scene.

2. *Number of bands*: the number of wavelengths band stored. For example, 1 (black/white photograph), 4 (SPOT multispectral image), or 256 (imaging spectroscopy data).

3. *Quantization*: the data format used to stored the energy measurements. Typically, for each measurement one (8 bits) byte is used, which represents the (discrete) values of 0–255. These values are known as Digital Numbers. Using sensor specific calibration parameters, the DN-values can be converted into measured energy (Watt).

4. *Ground pixel size*: the area coverage of a pixel on the ground. Usually this is a round figure (e.g., 20 m or 30 m). The ground pixel size of image data is related to, but not necessarily the same as the spatial resolution.

The image size, the number of bands and the quantification, allow the disk space required to be calculated. For example, a SPOT multispectral image requires: 3000 (columns) $\times$ 3000 (rows) $\times$ 4 (bands) $\times$ 1 byte = 36 Mbyte of storage.

3.5 Data selection criteria

Spatio-temporal characteristics

For the selection of the appropriate data type it is necessary to fully understand the information requirements for a specific application. Therefore, you have to analyse the spatio-temporal characteristics of the phenomena. It is obvious that you require a different type of image data for monitoring fast urban growth in a small area than for studying a slow desertification process in a large area.

In the case of urban area mapping much spatial detail (less than 0.5 m) is required. Aerial photography and airborne digital scanners can fulfill this requirement. Another consideration in your data selection process is the third dimension (height or elevation component). Stereo images or interferometric radar data can provide 3D information. The moment of image acquisition should also be given a thought. A low Sun angle causes long shadows due to elevated buildings. To avoid shadows, images should be taken around noon. Also the the type of cloud cover plays a role; under certain conditions no photos can be taken at all. A last issue is seasonal cycles: in countries with a temperate climate the trees have no leafs in autumn, winter and spring and therefore allow a clearer view of the infrastructure.

In the case of monitoring of desertification or studying the El Niño effect the temporal aspect is of prime importance and can be translated into conditions for data continuity and data quality. Optimally, image data of similar quality (spectral, radiometric and spatial resolution) should be available for long periods.

Availability of image data

Once the image data requirements have been determined, you have to investigate their availability and costs. The availability depends on the already acquired data stored in archives or data that need to be acquired at your request. The size and accessibility of image archives is growing at a fast rate. If up-to-date data are required, these need to be requested through an aerial survey company or from a remote sensing data distributor. Current operational spaceborne missions (July 2001) can be listed as follows :

- seven panchromatic missions with a spatial resolution between 1 and 10 m (3 SPOT, 2 IRS, 1 IKONOS,1 EO-1); at *ad hoc* basis there is SPIN-2 mission.
- ten multispectral missions with a spatial resolution between 4 and 30 m (2 Landsat, 3 SPOT, 2 IRS, 1 IKONOS, 1 Terra, 1 EO-1)
- two imaging radar missions with a spatial resolution between 20 and 100 m (1 ERS, 1 Radarsat)

Today, some 1400 aerial survey cameras are available and used to acquire (vertical) aerial photography. Per year, an estimated 30 aerial survey cameras are sold. Worldwide, a growing number of airborne laser scanners (more than 50) are available, with the majority being flown in North America. Also a growing number of airborne imaging spectrometers (more than 30) are available for data acquisition—these instruments are mainly owned by mining companies. The number of operational airborne radar systems is small: many experimental systems exist, but there are less than four commercially operated systems.

Costs of image data

It is rather difficult to provide indications about the costs of image data. Costs of different types of image data can only be compared when calculated for a specific project with specific data requirements. Existing data from the archive are cheaper than data that have to be specially ordered. Another reason to be cautious in giving prices is that different qualities of image data (processing level) exist.

The costs of vertical aerial photographs depend on the size of the area, the photo scale, the type of film and processing used and the availability of aerial reconnaissance companies. Under 'European' conditions the costs of the aerial photography is somewhere between 5–20 Euro/km^2. The costs of optical satellite data vary from free ('public domain') to 45 Euro/km^2. Low resolution data (NOAA/AVHRR) can be downloaded for free from the Internet. Medium resolution data (Landsat, SPOT, IRS) cost in the range of 0.01–0.70 Euro/km^2. High resolution satellite data (IKONOS, SPIN-2) cost between 15–45 Euro/km^2. For IKONOS derived information products prices can go up to 150-Euro/km^2.

Summary

This chapter has introduced the principle of remote sensing observations: a sensor attached to a moving platform. Aircraft and satellites are the main platforms used in remote sensing. Both types of platforms have their advantages and disadvantages. Two main categories of sensors are distinguished. Passive sensors depend on an external source of energy such as the Sun. Active sensors have their own source of energy. A sensor carries out measurements of reflected or emitted (EM) energy. The energy measured in specific wavelength bands is related to (Earth) surface characteristics. The measurements are usually stored as image data. The characteristics of image data are related to the 'resolutions' of the sensor-platform system (spatial, spectral and radiometric). Depending on the spatio-temporal phenomena of interest, the most appropriate remote sensing data can be determined. In practice, data availability and costs determine which remote sensing data are used.

Questions

The following questions can help you to study Chapter 3.

1. Think of an application, define the spatio-temporal characteristics of interest and determine the type of remote sensing image data required.

2. How many of the sensor types introduced in Section 3.2 were already known to you?

3. Which types of sensors are used in your discipline or field-of-interest?

4. Which aspects need to be considered to assess if the statement “RS data acquisition is a cost-effective method” is to be true?

The following are sample exam questions:

1. Explain the sensor-platform concept.

2. Mention two types of passive and two types of active sensor.

3. What is a typical application of a multispectral satellite image, and, what is a typical application of a very high spatial resolution satellite image?

4. Describe two differences between aircraft and satellite remote sensing and their implications for the data acquired.

5. Which two types of satellite orbits are mainly used for Earth observation?

6. List and describe four characteristics of image data.

Chapter 4

Aerial cameras

4.1 Introduction

Aerial photography has been used since the early 20th century to provide spatial data for a wide range of applications. It is the oldest, yet most commonly applied remote sensing technique. Photogrammetry is the science and technique of making measurements from photos or image data. Nowadays, almost all topographic maps are based on aerial photographs. Aerial photographs also provide the accurate data required for many cadastral surveys and civil engineering projects. Aerial photography is a useful source of information for specialists such as foresters, geologists and urban planners. General references for aerial photography are [10, 16, 17].

Two broad categories of aerial photography can be distinguished: *vertical photography* and *oblique photography* (Figure 4.1). In most mapping applications, vertical aerial photography is required. Vertical aerial photography is produced with a camera mounted in the floor of an aircraft. The resulting image is rather similar to a map and has a scale that is approximately constant throughout the image area. Usually, vertical aerial photography is also taken in stereo, in which successive photos have a degree of overlap to enable stereo-interpretation and stereo-measurements.

Oblique photographs are obtained when the axis of the camera is not vertical. Oblique photographs can be made using a hand-held camera and shooting through the (open) window of an aircraft. The scale of an oblique photo varies from the foreground to the background. This scale variation complicates the measurement of positions from the image and, for this reason, oblique photographs are rarely used for mapping purposes. Nevertheless, oblique images can be useful for purposes such as viewing sides of buildings and for inventories of wildlife.

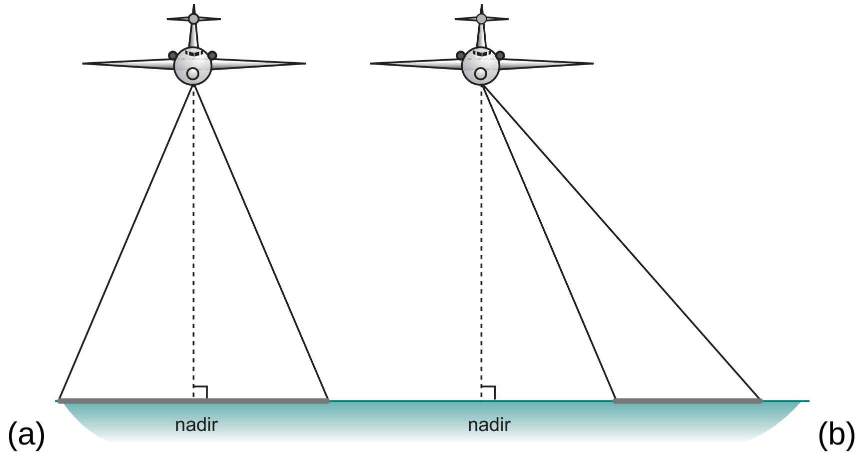
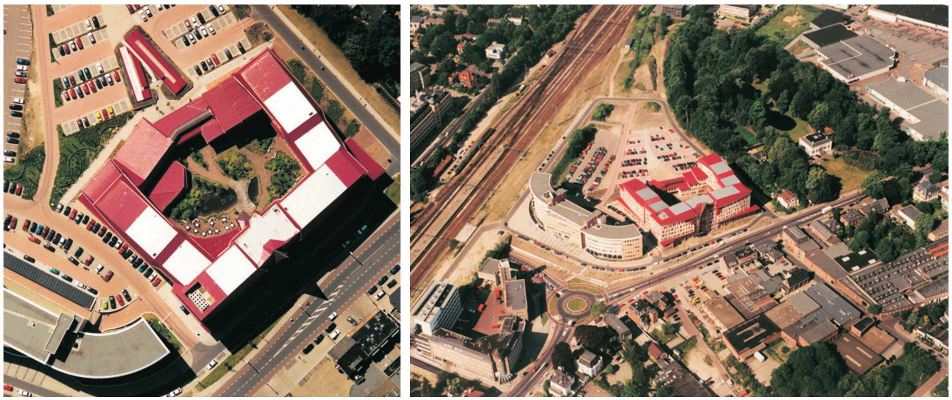


Figure 4.1: Vertical (a) and oblique (b) photography

This chapter focusses on the camera, films and methods used for vertical aerial photography. First of all, Section 4.2 introduces the aerial camera and its main components. Photography is based on exposure of a film, processing and printing. The type of film applied largely determines the spectral and radiometric characteristics of the printed products (Section 4.3). Section 4.4 focusses on the geometric characteristics of aerial photography. In Section 4.5 some aspects of aerial photography missions are introduced. In the advanced Section 4.6, some technological developments are discussed.



(a) (b)

Figure 4.2: Vertical (a) and oblique (b) aerial photo of the ITC building. Photos by Paul Hofstee, 1999.

4.2 Aerial camera

A camera used for vertical aerial photography for mapping purposes is called an aerial survey camera. In this section the standard aerial camera is introduced. At present, there are only two major manufacturers of aerial survey cameras, namely Zeiss and Leica. These two companies produce the RMK-TOP and the RC-30 respectively. Just like a typical hand-held camera, the aerial survey camera contains a number of common components as well as a number of specialized ones necessary for its specific role. Figure 4.3 shows a schematic drawing of an aerial camera. The large size of the camera results from the need to acquire images of large areas with a high spatial resolution. This is realized by using a large film size. Modern aerial survey cameras produce negatives measuring 23 cm $\times$ 23 cm. Up to 600 photographs may be recorded on a single roll of film.

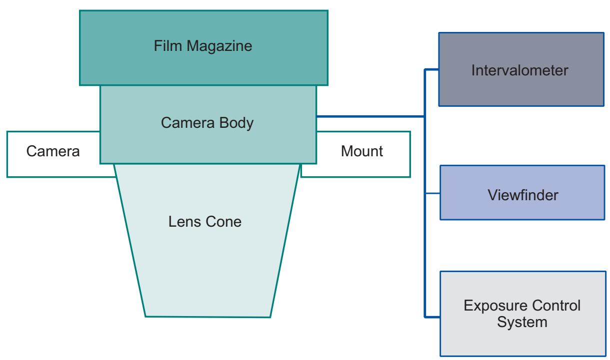


Figure 4.3: Major components of an aerial camera

4.2.1 Lens cone

Perhaps the most important (and expensive) single component within the camera is the lens cone. This is interchangeable, and the manufactures produce a range of cones, each of different focal length. *Focal length* is the most important property of a lens cone since, together with flying height, it determines the photo scale (Section 4.4.1). The focal length also determines the angle of view of the camera. The longer the focal length, the narrower the angle of view. Lenses are usually available in the following standard focal lengths, ranging from narrow angle (610 mm), to normal angle (305 mm) to wide angle (210 mm, 152 mm) to super-wide angle (88 mm). The 152 mm lens is the most commonly used lens.

The lens cone is responsible for 'projecting' an optical image onto the film. The accuracy of this projection depends on the quality of the lens. Even with high quality lenses some distortions still take place. These distortions are imperceptible to the human eye, but aversely affect the photogrammetric operations where very precise measurements (at μm level) are required. Therefore, the distortion of lenses is measured on a regular basis and reported in a 'calibration report'. This report is required in many photogrammetric processes.

In the lens cone (Figure 4.4), the diaphragm and shutter respectively control the intensity and the duration of the exposure reaching the film. Typically, *optical filters* are placed over the lens to control two important aspects of the image quality.

- Image contrast. A yellow filter is often used (in black and white photography) to absorb the ultra-violet and blue wavelengths, which are the most highly scattered within the atmosphere (Section 2.3.2). This scattering effect, if not corrected, normally leads to a reduction in image contrast.

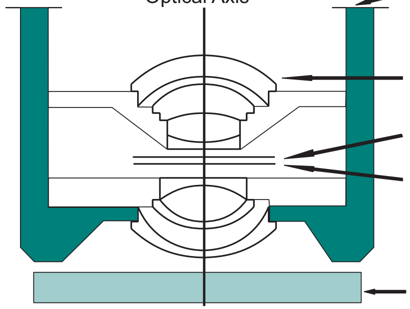


Figure 4.4: The lens cone comprises a lens system, focal plane frame, shutter, diaphragm, anti-vignetting and coloured filters.

- Evenness of illumination. Any image formed by a lens is brightest at the centre and darkest in the corners. This is known as light fall-off or vignetting, and is related to the angle a light ray makes with the optical axis, the greater the angle the darker the image. In wide angle lenses, this effect can be severe and images appear dark in the corners. The effect can be partially corrected by an anti-vignetting filter, which is a glass plate in which the centre transmits less light than the corners. In this way the image illumination is made more even over the whole image area, and the dark corners are avoided.

4.2.2 Film magazine and auxiliary data

The aerial camera is fitted with a system to record various items relevant information onto the side of the negative: mission identifier, date and time, flying height and the frame number (Figure 4.5).

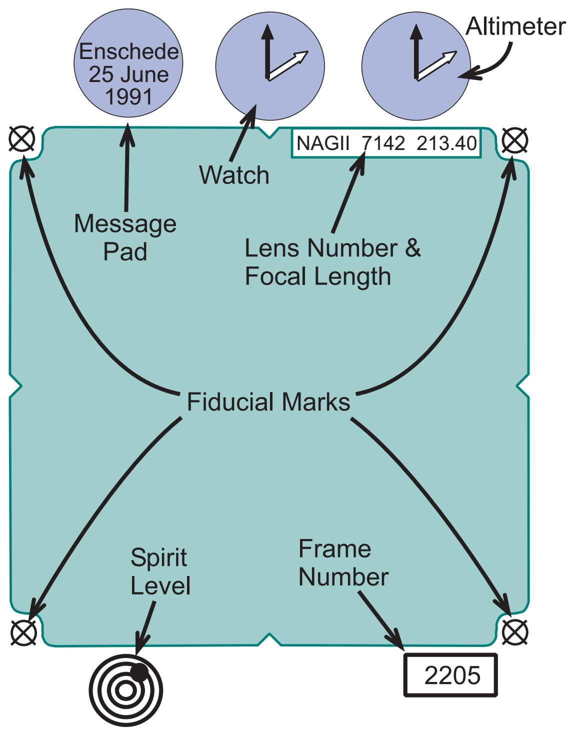


Figure 4.5: Auxiliary data annotation on an aerial photograph

A vacuum plate is used for flattening the film at the instant of exposure. So-called *fiducial marks* are recorded in all corners of the film. The fiducial marks are required to determine the optical centre of the photo needed to align photos for stereoviewing. The fiducials are also used to record the precise position of the film in relation to the optical system, which is required in photogrammetric processes (interior orientation, Section 9.4.3). Modern cameras also have an image motion compensation facility (Section 4.6).

4.2.3 Camera mounting

The camera mounting enables the camera to be levelled by the operator. This is usually performed at the start of a mission once the aircraft is in a stable configuration. Formal specifications for aerial survey photography usually require that the majority of images are maintained within a few degrees (less than 3°) of true vertical. Errors exceeding this cause difficulties in photogrammetric processing (Section 9.4.3).

4.3 Spectral and radiometric characteristics

Photographic recording is a multi-stage process that involves film exposure and chemical processing ('development'). It is usually followed by printing.

Photographic film comprises a light sensitive emulsion layer coated onto a base material (Figure 4.6). The emulsion layer contains silver halide crystals, or 'grains', suspended in gelatine. The emulsion is supported on a stable polyester base. Light changes the silver halide into silver metal, which, after processing of the film, appears black. The exposed film, before processing, contains a *latent image*.

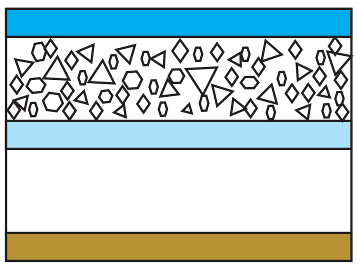


Figure 4.6: Section of a photographic film showing the emulsion layer with the silver halide crystals

The film emulsion type applied determines the spectral and radiometric characteristics of the photograph. Two terms are important in this context:

- *Spectral sensitivity* describes the range of wavelengths to which the emulsion is sensitive. For the study of vegetation the near-infrared wavelengths yield much information and should be recorded; for other purposes a standard colour photograph normally can yield the optimal basis for interpretation.

- *General sensitivity* is a measure of how much light energy is required to bring about a certain change in film density. Given specific illumination conditions, the general sensitivity of a film can be selected, for example, to minimize exposure time.

For reasons of understanding, first the general sensitivity is explained, followed by an explanation of the spectral sensitivity. Subsequently black-white, colour and colour infrared photography are explained. The last section presents some remarks about the scanning of photos.

4.3.1 General sensitivity

The energy of a light photon is inversely proportional to the light wavelength. In the visible range, therefore, the blue light has the highest energy. For a normal silver halide grain, only blue light photons have sufficient energy to form the latent image, and hence a raw emulsion is only sensitive to blue light.

The sensitivity of a film can be increased by increasing the mean size of the silver grains: larger grains produce more metallic silver per input light photon. The mean grain size of aerial films is in the order of a few μm. There is a problem related to increasing the grain size: larger grains are unable to record small details, i.e., the spatial resolution is decreased (Section 4.4.2). The other technique to improve the general sensitivity of a film is to perform a sensitization of the emulsion by adding small quantities of chemicals, such as gold or sulphur.

General sensitivity is often referred to as *film speed*. For a scene of given average brightness, the higher the film speed, the shorter the exposure time required to record the optical image on the film. Similarly, the higher the film speed, the less bright an object needs to be in order to be recorded upon the film.

4.3.2 Spectral sensitivity

Sensitization techniques are used not only to increase the general sensitivity but also to produce films that are sensitive to longer wavelengths. By adding sensitizing dyes to the basic silver halide emulsion, the energy of longer light wavelengths becomes sufficient to produce latent images. In this way a monochrome film can be made sensitive to green, red or infrared wavelengths (Section 3.3.2).

A black–and–white (monochrome) type of film has one emulsion layer. Using sensitization techniques, different types of monochrome films are available. Most common are the panchromatic and infra-red sensitive film. The sensitivity curves of these films are shown in Figure 4.7.

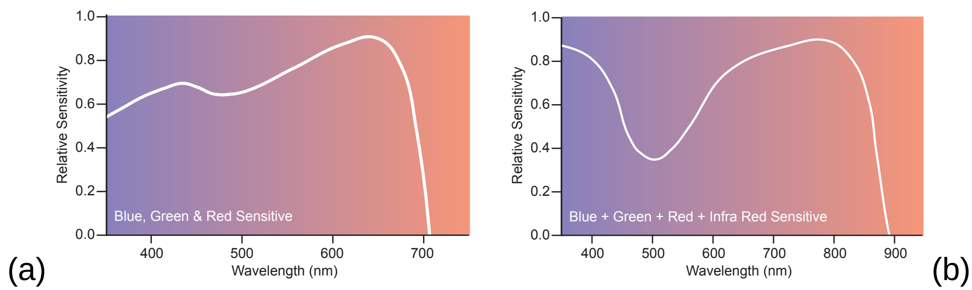


Figure 4.7: Spectral sensitivity curves of a panchromatic film (a) and a black/white infrared film (b). Note the difference in scaling on the x-axis

Colour photography uses an emulsion with three sensitive layers to record three wavelength bands corresponding to the three primary colours of the spectrum, i.e., blue, green and red. There are two types of colour photography: true colour and false colour infrared.

4.3.3 Monochrome photography

Photography is a multi-stage process involving exposure, chemical processing and, finally, printing.

During exposure in the camera an optical image of the original scene is projected by the lens onto the film. This image is effectively a two-dimensional pattern of light with varying intensity. The illumination of the film causes photons to be absorbed by the crystals of the emulsion. This energy absorption causes a number of the silver ions close to the surface of the crystal to be converted into atoms of silver metal. The greater the number of absorbed photons, the greater the number of silver atoms that are created. However, in the brief period of the exposure, insufficient atoms of silver are created to cause a visible change in the appearance of the crystals. This process is known as the formation of the 'latent image'.

When the film is later developed, in crystals containing the latent image, a much greater quantity of metallic silver is formed by the development process. Consequently, these areas of the film appear much darker than those where no exposure was received. Undeveloped silver halide is next dissolved by a 'fixing' solution, and is then washed out of the emulsion layer. The resulting photographic image is referred to as a negative because bright areas in the terrain appear dark on the film and *vice versa*.

In order to obtain a positive image, the negative must be photographically printed. This may be done by either a 'contact' process or by optical enlarging. The photographic printing material contains a similar emulsion layer to the film, but this is usually coated onto a paper base. In the printing process, the negative is illuminated by a light source and, depending upon the amount of silver metal in each area, either more or less light is allowed to pass through. In the dark areas of the negative, (corresponding to bright parts of the subject), little light passes

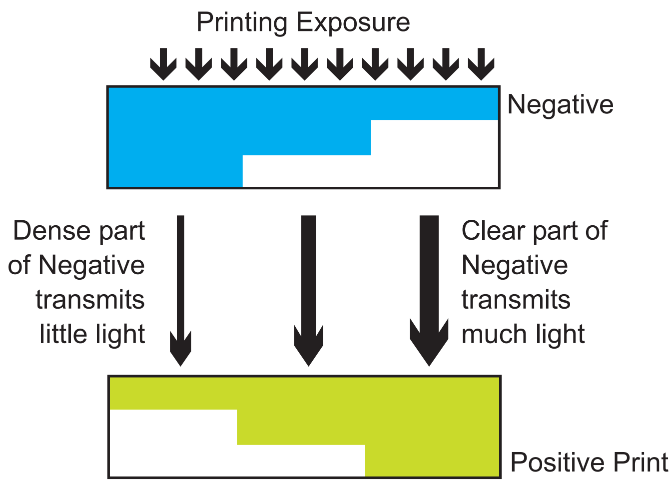


Figure 4.8: Negative film and positive print. Bright areas (white) on the negative result in dark areas (green) on the positive.

and forms only a small amount of silver in the positive image. Conversely, in clear areas of the negative, much light passes and forms large quantities of silver in the print (see Figure 4.8).

4.3.4 True colour and colour infrared photography

Colour photography uses the *subtractive principle of colours*, which is explained in Section 10.2.2. According to this principle, an original colour is reproduced by combinations of the three secondary colours: yellow, magenta and cyan.

A *true colour* film is made sensitive to the primary colours by coating three separate colour sensitive layers (blue, green, red) on a common base. Each layer thus records only one of the primary colours and produces quantities of complimentary coloured dye. The upper layer is sensitive to blue light only and contains a yellow dye forming coupler. A yellow filter layer absorbs any remaining blue light and prevents it from exposing lower layers. The middle layer is sensitive to green light and contains a magenta dye forming coupler. Finally, the lowest layer is sensitive to red and contains a cyan dye forming coupler (Figure 4.9).

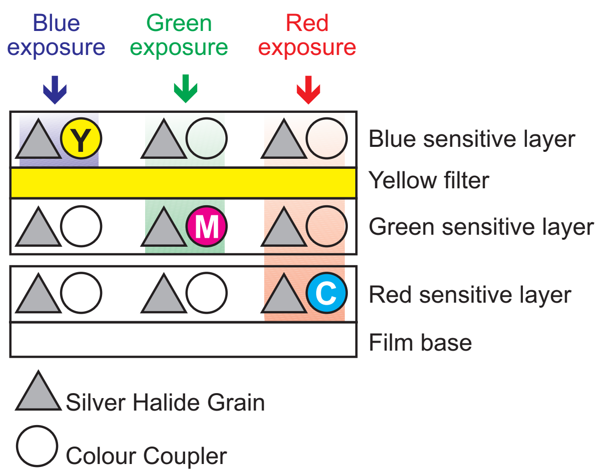


Figure 4.9: Emulsion layers of a true colour film

In the *development process*, the developing agent reduces the exposed silver halide to silver metal and the coloured dyes are produced. During development, subtractive dyes are produced in each of the three layers. These layers produce a quantity of the complimentary coloured dye in each layer that is proportional to the brightness of the original primary colour. This means that exposure of a high amount of green light produces much magenta dye in the film.

Colour negative *printing* uses paper that has a similar emulsion structure to that of colour negative film, and is processed in a similar manner. For example, a red colour in the scene produces cyan dye in the negative. The cyan colour in the negative, when illuminated with white light, absorbs red light but transmits blue and green light. This results in the formation of magenta and yellow dyes in the print, which then appears red. Bleaching and fixing are performed to remove all metallic silver and any remaining silver halide (Figure 4.10).

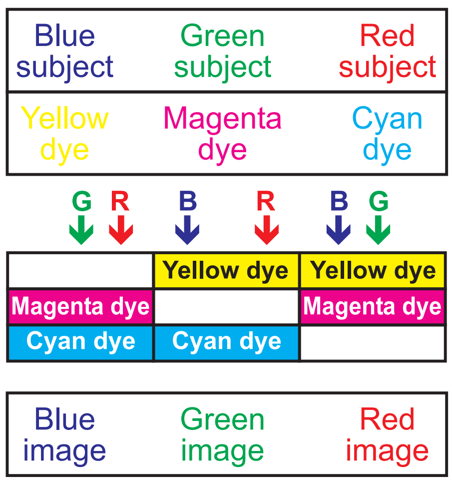


Figure 4.10: True world, colour negative and corresponding colour positive image

The film used in a *False Colour Infrared* film comprises three layers: a green sensitive layer containing a yellow dye forming coupler; a red sensitive layer containing a magenta dye forming coupler; a near-infrared sensitive layer containing a cyan dye forming coupler. False colour infrared film, is sensitive to the near-infrared wavelengths and is therefore particularly useful for recording information about vegetation. While true colour is used mostly for topographic and urban mapping, false colour photography is mostly used for mapping of (semi-) natural vegetation and forest types.

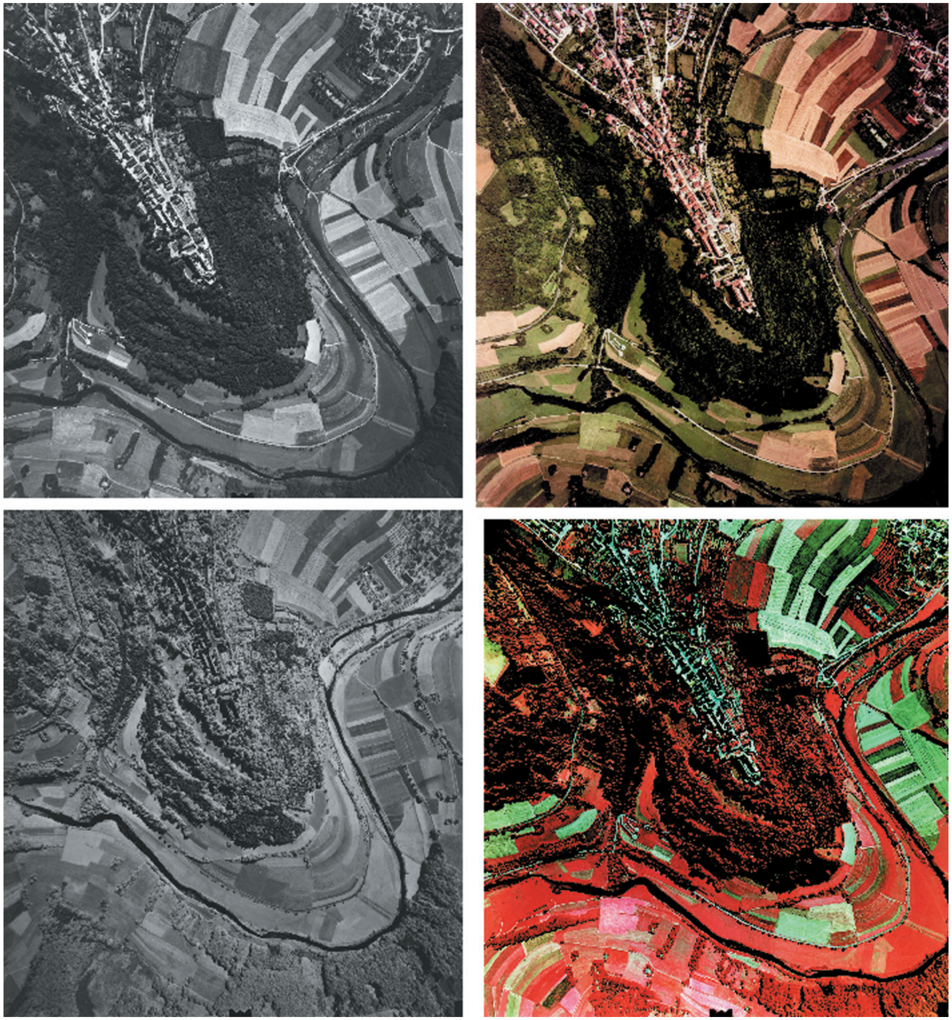


Figure 4.11: Comparison of panchromatic (upper left), black-and-white infrared (lower left), true colour (upper right) and false colour infrared (lower right) photographs of the same area.

4.3.5 Scanning

Classical photogrammetric techniques as well as visual photo-interpretation generally employ hard-copy photographic images. These can be the original negatives, positive prints or diapositives. Digital photogrammetric systems, as well as geographic information systems, require digital photographic images. A scanner is used to convert a film or print into a digital form. The scanner samples the image with an optical detector and measures the brightness for small areas (pixels). The brightness value are then represented as a digital number (DN) on a given scale. In the case of a monochrome image, a single measurement is made for each pixel area. In the case of a coloured image, separate red, green and blue values are measured. For simple visualization purposes, a standard office scanner can be used; but high metric quality scanners are required if the digital photos are to be used in precise photogrammetric procedures.

In the scanning process is the setting of the size of the scanning aperture is most relevant. This is also referred to as the *scanning density* and is expressed in dots per inch (dpi; 1 inch = 2.54 cm). The dpi-setting depends on the detail required for the application and is usually limited by the scanner. Office scanners permit around 600 dpi (43 μm) whilst photogrammetric scanners may produce 3600 dpi (7 μm).

For a monochrome 23×23 cm negative, 600 dpi scanning results in a file size of $9 \times 600 = 5,400$ rows and the same number of columns. Assuming that 1 byte is used per pixel (i.e., there are 256 grey levels), the resulting files requires 29 Mbyte of disk space. When the scale of the negative is given, the ground pixel size of the resulting image can be calculated. Assuming a photo scale of 1:18,000, the first step is to calculate the size of one dot: 25.4 mm / 600 dots = 0.04 mm per dot. The next step is to relate this to the scale: 0.04 mm $\times$ 18,000 = 720 mm in the terrain. The ground pixel size of the resulting image is therefore 0.72 metre.

4.4 Spatial characteristics

Two important properties of an aerial photograph are scale and spatial resolution. These properties are determined by sensor (lens cone and film) and platform (flying height) characteristics. Lens cones are produced with different focal length. *Focal length* is the most important property of a lens cone since, together with flying height, it determines the photo scale. The focal length also determines the angle of view of the camera. The longer the focal length, the narrower the angle of view. The 152 mm lens is the most commonly used lens.

4.4.1 Scale

The relationship between the photo *scale factor*, s, flying height, H, and lens focal length, f, is given by

$$s = \frac{H}{f}. \tag{4.1}$$

Hence, the same scale can be achieved with different combinations of focal length and flying height. If the focal length of a lens is decreased whilst the flying height remains constant, then (also refer to Figure 4.12):

- The *image scale factor* will increase and the size of the individual details in the image becomes smaller. In the example shown in Figure 4.12, using a 150 mm and 300 mm lens at H=2000 m results in a scale factor of 13,333 and 6,666 respectively;

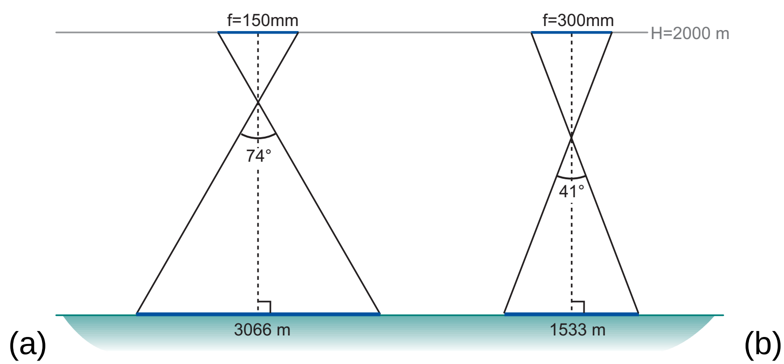


Figure 4.12: Effects of using a different focal length (a: 150 mm, b: 300 mm) when operating a camera from the same height

- The *ground coverage* increases. A 23 cm negative covers a length (and width) of respectively 3066 m and 1533 m using a 150 mm and 300 mm lens. This has implications on the number of photos required for the mapping of a certain area, which, in turn, affects the subsequent processing (in terms of labour) of the photos.

- The *angular field of view* increases and the image perspective changes. The total field of view in situations A and B respectively is 74 degrees and 41 degrees. When wide-angle photography is used for mapping, the measurement of height information (z dimension) in a stereoscopic model is more accurate than when long focal length lenses are used. The combination of a low flying height with a wide-angle lens can be problematic when there are large terrain height differences or high man-made objects in the scene. Some areas may become 'hidden' behind taller objects. This phenomenon is called the *dead ground effect*.

4.4.2 Spatial resolution

While scale is a generally understood and applied term, the use of 'spatial resolution' in aerial photography is quite difficult. Spatial resolution refers to the ability to record small adjacent objects in an image. The spatial resolution of monochrome aerial photographs ranges from 40 to 800 line pairs per mm. The better the resolution of a recording system, the more easily the structure of objects on the ground can be viewed in the image. The spatial resolution of an aerial photograph depends on:

- the image scale factor: spatial resolution decreases as the scale factor increases;
- the quality of the optical system: expensive high quality aerial lenses give much better performance than the inexpensive lenses on amateur cameras;
- the grain structure of the photographic film—the larger the grains, the poorer the resolution;
- the contrast of the original objects—the higher the target contrast, the better the resolution;
- atmospheric scattering effects—this leads to loss of contrast and resolution;
- image motion—the relative motion between the camera and the ground causes blurring and loss of resolution.

From the above list it can be concluded that the physical value of resolution in an aerial photograph depends on a number of factors. The most variable factor is the atmospheric condition, which can change from mission to mission, and, during a mission.

4.5 Aerial photography missions

Mission planning

When a mapping project requires aerial photographs, one of the first tasks is to select the required photo scale factor, the type of lens to be used, the type of film to be used and the required percentage of overlap. Forward overlap usually is around 60%; sideways overlap typically is around 20%. Figure 4.13 shows a survey area that is covered by a number of flight lines. The date and time of acquisition should be considered with respect to growing season, light conditions and shadowing effects.

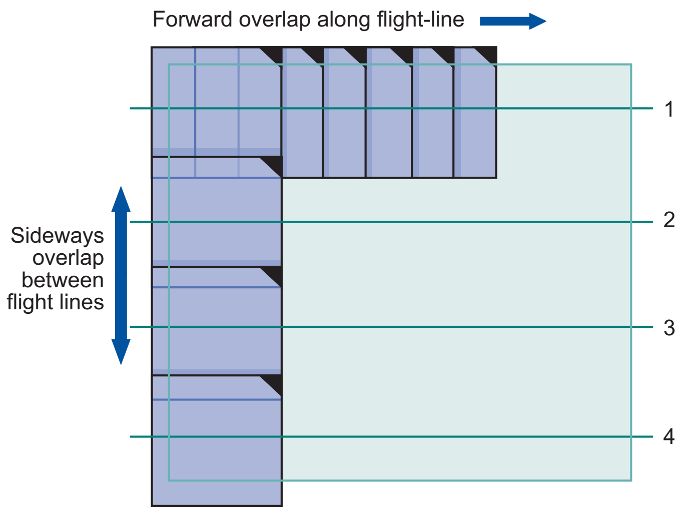


Figure 4.13: Example survey area for aerial photography. Note the forward and sideways overlap of the photographs.

If the required scale is defined the following parameters can be determined:

- the flying height required above the terrain;
- the ground coverage of a single photograph;

- the number of photos required along a flight line;
- the number of flight lines required.

After completion of the necessary calculations, mission maps are prepared for use by the survey navigator in flight.

Mission execution

During the execution of the mission, the navigator alternatively observes the terrain (through a navigation sight) and the flight map and advises small corrections to the pilot. The two main corrections required are:

- Aircraft heading to compensate for the effect of wind. The wind vector continuously changes in direction and magnitude. The navigator needs to monitor the track of the aircraft and apply small corrections in heading to ensure that the aircraft stays accurately on the required flight line. If the heading of the aircraft is corrected in a given direction in order to maintain track, then the orientation of the camera relative to the track is also changed. Whilst the aircraft may still accurately follow the desired track, the camera will no longer be correctly orientated along it. The photos would thus be skewed. The camera must, therefore, be rotated in the opposite direction to the wind correction angle so that the photos remain parallel to the track.

- Aircraft speed or exposure interval to maintain the required forward overlap. The camera viewfinder or navigation sight employs a system to regulate the overlap by reference to the ground. By keeping a series of moving lines in the viewfinder synchronized with ground features, the forward overlap is maintained.

4.6 Advances in aerial photography

The most significant improvements to standard aerial photography made during the last decade can be summarized as follows:

- Global Navigation Satellite Systems (GPS-USA, Glonass-SU and proposed Galileo-EU) provide a means of achieving accurate navigation. They offer precise positioning of the aircraft along the survey run, ensuring that the photographs are taken at the correct points. This method of navigation is especially important in survey areas where topographic maps do not exist, are old, are of small scale or of poor quality. It is also helpful in areas where the terrain has few features (deserts, forests etc) because in these cases conventional visual navigation is particularly difficult. The major aerial camera manufacturers (as well as some independent suppliers), now offer complete software packages that enable the flight crew to plan, execute and evaluate an entire aerial survey mission. Before the mission, the boundaries of the survey area to be photographed are first entered into the software (either from a keyboard or digitizing table), along with basic mission parameters such as the required scale, lens focal length, overlap requirements, terrain elevation *et cetera*. The program then calculates the optimum positions of the required flight lines as well as the positions of the individual photo-centres. This pre-calculated information is then stored on a diskette and taken into the aircraft in order to execute the mission. During the flight, the crew is presented with a moving map display showing the position of the aircraft, the required flight lines and the individual photo-centres. Either flying the aircraft manually or on auto-pilot, the display is followed and the software instructs the camera precisely where to take the photographs. After the mission the co-ordinates of the photographs taken are downloaded and an instant index of the photography produced, even showing parts of the area that need to be repeated due to cloud cover or other reasons.

- Gyroscopically stabilized camera mounting. The gyroscopically stabilized camera mounting enables the camera to be maintained in an accurate level position so that it is continuously pointing vertically downward. It compensates rapid oscillations of the aircraft and reduces effects of aircraft and camera vibration/movement. As a result it gives a more precise photo-positioning.

- Forward motion compensation in aerial survey cameras causes the camera to be displaced across the film during the time that the shutter is open. Forward motion is potentially the most damaging of the effects, and occurs as a direct result of the relative forward motion of the aircraft over the terrain during exposure. If the displacement is such that adjacent grains become exposed, then the resolution of fine detail will be degraded. Forward motion compensation permits the use of (slower) finer grained films and results in images of improved spatial resolution.

A recent significant development concerns the new type of digital cameras. Since the middle of the year 2000 the first examples of these modern sensors have been on the market. These include the Airborne Digital Sensor (ADS) of LH Systems and the Digital Modular Camera (DMC) of Z/I Imaging. The sensors developed have characteristics that relate both to a camera and to a multispectral scanner. Charged Coupled Devices (CCD) are used to record the electromagnetic energy. The design of these cameras is such that it enables multispectral data acquisition and means that overlapping (multi-angle) images are taken along track enabling direct generation of digital elevation models.

Compared to films, CCD recording allows a larger spectral domain and smaller wavelengths bands to be defined. In addition, the general sensitivity is higher and more flexible when compared to film-based systems. The other advantage of a digital camera is that the resulting data are already in digital format, which avoids the need of scanning of the film or prints.

One of the main disadvantages of the digital camera is that its spatial resolution is still somewhat lower than that achieved by film-based systems.

Summary

The characteristics of oblique and vertical aerial photography are distinguished. Vertical aerial photography requires a specially adapted aircraft. The execution of photo flights is performed mainly by specialized companies or governmental units. The main components of an aerial camera system are the lens and film. The lens, in combination with the flying height, determines the photo scale factor. The film type used determines which wavelengths bands are recorded. The most commonly used film types are panchromatic, black-and-white infrared, true-colour and false-colour infrared. Another characteristic of the film is the general sensitivity, which is related to the size of the grains. After exposure, the film is developed and printed. The printed photo can be scanned to use the photo in a digital environment.

There have been many technological developments to improve mission execution as well the image quality itself. Most recent in this development is the digital camera, which directly yields digital data.

Questions

The following questions can help you to study Chapter 4.

1. Consider an area of 500 km^2 that needs aerial photo coverage for topographic mapping at 1:50,000. Which specifications would you give on film, photo scale, overlap, *et cetera*?

2. Go to the Internet and locate three catalogues (archives) of aerial photographs. Compare the descriptions and specifications of the photographs (in terms of scale, resolution, format, . . .).

The following are typical exam questions:

1. Calculate the scale factor for an aerial photo taken at 2500 m height by a camera with a focal length of 88 mm.

2. Consider a (monochrome) black-and-white film. What determines the general sensitivity of this film and why is it important?

3. Explain spectral sensitivity and name three types of films.

4. A hard copy aerial photograph is to be scanned using a flat bed scanner. List three factors that influence the choice of the scanner resolution setting, and explain their significance.

5. Make a drawing to explain the dead ground effect.

Chapter 5

Multispectral scanners

5.1 Introduction

Multispectral scanners measure reflected electromagnetic energy by scanning the Earth's surface. This results in digital image data, of which the elementary unit is a picture element: pixel. As the name *multispectral* suggests, the measurements are made for different ranges of the EM spectrum. Multispectral scanners have been used in remote sensing since 1972 when the first Landsat satellite was launched. After the aerial camera it is the most commonly used sensor. Applications of multispectral scanner data are mainly in the mapping of land cover, vegetation, surface mineralogy and surface water.

Two types of multispectral scanners are distinguished: the whiskbroom scanner and the pushbroom scanner. The principles of these scanners and their characteristics are explained in Section 5.2 and Section 5.3 respectively. Multispectral scanners are mounted on airborne and spaceborne platforms. Section 5.4 describes the most widely used satellite based scanners.

5.2 Whiskbroom scanner

A combination of a single detector plus a rotating mirror can be arranged in such a way that the detector beam sweeps in a straight line over the Earth across the track of the satellite at each rotation of the mirror. In this way, the Earth's surface is scanned systematically line by line as the satellite moves forward. Because of this sweeping motion, the *whiskbroom scanner* is also known as the *across-track scanner* (Figure 5.1). The first multispectral scanners applied the whiskbroom principle. Today, many scanners are still based on this principle: NOAA/AVHRR and Landsat/TM, for instance (Recall: platform/sensor).

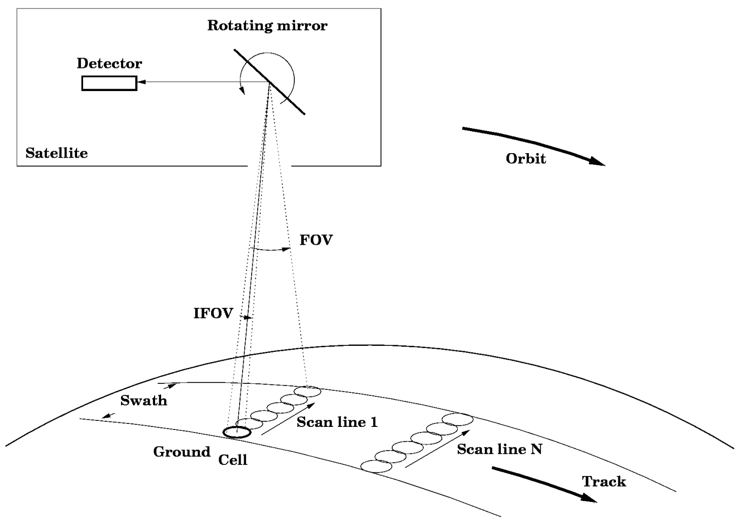


Figure 5.1: Principle of the whiskbroom scanner

5.2.1 Spectral characteristics

Whiskbroom scanners use *solid state detectors* for measuring the energy transferred by the optical system to the sensor. This optical system focusses the incoming radiation at the surface of the detector. Various techniques (prism, gratings) are used to split the incoming radiation into spectral components that each have their own detector.

The detector transforms the electromagnetic radiation (photons) into electrons. The electrons are input to an electronic device that quantifies the level of energy into the required units. In digital imaging systems, a discrete value is used to store the level of energy. These discrete levels are referred to as Digital Number values or DN-values. The fact that the input is measured in discrete levels is also referred to as *quantization*. Using the expression introduced in Chapter 2, one can calculate the amount of energy of a photon corresponding to a specific wavelength, using

$$Q = h \times v \tag{5.1}$$

where Q is the energy of a photon (J), h is Planck's constant ($6.6262 \cdot 10^{-34}$ J s), and, v is the frequency (Hz). The solid state detector measures the amount of energy (J) during a specific time period, which results in J/s = Watt (W).

The range of input radiance, between a maximum and a minimum level, that a detector can handle is called the *dynamic range*. This range is converted into the range of a specified data format. Typically an 8-bit, 10-bit or 12-bit data format is used. The 8-bit format allows $2^8 = 256$ levels or DN-values; similarly 12-bit format allows $2^{12} = 4096$ DN-values. The difference in input level that can be distinguished is called the *radiometric resolution*. Consider a dynamic range of energy between 0.5–3 W. Using 100 or 250 DN-values results in a radiometric resolution of 25 mW and 10 mW respectively.

The other main characteristic of a detector is the *spectral sensitivity*, which is similar to film sensitivity, as explained in Section 4.3.2. Each detector has a characteristic graph that is called the spectral response curve (Figure 5.2). The bandwidth is usually determined by the difference between the two wavelengths where the curve is at 50% of its maximum value. A multispectral scanner uses a number of detectors to measure a number of bands; each band having its own detector.

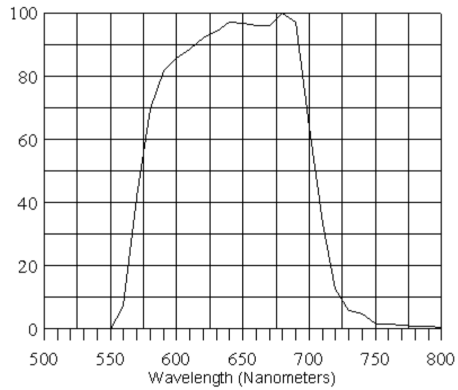


Figure 5.2: Normalized spectral response curve of a specific sensor. It shows that the setting of this band ranges from approximately 570 to 710 nm.

5.2.2 Geometric characteristics

At any instant the mirror of the whiskbroom scanner 'sees' a circle-like area on the ground. Directly below the platform (at nadir), the diameter, D, depends on the viewing angle of the system, β, and the height, H:

$$D = \beta \times H. \tag{5.2}$$

D should be expressed in metres, β in radians and H in metres. Consider a scanner with β = 2.5 mrad that is operated at 4000 m. Using Formula 5.2 one can easily calculate that the diameter of the area observed under the platform is 10 m.

The viewing angle of the system (β) is also referred to as *Instantaneous Field of View*, abbreviated as IFOV. The IFOV determines the spatial resolution of a scanner. The *Field of View* (FOV) describes the total angle that is scanned. For aircraft scanners it is usually expressed as an angle; for satellite based scanners with a fixed height the effective image width is used.

Consider a single line scanned by a whiskbroom scanner mounted to a static platform. This results in a series of measurements going from the left to the right side. The values for single pixels are calculated by integrating over a carefully selected time interval.

5.3 Pushbroom scanner

The pushbroom scanner is based on the use of *Charged Coupled Devices* (CCDs) for measuring the electromagnetic energy (Figure 5.3). A CCD-array is a line of photo-sensitive detectors that function similar to solid state detectors. A single element can be as small as 5 μm. Today, two-dimensional CCD-arrays are used in digital cameras and video recorders. The CCD-arrays used in remote sensing are more sensitive and have larger dimensions. The first satellite sensor using this technology was SPOT-1 HRV. High resolution sensors such as IKONOS and Orbview3 also apply the pushbroom principle.

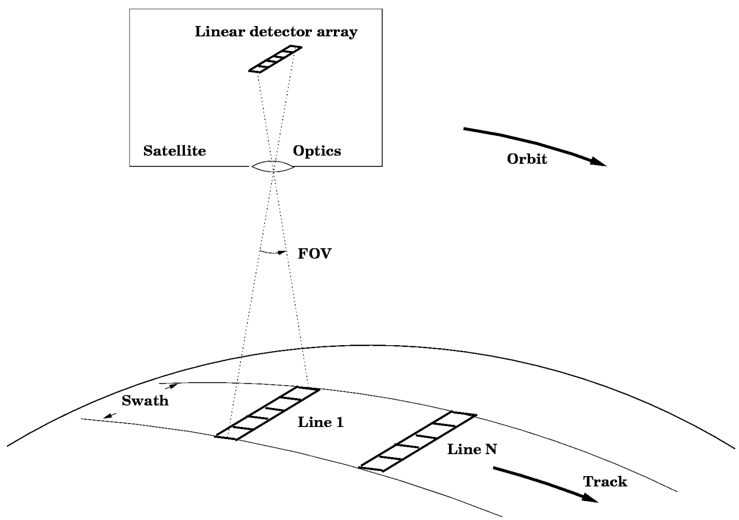


Figure 5.3: Principle of a pushbroom scanner

The pushbroom scanner records one entire line at a time. The principal advantage over the whiskbroom scanner is that each position (pixel) in the line has its own detector. This enables a longer period of measurement over a certain area, resulting in less noise and a relatively stable geometry. Since the CCD elements continuously measure along the direction of the platform, this scanner is also referred to as *along-track scanner*.

5.3.1 Spectral characteristics

To a large extent, the characteristics of a solid state detector are also valid for a CCD-array. In principle, one CCD-array corresponds to a spectral band and all the detectors in the array are sensitive to a specific range of wavelengths. With current state-of-the-art technology, CCD-array sensitivity stops at 2.5 μm wavelength. If longer wavelengths are to be measured, other detectors need to be used.

One drawback of CDD arrays is that it is difficult to produce an array in which all the elements have similar sensitivity. Differences between the detectors may be visible in the recorded images as vertical banding.

5.3.2 Geometric characteristics

For each single line, pushbroom scanners have a geometry similar to that of aerial photos (which have a 'central projection'). In the case of flat terrain, and a limited total field of view (FOV), the scale is the same over the line, resulting in equally spaced pixels. The concept of IFOV cannot be applied to pushbroom scanners.

Typical for most pushbroom scanners is the ability for *off-track viewing*. In such a situation, the scanner is pointed towards areas to the left or right of the orbit track (off-track) or to the back or forth (along-track). This characteristic has two advantages: it is used to produce stereo-images, and it can be used to image an area that is not covered by clouds at that particular moment. When applying off-track viewing, similar to oblique photography, the scale in an image varies and should be corrected for.

As with whiskbroom scanners, an integration over time takes place in pushbroom scanners. Consider a moving platform with a pushbroom scanner. Each element of the CCD-array measures the energy related to a small strip below the platform. Every n milliseconds the recorded energy (W) is averaged to determine a the DN-value for each pixel along the line.

5.4 Some operational spaceborne multispectral scanners

This section gives some details about specific spaceborne missions that carry multispectral scanners and describes some of their applications.

5.4.1 Meteosat-5

Meteosat is a geostationary satellite that is used in the world meteorological programme. The programme comprises seven satellites in total. The first Meteosat satellite was placed in orbit in 1977. Meteosat satellites are owned by the European organisation Eumetsat. At this moment, Meteosat-5 is operational with Meteosat-6 as a back-up.

System	Meteosat-5
Orbit	Geo-stationary, 0° longitude
Sensor	VISSR (Visible and Infrared Spin Scan Radiometer)
Swath width	Full Earth disc (FOV = 18°)
Off-track viewing	Not applicable
Revisit time	30 minutes
Spectral bands (μm)	0.5–0.9 (VIS), 5.7–7.1 (WV), 10.5–12.5 (TIR)
Ground pixel size	2.5 km (VIS and WV), 5 km (TIR)
Data archive at	www.eumetsat.de

Table 5.1: Meteosat-5 VISSR characteristics

The spectral bands of the VISSR sensor are chosen for observing phenomena that are relevant to meteorologists: a panchromatic band (VIS), a mid-infrared band, which gives information about the water vapour (WV) present in the atmosphere, and a thermal band (TIR). In case of clouds, the thermal data relate to the cloud top temperature, which is used for rainfall estimates and forecasts. Under cloud-free conditions the thermal data relate to the surface temperature of land and sea.

5.4.2 NOAA-15

NOAA stands for National Oceanic and Atmospheric Administration, which is a US-government body. The sensor onboard of NOAA missions that is relevant for Earth Observation is the Advanced Very High Resolution Radiometer (AVHRR). Today, two NOAA satellites (-14, -15) are operational.

Table 5.2: NOAA-15 AVHRR characteristics

System	NOAA-15
Orbit	850 km, 98.8°, sun-synchronous
Sensor	AVHRR-3 (Advanced Very High Resolution Radiometer)
Swath width	2800 km (FOV = 110°)
Off-track viewing	No
Revisit time	2–14 times per day, depending on latitude
Spectral bands (μm)	0.58–0.68(1), 0.73–1.10 (2), 3.55–3.93(3), 10.3–11.3(4), 11.4–12.4(5)
Spatial resolution	1 km (at nadir), 6 km (at limb), IFOV=1.4 mrad
Data archive at	www.saa.noaa.gov

As the AVHRR sensor has a very wide FOV (110°) and is at a large distance from the Earth, the whiskbroom principle causes a large difference in the ground cell measured within one scanline (Figure 5.4). The standard image data products of AVHRR yield image data with equally sized ground pixels.

AVHRR data are used primarily in day-to-day meteorological forecasting where it gives more detailed information than Meteosat. In addition, there are many land and water applications.

AVHRR data are used to generate *Sea Surface Temperature maps* (SST maps), which can be used in climate monitoring, the study of El Niño, the detection of eddies to guide vessels to rich fishing grounds, *et cetera*. *Cloud cover maps* based on AVHRR data, are used for rainfall estimates, which can be input into crop growing models. Another derived product of AVHRR data are the *Normalized Difference Vegetation Index maps* (NDVI). These 'maps' give an indication about the quantity of biomass (tons/ha). NDVI data are used as input into crop growth models and also for climate change models. The NDVI data are, for instance, used by FAO in their food security early warning system (FEWS). AVHRR data are appropriate to map and monitor regional land cover and to assess the energy balance of agricultural areas.

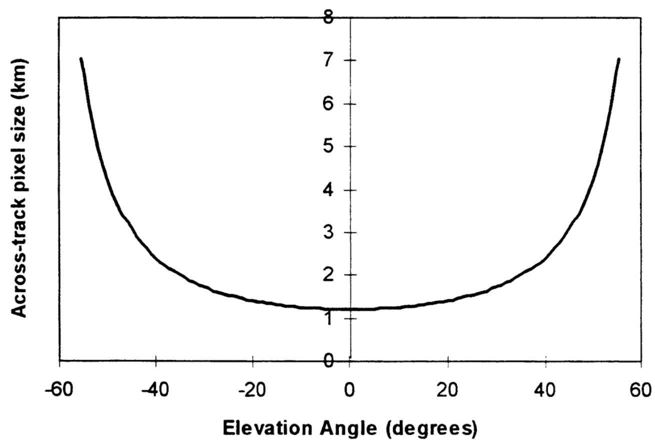


Figure 5.4: The NOAA/AVHRR sensor, which observes an area of 1×1 km^2 at the centre and 6×3 km^2 at the edge.

A new application of Meteosat and NOAA data is to track cloud-free areas to optimize the data acquisition of high resolution satellites.

5.4.3 Landsat-7

The Landsat programme is the oldest Earth Observation programme. It started in 1972 with the Landsat-1 satellite carrying the MSS multispectral sensor. After 1982, the Thematic Mapper (TM) replaced the MSS sensor. Both MSS and TM are whiskbroom scanners. In April 1999 Landsat-7 was launched carrying the ETM+ scanner. Today, only Landsat-5 and -7 are operational.

System	Landsat-7
Orbit	705 km, 98.2°, sun-synchronous, 10:00 AM crossing, 16 days repeat cycle
Sensor	ETM+ (Enhanced Thematic Mapper)
Swath width	185 km (FOV = 15°)
Off-track viewing	No
Revisit time	16 days
Spectral bands (μm)	0.45–0.52(1), 0.52–0.60(2), 0.63–0.69(3), 0.76–0.90(4), 1.55–1.75(5), 10.4–12.50(6), 2.08–2.34(7), 0.50–0.90(PAN)
Spatial resolution	15 m (PAN), 30 m (bands 1–5,7), 60 m (band 6)
Data archive at	earthexplorer.usgv.gov

Table 5.3: Landsat-7 ETM+ characteristics

There are many applications of Landsat Thematic Mapper data: land cover mapping, land use mapping, soil mapping, geological mapping, sea surface temperature mapping, *et cetera*. For land cover and land use mapping Landsat Thematic Mapper data are preferred, e.g., over SPOT multispectral data, because of the inclusion of middle infrared bands. Landsat Thematic Mapper is the only non-meteorological satellite that has a thermal infrared band. Thermal data are required to study energy processes at the Earth's surface, for instance, the crop temperature variability within irrigated areas. Table 5.4 gives the principle applications of the various TM bands.

Band	Wavelength (μm)	Principal Applications
1	0.45–0.52	Designed for water body penetration, making it useful for coastal water mapping. Also useful for soil-vegetation discrimination and forest type mapping.
2	0.52–0.60	Designed to measure green reflectance peak of vegetation for vegetation discrimination and vigour assessment.
3	0.63–0.69	Designed to sense in a chlorophyll absorption region aiding in plant species differentiation.
4	0.76–0.90	Useful for determining vegetation types, vigour, and bio-mass content, for delineating water bodies, and for soil moisture discrimination.
5	1.55–1.75	Indicative of vegetation moisture content and soil moisture. Also useful for differentiation of snow from clouds.
6	10.4–12.5	Useful in vegetation stress analysis, soil moisture discrimination, and thermal mapping applications.
7	2.08–2.35	Useful for discrimination of mineral and rock types. Also sensitive to vegetation moisture content.

Table 5.4: Principal applications of the Landsat TM bands (from [16])

5.4.4 SPOT-4

SPOT stands for *Systeme Pour l'Observation de la Terre*. SPOT-1 was launched in 1986. SPOT is owned by a consortium of French, Swedish and Belgium governments. It was the first operational pushbroom CCD sensor with off-track viewing capability to be put into space. At that time, the 10 m panchromatic spatial resolution was unprecedented. In March 1998 a significantly improved SPOT-4 was launched: the HRVIR sensor has 4 instead of 3 bands and the VEGETATION instrument was added. VEGETATION has been designed for frequent (almost daily) and accurate monitoring of the globe's landmasses.

System	SPOT-4
Orbit	835 km, 98.7°, sun-synchronous, 10:30 AM crossing, 26 days repeat cycle
Sensor	two HRVIR sensors (High Resolution Visible and Infrared)
Swath width	60 km (3000 pixels CCD-array)
Off-track viewing	Yes, ± 27° across-track
Revisit time	4–6 days (depending on latitude)
Spectral bands (μm)	0.50–0.59(1), 0.61–0.68(2), 0.79–0.89(3), 1.58–1.75(4), 0.61–0.68(PAN)
Spatial resolution	10 m (PAN), 20 m (bands 1–4)
Data archive at	sirius.spotimage.fr

Table 5.5: SPOT-4 HRVIR characteristics

5.4.5 IRS-1D

India puts much effort into remote sensing and has many operational missions and missions under development. The most important Earth Observation programme is the Indian Remote Sensing (IRS) programme. Launched in 1995 and 1997, two identical satellites, IRS-1C and IRS-1D, can deliver image data at high revisit times. IRS-1C and IRS-1D carry three sensors: the Wide Field Sensor (WiFS) designed for regional vegetation mapping, the Linear Imaging Self-Scanning Sensor 3 (LISS3), which yields multispectral data in four bands with a spatial resolution of 24 m, and the PAN.

In this subsection, the characteristics of the PAN sensor are given. For a number of years, up to the launch of IKONOS in September 1999, the IRS-1C and -1D were the civilian satellites with the highest spatial resolution. Applications are similar to those of SPOT and Landsat.

System	IRS-1D
Orbit	817 km, 98.6°, sun-synchronous, 10:30 AM crossing, 24 days repeat cycle
Sensor	PAN (Panchromatic Sensor)
Swath width	70 km
Off-track viewing	Yes, ± 26° across-track
Revisit time	5 days
Spectral bands (μm)	0.50–0.75
Spatial resolution	6 m
Data archive at	www.spaceimaging.com

Table 5.6: IRS-1D PAN characteristics

5.4.6 IKONOS

IKONOS was the first commercial high resolution satellite to be placed into orbit in space. IKONOS is owned by SpaceImaging, a USA based Earth observation company. The other commercial high resolution satellites foreseen are: Orbview-3 (OrbImage), Quickbird (EarthWatch), and EROS-A1 (West Indian Space). IKONOS was launched in September 1999 and regular data ordering has been taking place since March 2000.

The OSA sensor onboard is based on the pushbroom principle and can simultaneously take panchromatic and multispectral images. IKONOS delivers the highest spatial resolution so far achieved by a civilian satellite. Apart from the high spatial resolution it also has a high radiometric resolution using 11-bit quantization.

Many applications for the IKONOS data are foreseen. The owner expects that the application fields are able to pay for the commercially priced data. It is

System	IKONOS
Orbit	680 km, 98.2°, sun-synchronous, 10:30 AM crossing, 14 days repeat cycle
Sensor	Optical Sensor Assembly (OSA)
Swath width	11 km (12 μm CCD elements)
Off-track viewing	Yes, $\pm$ 50° omnidirectional
Revisit time	1–3 days
Spectral bands (μm)	0.45–0.52(1), 0.52–0.60(2), 0.63–0.69(3), 0.76–0.90(4), 0.45–0.90(PAN)
Spatial resolution	1 m (PAN), 4 m(bands 1–4)
Data archive at	www.spaceimaging.com

Table 5.7: IKONOS OSA characteristics

expected that, in the long term, 50% of the aerial photography will be replaced by high resolution imagery from space (digital airborne cameras will largely replace the remaining aerial photography.). IKONOS' first task will be to acquire imagery of all major USA cities. Until now, the mapping and monitoring of urban areas from space (not only in America) was possible only to a limited extent.

IKONOS data can be used for small to medium scale topographic mapping, not only to produce new maps, but also to update existing topographic maps.

Another potential application is 'precision agriculture'; this is reflected in the multispectral band setting, which includes a near-infrared band. Regular updates of the 'field situation' can help farmers to optimize the use of fertilizers and herbicides.

Applications of the 'picture' products are foreseen in business, the media, and tourism.

5.4.7 Terra

EOS (Earth Observing System) is the centerpiece of NASA's Earth Science mission. The EOS AM-1 satellite, later renamed to Terra, is the flagship of the fleet and was launched in December 1999. It carries five remote sensing instruments including MODIS and ASTER. ASTER, the Advanced Spaceborne Thermal Emission and Reflectance Radiometer, is a high resolution imaging spectrometer. The instrument is designed with three bands in the visible and near-infrared spectral range with a 15 m resolution, six bands in the short-wave infrared with a 30 m resolution, and five bands in the thermal infrared with a 90 m resolution. The VNIR and SWIR bands have a spectral bandwidth in the order of 10 nm. ASTER consists of three separate telescope systems, each of which can be pointed at selected targets. By pointing to the same target twice, ASTER can acquire high-resolution stereo images. The swath width of the image is 60 km and the revisit time is about 5 days.

MODIS, the Moderate-Resolution Imaging Spectroradiometer observes the entire surface of the Earth every 1–2 days with a whisk-broom scanning imaging radiometer. Its wide field of view (over 2300 km) provides images of daylight reflected solar radiation and day/night thermal emissions over the entire globe. Its spatial resolution ranges from 250–1000 m.

System	Terra
Orbit	705 km, 98.2°, sun-synchronous, 10:30 AM crossing, 16 days repeat cycle
Sensor	ASTER
Swath width	60 km
Off-track viewing	Yes, $\pm$ 8.5° SWIR and TIR $\pm$ 24° VNIR
Revisit time	5 days
Spectral bands (μm)	VNIR 0.0.56(1), 0.66(2), 0.81(3) SWIR 1.65(1), 2.17(2), 2.21(3), 2.26(4), 2.33(5), 2.40(6) TIR 8.3(1), 8.65(2), 9.10(3), 10.6(4), 11.3(5)
Spatial resolution	15 m (VNIR), 30 m(SWIR), 90 m(TIR)
Data archive at	terra.nasa.gov

Table 5.8: Terra characteristics

5.4.8 EO-1

The EO-1 mission is part of the NASA New Millennium Program and is focused on new sensor and spacecraft technologies that can directly reduce the cost of Landsat and related Earth monitoring Systems. The EO-1 satellite is in an orbit that covers the same ground track as Landsat 7, approximately one minute later. This enables EO-1 to obtain images of the same ground area at nearly the same time, so that direct comparison of results can be obtained from Landsat ETM+ and the three primary EO-1 instruments. The three primary instruments on the EO-1 spacecraft are the Hyperion and the Linear Etalon Imaging Spectrometer Array (LEISA), Atmospheric Corrector (LAC) and the Advanced Land Imager (ALI).

Hyperion is a grating imaging spectrometer with a 30 m ground sample distance over a 7.5 km swath, providing 10 nm (sampling interval) contiguous bands of the solar reflected spectrum from 400–2500 nm. LAC is an imaging spectrometer covering the spectral range from 900–1600 nm which is well-suited to monitor the atmospheric water absorption lines for correction of atmospheric effects in multispectral imagers such as ETM+ on Landsat.

The Earth Observing-1 (EO-1) Advanced Land Imager (ALI) is a technology verification instrument. Operating in a pushbroom fashion at an orbit of 705 km, the ALI will provide Landsat-type panchromatic and multispectral bands. These bands have been designed to mimic six Landsat bands with three additional bands covering 0.433–0.453, 0.845–0.890, and 1.20–1.30 μm. The ALI also contains wide-angle optics designed to provide a continuous $15° \times 1.625°$ field of view for a fully populated focal plane with 30 m resolution for the multispectral pixels and 10 m resolution for the panchromatic pixels.

System	EO-1
Orbit	705 km, 98.7°, sun-synchronous, 10:30 AM crossing, 16 days repeat cycle
Sensor	ALI (Advanced Land Imager)
Swath width	37 km
Off-track viewing	No
Revisit time	16 days
Spectral bands (μm)	As Landsat 7 + 0.433–0.453, 0.845–0.890, and 1.20–1.30
Spatial resolution	10 m (PAN), 30 m (other bands)
Data archive at	eo1.gsfc.nasa.gov

Table 5.9: EO-1 characteristics

Summary

The multispectral scanner is a sensor that collects data in various wavelength bands of the EM spectrum. The scanner can be mounted on an aircraft or on a satellite. There are two types of scanners: whiskbroom and pushbroom scanners. They use solid state detectors and CCD-arrays respectively for measuring the level of energy. The resulting image data store the level of energy as Digital Numbers, which are calculated during the quantization process. Multispectral scanners provide multi-band data.

In terms of geometrically reliable data the pushbroom scanner performs best. In terms of measuring many spectral bands (including thermal infrared) the current whiskbroom scanners are the best.

Questions

The following questions can help you to study Chapter 5.

1. Compare multispectral scanner data with scanned aerial photographs. Which similarities and differences can you identify?

2. Go to the Internet and check the availability of multispectral image data of your country (area of interest). First determine which range of spatial resolution you are interested in.

The following are typical exam questions:

1. Explain the principle of the whiskbroom scanner.

2. Explain the principle of the pushbroom scanner.

3. What does CCD stand for and what is it used for?

4. Consider a whiskbroom scanner at 5000 m height and with a Field of View (β) of 2 mrad. Calculate the diametre of the area observed on the ground.

5. Explain the difference between IFOV and FOV.

6. Explain off-nadir viewing. What is an advantage?

7. What is quantization and to which part of the scanning process does it relate?

8. Which range of spatial resolutions are encountered with today's multispectral (and panchromatic) scanners?

Chapter 6

RADAR

6.1 What is radar?

So far, you have learned about remote sensing using the visible and infrared part of the electromagnetic spectrum. Microwave remote sensing uses electromagnetic waves with wavelengths between 1 cm and 1 m (Figure 2.5). These relatively longer wavelengths have the advantage that they can penetrate clouds and are independent of atmospheric conditions, like haze. In microwave remote sensing there are active and passive sensors. Passive sensors operate similarly to thermal sensors by detecting naturally emitted microwave energy. They are used in meteorology, hydrology and oceanography. In active systems, the antenna transmits microwave signals from an antenna to the Earth's surface where they are *backscattered*. The part of the electromagnetic energy that is scattered into the direction of the antenna is detected by the sensor as illustrated in Figure 6.1. There are several advantages to be gained from the use of active sensors, which have their own energy source:

- It is possible to acquire data at any time including during the night (similar to thermal remote sensing).
- Since the waves are created actively, the signal characteristics are fully controlled (e.g., wavelength, *polarization*, incidence angle, *et cetera*) and can be adjusted according to the desired application.

Active sensors are divided into two groups: *imaging* and *non-imaging* sensors. *RADAR* sensors belong to the group of most commonly used active imaging microwave sensors. The term *RADAR* is an acronym for *RAdio Detection And Ranging*. *Radio* stands for the microwave and *range* is another term for distance. Radar sensors were originally developed and used by the military. Nowadays, radar

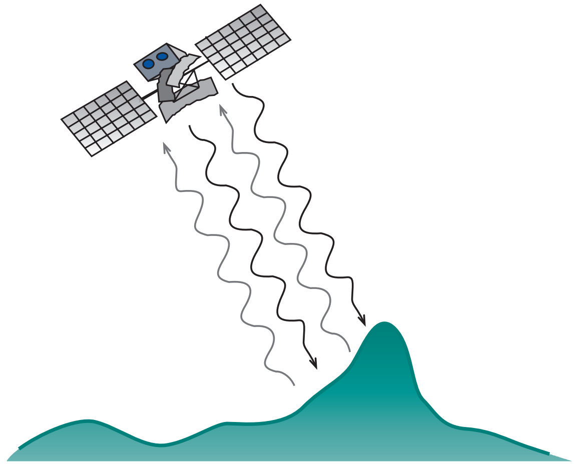


Figure 6.1: Principle of active microwave remote sensing

sensors are widely used in civil applications too, such as environmental monitoring. To the group of non-imaging microwave instruments belong *altimeters*, which collect distance information (e.g., sea surface height), and *scatterometers*, which acquire information about the object properties (e.g., wind speed).

This chapter will focus on the principles of imaging radar and its applications. The interpretation of radar imagery is less intuitive than that obtained from optical remote sensing. This is because of differences in physical interaction of the waves with the Earth's surface. The chapter will explain which interactions take place and how radar images can be interpreted.

6.2 Principles of imaging radar

Imaging radar systems include several components: a transmitter, a receiver, an antenna and a recorder. The transmitter is used to generate the microwave signal and transmit the energy to the antenna from where it is emitted towards the Earth's surface. The receiver accepts the backscattered signal as received by the antenna, filters and amplifies it as required for recording. The recorder then stores the received signal.

Imaging radar acquires an image in which each pixel contains a digital number according to the strength of the backscattered energy that is received from the ground. The energy received from each transmitted radar pulse can be expressed in terms of the physical parameters and illumination geometry using the so-called *radar equation*

$$P_r = \frac{G^2 \lambda^2 P_t \sigma}{(4\pi)^3 R^4}, \tag{6.1}$$

where

- P_r is the received energy,
- G is the antenna gain,
- λ is the wavelength,
- P_t is the transmitted energy,
- σ is the *radar cross section*, it is a function of the object characteristics and the size of the illuminated area,

- R is the range from the sensor to the object.

From this equation you can see that there are three main factors that influence the strength of the backscattered received energy:

- radar system properties, i.e., wavelength, antenna and transmitted power,
- radar imaging geometry, that defines the size of the illuminated area which is a function of i.e., beam-width,incidence angle and range,
- object characteristics in relation to the radar signal, i.e., surface roughness and composition, and terrain topography and orientation.

They are explained in the following sections in more detail.

What exactly does a radar system measure?

To interpret radar images correctly, it is important to understand what a radar sensor detects. Imagine the transmitter creates microwave signals, i.e., *pulses* of microwaves at regular intervals, the *Pulse Repetition Frequency* (PRF), that are bundled by the antenna into a *beam*. This beam travels through the atmosphere, illuminates a portion of the Earth's surface, is backscattered and passes through the atmosphere again to reach the antenna where the signal intensity is received. From the time interval the signal needs to pass twice the distance between object and antenna, and knowing the speed of light, the distance (*range*) between sensor and object can be derived.

To create an image, the return signal of a single pulse is sampled and these samples are stored in an image line. With the movement of the sensor, emitting pulses, a two-dimensional image is created (each pulse defines one line). The radar sensor, therefore, measures distances and detects backscattered signal intensities.

Commonly used imaging radar bands

Similarly to optical remote sensing, radar sensors operate with different bands. For better identification, a standard has been established that defines various wavelength ranges using letters to distinguish among the various bands (Figure 6.2). In the description of different radar missions you will recognise the different wavelengths used if you see the letters. The European *ERS* mission and the Canadian *Radarsat*, for example, use C-band radar. Just like multispectral bands, different radar bands provide information about different object characteristics.

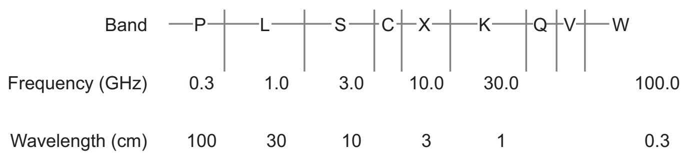


Figure 6.2: Microwave spectrum and band identification by letters

Microwave polarizations

The polarization of an electromagnetic wave is important in the field of radar remote sensing. Depending on the orientation of the transmitted and received radar wave, polarization will result in different images (Figure 6.3). It is possible to work with horizontally, vertically or cross-polarized radar waves. Using different polarizations and wavelengths, you can collect information that is useful for particular applications, for example, to classify agricultural fields. In radar system descriptions you will come across the following abbreviations:

- HH: horizontal transmission and horizontal reception,
- VV: vertical transmission and vertical reception,
- HV: horizontal transmission and vertical reception, and
- VH: vertical transmission and horizontal reception.

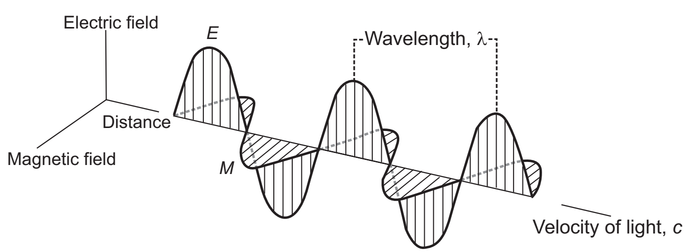


Figure 6.3: A vertically polarized electromagnetic wave; the electric field's variation occurs in the vertical plane in this example.

6.3 Geometric properties of radar

The platform carrying the radar sensor moves along the orbit in flight direction (Figure 6.4). You can see the ground track of the orbit/flight path on the Earth's surface at nadir. The microwave beam illuminates an area, or *swath*, on the Earth's surface, with an offset from the nadir. The direction along-track is called *azimuth*, the direction perpendicular (across-track) is called *range*.

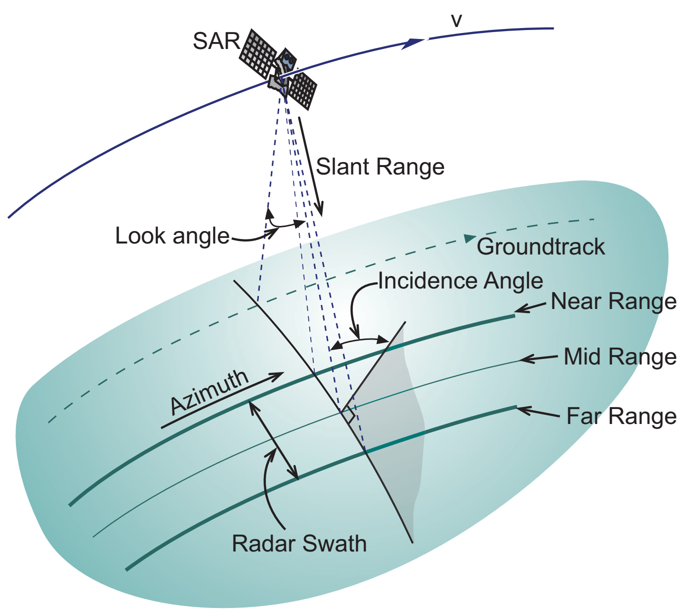


Figure 6.4: Radar remote sensing geometry

6.3.1 Radar viewing geometry

Radar sensors are side-looking instruments. The portion of the image that is closest to the nadir track of the satellite carrying the radar is called *near range*. The part of the image that is farthest from the nadir is called *far range* (Figure 6.4). The *incidence angle* of the system is defined as the angle between the radar beam and the local vertical. Moving from near range to far range, the incidence angle increases. It is important to distinguish between the incidence angle of the sensor and the *local incidence angle*, which differs depending on terrain slope and earth-curvature (Figure 6.5). It is defined as the angle between the radar beam and the local surface normal. The radar sensor measures the distance between antenna and object. This line is called *slant range*. But the true horizontal distance along the ground corresponding to each measure point in slant range is called *ground range* (Figure 6.6).

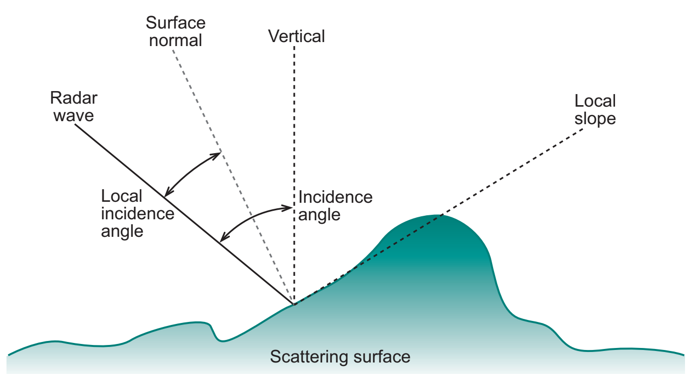


Figure 6.5: Radar incidence angle and local incidence angle

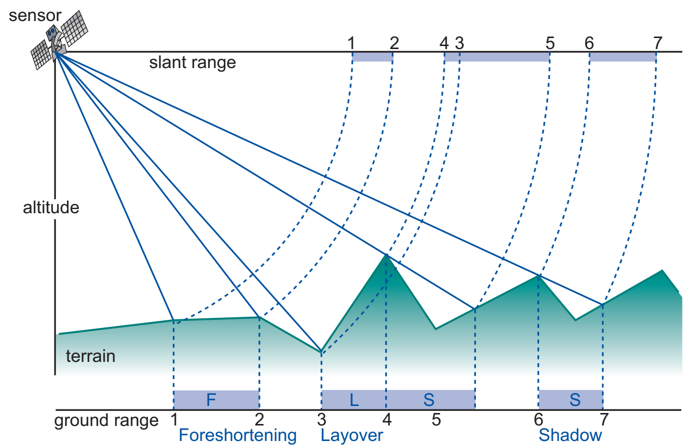


Figure 6.6: Geometric distortions in radar imagery due to terrain elevations

6.3.2 Spatial resolution

In radar remote sensing, the images are created from transmitted and backscattered signals. If each single transmitted pulse forms one element in the image the system is called *Real Aperture Radar*. The spatial resolutions in slant range and azimuth direction are defined by pulse length and antenna beam width, respectively. Due to the different parameters that determine the spatial resolution in range and azimuth resolution, it is obvious that the spatial resolution in the two directions is different. For radar image processing and interpretation it is useful to resample the image data to regular pixel spacing in both directions. In the case of ERS-1 SAR, this spacing is may be 30 m $\times$ 30 m or 12.5 m $\times$ 12.5 m depending on the parameter setting in the processing software.

Slant range resolution

In slant range the spatial resolution is defined as the distance that two objects on the ground have to be apart to give two different echoes in the return signal. In other words, two objects will be resolved in range direction if they are separated by at least half a pulse length. The slant range resolution is independent of the range. However, in ground range geometry the resolution will depend on the incidence angle.

Azimuth resolution

The spatial resolution in azimuth direction depends on the beam width and the range. The radar beam width is proportional to the wavelength and inversely proportional to the antenna length, i.e., *aperture*; this means the longer the antenna, the narrower the beam and the higher the spatial resolution in azimuth direction.

6.3.3 Synthetic Aperture Radar (SAR)

Obviously there is a physical limit to the length of the antenna, the aperture that can be carried on an aircraft or satellite. On the other hand, shortening the wavelength has its limitations in penetrating clouds. Therefore, an approach in which the aperture is increased synthetically is applied. Systems using this approach are called *Synthetic Aperture Radar (SAR)*. The synthesization of the antenna length is achieved by taking advantage of the forward motion of the platform and using several backscattered signals including the same object to simulate a very long antenna. Most airborne and spaceborne radar systems use this type of radar.

6.4 Distortions in radar images

Due to the side-looking viewing geometry, radar images suffer from serious geometric and radiometric distortions. In radar imagery, you encounter variations in scale (slant range to ground range conversion), *foreshortening*, *layover* and *shadows* (terrain elevation). Interference due to the coherency of the signal causes *speckle* effects.

6.4.1 Scale distortions

Radar measures ranges to objects in slant range rather than true horizontal distances along the ground. Therefore, the image has different scales moving from near to far range (Figure 6.4). This means that objects in near range are compressed with respect to objects at far range. For proper interpretation, the image has to be corrected and transformed into ground range geometry.

6.4.2 Terrain-induced distortions

Similarly to optical sensors that can operate in an oblique manner (e.g., SPOT) radar images are subject to relief displacements (Figure 6.6). In the case of radar, these distortions can be severe. There are three effects that are typical for radar: *foreshortening*, *layover* and *shadow*.

Foreshortening

Radar measures distance in slant range. The slope area is compressed in the image. Depending on the angle that the slope forms in relation to the incidence angle of the radar beam the slope will be shortened more or less. The distortion is at its maximum if the radar beam is almost perpendicular to the slope. Foreshortened areas in the radar image are very bright.

Layover

If the radar beam reaches the top of the slope earlier than the bottom, the slope is imaged upside down, i.e., the slope 'lays over'. As you can understand from the definition of foreshortening, layover is an extreme case of foreshortening. Layover areas in the image are very bright.

Shadow

In the case of slopes that are facing away from the sensor, the radar beam cannot illuminate the area. Therefore, there is no energy that can be backscattered to the sensor and those regions remain dark in the image. Radar shadow has to be distinguished from shadow areas in optical images where the radiometric values are altered due to the difference in Sun illumination.

6.4.3 Radiometric distortions

The above-mentioned geometric distortions also have an influence on the received energy. Since the backscattered energy is collected in slant range the received energy coming from a slope facing the sensor is stored in a reduced area in the image, i.e., it is compressed into fewer image pixels than should be the case if obtained in ground range geometry. This results in high digital numbers because the energy collected from different objects is combined. Unfortunately this effect cannot be corrected. This is why especially layover and shadow areas in radar imagery cannot be used for interpretation. However, they are useful in the sense that they contribute to a three-dimensional look of the image and therefore help the understanding of the terrain structure and topography.

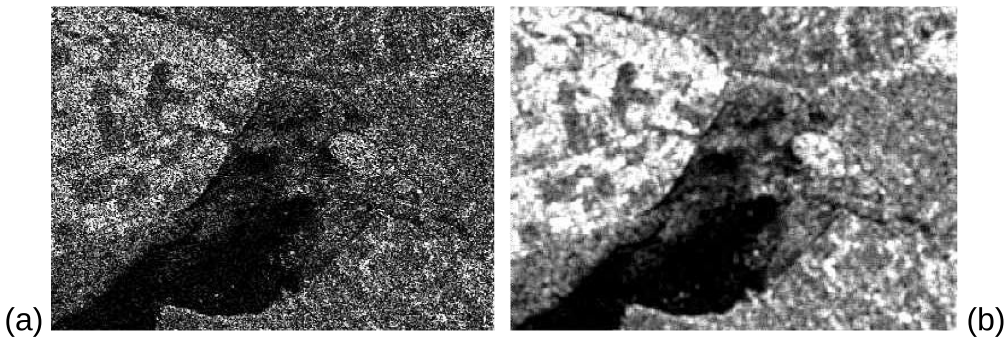


Figure 6.7: Original (a) and speckle filtered (b) radar image

A typical property of radar images is the so-called *speckle*. It appears as grainy 'salt and pepper' effects in the image (Figure 6.7). Speckle is caused by the interaction of the different microwaves backscattered from the object area. The wave interactions are called *interference*. Interference causes the return signals to be extinguished or amplified resulting in dark and bright pixels in the image even when the sensor observes a homogenous area. Speckle degrades the quality of the image and makes the interpretation of radar imagery difficult.

Multi-look processing

It is possible to reduce speckle by means of multi-look processing or spatial filtering. In the case of multi-look processing, the radar beam is divided into several narrower beams. Each beam provides a *look* at the object. Using the average of these multiple looks, the final image is obtained. Multi-look processing reduces the spatial resolution. If you purchase an ERS SAR scene in PRI-format you will receive a 3-look image.

Speckle filters

Another way to reduce speckle is to apply spatial filters on the images. Speckle filters are designed to adapt to local image variations in order to smooth the values to reduce speckle but to enhance lines and edges to maintain the sharpness of the imagery.

6.5 Interpretation of radar images

The brightness of features in a radar image depends on the strength of the backscattered signal. In turn, the amount of energy that is backscattered depends on various factors. An understanding of these factors will help you to interpret radar images properly.

6.5.1 Microwave signal and object interactions

The amount of energy that is received at the radar antenna depends on the illuminating signal (radar system parameters such as wavelength, polarization, viewing geometry, *et cetera*) and the characteristics of the illuminated object (roughness, shape, orientation, dielectric constant, *et cetera*).

Influence of the illuminating signal

To a certain degree, the wavelength of a radar system influences the depth to which the waves penetrate into the object surface. In addition, it determines the size of the objects that the waves interact with. For example, a short microwave will only penetrate the leaves on top of the trees (e.g., X-band = 3 cm) whereas in the case of L-band (23 cm), the radiation penetrates into the canopy. The polarization of the microwave plays an important role in the interpretation of the form and the orientation of small scattering elements that compose the surface object. Therefore, the use of microwaves with different polarizations yields different images that might help in the identification of objects.

Influence of the illuminated surface

An absolute measure for the backscatter behaviour of an object—similar to reflectance in optical systems—is calculated from the ratio of the emitted and received signal, taking into account the range to the object. This is called the radar cross section *sigma* (σ) and it is expressed in *decibels* (db). The amount of energy backscattered from an object depends on its characteristics such as surface roughness, moisture content (electrical properties of the object), its orientation with respect to the illuminating signal (local incidence angle) and its shape. Apart from topography, the surface roughness is the terrain property that most strongly influences the strength of the radar return. It is a relative aspect, depending upon wavelength and incidence angle. A surface is considered 'rough' if it has height variations of dimensions that are close to the radar wavelength, for example, the size of a leave. In radar images, rough surfaces appear bright, smooth surfaces appear dark in radar images. This is a result of the scattering behaviour of radar waves. Another important parameter that influences the microwave backscatter behaviour is the *dielectric constant*, which describes the electrical properties of the surface material. The moisture content of the object affects the electrical properties and therefore the dielectric constant.

6.5.2 Scattering patterns

The radar is scattered in different ways depending on the above-mentioned characteristics of the signal and object. Changes in the electrical properties influence the absorption, transmission and reflection of microwaves. This means that the moisture content of the surface is an essential component of the total scattering. If an object is wet, surface scattering takes place. The type of reflection (ranging from specular to diffuse) and the strength depend on the roughness of the material. Generally, reflectivity, and therefore image brightness, increases with increasing moisture content.

6.5.3 Applications of radar

There are many useful applications of radar images. Radar data provide complementary information to visible and infrared remote sensing data. In the case of forestry, radar images can be used to obtain information about forest canopy, biomass and different forest types. Radar images also allow the differentiation of different land cover types such as urban areas, agricultural fields, water bodies, *et cetera*. In agricultural crop identification, the use of radar images acquired using different polarization (mainly airborne) is quite effective. It is crucial for agricultural applications to acquire data at a certain point in time (season) to obtain the necessary parameters. This is possible because radar can operate independently of weather or daylight conditions. In geology and geomorphology the fact that radar provides information about surface texture and roughness plays an important role in lineament detection and geological mapping. Other successful applications of radar include hydrological modelling and soil moisture estimation, based on the sensitivity of the microwave to the dielectric properties of the observed surface. The interaction of microwaves with ocean surfaces and ice provides useful data for oceanography and ice monitoring. Operational systems use data from the European SAR system ERS-2 and the Canadian Radarsat programme. In this framework the data is also used for oil slick monitoring and environmental protection. Looking at SAR interferometry there are plenty of interesting examples in the field of natural disaster monitoring and assessment, i.e., earthquakes, volcano eruptions, flooding, *et cetera*.

6.6 Advanced radar processing techniques

Radar data provide a wealth of information that is not only based on a derived intensity image but also on other data properties that measure characteristics of the objects. One example is *radar interferometry*, an advanced processing method that takes advantage of the phase information of the microwave. If you look at two waves that are emitted with a slight offset you obtain a phase difference between the two waves. The offset can be based on two antennas mounted on the same platform (e.g., aircraft or space shuttle) or based on two different orbits/passes. The range difference is calculated by measuring the phase difference between the two backscattered waves received at the antenna. With the knowledge of the position of the platform with respect to the Earth surface the elevation of the object is determined. Phase differences are displayed in so-called *interferograms* where different colours represent variations in height. The interferogram is used to produce digital elevation models. Differential interferometry is based on the creation of two interferograms of successive radar data acquisitions. These interferograms are subtracted from each other in order to illustrate the changes that have occurred. These are useful for change detection, e.g., earthquake damage assessment.

Summary

In this chapter, the principles of imaging radar and its applications have been introduced. The microwave interactions with the surface have been explained to illustrate how radar images are interpreted. Radar sensors measure distances and detect backscattered signal intensities. In radar processing, special attention has to be paid to geometric corrections and speckle reduction for improved interpretation. Radar data have many potential applications in the fields of geology, oceanography, hydrology, environmental monitoring, land use and land cover mapping and change detection.

Questions

The following questions can help to study Chapter 6.

1. List three major differences between optical and microwave remote sensing?

2. What type of information can you extract from imaging radar data?

3. What are the limitations of radar images in terms of visual interpretation?

4. What kind of processing is necessary to prepare radar images for interpretation? Which steps are obligatory and which are optional?

5. Search the Internet for successful applications of radar images from ERS-1/2, Radarsat and other sensors.

Chapter 7

Remote sensing below the ground surface

7.1 Introduction

The foregoing methods have relied on the electromagnetic spectrum in or near the wavelength of visible light and are largely confined in their application to the investigation of reflections and emissions from the Earth's surface. To probe more deeply into the ground, a range of methods that have their origin in the physical or chemical properties of the buried rocks themselves may be employed. These methods are often called 'geophysical methods'. While the methods of data collection may, of necessity, differ from the methods employed in others types of remote sensing, the presentation of geophysical data nowadays uses much of the technology originally developed for remote sensing data *sensu stricto*. To maximize our ability to probe into the depth dimension, the integrated interpretation of all available data is to be recommended, so much is to be gained by a closer integration of remote sensing and geophysics, while appreciating the important physical differences in the origins of the 'images' obtained. General references to this subject are [9, 22].

7.2 Gamma-ray surveys

Gamma radiation (electromagnetic radiation of very short wavelength) arises from the spontaneous radioactive decay of certain naturally occurring isotopes. These gamma rays have sufficient energy to penetrate a few hundred metres of air and so may be detected conveniently from a low-flying aircraft. Their ability to penetrate rock and soil is modest, so only gamma rays from radioactive sources within a few tens of centimetres of the ground surface ever reach the air in significant numbers. As a result, mapping gamma radiation is confined to mapping the shallowest sub-surface, but where the soils are derived directly from the underlying bedrock and where bedrock outcrops directly, gamma rays are useful in mapping large areas for their geology. Where the soil has been deposited from distant origins, gamma radiation from the underlying bedrock is obscured but the method can still reveal interesting features of soil composition and origin that have, so far, been little used. Only three isotopes lead to the emission of gamma rays when they undergo their radioactive decay chain. These are isotopes of the elements Thorium (Th), Uranium (U) and Potassium (K). While potassium is often present in rocks at the level of a few per cent, the abundance of Th and U is usually measured in only parts per million. The energy of a gamma-ray is characteristic of its elemental source. A gamma-ray spectrometer can, then, not only count the number of incoming rays (counts per second) but also, through analysing the energy spectrum of all incoming gamma rays attribute gamma-rays to their source elements and so estimate the abundances of Th, U and K in the source area. This requires suitable precautions and careful calibration. While the abundance of the radio-elements is itself of little interest (except, of course, in uranium exploration), in practice it is found that each rock unit has a relative abundance of Th, U and K that is distinct from that of adjacent rock units. Hence, if the abundance of each of the three elements is imaged as a primary colour (say, Th = green, U = blue and K = red) with appropriate contrast-stretching and the three colours are combined in a visual display, each rock unit appears with its own characteristic hue. The changes in the hue evident in such an image correspond to geological boundaries and so, under favourable circumstances, gamma-ray spectrometer surveys can lead to a kind of 'instant' geological map (Figure 7.1).

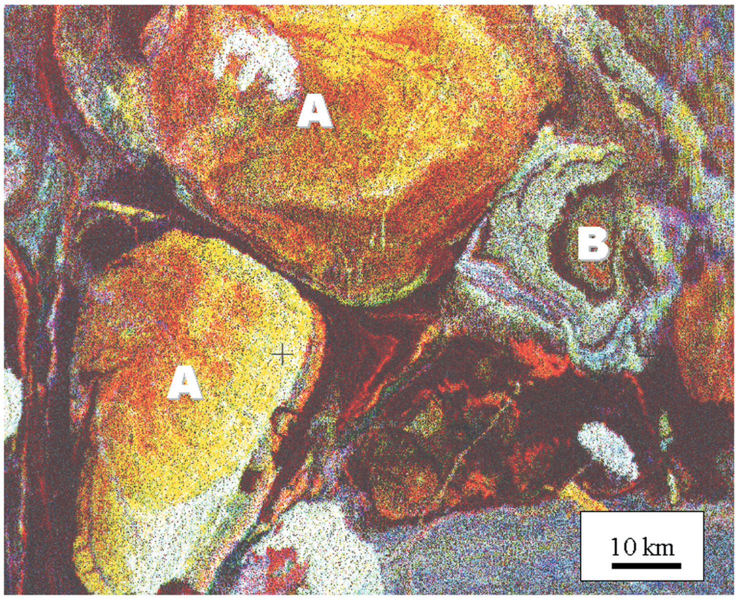


Figure 7.1: Ternary image (K=red, Th=green, U=blue) of an area of Archean geology in NW Australia. Domes of gneisses (A) show up as bright and red-orange in colour largely on account of their potassium content. The layers in an old sedimentary basin or syncline, cut through horizontally by erosion, are visible around B, where each layer has a different hue. Courtesy of the Australian Geological Survey Organization

7.3 Gravity and magnetic anomaly mapping

The Earth has a gravity field and a magnetic field. The former we experience as the weight of any mass and its tendency to accelerate towards the centre of the Earth when dropped. The latter is comparatively weak but is exploited, for example, in the design of the magnetic compass that points towards magnetic north when used in the field. Rocks that have abnormal density or magnetic properties—particularly rocks lying in the uppermost few kilometres of the Earth's crust—distort the broad gravity and magnetic fields of the main body of the Earth by tiny but perceptible amounts, producing local gravity and magnetic *anomalies*. Careful and detailed mapping of these anomalies over any area reveals complex patterns that are related to the structure and composition of the bedrock geology. Both methods therefore provide important 'windows' on the geology, even when it is completely concealed by cover formations such as soil, water, younger sediments and vegetation. The unit of measurement in gravimetry is the milli-Gal (mGal), an acceleration of 10^{-5} m/s^2. The normal acceleration due to gravity (g) is about 9.8 m/s^2 (980,000 mGal) and, to be useful, a gravity survey must be able to detect changes in g as small as 1 mGal or about 1 part per million (ppm) of the total acceleration. This may be achieved easily by reading a gravimeter at rest on the ground surface, but is still at the limit of technical capability from a moving vehicle such as an aircraft.

Conventional gravity surveys are ground- based and therefore slow and costly; the systematic scanning of the Earth's surface by gravity survey is still confined largely to point observations that lack the continuity of coverage achievable with other geophysical methods. An exception is over the world's oceans where radar altimetry of the sea- level surface from a satellite has been achieved with a precision of better than 10 cm. The sea-surface is an equipotential surface with undulations of a few metres in height attributable to gravity anomalies. These arise from density variations in the subsurface, which, at sea, are due mainly to the topography of the ocean floor. Mapping sea-level undulations has therefore made possible the mapping of sea-floor topography at the scale of a 5 km pixel for all the world's oceans (Figure 7.2).

Mapping of magnetic anomalies from low-flying aircraft has been widely used in commercial exploration for 50 years. The Earth's main field has a value that varies between 20,000 and 80,000 nano-Teslas (nT) over the Earth's surface. At a ground clearance of 50 to 100 metres, magnetic anomalies due to rocks are usually no more than a few hundred nT in amplitude. Modern airborne magnetometers can reliably record variations as small as 0.1 nT in an airborne profile, about 2 parts per million of the total field. Each reading takes only 0.1 second,

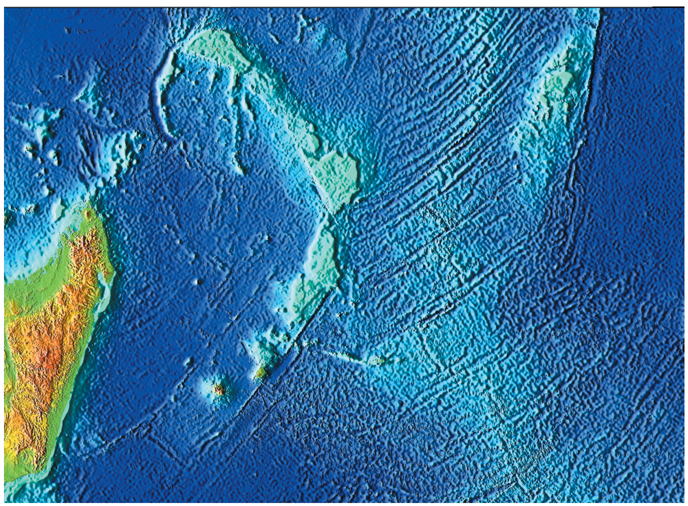


Figure 7.2: Sea floor topography in the western part of the Indian Ocean, revealed by satellite altimetry of the sea surface (GEOSAT). Note the mid-ocean ridges and the transform faults either side of them. Note also the (largely sub-marine) chains of volcanic islands between India and Madagascar (from [26])

corresponding to an interval of about 6 metres on the ground, and a normal survey flight of six hours duration can collect 1500 km of profile, keeping costs low. When ground clearance is only 50 m and flight-lines are closely spaced (200 m), a great deal of geological detail may be revealed by an aeromagnetic survey (Figure 7.3). In the exploration of ancient terrains that have been levelled by weathering and erosion and consequently rendered difficult to map by conventional means, aeromagnetic surveys are invaluable in directing ground exploration to the most promising location for mineral occurrences.

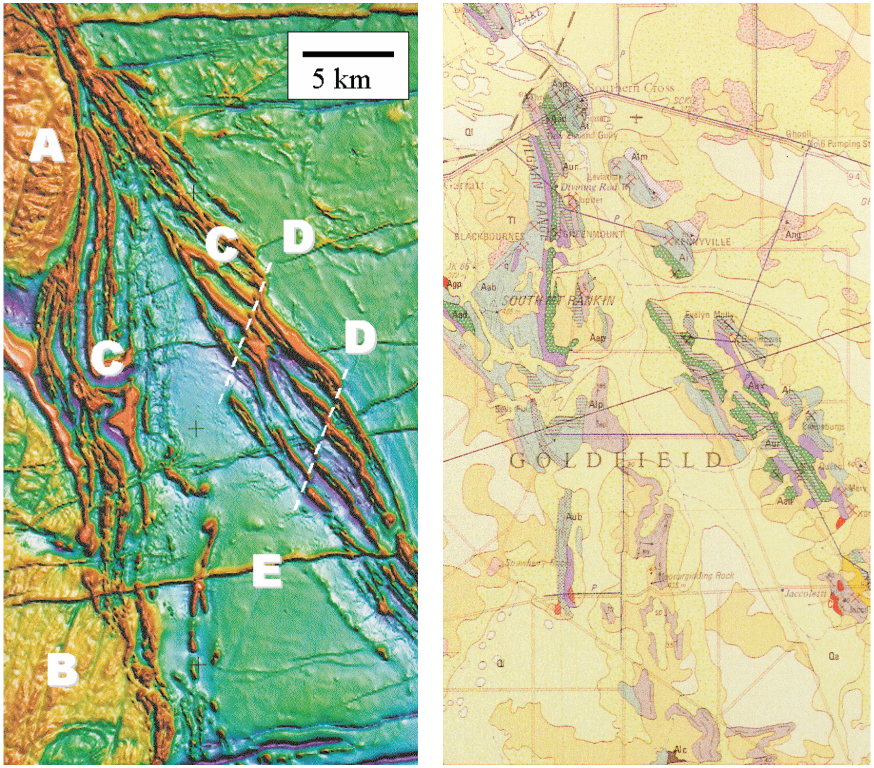
(a)  (b)

Figure 7.3: (a) Magnetic anomaly image of an area of Western Australia. Red = high magnetic values, Blue = low. Granitic bodies such as A and B are distinct from the tightly-folded greenstone rocks seen at C that have some highly magnetic and almost vertical layers within them. Faults offset the greenstones at D. Approximately E–W striking dykes (e.g., E) cut all formations. Note that faults such as D predate the emplacement of the dykes. Courtesy of Tesla Exploration Geophysics; (b) Conventional geological map of the same area. Courtesy of AGSO and GSWA

7.4 Electrical imaging

Solid rocks are normally rather resistive to the passage of electricity. The presence of water (groundwater) in pores, cracks and fissures and the electrical properties of certain minerals nevertheless allow applied currents to flow through the large volume of the subsurface. This has been exploited in methods developed to permit the mapping of subsurface electrical conductivity in two and three dimensions. While seldom of such regional (geological mapping) application as gravity and magnetic methods, electrical methods have found application both in the search for groundwater and in mineral exploration where certain ore minerals have distinctive electrical properties.

Where the ground is stratified an *electrical sounding* can be interpreted to reveal the layering in terms of the resistivity or conductivity of each layer. *Electrical profiling* can be used to reveal lateral variations in rock resistivity, such as often occur across fissures and faults. Ground- based methods that require physical contact between the apparatus and the ground by way of electrodes are supplemented by so-called electromagnetic (EM) methods where current is induced to flow in the ground by the passage of an alternating current (typically of low audio frequency) through a transmitter coil. EM methods require no electrical contact with the ground and can therefore also be operated from an aircraft, increasing the speed of survey and the uniformity of the data coverage. Airborne EM surveys have been developed largely by the mineral exploration community since many important ore bodies—such as the massive sulphide ores of the base metals—are highly conductive and stand out clearly from their host rocks through electrical imaging (Figure 7.4). Other important ore bodies are made up of disseminated sulphides that display an electrochemical property known as chargeability. Mapping of chargeability variations is the objective in *induced* *polarization* (IP) surveys.

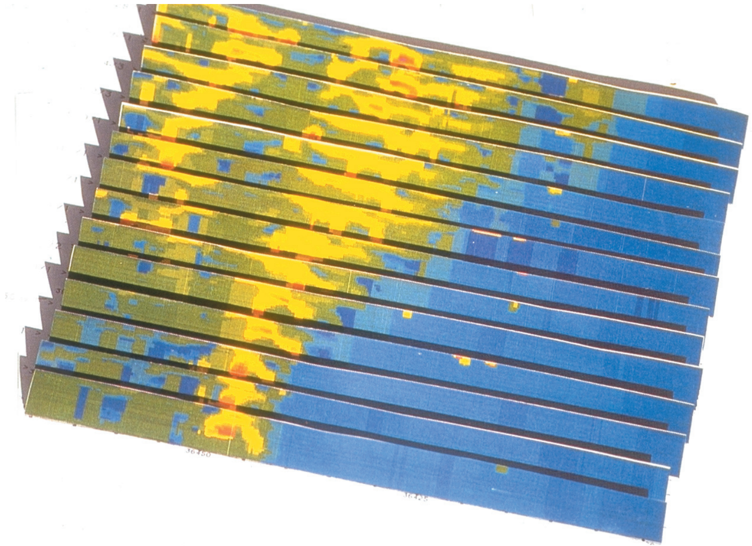


Figure 7.4: Conductivity cross-sections through the Bushman copper sulphide ore body in Botswana derived from airborne EM traverses flown east-west across the strike of the body. Red and yellow colours = highly conductive zones, blue = non-conductive zones. Courtesy of Fugro Airborne Surveys

7.5 Seismic surveying

Virtually all new discoveries of oil and gas are these days made possible by *seismic imaging* of the Earth's subsurface. Such surveys probably account for over 90 per cent of the expenditure on geophysical surveys for all exploration purposes. Seismic waves are initiated by a small explosion or a vibratory source at the surface, in a shallow borehole or in the water above marine areas. Energy in a typically sub-audio frequency range (10 to 100 Hz) radiates from the source and is reflected off changes in acoustic properties of the rock, typically changes in lithology from one stratum to the next, and are detectable from depths of many kilometres.

By deploying a suitable array of seismic sources and receiving reflected energy at a large number of receiving stations known as geophones, an image of the surface may be built up in three dimensions. This involves processing an enormous amount of data to correct for multiple reflections and the geometry of the source-receiver configurations. To achieve the detail necessary for the successful siting of expensive, deep exploratory wells, most surveys now carried out are known as *3D surveys*, though isolated lines of 2D survey, typical of earlier decades, are still carried out for reconnaissance purposes in new areas. The accuracy and precision of the seismic method in mapping the subsurface (Figure 7.5) is now sufficient not only to find trapped oil and gas but also to assess the volume and geometry of the reservoir to plan optimum extraction strategies. Repeated surveys during the production lifetime of a given field (time lapse seismic) permit the draw-down to be monitored and so maximize the recovery of the oil and gas in a field. Similar seismic technology, adapted to more modest scales of exploration, can be applied for shallow investigations (depths of a few tens of metres), useful in groundwater exploration and site investigation.

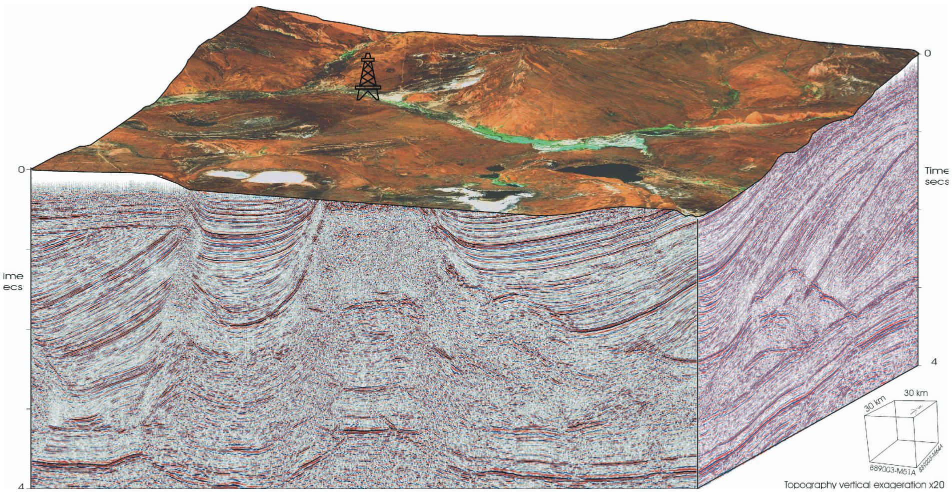


Figure 7.5: 3D seismic surveys map the layering in the subsurface, vital in oil exploration. The top surface of the cube is a satellite image draped on topography (20 times vertical exaggeration). Side faces show seismic sections through the underlying strata.

Summary

Geophysical methods therefore provide a wide range of possible methods of imaging the subsurface. Some are used routinely, others only for special applications. All are potentially useful to the alert geoscientist.

Gravity and magnetic anomaly mapping has been carried out for almost 50 years. While most countries have national programmes, achievements to date are somewhat variable from country to country. The data are primarily useful for *geological reconnaissance* at scales from 1:250,000 to 1:1,000,000. Gamma-ray spectrometry, flown simultaneously with aeromagnetic surveys, has joined the airborne geophysical programmes supporting geological mapping in the past decade. All three methods are therefore used primarily by national geological surveys to support basic geoscience mapping, alongside conventional field and photo-geology, and to set the regional scene for dedicated mineral and oil exploration. It is normal that the results are published at nominal cost for the benefit of all potential users.

Geophysical surveys for *mineral exploration* are applied on those more limited areas (typically at scales 1:50,000 to 1:10,000) selected as being promising for closer (and more expensive!) examination. Typically this might start with an airborne EM and magnetometer survey that would reveal targets suitable for detailed investigation with yet more expensive methods (such as EM and IP) on the ground. Once accurately located in position (x, y) and depth, the most promising anomalies can be tested further by drilling.

Groundwater exploration has historically relied on electrical sounding and profiling but has been supplemented in some cases by EM profiling and sounding and shallow seismic surveys. Regrettably, poor funding usually dictates that such surveys are less thorough and systematic than is the case in mineral exploration, despite the fact that drilling (especially the drilling of non-productive boreholes!) is such an expensive item.

Oil exploration relies almost entirely on detailed seismic surveys, once their location has been selected on the basis of all available geological and regional geophysical data. The surveys are carried out by highly specialized contractors, up to date with the latest technology in this complex and sophisticated industry.

Questions

The following questions can help you to study Chapter 7.

1. Make a list of geophysical maps (and their scales) that you are aware of in your own country (or that part of it you are familiar with).

2. Trace the geophysical features revealed in Figure 7.3(a) on a transparent overlay and compare your result with the geological map in Figure 7.3(b).

The following are typical exam questions:

1. Why is it necessary to use geophysical methods to explore the subsurface? What are the limitations of 'visual' RS methods in this respect?

2. Make a list of the physical properties of rocks that have been used as the basis of geophysical mapping methods.

3. In the process of systematic exploration for Earth resources, why is it important to use inexpensive methods for the reconnaissance of large areas before using more expensive methods over much smaller ones?

Chapter 8

Radiometric aspects

8.1 Introduction

The previous chapters have examined remote sensing as a means of producing image data for a variety of purposes. The following chapters deal with processing of the image data for rectification, visualization and interpretation. The first step in the processing chain, often referred to as pre-processing, involves radiometric and geometric corrections (Figure 8.1). The radiometric aspects are dealt with in this chapter and the geometric aspects in the following. Two groups of radiometric corrections are identified:

- the 'cosmetic' rectification to compensate for data errors (Section 8.2), and
- the atmospheric corrections to compensate for the effect of atmospheric and illumination parameters, such as haze, sun angle and skylight on the image data (Section 8.3).

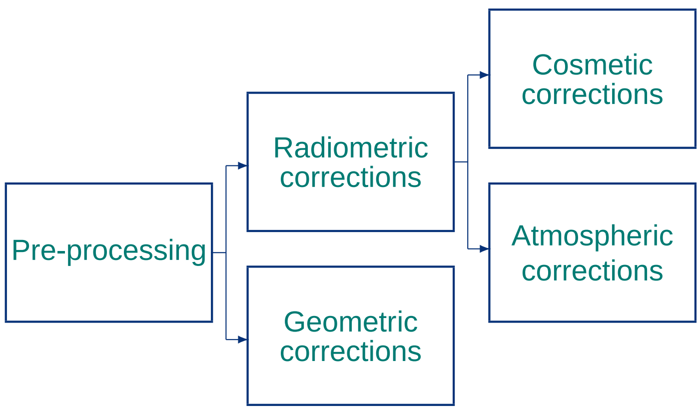
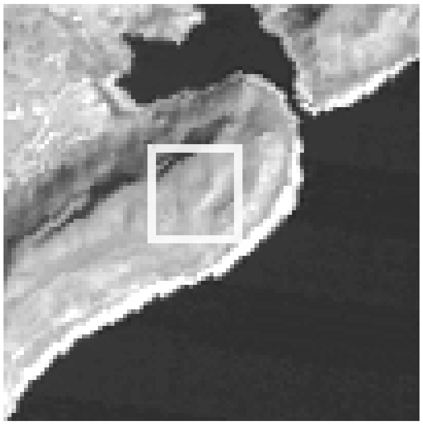


Figure 8.1: Image pre-processing steps

8.2 Cosmetic corrections

Cosmetic corrections involve all those operations that are aimed at correcting visible errors and noise in the image data. Defects in the data may be in the form of periodic or random missing lines (line dropouts), line striping, and random or spike noise. These effects can be identified visually and automatically.

(a) 

34	27	20	17	17	19	20	21	22	21	19	18	16	17	22
28	21	16	15	17	18	18	18	20	19	16	14	14	17	24
23	17	14	14	15	16	15	15	17	16	13	12	14	19	28
17	14	13	12	13	12	11	12	14	13	11	12	17	25	34
13	12	12	11	11	11	10	12	16	15	13	16	23	31	40
11	11	11	11	11	11	12	15	22	23	21	24	31	39	46
10	11	11	11	11	12	16	20	28	31	29	32	39	46	51
10	11	11	11	13	16	22	28	36	39	39	42	47	52	55
10	11	11	13	17	23	30	36	42	45	45	48	51	55	57
10	11	13	17	25	33	40	45	47	48	48	50	53	55	57
15	18	21	26	33	40	47	50	50	50	50	51	54	55	57
25	31	36	40	42	46	49	50	50	50	50	51	54	55	57
33	40	45	47	48	49	51	51	50	50	50	52	54	55	57
39	44	47	49	50	51	53	53	51	50	50	52	54	55	57
43	47	49	50	51	52	54	54	52	51	51	51	53	55	57
46	48	50	51	53	54	55	54	52	51	51	51	53	55	57
48	50	52	53	55	56	56	55	53	51	50	50	52	55	57
50	52	54	55	57	58	58	56	54	51	50	50	52	55	58
51	53	55	57	58	59	59	57	55	52	49	49	51	54	59
52	54	56	58	58	59	59	57	55	52	49	49	51	54	59
53	55	57	59	58	58	58	57	55	52	48	49	52	55	61

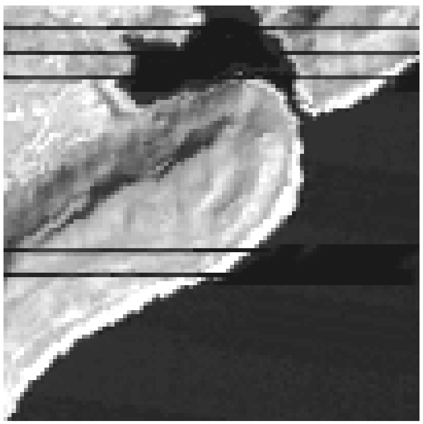
(b)

Figure 8.2: Simulated Landsat MSS image of a coastal area (a) and the corresponding Digital Numbers (DN) of a subset of it (b).

8.2.1 Periodic line dropouts

Periodic line dropouts occur due to recording problems when one of the detectors of the sensor in question either gives wrong data or stops functioning. The Landsat Thematic Mapper, for example, has 16 detectors in all its bands except the thermal band. A loss of one of the detectors would result in every sixteenth scan line being a string of zeros that would plot as a black line on the image. For the Landsat MSS, this defect would occur in every sixth line (see Figure 8.3).

(a)



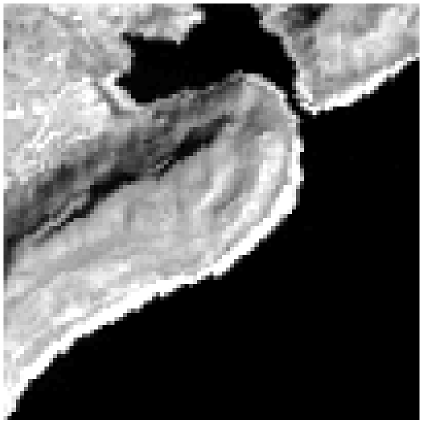
34	27	20	17	17	19	20	21	22	21	19	18	16	17	22
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	17	14	14	15	16	15	15	17	16	13	12	14	19	28
17	14	13	12	13	12	11	12	14	13	11	12	17	25	34
13	12	12	11	11	11	10	12	16	15	13	16	23	31	40
11	11	11	11	11	11	12	15	22	23	21	24	31	39	46
10	11	11	11	11	12	16	20	28	31	29	32	39	46	51
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	11	11	13	17	23	30	36	42	45	45	48	51	55	57
10	11	13	17	25	33	40	45	47	48	48	50	53	55	57
15	18	21	26	33	40	47	50	50	50	50	51	54	55	57
25	31	36	40	42	46	49	50	50	50	50	51	54	55	57
33	40	45	47	48	49	51	51	50	50	50	52	54	55	57
39	44	47	49	50	51	53	53	51	50	50	52	54	55	57
43	47	49	50	51	52	54	54	52	51	51	51	53	55	57
46	48	50	51	53	54	55	54	52	51	51	51	53	55	57
48	50	52	53	55	56	56	55	53	51	50	50	52	55	57
50	52	54	55	57	58	58	56	54	51	50	50	52	55	58
51	53	55	57	58	59	59	57	55	52	49	49	51	54	59
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
53	55	57	59	58	58	58	57	55	52	48	49	52	55	61

(b)

Figure 8.3: The image with line dropouts (a) and the DN-values (b).

The first step in the restoration process is to calculate the average DN-value per scan line for the entire scene. The average DN-value for each scan line is then compared with this scene average. Any scan line deviating from the average by more than a designated threshold value is identified as defective. In regions of very diverse land cover, better results can be achieved by considering the histogram for sub-scenes and processing these sub-scenes separately.

The next step is to replace the defective lines. For each pixel in a defective line, an average DN is calculated using DNs for the corresponding pixel in the preceding and succeeding scan lines. The average DN is then substituted for the defective pixel. The resulting image is a major improvement, although every sixteenth scan line (or every sixth scan line, in case of Landsat MSS data) consists of artificial data (see Figure 8.4). This restoration program is equally effective for random line dropouts that do not follow a systematic pattern.

(a)


34	27	20	17	17	19	20	21	22	21	19	18	16	17	22
28	22	17	15	16	17	17	18	19	18	16	15	15	18	25
23	17	14	14	15	16	15	15	17	16	13	12	14	19	28
17	14	13	12	13	12	11	12	14	13	11	12	17	25	34
13	12	12	11	11	11	10	12	16	15	13	16	23	31	40
11	11	11	11	11	11	12	15	22	23	21	24	31	39	46
10	11	11	11	11	12	16	20	28	31	29	32	39	46	51
10	11	11	12	14	17	23	28	35	38	37	40	45	50	54
10	11	11	13	17	23	30	36	42	45	45	48	51	55	57
10	11	13	17	25	33	40	45	47	48	48	50	53	55	57
15	18	21	26	33	40	47	50	50	50	50	51	54	55	57
25	31	36	40	42	46	49	50	50	50	50	51	54	55	57
33	40	45	47	48	49	51	51	50	50	50	52	54	55	57
39	44	47	49	50	51	53	53	51	50	50	52	54	55	57
43	47	49	50	51	52	54	54	52	51	51	51	53	55	57
46	48	50	51	53	54	55	54	52	51	51	51	53	55	57
48	50	52	53	55	56	56	55	53	51	50	50	52	55	57
50	52	54	55	57	58	58	56	54	51	50	50	52	55	58
51	53	55	57	58	59	59	57	55	52	49	49	51	54	59
52	54	56	58	58	58	58	57	55	52	48	49	51	54	60

(b)

Figure 8.4: The image after correction for line dropouts (a) and the DN-values (b).

8.2.2 Line striping

Line striping is far more common than line dropouts are. Line striping often occurs due to non-identical detector response. Although the detectors for all satellite sensors are carefully calibrated and matched before the launch of the satellite, with time the response of some detectors may drift to higher or lower levels. As a result, every scan line recorded by that detector is brighter or darker than the other lines (see Figure 8.5). It is important to understand that valid data are present in the defective lines, but these must be corrected to match the overall scene.

(a)

34	27	20	17	17	19	20	21	22	21	19	18	16	17	22
96	92	92	90	94	88	86	86	94	98	92	88	87	84	88
23	17	14	14	15	16	15	15	17	16	13	12	14	19	28
17	14	13	12	13	12	11	12	14	13	11	12	17	25	34
13	12	12	11	11	11	10	12	16	15	13	16	23	31	40
11	11	11	11	11	11	12	15	22	23	21	24	31	39	46
10	11	11	11	11	12	16	20	28	31	29	32	39	46	51
20	22	22	22	26	32	44	56	72	78	78	84	94	95	96
10	11	11	13	17	23	30	36	42	45	45	48	51	55	57
10	11	13	17	25	33	40	45	47	48	48	50	53	55	57
15	18	21	26	33	40	47	50	50	50	50	51	54	55	57
25	31	36	40	42	46	49	50	50	50	50	51	54	55	57
33	40	45	47	48	49	51	51	50	50	50	52	54	55	57
78	88	93	95	95	97	98	98	96	94	94	96	98	99	99
43	47	49	50	51	52	54	54	52	51	51	51	53	55	57
46	48	50	51	53	54	55	54	52	51	51	51	53	55	57
48	50	52	53	55	56	56	55	53	51	50	50	52	55	57
50	52	54	55	57	58	58	56	54	51	50	50	52	55	58
51	53	55	57	58	59	59	57	55	52	49	49	51	54	59
89	91	99	99	99	99	99	97	96	94	98	98	96	94	94
53	55	57	59	58	58	58	57	55	52	48	49	52	55	61

(b)

Figure 8.5: The image with line striping (a) and the DN-values (b). Note that the destriped image would look almost similar to the original image.

Though several procedures can be adopted to correct this effect, the most popular is the histogram matching. Separate histograms corresponding to each detector unit are constructed and matched. Taking one response as standard, the gain (rate of increase of DN) and offset (relative shift of mean) for all other detector units are suitably adjusted and new DN-values are computed and assigned.

This yields a destriped image in which all DN-values conform to the reference level and scale.

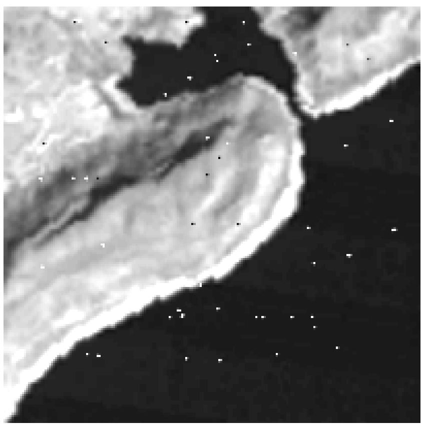
8.2.3 Random noise or spike noise

The periodic line dropouts and striping are forms of non-random noise that may be recognised and restored by simple means. Random noise, on the other hand, requires a more sophisticated restoration method such as digital filtering.

Random noise or spike noise may be due to errors during transmission of data or to a temporary disturbance. Here, individual pixels acquire DN-values that are much higher or lower than the surrounding pixels (Figure 8.6). In the image, these pixels produce bright and dark spots that interfere with information extraction procedures.

A spike noise can be detected by mutually comparing neighbouring pixel values. If neighbouring pixel values differ by more than a specific threshold margin, it is designated as a spike noise and the DN is replaced by an interpolated DN-value.

(a)



34	27	20	17	17	19	20	21	22	21	19	18	16	17	22
28	21	16	15	17	18	18	18	20	204	16	14	14	17	24
23	17	14	14	15	16	15	15	17	16	13	12	14	19	28
17	14	13	12	13	12	11	12	14	13	11	12	17	25	34
13	12	12	11	11	11	10	12	16	15	13	16	23	31	40
11	11	11	11	11	11	12	15	22	23	21	24	31	39	46
10	11	180	11	11	12	16	20	28	31	29	32	39	46	51
10	11	11	11	13	16	22	28	36	39	39	42	47	52	55
10	11	11	13	17	23	30	36	42	45	45	48	51	55	57
10	11	13	17	25	33	40	45	47	48	48	50	53	55	57
15	18	21	26	33	40	47	50	50	50	50	51	54	55	57
25	31	36	40	42	46	49	50	50	50	50	51	54	55	57
33	40	45	47	48	49	51	51	50	50	8	52	54	55	57
39	44	47	49	50	51	53	53	51	50	50	52	54	55	57
43	47	49	50	51	52	54	54	52	51	51	51	53	55	57
46	48	50	51	53	54	55	54	52	51	51	51	53	55	57
48	50	52	53	55	0	56	55	53	51	50	50	52	55	57
50	52	120	55	57	58	58	56	54	51	50	50	52	55	58
51	53	55	57	58	59	59	57	55	52	49	49	51	54	59
52	54	56	58	58	59	59	57	55	52	49	49	51	2	59
53	55	57	59	58	58	58	57	55	52	48	49	52	55	61

(b)

Figure 8.6: The image with spike errors (a) and the DN-values (b).

8.3 Atmospheric Corrections

All reflected and emitted radiations leaving the Earth's surface are attenuated mainly due to absorption and scattering by the constituents in the atmosphere (refer to Section 2.1). The atmospheric induced distortions occur twice in case of sunlight reflection and once in case of emitted radiation. These distortions are wavelength dependent. Their effect on remote sensing data can be reduced by applying 'atmospheric correction' techniques. These corrections are related to the influence of

- haze,
- sun angle, and
- skylight.

8.3.1 Haze correction

Light scattered by the atmospheric constituents that reaches the sensor constitutes the 'haze' in remote sensing image data. Haze has an additive effect resulting in higher DN-values and a decrease in the overall contrast in the image data. The effect is wavelength dependent, being more pronounced in the shorter wavelength range and negligible in the infrared.

Haze corrections are based on the assumption that the infrared bands are essentially free of atmospheric effects and in these bands black bodies, such as large clear water bodies and shadow zones, will have zero DN-value. The DN-values in other bands for the corresponding pixels can be attributed to haze and should be subtracted from all pixels of the corresponding band.

8.3.2 Sun angle correction

The position of the sun, relative to the earth, changes depending on the time of the day and the day of the year. In the northern hemisphere, the solar elevation angle is smaller in winter than in summer (see Figure 8.7). As a result, the image data of different seasons are acquired under different solar illumination. Sun angle correction becomes more important when one wants to generate mosaics taken at different times or perform change detection studies.

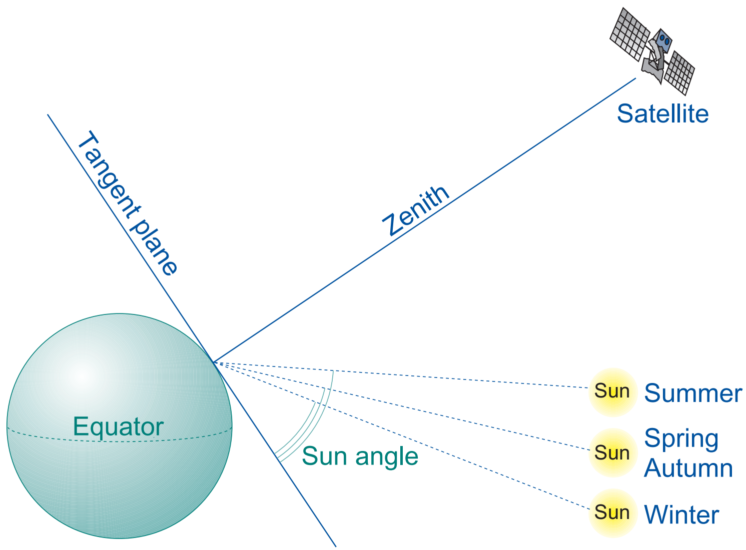


Figure 8.7: Effects of seasonal changes on the solar elevation angle in the northern hemisphere (after Lillesand and Kiefer, 1994).

An absolute correction involves dividing the DN-values in the image data by the sine of the solar elevation angle (the size of the angle is given in the header of the image data), as per the following formula.

$$DN' = \frac{DN}{\sin(\alpha)} \tag{8.1}$$

Here, DN is the input pixel value, DN' is the output pixel value, and α is the solar elevation angle. Note that since the angle is smaller than 90°, the sine will be smaller than 1 and DN' will be larger than DN.

When multitemporal data sets of the same area are available, a 'relative' sun angle correction can be performed. In such cases, the image with higher sun elevation angle is taken as a reference and the radiometric values of the other image are adjusted to it.

8.3.3 Skylight correction

Scattered light reaching the sensor after being reflected from the Earth's surface constitutes the 'skylight' or 'sky irradiance'. This also causes reduced contrast of image data. Correcting for this effect requires additional information that cannot be extracted from the image data itself. This information (e.g., aerosol distribution, gas composition) is difficult to obtain and needs to be later input into a numerical model, the details of which are beyond the scope of this chapter.

Caution: Some researchers use published data for 'standard atmospheres' as input into the numerical model to perform atmospheric corrections. Such corrections may introduce inaccuracies, the extent of which may not be assessable.

Summary

Radiometric corrections constitute an important step in the pre-processing of remotely sensed data. They comprise of cosmetic corrections to reduce the influence of atmospheric and illumination parameters. Atmospheric corrections are particularly important for generating image mosaics and for comparing multitemporal remote sensing data. However, such corrections should be applied with care, after understanding the physical principles behind these corrections.

Questions

The following questions can help to study Chapter 8.

1. Should radiometric corrections be performed before or after geometric corrections, and why?

2. Why is the effect of haze more pronounced in shorter wavelength bands?

3. In 'relative' sun angle correction, why is it preferred to take the image with a larger sun angle as the reference image?

4. In a change detection study, if there were images from different years but from the same season, would it still be necessary to perform atmospheric corrections? Why or why not?

The following are typical exam questions:

1. Which are the radiometric errors that can be introduced due to malfunctioning of satellite sensors?

2. What are the differences between line dropouts and line striping?

3. Explain the procedure for haze correction on remotely sensed data.

Chapter 9

Geometric aspects

9.1 Introduction

The discussions of the various sensors in chapters 4, 5, and 6 have indicated some of the geometric characteristics inherent to different type of image data. These geometric characteristics have to be taken into consideration when the data are used:

- To derive two-dimensional (x, y) and three-dimensional (x, y, z) coordinate information. 2D geometric descriptions of objects (points, lines, areas) can be derived from a single image or photo. 3D geometric descriptions (2.5D terrain relief, 3D objects as volumes) can be derived from stereo pairs of images or photos. Extraction of 3D information from images requires a specific process called orientation.

- To merge different types of image data for integrated processing and analysis. Consider a land cover classification based on multispectral Landsat and SPOT data. Both data sets need to be converted into the same geometric grid before they can be processed simultaneously. This can be achieved by a geocoding process.

- To visualize the image data in a GIS environment. There is a growing amount of image data available that are used as a backdrop for other (vector stored) data. To enable such integration, the image data need to be georeferenced to the coordinate system of the vector stored data.

Traditionally, the subjects of georeferencing, geocoding and orientation have been addressed by the discipline of *photogrammetry*. Today, not only aerial photographs but also various other types of image data are used in the field of *digital* *photogrammetry* [25]. Deriving 3D measurements from radar data is in the discipline of *interferometry and radargrammetry*. This chapter deals with a limited number of concepts and topics. In Section 9.2 relief displacement is introduced. Section 9.3 and Section 9.4 introduce 2D and 3D approaches respectively. 2D approaches relate to methods and techniques that neglect terrain elevation differences. In 3D approaches, methods for correction of relief displacement are introduced. If a DTM is not available, 3D coordinates can be measured directly from stereo pairs. Note that spatial referencing in general and an introduction to map projections have been introduced in *Principles of GIS* [6].

9.2 Relief displacement

A characteristic of most sensor systems is the distortion of the geometric relationship between the image data and the terrain caused by relief differences on the ground. This effect is most apparent in aerial photographs and airborne scanner data. The effect of relief displacement is illustrated in Figure 9.1. Consider the situation on the left in which a true vertical aerial photograph is taken of a flat terrain. The distances $(A - B)$ and $(a - b)$ are proportional to the total width of the scene and its image on the negative, respectively. In the left hand situation, by using the scale factor, we can compute $(A-B)$ from a measurement of $(a - b)$ in the negative. In the right hand situation, there is significant terrain relief difference. As you can now observe, the distance between a and b in the

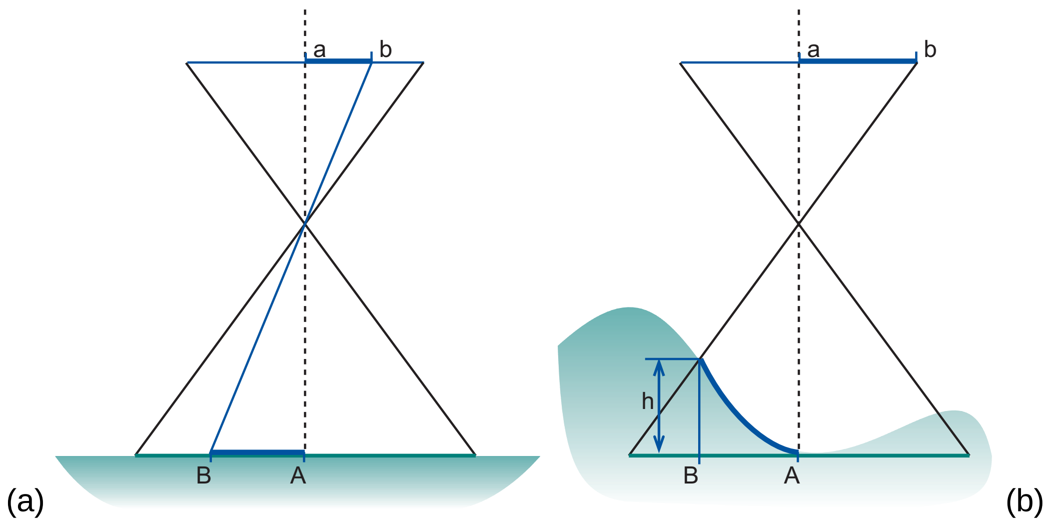


Figure 9.1: Illustration of the effect of terrain topography on the relationship between A–B (on the ground) and a–b (on the photograph). Flat terrain (a); significant height difference (b).

negative has become larger, although when measured in the terrain system, it is still the same as in the left hand situation. This phenomenon does not occur in the centre of the photo but becomes increasingly prominent towards the edges of the photo. This effect is called *relief displacement*: terrain points whose elevation is above or below the reference elevation are displaced respectively away from or towards the nadir point. The magnitude of displacement, δr (mm), is approximated by:

$$\delta r = \frac{r \times h}{H} \tag{9.1}$$

In this equation, r is the radial distance (mm) from the nadir, h (m) is the height of the terrain above the reference plane and H (m) is the flying height above the reference plane (where nadir intersects the terrain). The equation shows that the amount of relief displacement is zero at the nadir ($r = 0$), greatest at the corners of the photograph and, is inversely proportional to the flying height.

In addition to relief displacement you can imagine that also buildings and other tall objects can cause displacement (height displacement). This effect is, for example, encountered when dealing with large scale photos of urban or forest areas (Figure 9.2). In this chapter the subject of height displacement is not further elaborated.

The main effect of relief displacement is that inaccurate or wrong coordinates might be determined when for example digitizing from image data. Whether relief displacement should be considered in the geometric processing of the image data depends on its impact on the required accuracy of the geometric information derived from the images. Relief displacement can be corrected for if information on the terrain topography is available (in the form of a DTM). The procedure is explained in Section 9.4.

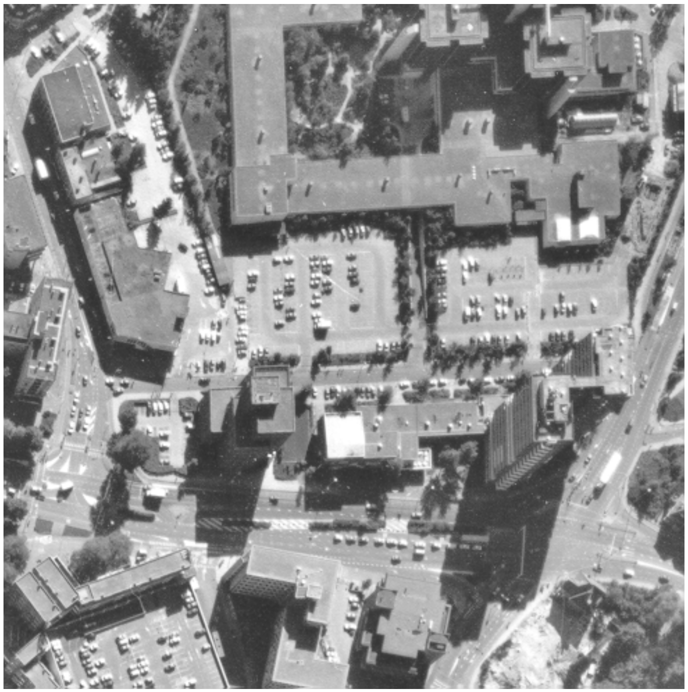


Figure 9.2: Fragment of a large scale aerial photograph of the centre of Enschede. Note the effect of height displacement on the higher buildings.

9.3 Two-dimensional approaches

In this section we consider the geometric processing of image data in situations where relief displacement can be neglected. Examples of such image data are a digitized aerial photograph of a flat area (flat being defined by $h/H < 1/1000$). For images taken from space with only a medium resolution, the relief displacement usually is less than a few pixels and thus not important. These data are stored in a column-row system in which columns and rows are indicated by index i and j respectively. The objective is to relate the image coordinate system to a specific map coordinate system (Figure 9.3).

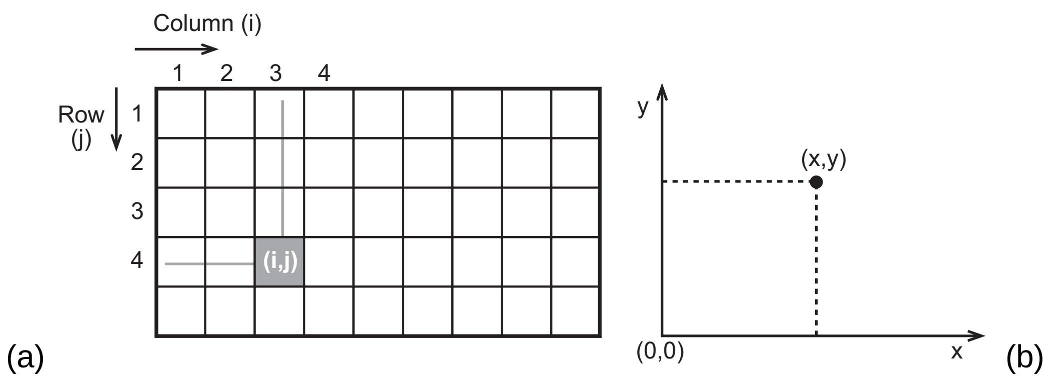


Figure 9.3: Coordinate system of the image defined by rows and columns (a) and map coordinate system with x- and y-axes (b).

9.3.1 Georeferencing

The simplest way to link an image to a map projection system is to use a *geometric transformation*. A transformation is a function that relates the coordinates of two systems. A transformation, relating (x, y) to (i, j) is typically defined by linear equations, such as: $x = 3 + 5i$, and, $y = -2 + 2.5j$.

Using the above transformation, for example, image position $(i = 5, j = 8)$ relates to map coordinates $(x = 28, y = 18)$. Once such a transformation has been determined, the map coordinates for each image pixel can be calculated. The resulting image is called a *georeferenced image*. It allows the superimposition of vector data and the storage of the data in map coordinates when applying on-screen digitizing. Note that the image as such remains stored in the original (i, j) raster structure, and that its geometry is not altered.

The process of georeferencing includes two steps: selection of the appropriate type of transformation, and determination of the transformation parameters. The type of transformation depends mainly on the sensor-platform system used. For aerial photographs (of a flat terrain) usually a so-called *perspective transformation* is used to correct for the effect of tilt and roll (Section 3.2). A more general transformation is the polynomial transformation, which enables 1^{st}, 2^{nd} to n^{th} order transformations. In many situations a 1^{st} order transformation is adequate. A 1^{st} order transformation relates map coordinates (x, y) with image coordinates (i, j) as follows:

$$x = a + bi + cj \tag{9.2}$$

$$y = d + ei + fj \tag{9.3}$$

Equations 9.2 and 9.3 require that six parameters $(a \ldots f)$ be determined. The transformation parameters can be determined by means of *ground control points*

GCP	i	j	x	y	x_c	y_c	d_x	d_y
1	254	68	958	155	958.552	154.935	0.552	-0.065
2	149	22	936	151	934.576	150.401	-1.424	-0.599
3	40	132	916	176	917.732	177.087	1.732	1.087
4	26	269	923	206	921.835	204.966	-1.165	-1.034
5	193	228	954	189	954.146	189.459	0.146	0.459

Table 9.1: A set of five ground control points, which are used to determine a 1^{st} order transformation. x_c and y_c are calculated using the transformation, d_x and d_y are the residual errors.

(GCPs). GCPs are points that can be clearly identified in the image and in a source, which is in the required map projection system. One possibility is to use topographical maps of an adequate scale. The operator then needs to identify identical points on both sources e.g., using road crossings, waterways, typical morphological structures, *et cetera*. Another possibility is to identify points in the image and to measure the coordinates in the field by satellite positioning. The result of GCP selection is a set of related points (Table 9.1). To solve the above equations, at least three GCPs are required. However, to calculate the error of the transformation more points are required. Based on the set of GCPs the computer calculates the optimal transformation parameters using a best fit procedure. The errors that remain after transformation are called residual errors. Their magnitude is an indicator of the quality of the transformation.

Table 9.1 shows an example in which five GCPs have been used. Based on the given (i, j) and (x, y), one can determine the following transformation:

$$x = 902.76 + 0.206i + 0.051j,$$

and

$$y = 152.579 - 0.044i + 0.199j.$$

Using these equations, you can calculate the image coordinates for each individual GCP. In doing so, you will not exactly find the x and y values in the table because you calculate with three decimals only. Furthermore, a difference remains between the calculated values, (x_c, y_c), and true values, (x, y), because the transformation has been optimized to give the best fit for the whole set of GCPs. The resulting differences (residual errors) are listed by d_x and d_y. The residual errors can be used to analyse which GCPs have the largest contribution to the errors. This may indicate, for example, a GCP that has been inaccurately identified.

The overall accuracy of a transformation is usually expressed by the *Root Mean Square Error* (RMS error), which calculates a mean value from the individual residuals. The RMS error in x-direction, m_x, is calculated using the following equation:

$$m_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \delta {x_i}^2} \tag{9.4}$$

For the y-direction, a similar equation can be used to calculate m_y. The overall error, m_{total}, is calculated by:

$$m_{total} = \sqrt{{m_x}^2 + {m_y}^2} \tag{9.5}$$

For the example data set given in Table 9.1, the residuals have been calculated. The respective m_x, m_y and m_{total} are 1.159, 0.752 and 1.381. The RMS error is a quantitative method to check the accuracy of the transformation. However, the RMS error does not take into account the spatial distribution of the GCPs. The RMS error is only valid for the area that is bounded by the GCPs. In the selection of GCPs, therefore, you should preferably include points located near the edges of an image.

9.3.2 Geocoding

The previous section explained that two-dimensional coordinate systems, for example an image system and a map projection system, can be related using geometric transformations. This georeferencing approach is useful in many situations. However, in other situations a *geocoding* approach, in which the image grid is also transformed, is required. Geocoding is required when different images need to be combined or when the image data are used in a GIS environment that requires all data to be stored in the same map projection. The effect of georeferencing and geocoding is well illustrated by Figure 9.5.

Geocoding is georeferencing with subsequent *resampling* of the image raster. This means that a new image raster is defined along the xy-axes of the selected map projection. The geocoding process comprises two main steps: first each new raster element is projected (using the transformation parameters) onto the original image, secondly a (DN) value for the new pixel is determined and stored.

As the orientation and size of original (input) and required (output) raster are different, there is no exclusive one-to-one relationship between elements (pixels) of these rasters. Therefore, *interpolation methods* are required to make a decision regarding the new value of each pixel. Various resampling algorithms are available (Figure 9.4). The main methods are: nearest neighbour, bilinear interpolation and cubic convolution. In *nearest neighbour*, the pixel value is assigned the nearest value in the original image. Using *bilinear interpolation* the weighted mean is calculated over the four nearest pixels in the original image. *Cubic convolution* applies a polynomial approach based on the values of 16 surrounding pixels. The choice of the resampling algorithm depends, among others, on the ratio between input and output pixel size and the purpose of the resampled image data.

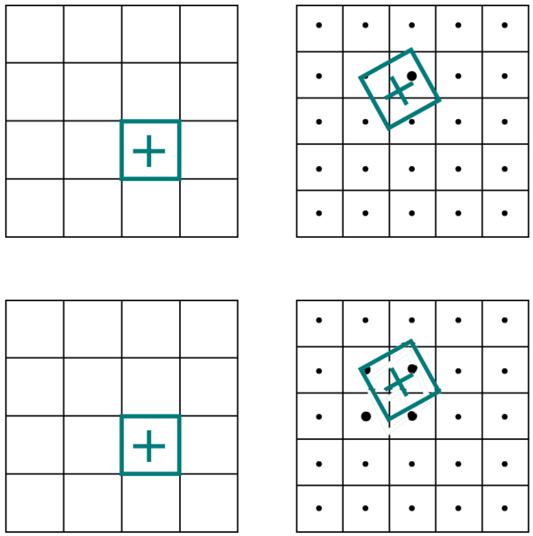


Figure 9.4: Principle of resampling: nearest neighbour (upper) and bilinear interpolation (lower).

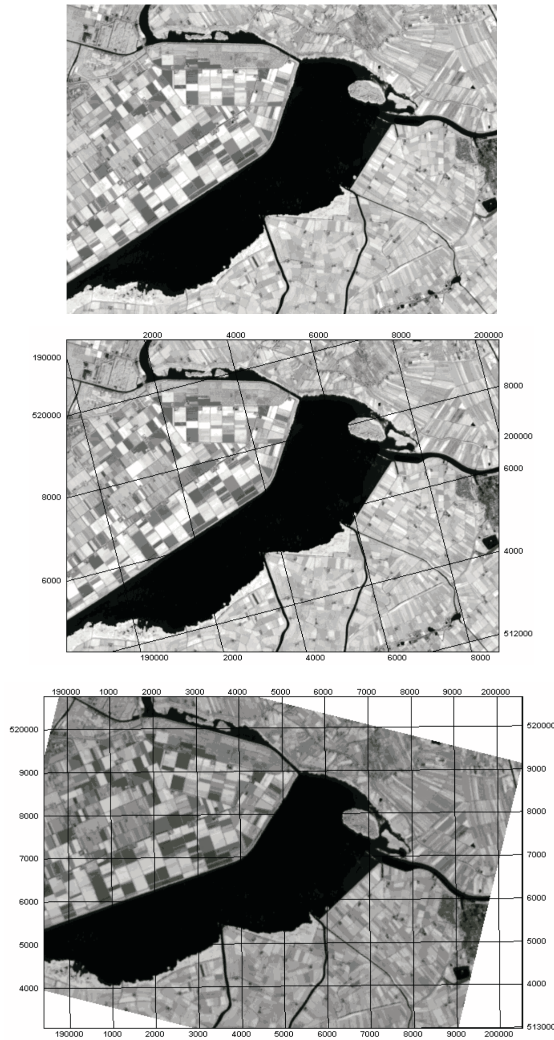


Figure 9.5: Original, georeferenced and geocoded satellite image of a part of Flevoland

9.4 Three-dimensional approaches

Unlike the previous section, which was concerned only with 2D coordinates, in this section also the 3^{rd} dimension is relevant. The following processes will be explained in the subsequent sections:

- *Monoplotting*, which is an approach to correct for terrain relief *during* digitizing of terrain features from aerial photos that results in relatively accurate (x, y) coordinates.

- *Orthoimage production*, which is an approach to correct image data for terrain relief and store the image in a specific map projection. Orthoproducts can be used as a backdrop to other data or, used to directly determine the (x, y) geometry of the features of interest. Today, many organizations and companies sell orthoproducts.

- *Stereoplotting*, which is used to extract 3D information from stereo pairs. Stereo pairs allow for the determination of (x, y, z) positions. 3D information extraction requires an orientation process. Examples of 3D products are a Digital Terrain Model (DTM) and large-scale databases, for instance, related to urban constructions, roads, or cadastral parcels.

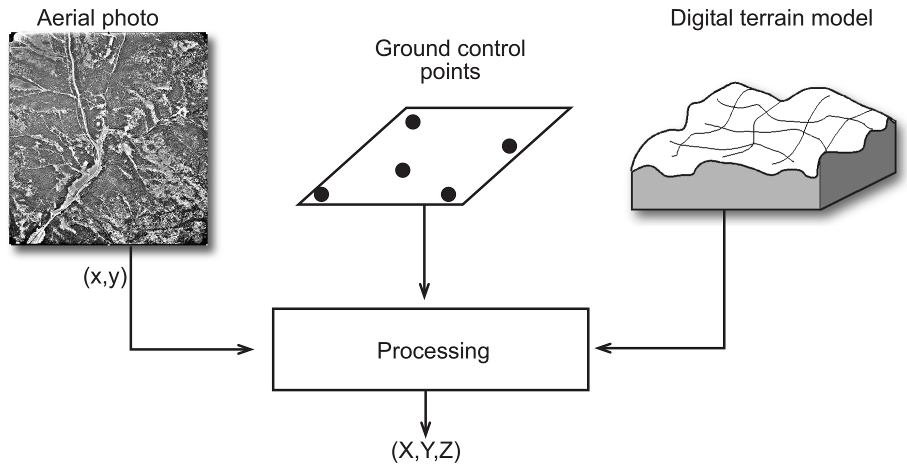


Figure 9.6: The process of digital monoplotting enables accurate determination of terrain coordinates from a single aerial photograph.

9.4.1 Monoplotting

Suppose you need to derive accurate positions from an aerial photograph expressed in a specific map projection. This can be achieved for a flat terrain using a true vertical photograph and a georeferencing approach. However, if there are significant terrain relief differences you need to correct for relief displacement. For this purpose the method of *monoplotting* has been developed.

Monoplotting is based on reconstruction of the position of the camera at the moment of exposure relative to a Digital Terrain Model (DTM) of the terrain. This is achieved by identification of a number (at least four) of ground control points (GCPs) for which both the photo and map coordinates must be known. The applied DTM should be stored in the required map projection system and the heights should be expressed in an adequate vertical reference system. When digitizing features from the photograph, the computer uses the DTM to calculate the effect of relief displacement and corrects for it. A monoplotting approach is possible by using a hard-copy image on a digitizer tablet, or by on-screen digitizing from a digital image. In the latter situation, vector-stored information can be superimposed over the image to update the changed features. Note that monoplotting is a (real time) correction procedure and does not yield new image data.

9.4.2 Orthoimage production

Monoplotting can be considered a georeferencing procedure that incorporates corrections for relief displacement. For some applications, however, it is useful to correct the photograph itself. In such cases the photo should be transformed and resampled into a product with the geometric properties of a specific map projection. Such a photo is called an *orthophoto*. Similarly, other types of image data can be converted into *orthoimages*.

The production of orthophotos is quite similar to the process of monoplotting. Consider a digitised aerial photograph. First, the photo is oriented using ground control points. The terrain height differences are modelled by a DTM. The computer then calculates the position in the original photo for each output element (pixel). Using a resampling algorithm, the output value is determined and stored in the required raster.

In the past, special optical instruments were employed for orthophoto production. Their use required substantial effort. Nowadays, application of digital image data together with digital photogrammetric software enables easy production of orthophotos.

9.4.3 Stereoplotting

The basic process of stereoplotting is to form a *stereo model* of the terrain and to digitize features by measurements made in this *stereo model*. A stereo model is a special combination of two photographs of the same area taken from different positions. For this purpose, aerial photographs are usually flown with 60% overlap between subsequent photos. Stereo pairs can also be derived from other sensors such as multispectral scanners and imaging radar.

The *measurements* made in a stereo model refer to a phenomenon that is called *parallax*. Parallax refers to the fact that an object photographed from different positions has different relative positions in the two images. Since this effect is directly related to the relative height, measurement of these parallax differences yield height information. (Figure 9.7).

A stereo model enables (parallax) measurement using a special (3D) cursor. If the stereo model is appropriately oriented (see below) the parallax measurements yield (x, y, z) coordinates. To view and navigate in a stereo model, various hardware solutions may be used. So-called analogue and analytical plotters were used in the past. These instruments are called 'plotters' since the features delineated in the stereo model were directly plotted onto film (for reproduction). Today, digital photogrammetric workstations (DPWs) are increasingly used. Stereovision, the impression of depth, in a DPW is realized using a dedicated combination of monitor and special spectacles (e.g., polarized).

To form a stereo model for 3D measurements, the stereo model needs to be oriented. The orientation process involves three steps:

1. First, the relation between the film-photo and the camera system is defined. This is the so-called *inner orientation*. It requires identification of the fiducial marks on the photos and the exact focal length.

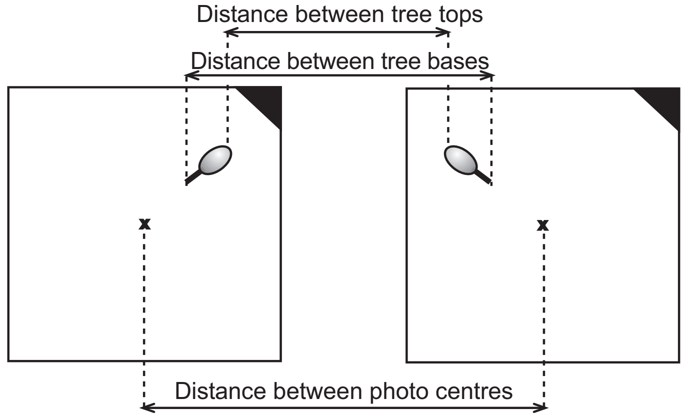


Figure 9.7: The same tree is present in two (overlapping) photographs. Because of the height of the tree the positions of the tree top and base relative to photo centres are different. This difference (parallax) can be used to calculate its height.

2. Second, the relative tilts of the two photographs is determined. This is called the *relative orientation* and requires identification of identical points ('tie points') in both photographs.

3. After the inner and relative orientations, a geometrically correct three-dimensional model is formed. This model must be brought to a known scale and levelled with respect to the horizontal reference datum of the terrain coordinate system. For this purpose 3D ground control points (i.e., x, y and elevation) are identified. This third step is known as the *absolute orientation*.

One of the possibilities in stereoplotting is to superimpose vector data into a stereo model for updating purposes. In this way the operator can limit his/her work to the changed features rather than recording all information.

Summary

This chapter has introduced some general geometric aspects of dealing with image data. A basic principle in dealing with remote sensing sensors is 'terrain relief', which can be neglected (2D approaches) or taken into account (3D approaches). In both approaches there is a possibility to keep the image data stored in their (i, j) system and relate it to other data through coordinate transformations (georeferencing and monoplotting). The other possibility is to change the image raster into a specific map projection system using resampling techniques (geocoding and orthoimage production). A true 3D approach is that of stereoplotting, which applies parallax differences as observed in stereo pairs to measure (x, y, z) coordinates of terrain and objects.

Questions

The following questions can help you to study Chapter 9.

1. Suppose your organization develops a GIS application for road maintenance. What would be the consequences of using georeferenced versus geocoded image data as a backdrop?

2. Think of two situations in which image data are applied and in which you need to take relief displacement into account.

3. For a transformation of a specific image into a specific coordinate system, an m_{total} error of two pixels is given. What additional information do you need to assess the 'quality' of the transformation?

The following are typical exam questions:

1. Compare an image and map coordinate system (give figure with comment).

2. What is the purpose of acquiring stereo pairs of image data?

3. What are ground control points used for?

4. Calculate the map position (x, y) for image position $(10, 20)$ using the two following equations: $x = -10 + 5i - j$ and $y = 5 + 2i + 2j$

5. Explain the purpose of monoplotting. What inputs do you need?

Chapter 10

Image enhancement and visualisation

10.1 Introduction

Many of the figures in the previous chapters have presented examples of remote sensing image data. There is a need to visualize image data at most stages of the remote sensing process. For example, the procedures for georeferencing, explained in Chapter 8, cannot be performed without visual examination to measure the location of ground control points on the image. However, it is in the process of information extraction that visualization plays the most important role. This is particularly so in the case of visual interpretation (Chapter 11), but also during automated classification procedures (Chapter 12).

Because many remote sensing projects make use of multispectral data, this chapter focuses on the visualization of colour imagery. An understanding of how we perceive colour is required at two main stages in the remote sensing process. In the first instance, it is required in order to produce optimal 'pictures' from (multispectral) image data on the computer screen or as a (printed) hard-copy. Thereafter, the theory of colour perception plays an important role in the subsequent interpretation of these pictures. To understand how we perceive colour, Section 10.2 deals with the theory of colour perception and colour definition. Section 10.3 gives the basic principles you need to understand and interpret the colours of a displayed image. The last section (Section 10.5) introduces some filter operations for enhancing specific characteristics of the image.

10.2 Perception of colour

Colour perception takes place in the human eye and the associated part of the brain. Colour perception concerns our ability to identify and distinguish colours, which, in turn, enables us to identify and distinguish entities in the real world. It is not completely known how human vision works, or what exactly happens in the eyes and brain before someone decides that an object is (for example) light blue. Some theoretical models, supported by experimental results are, however, generally accepted. Colour perception theory is applied whenever colours are reproduced, for example in colour photography, TV, printing and computer animation.

10.2.1 Tri-stimuli model

The eye's *general sensitivity* is to wavelengths between 400–700 nm. Different wavelengths in this range are experienced as different colours.

The retinas in our eyes have *cones* (light-sensitive receptors) that send signals to the brain when they are hit by photons with energy levels that correspond to different wavelengths in the visible range of the electromagnetic spectrum. There are three different kinds of cones, responding to blue, green and red wavelengths (Figure 10.1). The signals sent to our brain by these cones, and the differences between them, give us colour-sensations. In addition to cones, we have *rods*, which do not contribute to colour vision. The rods can operate with less light than the cones. For this reason, objects appear less colourful in low light conditions.

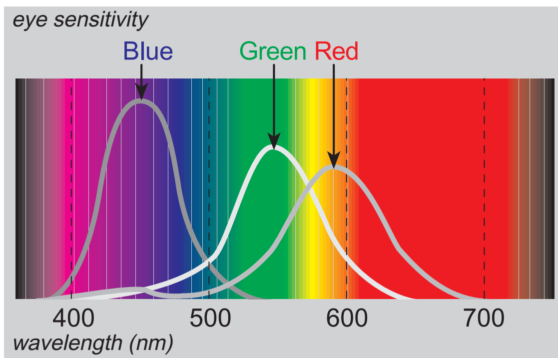


Figure 10.1: Visible range of electromagnetic spectrum including the sensitivity curves of cones in the human eye.

Screens of colour television sets and computer monitors are composed of a large number of small dots arranged in a regular pattern of groups of three: a red, a green and a blue dot. At a normal viewing distance from the screen we cannot distinguish the individual dots Electron-guns for red, green and blue are positioned at the back-end of the tube. The number of electrons fired by these guns at a certain position on the screen determines the amount of (red, green and blue) light emitted from that position. All colours visible on such a screen are therefore created by mixing different amounts of red, green and blue. This mixing takes place in our brain. When we see monochromatic yellow light (i.e., with a distinct wavelength of, say, 570 nm), we get the same impression as when we see a mixture of red (say, 700 nm) and green (530 nm). In both cases, the cones are stimulated in the same way. According to the *tri-stimuli model*, therefore, three different kinds of dots are necessary and sufficient.

10.2.2 Colour spaces

The tri-stimuli model of colour perception is generally accepted. This states that there are three degrees of freedom in the description of a colour. Various three-dimensional spaces are used to describe and define colours. For our purpose the following three are sufficient.

1. Red Green Blue (RGB) space based on the additive principle of colours.
2. Intensity Hue Saturation (IHS) space, which is most related to our, intuitive, perception of colour.
3. Yellow Magenta Cyan (YMC) space based on the subtractive principle of colours.

RGB

The RGB definition of colours is directly related to the way in which computer and television monitors function. Three channels (RGB) directly related to the red, green and blue dots are input to the monitor. When we look at the result, our brain combines the stimuli from the red, green and blue dots and enables us to perceive all possible colours from the visible part of the spectrum. During the combination, the three colours are added. When green dots are illuminated in addition to red ones, we see yellow. This principle is called the *additive colour scheme*. Figure 10.2 illustrates the additive colours caused by bundles of light from red, green and blue spotlights shining on a white wall in a dark room. When only red and green light occurs, the result is yellow. In the central area there are equal amounts of light from all three the spotlights, and we experience 'white'.

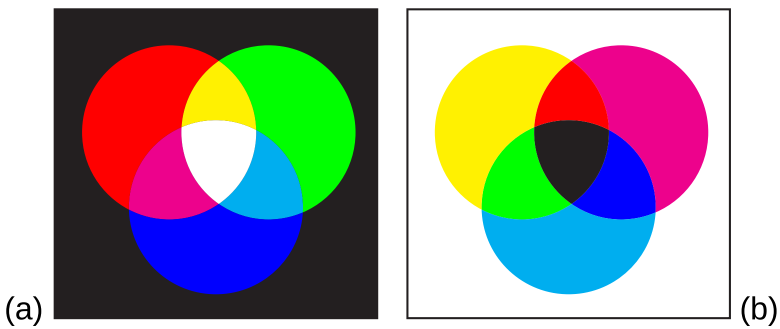


Figure 10.2: Comparison of the (a) additive and (b) subtractive colour schemes

In the additive colour scheme, all visible colours can be expressed as combinations of red, green and blue, and can therefore be plotted in a three-dimensional space with R, G and B along the axes. The space is bounded by minimum and maximum values (intensities) for red, green and blue, defining the so-called

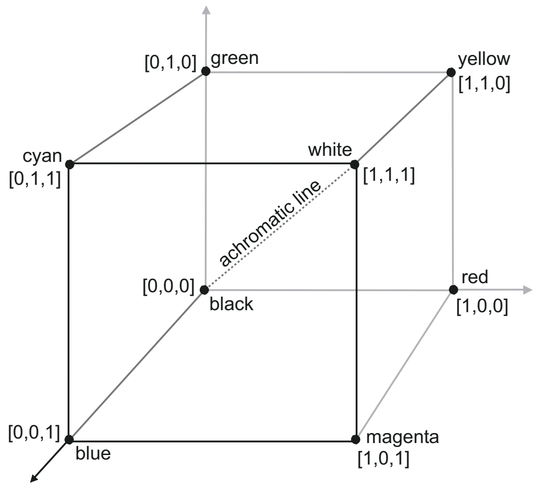


Figure 10.3: The RGB cube; note the red, green and blue corner points

colour cube (Figure 10.3).

IHS

In daily speech we do not express colours in the red, green and blue of the RGB system. The IHS system, which refers to intensity, hue and saturation, more naturally reflects our sensation of colour. *Intensity* describes whether a colour is light or dark. *Hue* refers to the names that we give to colours: red, green, yellow, orange, purple, *et cetera*. *Saturation* describes a colour in terms of pale versus vivid. Pastel colours have low saturation; grey has zero saturation. As in the RGB model, three degrees of freedom are sufficient to describe any colour.

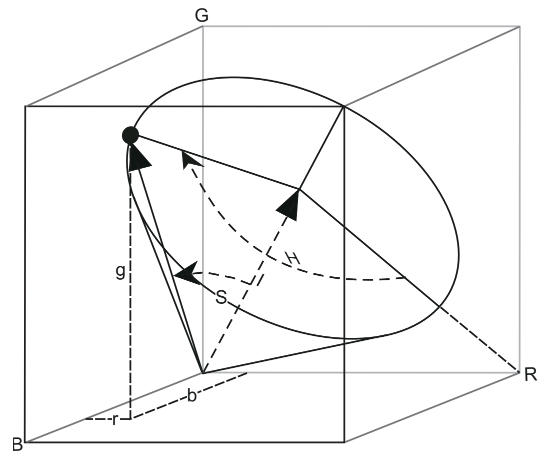


Figure 10.4: Relationship between RGB and IHS colour spaces

Figure 10.4 illustrates the correspondence between the RGB and the IHS system. Although the mathematical model for this description is tricky, the description is, in fact, more natural. For example, 'light, pale red' is easier to imagine than 'a lot of red with considerable amounts of green and blue'. The result, however, is the same. Since the IHS scheme deals with colour perception, which is somewhat subjective, complete agreement about the definitions does not exist. It is safe to define intensity as the sum of the R, G and B values. On the main diagonal of the colour cube (R = G = B), running from black to white, where we find all the grey tones, the saturation equals 0. At the red-green, green-blue and blue-red sides of the cube (these are the sides that contain the origin), the saturation is maximum (100%). If we regard colours as (additive) mixtures of a (saturated) colour and grey, then saturation is the percentage of the colour in the mixture. For example, (R,G,B) = (100, 40, 40) is a mixture of the saturated colour (60, 0, 0) and the grey tone (40, 40, 40). Therefore, its saturation equals 60 / 100 = 60%. To define the hue of a colour, we look in the plane, perpendicular to the main diagonal, that contains the colour. In that plane, we can turn around through 360 degrees, respectively looking towards red, yellow, green, cyan, blue, magenta and back to red. The angle at which we find our colour is its hue.

YMC

Whereas RGB is used in computer and TV display, the YMC colour description is used in colour definition on hard copy, for example printed pictures but also photographic films and paper. The principle of the YMC colour definition is to consider each component as a coloured filter. The filters are yellow, magenta and cyan. Each filter subtracts one primary colour from the white light: the magenta filter subtracts green, so that only red and blue are left; the cyan filter subtracts red, and the yellow one blue. Where the magenta filter overlaps the cyan one, both green and red are subtracted, and we see blue. In the central area, all light is filtered away and the result is black. Colour printing, which uses white paper and yellow, magenta and cyan ink, is based on the subtractive colour scheme. When white light falls on the document part is filtered out by the ink layers and the remainder is reflected from the underlying paper (Figure 10.2).

10.3 Visualization of image data

In this section, various ways for visualizing single and multi-band image data are introduced. The section starts with an explanation of the concept of an image histogram. The histogram has a crucial role in realizing optimal contrast of images. An advanced section deals with the application of RGB-IHS transformation to integrate different types of image data.

10.3.1 Histograms

A number of important characteristics of a single-band image, such as a panchromatic satellite image, a scanned monochrome photograph or a single band from a multi-band image, are found in the histogram of that image. The *histogram* describes the distribution of the pixel values (Digital Numbers, DN) of that image. In the usual case, the DN-values range between 0–255. A histogram indicates the number of pixels for each value in this range. In other words, the histogram contains the frequencies of DN-values in an image. Histogram data can be represented either in tabular form or graphically. The *tabular representation* (Table 10.1) normally shows five columns. From left to right these are:

- DN: Digital Numbers, in the range [0...255]
- Npix: the number of pixels in the image with this DN (frequency)
- Perc: frequency as a percentage of the total number of image pixels
- CumNpix: cumulative number of pixels in the image with values less than or equal to DN
- CumPerc: cumulative frequency as a percentage of the total number of image pixels

Histogram data can be further summarized in some characteristic statistics: mean, standard deviation, minimum and maximum, as well as the 1% and 99% values (Table 10.2). Standard deviation is a statistical measure of the spread of the values around the mean. The 1% value, for example, defines the cut-off value below which only 1% of all the values are found. 1% and 99% values can be used to define an optimal stretch for visualization.

DN	Npix	Perc	CumNpix	CumPerc
0	0	0.00	0	0.00
13	0	0.00	0	0.00
14	1	0.00	1	0.00
15	3	0.00	4	0.01
16	2	0.00	6	0.01
51	55	0.08	627	0.86
52	59	0.08	686	0.94
53	94	0.13	780	1.07
54	138	0.19	918	1.26
102	1392	1.90	25118	34.36
103	1719	2.35	26837	36.71
104	1162	1.59	27999	38.30
105	1332	1.82	29331	40.12
106	1491	2.04	30822	42.16
107	1685	2.31	32507	44.47
108	1399	1.91	33906	46.38
109	1199	1.64	35105	48.02
110	1488	2.04	36593	50.06
111	1460	2.00	38053	52.06
163	720	0.98	71461	97.76
164	597	0.82	72058	98.57
165	416	0.57	72474	99.14
166	274	0.37	72748	99.52
173	3	0.00	73100	100.00
174	0	0.00	73100	100.00
255	0	0.00	73100	100.00

Table 10.1: Example histogram in tabular format

Mean	StdDev	Min	Max	1%-value	99%-value
113.79	27.84	14	173	53	165

Table 10.2: Summary statistics for the example histogram given above.

In addition to the frequency distribution, the *graphical representation* shows the cumulative frequency (see Figure 10.5). The cumulative frequency curve shows the percentage of pixels with DN that are less than or equal to a given value.

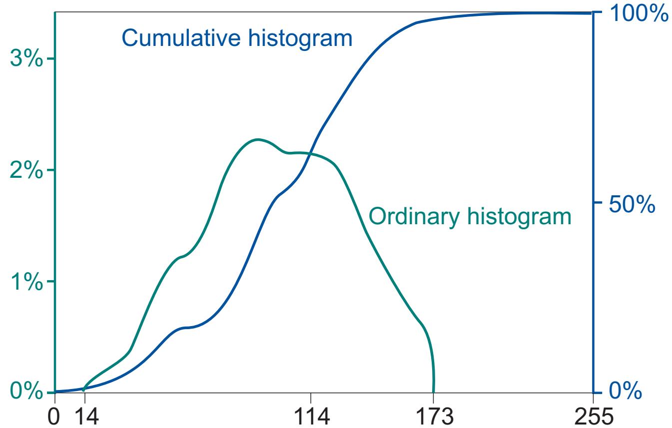


Figure 10.5: Standard histogram and cumulative histogram corresponding with Table 10.1.

10.3.2 Single band image display

The histogram is used to obtain optimum display of single band images. Single band images are normally displayed using a grey scale. Grey shades of the monitor typically range from black (value 0) to white (value 255). When applying grey shades, the same signal is input to each of the three (RGB) channels of the computer monitor (Figure 10.6).

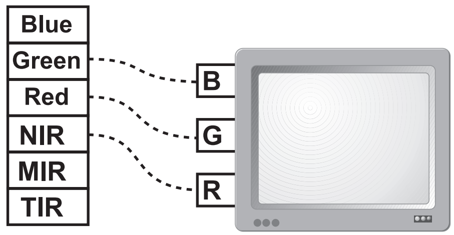


Figure 10.6: Multi-band image displayed on a monitor using the monitor's Red, Green and Blue input channels.

Using the original image values to control the monitor values usually results in an image with little contrast since only a limited number of grey values are used. In the example introduced in the previous section (Table 10.2) only $173 - 14 = 159$ out of 255 grey levels would be used. To optimize the range of grey values, a *transfer function* maps DN-values into grey shades on the monitor (Figure 10.7). The transfer function can be chosen in a number of ways. *Linear contrast stretch* is obtained by finding the DN-values where the cumulative histogram of the image passes 1% and 99%. DNs below the 1% value become black (0), DNs above the 99% value are white (255), and grey levels for the intermediate values are found by linear interpolation. *Histogram equalization*, or histogram stretch, shapes the transfer function according to the cumulative histogram. As a result, the DNs in the image are distributed as equally as possible over the available grey levels (Figure 10.7).

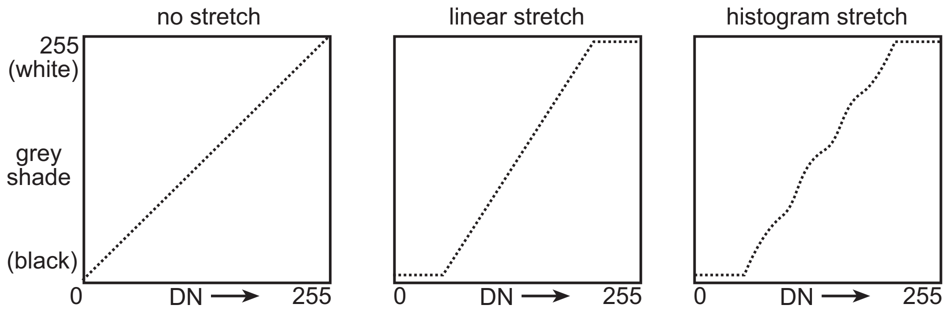


Figure 10.7: The relationship between image value and monitor grey value is defined by the transfer function, which can have different shapes.

Transfer function manipulations are usually implemented as modifications of the values in the colour lookup table (in the memory of the monitor). Using a transfer function for visualization, therefore, does not result in a new image file on hard disk. You should be aware that transfer functions may already have been applied when you want to interpret the intensity or colour observed in a picture.

An alternative way to display single band data is to make use of a *pseudo-colour* lookup table that assigns colours ranging from blue via cyan, green and yellow to red. The use of pseudo-colour is especially useful for displaying data that are not reflection measurements. With thermal infrared data, for example, the association of cold-warm with blue-red is more intuitive than with dark-light.

Figure 10.8: Single band of a Landsat TM image of a polder area in The Netherlands. Image without contrast enhancement (left); histogram equalization (middle) and pseudo-colour representation (right).

10.4 Colour composites

The previous section explained visualization of single band images. When dealing with a multi-band image any combination of three bands can, in principle, be used as input to the RGB channels of the monitor. The choice should be made based on the application of the image data. To increase contrast, the three bands can be subjected to linear contrast stretch or histogram equalization.

Sometimes a *true colour* composite, where the RGB channels relate to the red, green and blue wavelength bands of a scanner, is made. A popular choice is to link RGB to the near-infrared, red and green bands respectively to yield a *false colour composite* (Figure 10.9). The results look similar to prints of colour-infrared photography (CIR). As explained in Chapter 4, the three layers in a *false colour infrared film* are sensitive to the NIR, R, and G parts of the spectrum and made visible as R, G and B respectively in the printed photo. The most striking characteristic of false colour composites is that vegetation appears in a red-purple colour. In the visible part of the spectrum, plants reflect mostly green light (this is why plants appear green), but their infrared reflection is even higher. Therefore, vegetation in a false colour composite is shown as a combination of some blue but even more red, resulting in a reddish tint of purple.

Depending on the application, band combinations other than true or false colour may be used. Land-use categories can often be distinguished quite well by assigning a combination of Landsat TM bands 5–4–3 or 4–5–3 to RGB. Combinations that display near-infrared as green show vegetation in a green colour and are, therefore, called *pseudo-natural colour composites* (Figure 10.9).

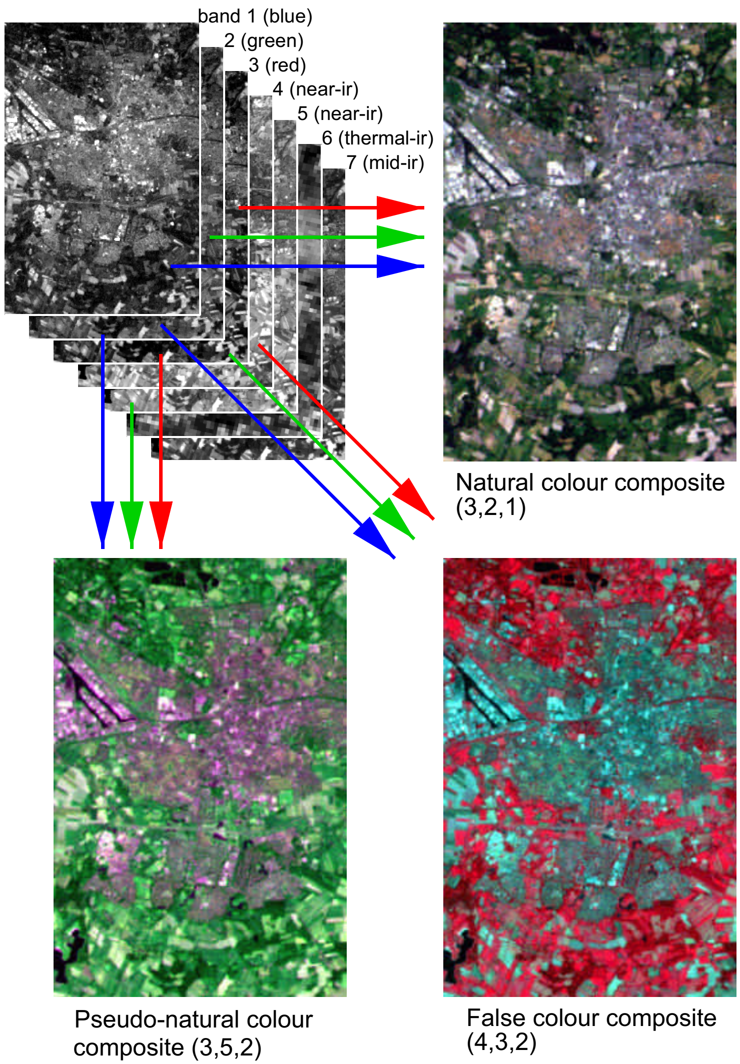


Figure 10.9: Landsat TM false colour composite of Enschede and surrounding. Three different colour composites are shown: natural colour, pseudo-natural colour and false colour composite. Other band combinations are possible.

10.4.1 Application of RGB and IHS for image fusion

One of the practical applications of (switching) between different colour spaces is in the fusion of image data with different spectral and spatial resolutions. A colour image with three bands for red, green and blue can actually be transformed into the IHS domain and back again. This yields a new image, again with three components. When looking at these components separately, the intensity image gives a proper grey scale representation of the colour image. The hue image and the saturation image are not very informative, although they can be manipulated. For example, when all saturation values appear low, it is possible to increase the saturation by a certain amount. When transforming the hue, intensity and the new saturation image back to the RGB domain, saturation enhancement is achieved. An example application of this method in *image fusion* is described below.

Consider the following two images:

- SPOT XS multispectral image with 20 m pixels.
- SPOT panchromatic image with 10 m pixels.

In certain applications, we would like to have colour information at a high (10 m) resolution. It is, however, not available from SPOT XS. The closest we can get is to create an image that combines hue and saturation from XS with intensity from Pan (Figure 10.10). First step is to resample the SPOT XS (pixels of 20 m) into the image raster of the SPOT Pan (pixels of 10 m). The SPOT XS is then converted from (RGB) to (IHS). Subsequently, 'I' is replaced by the Pan image, and (Pan, H, S) converted back to (R'G'B'). The result is a high-resolution (10 m) colour composite, which is potentially more suitable for image interpretation purposes than the two separate originals. The same technique can be used to fuse other combinations of data, for example TM with SPOT PAN, SPOT XS with ERS-SAR, multispectral satellite imagery with B/W aerial photography, *et cetera*.

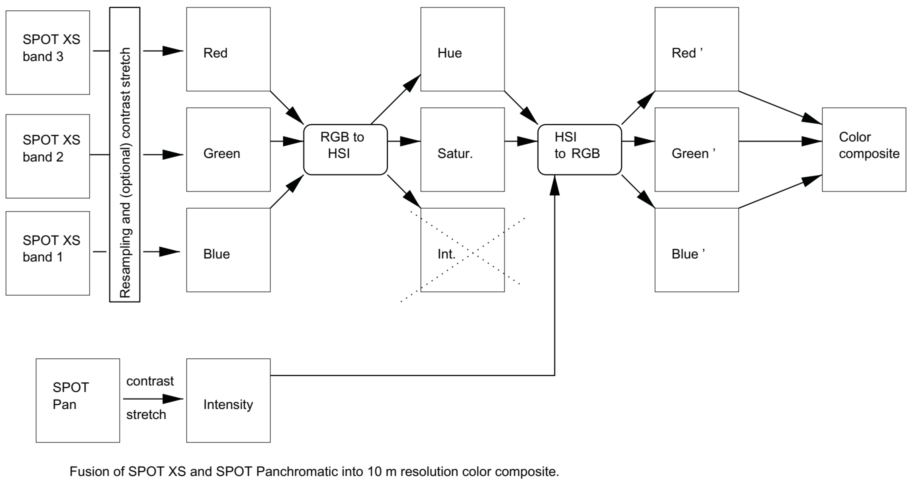


Figure 10.10: Procedure to merge SPOT panchromatic and multispectral data using RGB to IHS conversion and *vice versa*.

10.5 Filter operations

A further step in producing optimal images for interpretation is the use of filter operations. Filter operations are local image transformations: a new image is calculated and the value of a pixel depends on the values of its former neighbours. Filter operations are usually carried out on a single band. Filters are used for spatial image enhancement, for example to reduce noise or to sharpen blurred images. Filter operations are extensively used in various semi-automatic procedures that are outside the scope of this chapter.

To define a filter, a *kernel* is used. A kernel defines the output pixel value as a linear combination of pixel values in a neighbourhood around the corresponding position in the input image. For a specific kernel, a so-called *gain* can be calculated as follows:

$$gain = \frac{1}{\Sigma k_i} \tag{10.1}$$

The gain sums all kernel coefficients (k_i). In general, the sum of the kernel coefficient, after multiplication by the gain, should be equal to 1 to result in an

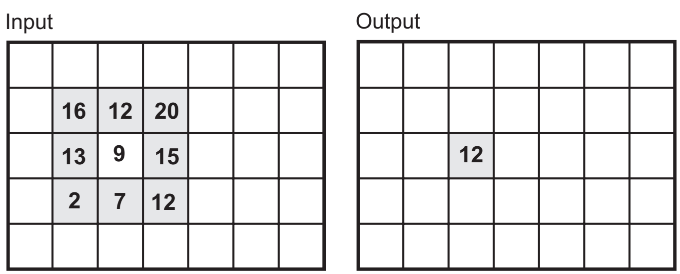


Figure 10.11: Input and output result of a filtering operation: the neighbourhood in the original image determines the value of the output. In this situation a smoothing filter was applied.

image with approximately the same range of grey values. The effect of using a kernel is illustrated in Figure 10.11, which shows how the output value is calculated in terms of average filtering.

The significance of the gain factor is explained in the next two subsections. In these examples only small neighbourhoods of 3×3 kernels are considered. In practice other kernel dimensions may be used.

10.5.1 Noise reduction

Consider the kernel shown in Table 10.3 in which all coefficients equal 1. This means that the values of the nine pixels in the neighbourhood are summed. Subsequently, the result is divided by 9 to achieve that the overall pixel values in the output image are in the same range as the input image. In this situation the gain is $1/9 = 0.11$. The effect of applying this *averaging filter* is that image will become blurred or smoothed. When dealing with speckle effect in radar imagery the result of applying this filter is to reduce the speckle.

1	1	1
1	1	1
1	1	1

Table 10.3: Filter kernel for smoothing.

In the above kernel, all pixels have equal contribution in the calculation of the result. It is also possible to define a weighted average. To emphasize the value of the central pixel, a larger value can be put in the centre of the kernel. As a result, less drastic blurring takes place. In addition, it is necessary to take into account that the horizontal and vertical neighbours influence the result more strongly that the diagonal ones. The reason for this is that the direct neighbours are closer to the central pixel. The resulting kernel, for which the gain is $1/16 = 0.0625$, is given in Table 10.4.

1	2	1
2	4	2
1	2	1

Table 10.4: Filter kernel for weighted smoothing.

10.5.2 Edge enhancement

Another application of filtering is to emphasize local differences in grey values, for example related to linear features such as roads, canals, geological faults, *et cetera*. This is done using an *edge enhancing* filter, which calculates the difference between the central pixel and its neighbours. This is implemented using negative values for the non-central kernel coefficients. An example of an edge enhancement filter is given in Table 10.5.

-1	-1	-1
-1	16	-1
-1	-1	-1

Table 10.5: Filter kernel used for edge enhancement.

The gain is calculated as follows: $1/(16-8) = 1/8 = 0.125$. The sharpening effect can be made stronger by using smaller values for the centre pixel (with a minimum of 9). An example of the effect of using smoothing and edge enhancement is shown in Figure 10.12.

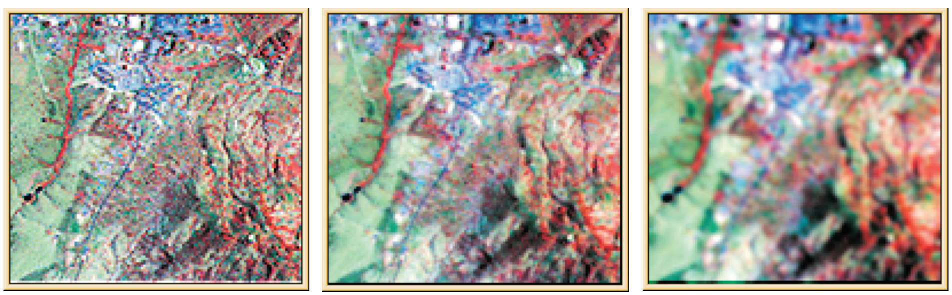


Figure 10.12: Original image (middle), edge enhanced image (left) and smoothed image (right).

Summary

The way we perceive colour is most intuitively described by the Hue component of the IHS colour space. The colour space used to describe colours on computers monitor is the RGB space.

When displaying an image on a screen (or on hard copy) many choices need to be made. The selection of bands, the sequence in which these are linked to the Red-Green-Blue channels of the monitor, the use of stretching techniques and the possible use of (spatial) filtering techniques.

The histogram, and the derived cumulative histogram, is the basis for all stretching methods. Stretching, or contrast enhancement, is realized using transfer functions.

Filter operations are based on the use of a kernel. The weights of the coefficients in the kernel determine the effect of the filter which can be, for example, to smooth or sharpen the original image.

Questions

The following questions can help you to study Chapter 10.

1. How many possibilities are there to visualize a 4 band image using a computer monitor?

2. You are shown a picture in which grass looks green and houses are red—what is your conclusion? Now, you are shown a picture in which grass shows as purple and houses are black—what is your conclusion now?

3. What would be a reason for not using the default application of 'histogram equalization' for all image data?

4. Can you think of a situation in your own context where you would probably use filters to optimize interpretation of image data?

The following are typical exam questions:

1. List the three colour spaces used in the context of remote sensing and visualization.

2. Which colour space should be applied when using computer monitors? How is the colour 'white' produced?

3. What information is contained in a histogram of image data?

4. Which technique is used to maximize the range of colours (or grey values) when displaying an image?

5. Using an example, explain how a filter works.

Chapter 11

Visual image interpretation

11.1 Introduction

Up to now, we have been dealing with acquisition of image data. The data acquired still needs to be interpreted (or analysed) to extract the required information. In general, information extraction methods from remote sensing imagery can be subdivided into two groups:

- Information extraction based on visual analysis or interpretation of the data. Typical examples of this approach are visual interpretation methods for land use or soil mapping. Also the generation/updating of topographic maps from aerial photographs is based on visual interpretation. *Visual image interpretation* is introduced in this Chapter.

- Information extraction based on semi-automatic processing by the computer. Examples include automatic generation of DTMs, image classification and calculation of surface parameters. *Image classification* is introduced in Chapter 12.

The most intuitive way to extract information from remote sensing images is by *visual image interpretation*, which is based on man's ability to relate colours and patterns in an image to real world features. Chapter 10 has explained different methods used to visualize remote sensing image data.

In some situations pictures are studied to find evidence of the presence of features, for example, to study natural vegetation patterns. Most often the result of the interpretation is made explicit by digitizing the geometric and thematic data of relevant objects ('mapping'). The digitizing of 2D features (points, lines and areas) is carried out using a digitizer tablet or on-screen digitizing. 3D features interpreted in stereopairs can be digitized using stereoplotters or digital photogrammetric workstations.

In Section 11.2 some theory about image understanding is explained. Visual image interpretation is used to produce spatial information in all of ITC's fields of interest: urban mapping, soil mapping, geomorphological mapping, forest mapping, natural vegetation mapping, cadastral mapping, land use mapping and many others. As visual image interpretation is application specific, it is illustrated by two examples (soil mapping, land cover mapping) in (Section 11.3). The last section (11.4) addresses some aspects of quality.

11.2 Image understanding and interpretation

11.2.1 Human vision

In Chapter 10 human's perception of colour was explained. *Human vision* goes a step beyond perception of colour: it deals with the ability of a person to draw conclusions from visual observations. In analysing a picture, typically you are somewhere between the following two situations: direct and spontaneous recognition or using several clues to draw conclusions by a reasoning process (logical inference).

Figure 11.1: Satellite image of Antequera area in Spain, the circular features are pivot irrigation systems

Spontaneous recognition refers to the ability of an interpreter to identify objects or phenomena at a first glance. Consider Figure 11.1. An agronomist would immediately recognize the pivot irrigation systems with their circular shape. S/he would be able to do so because of earlier (professional) experience. Similarly, most people can directly relate an aerial photo to their local environment. The quote from people that are shown an aerial photograph for the first time "I see because I know" refers to *spontaneous recognition*.

Logical inference means that the interpreter applies reasoning. In the reasoning the interpreter will use his/her professional knowledge and experience. Logical inference is, for example, concluding that a rectangular shape is swimming pool because of its location in a garden and near to a house. Sometimes, logical inference alone cannot help you in interpreting images and field observations are required (Section 11.4). Consider the aerial photograph in Figure 11.2. Would you be able to interpret the material and function of the white mushroom like objects? A field visit would be required for most of us to relate the different features to elements of a house or settlement.

Figure 11.2: Mud huts of Labbezanga near the Niger river. Photo by Georg Gerster, 1972.

11.2.2 Interpretation elements

When dealing with image data, visualized as pictures, a set of terms is required to express and define characteristics present in a pictures. These characteristics are called *interpretation elements* and are used, for example, to define *interpretation keys*, which provide guidelines on how to recognize certain objects.

The following interpretation elements are distinguished: tone/hue, texture, shape, size, pattern, site and association.

- *Tone* is defined as the relative brightness of a black/white image. *Hue* refers to the colour on the image as defined in the intensity-hue-saturation (IHS) system. Tonal variations are an important interpretation element in an image interpretation. The tonal expression of objects on the image is directly related to the amount of light (energy) reflected from the surface. Different types of rock, soil or vegetation most likely have different tones. Variations in moisture conditions are also reflected as tonal differences in the image: increasing moisture content gives darker grey tones. Variations in hue are primarily related to the spectral characteristics of the measured area and also to the bands selected for visualization (see Chapter 10). The advantage of hue over tone is that the human eye has a much larger sensitivity for variations in colour (approximately 10,000 colours) as compared to tone (approximately 200 grey levels).

- *Shape* or form characterizes many terrain objects visible in the image. Shape also relates to (relative) height when dealing with stereo-images, which we discuss in Section 11.2.3). Height differences are important to distinguish between different vegetation types and also in geomorphological mapping. The shape of objects often helps to determine the character of the object (built-up areas, roads and railroads, agricultural fields, *et cetera*).

- *Size* of objects can be considered in relative or absolute sense. The width of a road can be estimated, for example, by comparing it to the size of the cars, which is generally known. Subsequently this width determines the road type, e.g., primary road, secondary road, *et cetera*.

- *Pattern* refers to the spatial arrangement objects and implies the characteristic repetition of certain forms or relationships. Pattern can be described by terms such as concentric, radial, checkerboard. Some land uses, however, have specific and characteristics patterns when observed on aerospace data. You may think of different irrigation types but also different types of housing in the urban fringe. Other typical examples include the hydrological system (river with its branches) and patterns related to erosion.

- *Texture* relates to the frequency of tonal change. Texture may be described by terms as coarse or fine, smooth or rough, even or uneven, mottled, speckled, granular, linear, woolly, *et cetera*. Texture can often be related to terrain roughness. Texture is strongly related to the spatial resolution of the sensor applied. A pattern on a large scale image may show as texture on a small scale image of the same scene.

- *Site* relates to the topographic or geographic location. A typical example of this interpretation element is that backswamps can be found in a floodplain but not in the centre of a city area. Similarly, a large building at the end of a number of converging railroads is likely to be a railway station—we do not expect a hospital at this site.

- *Association* refers to the fact that a combination of objects makes it possible to infer about its meaning or function. An example of the use of 'association' is an interpretation of a thermal power plant based on the combined recognition of high chimneys, large buildings, cooling towers, coal heaps and transportation belts. In the same way the land use pattern associated with small scale farming will be characteristically different to that of large scale farming.

Having introduced these seven interpretation elements you may have noticed a relation with the spatial extent of the feature to which they relate. Tone or hue can be defined for a single pixel; texture is defined for a neighbouring group of pixels, not for single pixels. The other interpretation elements relate to individual objects or a combination of objects. The simultaneous and often implicit use of all these elements is the strength of visual image interpretation. In standard image classification (Chapter 12) only 'hue' is applied, which explains the limitations of automated methods compared to visual image interpretation.

11.2.3 Stereoscopic vision

The impression of depth encountered in the real world can also be realized by images of the same object that are taken from different positions. Such a pair of images, photographs or digital images that are separated and observed at the same time by the two eyes. These give images on the retinas in which objects at different positions in space are projected on relative different positions. We call this *stereoscopic vision*. Pairs of images that can be viewed stereoscopically are called *stereograms* . Stereoscopic vision is explained here because the impression of height and height differences is important in the interpretation of both natural and man-made features from image data. Note that in Chapter 9 we explained that under specific conditions stereo-models can be used to derive 3D coordinates.

Under normal conditions we can focuss on objects between 150 mm distance and infinity. In doing so we direct both eyes to the object (point) of interest. This is known as *convergence*. To view the stereoscopic model formed by a pair of overlapping photographs, the two images have to be separated so that the left and right eyes see the left and right photographs. In addition one should not focus on the photo itself but at infinity. Some experienced persons can experience 'stereo' by putting the two photos at a suitable distance from their eyes. Most of us need some help and different methods have been developed.

Pocket and mirror stereoscopes, and also the photogrammetric plotters, use a system of lenses and mirrors to 'feed' one image into one eye. Pocket and mirror stereoscopes are mainly applied in mapping applications related to vegetation, forest, soil and geomorphology (Figure 11.3) Photogrammetric plotters are used in topographic and large scale mapping activities.

Another way of achieving stereovision is to project the two images in two colours. Most often red and green colours are applied; the corresponding spec-

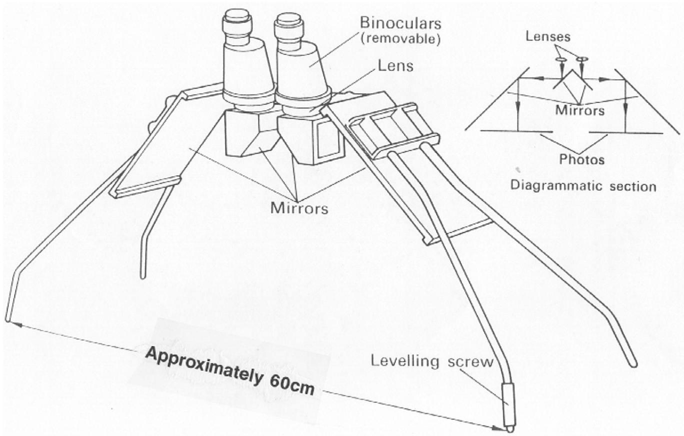


Figure 11.3: The mirror stereoscope enables stereoscopic vision of stereograms. Each photo is projected onto one eye.

tacles comprise one red and one green glass. This method is known as the anaglyph system and is particularly suited to viewing overlapping images on a computer screen. An approach used in digital photogrammetric systems is to apply polarization for the left and right images. Polarized spectacles make the left image visible to the left eye and the right image to the right eye.

11.3 Application of visual image interpretation

11.3.1 Soil mapping with aerial photographs

Semi-detailed soil mapping is a typical application of visual interpretation of panchromatic stereopairs. The result is a soil map at scale 1:50,000. This type of mapping is carried out either systematically or on a project basis, and is rarely updated unless the soil landscape is radically changed, e.g., by natural disasters or land reclamation works.

This section explains a method of soil mapping with the aid of aerial photos that can be used in many areas of the world, the so-called 'geo-pedologic' or soil-landscape approach [7, 23, 28]. It starts from the assumption that the type of soil is highly dependent on the position on the landscape (the 'relief'), the type of landscape itself (indicating the 'time' at which the present soil formed), and the underlying geology (so-called 'parent material' in which the soil formed). Local effects of climate and organisms (vegetation or animals) can also be seen on aerial photos.

There are some soil landscapes where this method is not satisfactory. At more detailed scales, field observations become more important than photo interpretation, and at more general scales, monoscopic interpretation of satellite imagery is often more appropriate.

The overall process for semi-detailed soil mapping is as follows:

- In the first phase, a transparency is placed over the working area of one photo of a stereo pair and Aerial Photo Interpretation (API) units are delineated on it. The minimum size of the delineated units is 10 ha (0.4 cm^2 on the final map). Delineation is mainly based on geomorphological characteristics, using a structured legend. Stereopairs allow observation of height and terrain form. In addition, the usual interpretation elements (colour, texture, shape, ...) may be used. The interpretation is hierarchical: first

Figure 11.4: Panchromatic photograph to be interpreted.

the interpreter finds and draws 'master lines' dividing major landscapes (mountains, hill land, plateau, valley, ...). Each landscape is then divided into relief types (e.g., 'sharply-dissected plateau'), each of which is further divided by lithology (e.g., 'fine-bedded shales and sandstones'), and finally by detailed landform (e.g., 'scarp slope'). The landform consists of a topographic form, a geomorphic position, and a geochronological unit, which together determine the environment in which the soil formed. A legend category usually comprises many areas (polygons) with the same photo-interpretation characteristics. Figure 11.4 shows a photograph and

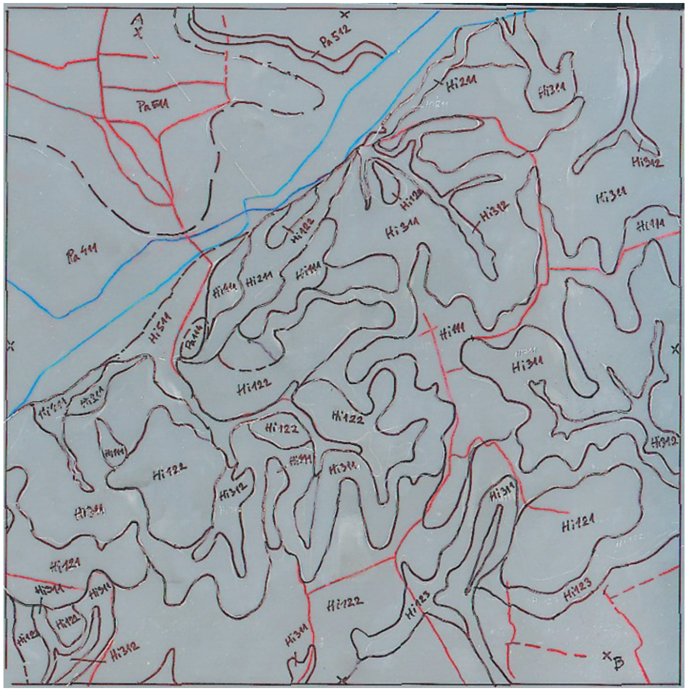


Figure 11.5: Photo-interpretation transparency related to the aerial photo shown in Figure 11.4.

Figure 11.5 shows the interpretation units that resulted from its stereo interpretation.

- In the next phase, a sample area of the map is visited in the field to study the soil. The sampled area is between 10–20% of the total area and comprises all legend classes introduced in the previous stage. The soils are described in the field, and samples are taken for laboratory analysis, to determine their characteristics (layering, particle-size distribution, density, chemical composition, . . .). The observations are categorized into a local or more general soil classification system. In this way, each photo-interpretation unit is characterized by the soil types it contains, their relative abundance in the unit, and their detailed landscape relation. Some API units have only one dominant soil in a 10 ha area, but in complex terrain, there may be several. The field check may also reveal a deficiency in photo-interpretation, in which case the lines are adjusted, both in the sample area, and according to the same criteria in extrapolation areas.

- The final legend of such a map shows two linked tables: the geopedologic legend (Table 11.1) and the soil legend (Table 11.2). Each map unit belongs to one geopedologic class, which can be linked to more than one soil class (1-to-n relationship). It is possible for the same soil class to occur in several geopedologic classes, in which case the relationship is n-to-m.

- Following the field check and possible adjustment of the API lines, the transparencies are digitized, then individually corrected by ground control points and, preferably, a DEM (for ortho-correction), and merged into one map. Without ortho-correction, it is necessary to manually edge-match the georeferenced transparencies and correct for the effects of relief displacement. Another possibility is to manually re-compile lines on either a topographic map with sufficient detail, or on orthophotos made independently.

- Validation of these maps is typically carried out by transects that cross the most variation over the shortest distance. Soils are observed at points located by photo-reading or GPS at pre-determined intervals of the transect (for example every 100 m), and the predicted soil class is compared with the actual class. This provides a quantitative measure of the map accuracy.

Soil mapping is an example of geo-information production with the aid of aerospace survey, which is based on complementary roles for remote sensing and ground observations, as well as prior knowledge of the survey area (geology, geomorphology, climate history, ecology, ...) and a sound conceptual framework about the object of study (here, soil forming factors). Remote sensing is used to provide a synoptic view, based on a mental model of soil-landscape relationships, and to limit the amount of work in the field by stratification.

Landscape	Relief	Lithology	Landform	API Code
Hilland	Dissected ridge	Loess	Summit	Hi111
			Shoulder & backslope	Hi112
	Escarpment	Loess over basalt	Scarp	Hi211
		Colluvium from loess / basalt	Toe slope	Hi212
	Vales	Alluvium from loess	Slope	Hi311
			Bottom	Hi312
	Glacis	Colluvium from loess	Slope	Hi411
Plain	High terrace	Loess over old river alluvium	Tread	Pl311
			Abandoned channel	Pl312
	Old floodplain	Old alluvium	Abandoned floodplain (channelized)	Pl411

Table 11.1: Geopedologic legend, which results in (hierarchical) API-codes.

API Code	Dominant Soil Type	%
Hi111	Eutri-chromic Cambisols, fine silty	
Hi112	Eutri-chromic Cambisols, coarse silty	60%
	Calcaric Regosols, coarse silty	40%
Hi211	Calcaric Regosols, coarse silty	
Hi212	Stagnic Cambisols, silty skeletal	
Hi311	Eutri-chromic Cambisols, fine silty	
Hi312	Eutric Gleysols, fine silty	
Hi411	Eutric Luvisols, coarse silty	
Pl311	Silti-calcic Kastanozems, fine silty	
Pl312	Mollic Gleysols, fine loamy	
Pl411	Gleyic Fluvisols, fine loamy	50%
	Eutric Fluvisols, fine loamy	50%

Table 11.2: Relationship between the soil types and the geo-pedologic legend is made through the API-codes.

11.3.2 Land cover mapping from multispectral data

This second example refers to the CORINE land cover project in which a land cover database for Europe is being established. CORINE is based on the work done in different countries of the European Union. This example is to illustrate a situation in which different organizations and individuals are involved in the production of the total data set. To enable such approach, research activities have been carried out that resulted in a technical guide for producing the land cover data from spaceborne multispectral data [5, 20].

At this point, the terms land cover and land use should be defined since they both are often used in the context of image interpretation. *Land cover* refers to the type of feature present on the surface of the land. It refers to a physical property or material, e.g., water, sand, potato crop, and asphalt. *Land use* relates to the human activity or economic function for a specific piece of land, e.g., urban use, industrial use or nature reserve. Another way to put it is that land cover is more dynamic than land use. Most vegetated areas (land cover) change over time. Land use as it is related to the human activity is more stable: land is being used for irrigated crops or extensive cattle breeding will not change within a year. The difference between land cover and land use is also explained in Figure 11.6 showing that different land use classes can be composed of the same land cover classes. Principally, remote sensing image data give information about land cover. Using context information the land use sometimes can be deduced. The distinction between land cover and land use becomes highly relevant in dealing with, for example, environmental impact assessment studies. In such studies, not only the presence of a certain type of crop is relevant but also the way in which this crop is grown and treated (in terms of manure, herbicides, pesticides etc) is important.

The overall CORINE mapping process is as follows:

- The process starts with the selection of cloud free multispectral satellite images of an area. Hard copy prints are produced from the image data. Guidelines are given to yield prints that are similar in terms of 'scale' and colour enhancement. The prints are the basis for the interpretation. (Today, also on-screen digitizing procedures are described.) Additional information in the form of topographical maps, photo-atlases, soil maps *et cetera* is collected to aid the interpretation.

- The interpretation of the land cover units is made on transparencies, which are overlaid on the hard copy images and which are digitized at a later stage. In the CORINE project the minimum mapping unit is set at 25 ha; the minimum width of mapping units is set at 100 m. The interpretation is based on the 'nomenclature', which gives the names of the land cover classes. CORINE has a three level categorization (Table 11.3 and Table 11.4). The reason for defining the land cover at three different levels is ensure consistent aggregation of the different classes. At the same

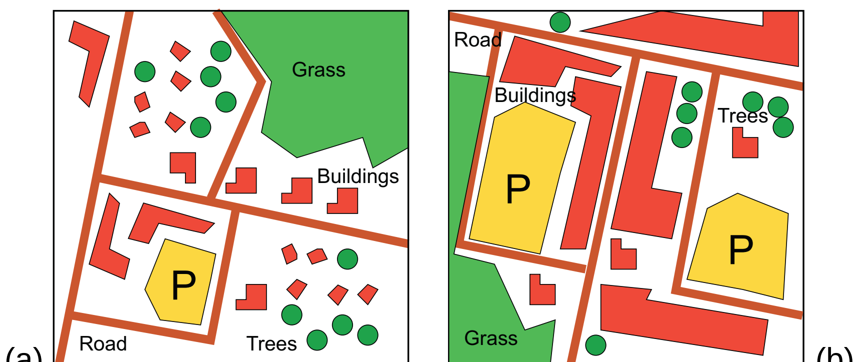


Figure 11.6: Residential land use (a) and industrial land use (b) may be composed of similar land cover classes.

time, one can imagine that a map of level-1 classes is more accurate than a map of level-3 classes. For each class, a more extended class description is given, followed by instructions on how to recognize the specific class (Table 11.5) together with some example images of the class. The instructions with respect to discriminating different classes or categories are called an *interpretation key*.

- The line patterns on the transparencies are digitized and the corresponding class codes are entered into the computer. The result is a polygon database. One of the attributes is the land cover code. A sample result of the interpretation is shown in Figure 11.7.

- The final activity of any mapping project is an independent assessment of accuracy (validation). In the CORINE project this is done by field check of a limited number of objects that are selected by a sampling strategy.

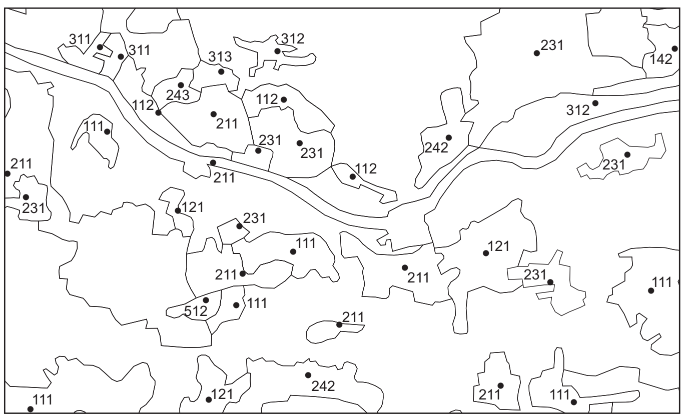


Figure 11.7: Example of the result of the CORINE Land Cover classification showing a part around one of the rivers in The Netherlands. Courtesy of Wageningen-UR

Level 1	Level 2	Level 3
1. **Artificial Surfaces**	1.1. Urban fabric	1.1.1. Continuous urban fabric
		1.1.2. Discontinuous urban fabric
	1.2. Industrial, commercial and transport units	1.2.1. Industrial or commercial units
		1.2.2. Road and rail networks and associated land
		1.2.3. Port areas
		1.2.4. Airports
	1.3. Mine, dump and construction sites	1.3.1. Mineral extraction sites
		1.3.2. Dump sites
		1.3.3. Construction sites
	1.4. Artificial non-agricultural vegetated areas	1.4.1 Green urban areas
		1.4.2. Sport and leisure facilities

Table 11.3: Part of the CORINE nomenclature for land cover.

Level 1		Level 2		Level 3	
2.	**Agricultural areas**	2.1.	Arable land	2.1.1.	Non-irrigated arable land
				2.1.2.	Permanently irrigated land
				2.1.3.	Rice fields
		2.2.	Permanent crops	2.2.1.	Vineyards
				2.2.2.	Fruit trees and berry plantations
				2.2.3.	Olive groves
		2.3.	Pastures	2.3.1.	Pastures
		2.4.	Heterogeneous agricultural areas	2.4.1.	Annual crops associated with permanent crops
				2.4.2.	Complex cultivation patterns
				2.4.3.	Land principally occupied by agriculture, with significant areas of natural vegetation
				2.4.4.	Agro-forestry
3.	**Forests and semi natural areas**	...	...	...	...
4.	**Wetlands**	...	...	...	...
5.	**Water bodies**	...	...	...	...

Table 11.4: Part (two) of the CORINE nomenclature for land cover.

Class 1.3.1 Mineral Extraction Sites
Extended Description: Areas with open-pit extraction of construction material (sandpits, quarries) or other minerals (open-cast mines). Includes flooded gravel pits, except for -bed extraction. *How to recognize*: Quarries are easily recognizable on satellite images (white patches) because they contrast with their surroundings. The same is true for working gravel pits. For open-cast mines, the difference with item 1.3.2 (dump sites) is not always obvious. In such cases, ancillary data will be needed to remove any doubt. Disused open-cast mines, quarries, sandpits, sludge quarries and gravel pits (not filled with water) are included in this category, However, ruins do not come under this heading. Sites being worked or only recently abandoned, with no trace of vegetation, come under this heading. Where vegetal colonization is visible, sites are classified under the appropriate vegetal cover category. This heading includes buildings and associated industrial infrastructure (e.g., cement factories) and small water bodies of less than 25 ha created by mining.

Table 11.5: CORINE's extended description for class 1.3.1 (Mineral Extraction Sites). Source:[5]

11.3.3 Some general aspects

Mapping, in general, requires an abstraction of the world. The most simple way to do this is by introducing classes or categories. The 'land' is divided into (discrete) objects, which are assigned to one class (only). Sometimes such abstraction is defined beforehand and determined by the required information. On other occasions, this abstraction is made during the process itself because of lack of experience with the image data and terrain at hand. For different reasons hierarchical systems are often used. Among others they provide an easy way to aggregate data into higher level classes [19]. One of the drawbacks of such a discrete approach is that the actual phenomenon that is 'mapped' is not discrete. You may imagine that even a qualified photo-interpreter has difficulties in discriminating between 'agricultural land including semi-natural areas' and 'semi-natural areas with limited agricultural activity'. In the context of type of objects, therefore, a distinction can be made made between objects with determinate and indeterminate boundaries. The latter category requires quite different methods and approaches [3].

Another aspect of the interpretation concerns geometric properties. The minimum size (area) or width of the objects (units) to be distinguished needs to be defined. Another variable is the degree to which boundaries are generalized. This aspect is directly linked to the characteristics of the objects mapped: only crisp objects (such as houses, roads, etc) can be delineated accurately. Drawing boundaries in natural vegetation areas can be highly problematic if not useless.

In general, all image based mapping processes require *field observations*. Field observations can be used:

- to gather local knowledge beforehand to guide the interpretation. When dealing with a new area some of the features observed on the images will not be understood (for example Figure 11.2). Field observations will help to interpret these features.

- to gather data about areas or features that cannot be studied from the image data. For example, when dealing with large scale photos of an urban environment parts of the area will be in 'dead ground' or hidden by other construction such as bridge. The only way to get information about these areas is to visit them. Possibly, additional information that cannot be derived from the images needs to be collected, e.g., the address of a house. This can only be done by walking through the streets.

- to evaluate the intermediate and final interpretation result. The evaluation of the final result is called validation. In a validation the accuracy of the established data is determined. For this purpose a limited number of objects or areas are selected (using a sampling approach) and visited in the field. The data collected in the field is referred to as *ground truth*. Comparison of the ground truth and the interpretation result then is used to calculate different measures of accuracy.

11.4 Quality aspects

The quality of the result of an image interpretation depends a number of factors: the interpreter, the image data used and the guidelines provided.

- The professional experience and the experience with image interpretation determine the skills of a photo-interpreter. A professional background is required: a geological interpretation can only be made by a geologist since s/he is able to related image features to geological phenomena. Local knowledge, derived by field visits, is required to help the interpretation.

- The image data applied limit the phenomena that can be studied, both in thematic and geometric sense. One cannot, for example, generate a reliable database on the tertiary road system using multispectral satellite data. Likewise, black–and–white aerial photos contain limited information about agricultural crops.

- Finally, the quality of the interpretation guidelines is of large influence. Consider, for example, a project in which a group of persons is to carry out a mapping project. Ambiguous guidelines will prevent a consistent mapping in which individual results form a seamless database of consistent quality.

Especially in large projects and monitoring programmes, all the above three points play an important role in ensuring the replicability of the work. *Replicability* refers to the degree of correspondence obtained by different persons for the same area or by the same person for the same area at different moments in time. In principle, replicability does not tell directly about the accuracy (the relation with the real world) but it does tell something about the quality of the class

Figure 11.8: Two interpretation results derived by two photo-interpreters analysing the same image. Note the overall differences but also differences in the generalization of the lines. (from [8])

definition (crisp or ambiguous) and the instructions and methods used. Two examples are given here to give you an intuitive idea. Figure 11.8 gives two interpretation results for the same area. Note that both results differ in terms of total number of objects (map units) and in terms of (line) generalization. Figure 11.9 compares 13 individual interpretation results of a geomorphological interpretation. Similar to the previous example, large differences are found along the boundaries. In addition to this, you also can conclude that for some objects (map units) there was no agreement on the thematic attribute.

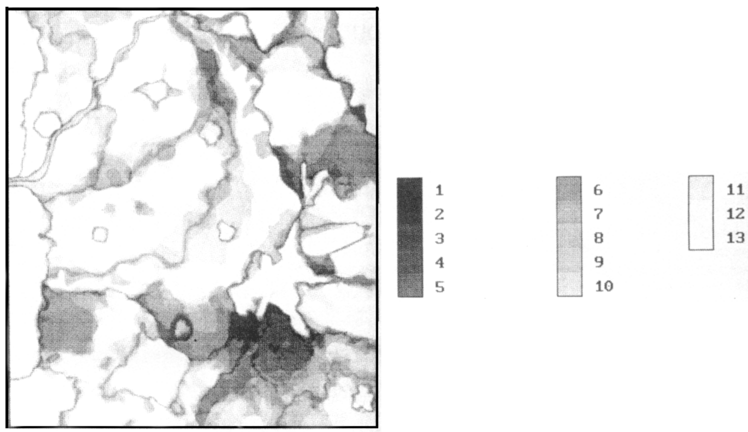


Figure 11.9: Comparison of 13 interpretations of the same image. The grey value represents the the degree of correspondence: white indicates agreement of all 13 interpreters, black indicates that all 13 interpreters disagreed on the thematic class for that location. (from [18])

Summary

Visual image interpretation is one of the methods to extract information from remote sensing image data. For that purpose, images need to be visualized on screen or in hard-copy. The human vision system is used to interpret the colours and patterns on the picture. Spontaneous recognition and logical inference (reasoning) are distinguished.

Interpretation keys or guidelines are required to instruct the image interpreter. In such guidelines, the (seven) interpretation elements can be used to describe how to recognize certain objects. Guidelines also provide a classification scheme, which defines the thematic classes of interest and their (hierarchical) relationships. Finally, guidelines give rules on the minimum size of objects to be included in the interpretation.

When dealing with a new area or a new application, no guidelines are available. An iterative approach is then required to establish the relationship between features observed in the picture and the real world.

In all interpretation and mapping processes the use of ground observations is essential to (i) acquire knowledge of local situation, (ii) gather data for areas that cannot be mapped from the images (iii) to check the result of the interpretation.

The quality of the result of visual image interpretation depends on the experience and skills of the interpreter, the appropriateness of the image data applied and the quality of the guidelines being used.

Questions

The following questions can help you to study Chapter 11.

1. What is the relationship between image visualization and image interpretation?

2. Describe (for a colleague) how to recognize a road on an aerial photo (make use of the interpretation elements).

3. Why is it necessary to have a sound conceptual model of how soils form in the landscape to apply the aerial photo-interpretation method presented in Section 11.3.1? What are the advantages of this approach in terms of efficiency and thematic accuracy, compared to interpretation element(only) analysis?

4. Describe a relatively simple method to check the quality (in terms of replicability) of visual image interpretation.

5. Which products in your professional environment are based on visual image interpretation?

6. Consider the CORINE nomenclature; identify three classes which can be accurately mapped; also identify three classes that can be expected to be exchanged (confused) with other classes.

The following are typical exam questions:

1. List the seven interpretation elements.

2. Give three reasons for field observations in the process of image interpretation.

3. Give definitions of land cover and land use. Give an example of each of them.

4. In what situations is visual image interpretation preferred to semi-automated interpretations?

5. Give two examples of cases where visual interpretation is not possible, even with good quality imagery.

Chapter 12

Digital image classification

12.1 Introduction

Chapter 11 explained the process of visual image interpretation. In this process, human vision plays a crucial role in extracting information from image data. Although computers may be used for visualization and digitization, the interpretation itself is carried out by the operator.

In this chapter, *digital image classification* is introduced. In this process the (human) operator instructs the computer to perform an interpretation according to certain conditions. These conditions are defined by the operator. Image classification is one of the techniques in the domain of digital image interpretation. Other techniques include automatic object recognition (for example, road detection) and scene reconstruction (for example, generation of 3D object models). Image classification, however, is the most commonly applied technique in the ITC context.

Application of image classification is found in many regional scale projects. In Asia, the Asian Association of Remote Sensing (AARS) is generating various land cover data sets based on (un)supervised classification of multispectral satellite data. In the Africover project (by FAO), image classification techniques are being used to establish a pan-African land cover data set. The European Commission requires national governments to verify the claims of farmers related to subsidized crops. These national governments employ companies to make a first inventory, using image classification techniques, which is followed by field checks.

Image classification is based on the different spectral characteristics of different materials on the Earth's surface, introduced in Section 2.4. This chapter focusses on classification of multispectral image data. Section 12.2 explains the concepts of image space and feature space. Image classification is a process that operates in the feature space. Section 12.3 gives an overview of the classification process, the steps involved and the choices to be made. The result of an image classification needs to be validated to assess its accuracy (Section 12.4). Finally, two major problems in image classification are addressed in Section 12.5

12.2 Principle of image classification

12.2.1 Image space

A digital image is a 2D-array of elements. In each element the energy reflected or emitted from the corresponding area on the Earth's surface is stored. The spatial arrangement of the measurements defines the image or *image space*. Depending on the sensor, data are recorded in n bands (Figure 3.11, repeated here as Figure 12.1). Digital image elements are usually stored as 8-bit DN-values (range: 0–255).

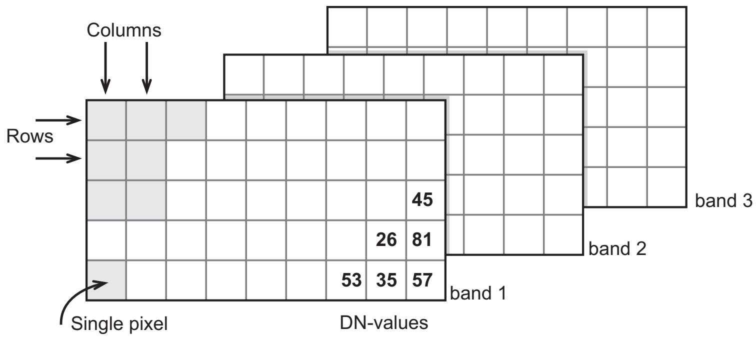


Figure 12.1: The structure of a multi-band image

12.2.2 Feature space

In one pixel, the values in (for example) two bands can be regarded as components of a two-dimensional vector, the *feature vector*. An example of a feature vector is $(13, 55)$, which tells that 13 DN and 55 DN are stored for band 1 and band 2 respectively. This vector can be plotted in a two-dimensional graph.

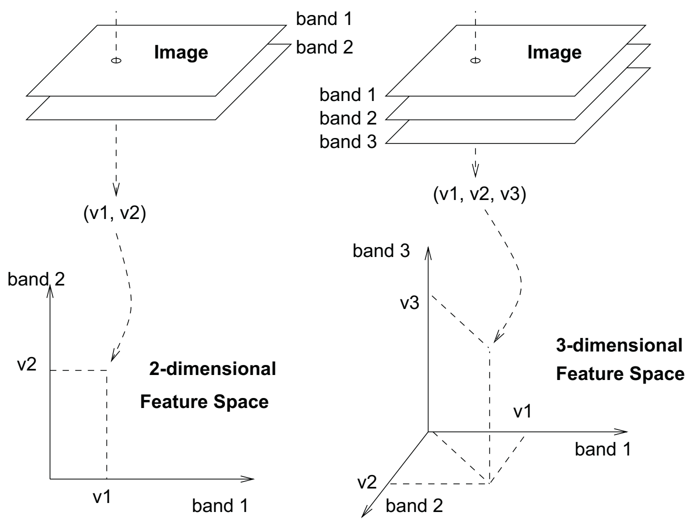


Figure 12.2: Plotting of the values of a pixel in the feature space for a two and three band image.

Similarly, this approach can be visualized for a three band situation in a three-dimensional graph. A graph that shows the values of the feature vectors is called a *feature space* or feature space plot. Figure 12.2 illustrates how a feature vector (related to one pixel) is plotted in the feature space for two and three bands. Usually we only find two axis feature space plots.

Note that plotting values is difficult for a four- or more-dimensional case. A practical solution when dealing with four or more bands is that all the possible

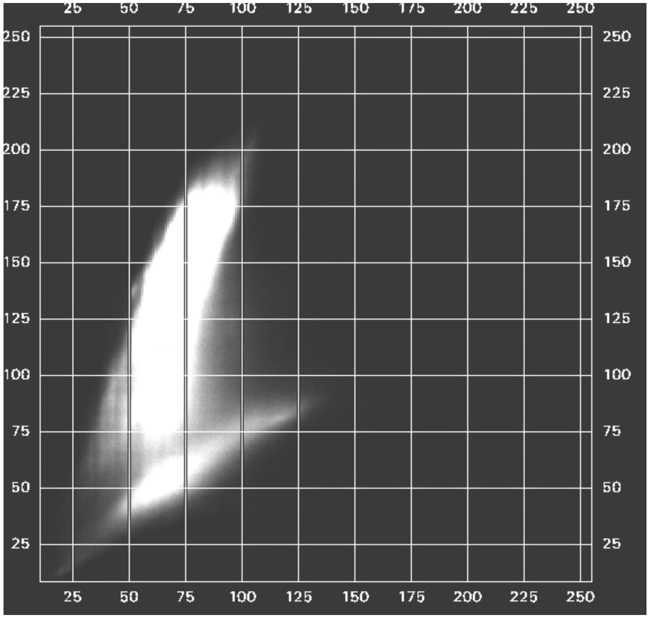


Figure 12.3: Scatterplot of two bands of a digital image. Note the units (DN-values) along the x- and y-axes. The intensity at a point in the feature space is related to the number of pixels at that point.

combinations of two bands are plotted separately. For four bands, this already yields six combinations: bands 1 and 2, 1 and 3, 1 and 4, bands 2 and 3, 2 and 4, and bands 3 and 4.

Plotting the combinations of the values of all the pixels of one image yields a large cluster of points. Such a plot is also referred to as a *scatterplot* (Figure 12.3). A scatterplot provides information about the combinations of pixel values that occur within the image. Note that some combinations will occur more frequently and can be visualized by using intensity or colour.

Distances and clusters in the feature space

Distance in the feature space is expressed as 'Euclidian distance' and the units are DN (as this is the unit of the axes). In a two-dimensional feature space the distance can be calculated according to Pythagoras' theorem. In the situation of Figure 12.4, the distance between $(10, 10)$ and $(40, 30)$ equals the square root of $(40 - 10)^2 + (30 - 10)^2$. For three or more dimensions, the distance is calculated in a similar way.

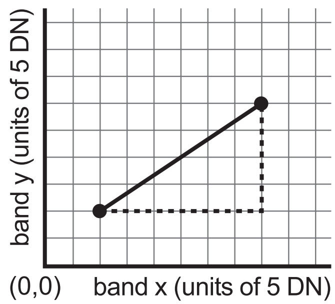


Figure 12.4: Euclidian distance between the two points is calculated using Pythagoras' theorem

12.2.3 Image classification

The scatterplot shown in Figure 12.3 gives information about the distribution of corresponding pixel values in two bands of an image. Figure 12.5 shows a feature space in which the feature vectors have been plotted for six specific land cover classes (grass, water, trees, etc). Each cluster of feature vectors (class) occupies its own area in the feature space. Figure 12.5 shows the basic assumption for image classification: a specific part of the feature space corresponds to a specific class. Once the classes have been defined in the feature space, each image pixel can be compared to these classes and assigned to the corresponding class.

Classes to be distinguished in an image classification need to have different spectral characteristics. This can, for example, be analyzed by comparing spectral reflectance curves (Section 2.4). Figure 12.5 also illustrates the limitation of image classification: if classes do not have distinct clusters in the feature space, image classification can only give results to a certain level of reliability.

The *principle of image classification* is that a pixel is assigned to a class based on its feature vector, by comparing it to predefined clusters in the feature space. Doing so for all image pixels results in a classified image. The crux of image classification is in *comparing it to predefined clusters*, which requires definition of the clusters and methods for comparison. Definition of the clusters is an interactive process and is carried out during the *training process*. Comparison of the individual pixels with the clusters takes place using *classifier algorithms*. Both are explained in the next section.

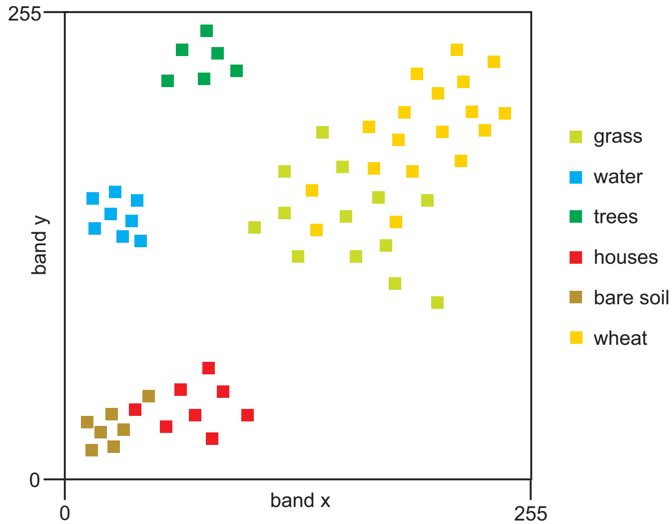


Figure 12.5: Feature space showing the respective clusters of six classes; note that each class occupies a limited area in the feature space.

12.3 Image classification process

The process of image classification (Figure 12.6) typically involves five steps:

1. Selection and preparation of the image data. Depending on the cover types to be classified, the most appropriate sensor, the most appropriate date(s) of acquisition and the most appropriate wavelength bands should be selected (Section 12.3.1).

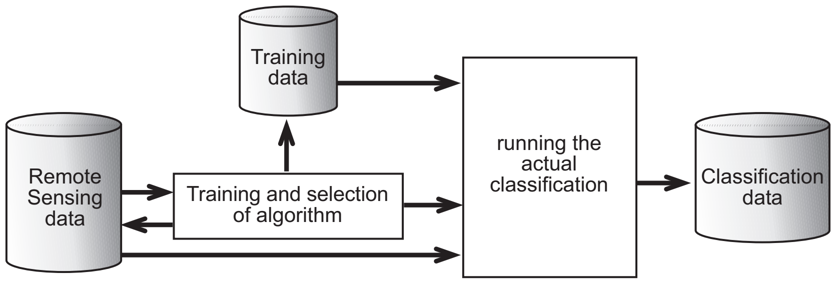


Figure 12.6: The classification process; most important component is the training in combination with selection of the algorithm.

2. Definition of the clusters in the feature space. Here two approaches are possible: *supervised classification* and *unsupervised classification*. In a supervised classification, the operator defines the clusters during the training process (Section 12.3.2); in an unsupervised classification a clustering algorithm automatically finds and defines a number of clusters in the feature space (Section 12.3.3).

3. Selection of classification algorithm. Once the spectral classes have been defined in the feature space, the operator needs to decide on how the pixels (based on their DN-values) are assigned to the classes. The assignment can be based on different criteria (Section 12.3.4).

4. Running the actual classification. Once the training data have been established and the classifier algorithm selected, the actual classification can be carried out. This means that, based on its DN-values, each individual pixel in the image is assigned to one of the defined classes (Figure 12.7).

5. Validation of the result. Once the classified image has been produced its quality is assessed by comparing it to reference data (ground truth). This requires selection of a sampling technique, generation of an error matrix, and the calculation of error parameters (Section 12.4).

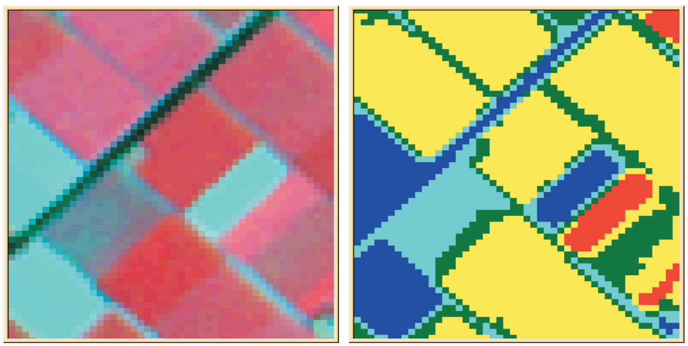
(a)  (b)

Figure 12.7: The result of classification of a multispectral image (a) is a raster in which each cell is assigned to some thematic class (b).

The above points are elaborated in the next sections. Most examples deal with a two-dimensional situation (two bands) for reasons of simplicity and visualization. In principle, however, image classification can be carried out on any n-dimensional data set. Visual image interpretation limits itself to an image that is composed of a maximum of three bands.

12.3.1 Preparation for image classification

Image classification serves a specific goal: converting image data into thematic data. In the application context, one is rather interested in thematic characteristics of an area (pixel) rather than in its reflection values. Thematic characteristics such as land cover, land use, soil type or mineral type can be used for further analysis and input to models. In addition, image classification can also be considered as data reduction: the n multispectral bands result in a single valued raster file.

With the particular application in mind, the information classes of interest need to be defined and their spatio-temporal characteristics assessed. Based on these characteristics the appropriate image data can be selected. Selection of the adequate data set concerns the type of sensor, the relevant wavelength bands and the date(s) of acquisition.

The possibilities for the classification of land cover types depend on the date an image was acquired. This not only holds for crops, which have a certain growing cycle, but also for other applications. Here you may think of snow cover or illumination by the sun. In some situations, a multi-temporal data set is required. A non trivial point is that the required image data should be available at the required moment. Limited image acquisition and cloud cover may force you to make use of a less optimal data set.

Before starting to work with the acquired data a selection of the available spectral bands may be made. Reasons for not using all available bands (for example all seven bands of Landsat TM) lie in the problem of band correlation and, sometimes, in limitations of hard- and software. Band correlation occurs when the spectral reflection is similar for two bands. An example is the correlation between the green and red wavelength bands for vegetation: a low reflectance in green correlates with a low reflectance in red. For classification purposes, correlated bands give redundant information and might disturb the classification process.

12.3.2 Supervised image classification

One of the main steps in image classification is the 'partitioning' of the feature space. In *supervised classification* this is realized by an operator who defines the spectral characteristics of the classes by identifying sample areas (training areas). Supervised classification requires that the operator be familiar with the area of interest. The operator needs to know where to find the classes of interest in the area covered by the image. This information can be derived from 'general area knowledge' or from dedicated field observations (Section 11.3.1).

A sample of a specific class, comprising of a number of training pixels, forms a cluster in the feature space (Figure 12.5). The clusters, as selected by the operator:

- should form a representative data set for a given class; this means that the variability of a class within the image should be taken into account. Also, in absolute sense, a minimum number of observations per cluster is required. Although it depends on the classifier algorithm to be used, a useful rule of thumb is $30 \times n$ (n = number of bands).

- should not or limitedly overlap with the other clusters, otherwise, a reliable separation is not possible. Using a specific data set, some classes may have significant spectral overlap, which, in principle, means that these classes cannot be discriminated by image classification. Solutions are to add other spectral bands, and/or, add image data acquired at other moments.

12.3.3 Unsupervised image classification

Supervised classification requires knowledge of the area at hand. If this knowledge is not sufficient available or the classes of interest are not yet defined, an unsupervised classification can be applied. In an *unsupervised classification*, clustering algorithms are used to partition the feature space into a number of clusters.

Several methods of unsupervised classification system exist, their main purpose being to produce spectral groupings based on certain similarities. In one of the most common approaches, the user has to define the maximum number of clusters in a data set. Based on this, the computer locates arbitrary mean vectors as the centre points of the clusters. Each pixel is then assigned to a cluster by the minimum distance to cluster centroid decision rule. Once all the pixels have been labelled, recalculation of the cluster centre takes place and the process is repeated until the proper cluster centres are found and the pixels are labelled accordingly. The iteration stops when the cluster centres do not change any more. At any iteration, however, clusters with less than a specified number of pixels are eliminated. Once the clustering is finished, analysis of the closeness or separability of the clusters will take place by means of inter cluster distance or divergence measure. Merging of clusters needs to be done to reduce the number of unnecessary subdivisions in the data set. This will be done using a pre-specified threshold value. The user has to define the maximum number of clusters/classes, the distance between two cluster centres, the radius of a cluster, and the minimum number of pixels as a threshold number for cluster elimination. Analysis of the cluster compactness around its centre point is done by means of the user-defined standard deviation for each spectral band. If a cluster is elongated, separation of the cluster will be done perpendicularly to the spectral axis of elongation. Analysis of closeness of the clusters is carried out by measuring the distance between the two cluster centres. If the distance between two cluster centres is less than the pre-specified threshold, merging of the clusters takes place. At each iteration, any cluster with less than a specified number of pixels is eliminated. The clusters that result after the last iteration are described by their statistics. Figure 12.8 shows the results of a clustering algorithm on a data set. As you can observe, the cluster centres coincide with the high density areas in the feature space.

The derived cluster statistics are then used to classify the complete image using a selected classification algorithm (similar to the supervised approach).

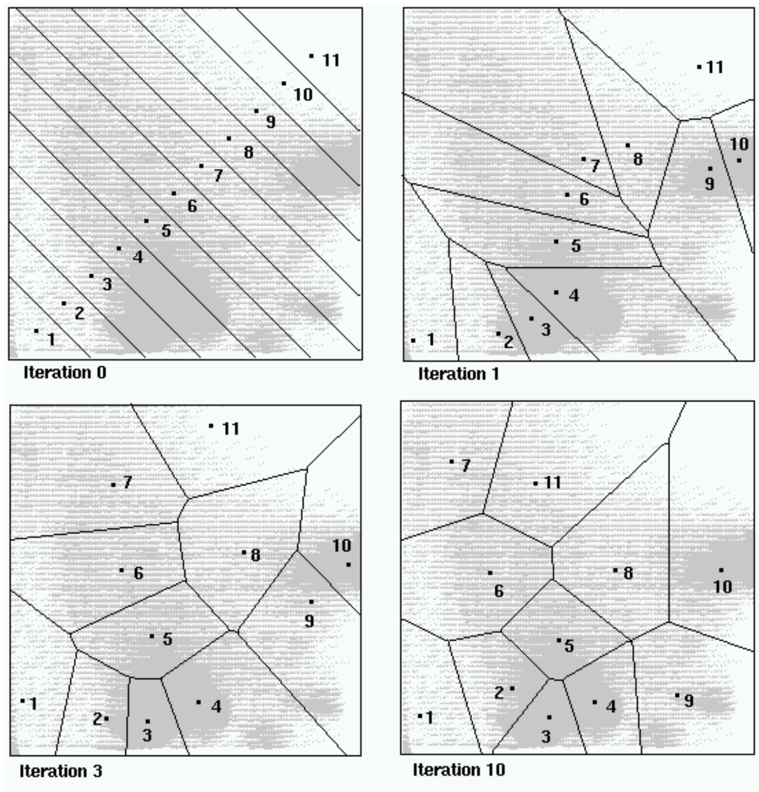


Figure 12.8: The subsequent results of a clustering algorithm on a sample data set

12.3.4 Classification algorithms

After the training sample sets have been defined, classification of the image can be carried out by applying a classification algorithm. Several classification algorithms exist. The choice of the algorithm depends on the purpose of the classification and the characteristics of the image and training data. The operator needs to decide if a ‘reject’ or ‘unknown’ class is allowed. In the following, three classifier algorithms are explained. First the *box classifier* is explained, for its simplicity to help you understanding the principle. In practice, the box classifier is hardly ever used. In practice the *Minimum Distance to Mean* and the *Maximum Likelihood* classifiers are used.

Box classifier

The box classifier is the most simple classification method. For this purpose, upper and lower limits are defined for each class. The limits may be based on the minimum and maximum values, or on the mean and standard deviation per class. When the lower and the upper limits are used, they define a box-like area in the feature space, which is why it is called box classifier. The number of boxes depends on the number of classes. Box classification is also known as parallelepiped classification since the opposite sides are parallel (Figure 12.9). During classification, an unknown pixel will be checked to see if it falls in any of the boxes. It is labelled with the class in which box it falls. Pixels that do not fall inside any of the boxes will be assigned the *unknown class*, sometimes also referred to as the *reject class*.

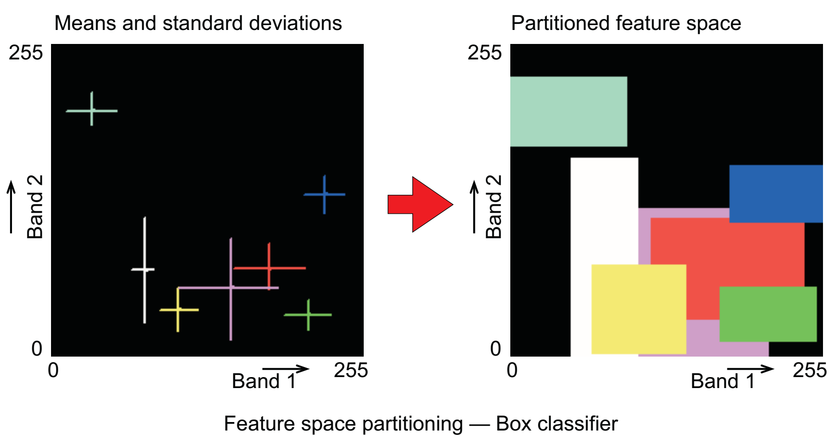


Figure 12.9: Principle of the box classification in a two-dimensional situation.

The disadvantage of the box classifier is the overlap between the classes. In such a case, a pixel is arbitrarily assigned the label of the first box it encounters.

Minimum Distance to Mean classifier

The basis for the Minimum Distance to Mean (MDM) classifier is the cluster centres. During classification the Euclidean distances from an unknown pixel to various cluster centres are calculated. The unknown pixel is assigned to that class to which the distance is least. Figure 12.10 illustrates how a feature space is partitioned based on the cluster centres. One of the flaws of the MDM classifier is that also pixels that are at a large distance from a cluster centre may be assigned to this centre. This problem can be overcome by defining a threshold value that limits the search distance. Figure 12.10 illustrates this effect, the threshold distance to the centre is shown as a circle.

A further disadvantage of the MDM classifier is that it does not take the class variability into account: some clusters are small and dense while others are large and dispersed. Maximum likelihood classification takes class variability into account.

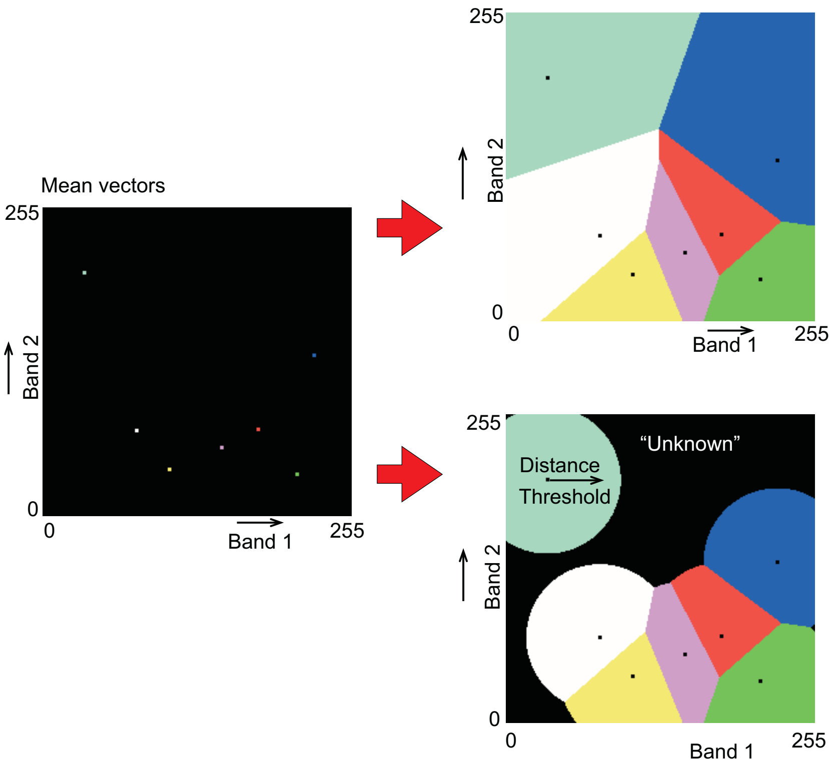


Figure 12.10: Principle of the minimum distance to mean classification in a two-dimensional situation. The decision boundaries are shown for a situation without threshold distance (upper right) and with threshold distance (lower right).

Maximum Likelihood classifier

The Maximum Likelihood (ML) classifier considers not only the cluster centre but also its shape, size and orientation. This is achieved by calculating a statistical distance based on the mean values and covariance matrix of the clusters. The statistical distance is a probability value: the probability that observation x belongs to specific cluster. The pixel is assigned to the class (cluster) to which it has the highest probability. The assumption of most ML classifiers is that the statistics of the clusters have a 'normal' (Gaussian) distribution.

For each cluster, so-called 'equiprobability contours' can be drawn around the centres of the clusters. Maximum likelihood also allows the operator to define a threshold distance by defining a maximum probability value. A small ellipse centred on the mean defines the values with the highest probability of membership of a class. Progressively larger ellipses surrounding the centre represent contours of probability of membership to a class, with the probability decreasing away from the centre. Figure 12.11 shows the decision boundaries for a situation with and without threshold distance.

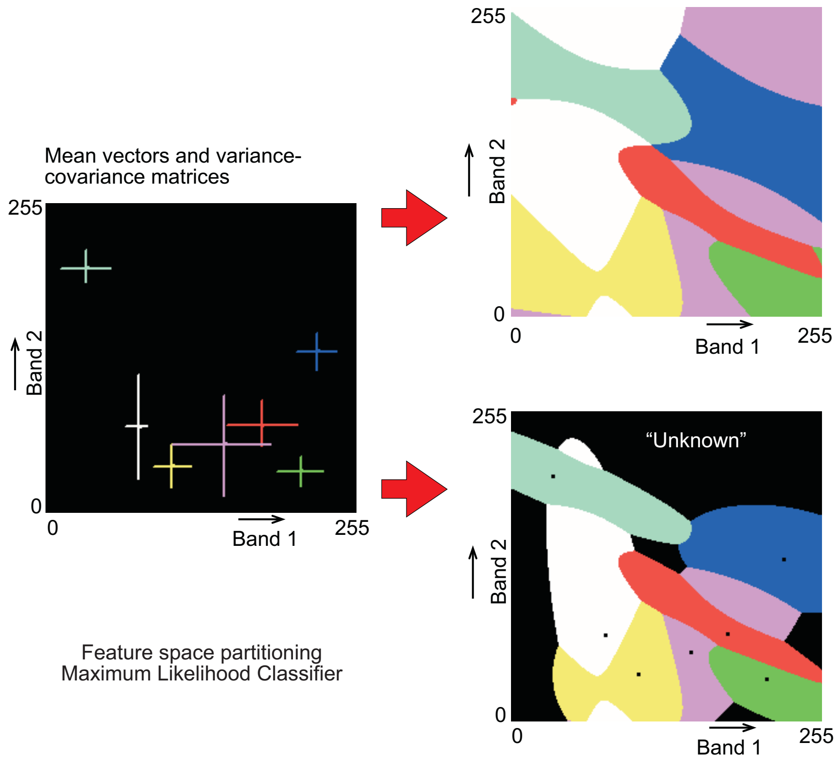


Figure 12.11: Principle of the maximum likelihood classification. The decision boundaries are shown for a situation without threshold distance (upper right) and with threshold distance (lower right).

12.4 Validation of the result

Image classification results in a raster file in which the individual raster elements are class labelled. As image classification is based on samples of the classes, the actual quality should be checked and quantified afterwards. This is usually done by a sampling approach in which a number of raster elements are selected and both the classification result and the true world class are compared. Comparison is done by creating an *error matrix* from which different accuracy measures can be calculated. The 'true world class' are preferable derived from field observations. Sometimes, sources of an assumed higher accuracy, such as aerial photos, are used as a reference.

Various *sampling schemes* have been proposed to select pixels to test. Choices to be made relate to the design of the sampling strategy, the number of samples required, and the area of the samples. Recommended sampling strategies in the context of land cover data are simple random sampling or stratified random sampling. The number of samples may be related to two factors in accuracy assessment: (1) the number of samples that must be taken in order to reject a data set as being inaccurate; or (2) the number of samples required to determine the true accuracy, within some error bounds, for a data set. Sampling theory is used to determine the number of samples required. The number of samples must be traded-off against the area covered by a sample unit. A sample unit can be a point but also an area of some size; it can be a single raster element but may also include the surrounding raster elements. Among other considerations the 'optimal' sample area size depends on the heterogeneity of the class.

Once the sampling has been carried out and the data collected, an error matrix can be established (Table 12.1). Other terms for this table are *confusion matrix* or *contingency matrix*. In the table, four classes (A, B, C, D) are listed. A total

	A	B	C	D	Total	Error of Commission	User Accuracy
a	35	14	11	1	61	43	57
b	4	11	3	0	18	39	61
c	12	9	38	4	63	40	60
d	2	5	12	2	21	90	10
Total	53	39	64	7	163		
Error of Omission	34	72	41	71			
Producer Accuracy	66	28	59	29			

Table 12.1: The error matrix with derived errors and accuracy expressed as percentages. A, B, C and D refer to the reference classes; a, b, c and d refer to the classes in the classification result. Overall accuracy is 53%.

of 163 samples were collected. From the table you can read that, for example, 53 cases of A were found in the real world ('reference') while the classification result yields 61 cases of a; in 35 cases they agree.

The first and most commonly cited measure of mapping accuracy is the *overall accuracy*, or Proportion Correctly Classified (PCC). Overall accuracy is the number of correctly classified pixels (i.e., the sum of the diagonal cells in the error matrix) divided by the total number of pixels checked. In Table 12.1 the overall accuracy is $(35+11+38+2)/163 = 53\%$. The overall accuracy yields one figure for the result as a whole.

Most other measures derived from the error matrix are calculated per class. *Error of omission* refers to those sample points that are omitted in the interpretation result. Consider class A, for which 53 samples were taken. 18 out of the 53 samples were interpreted as b, c or d. This results in an error of omission of

$18/53 = 34\%$. Error of omission starts from the reference data and therefore relates to the columns in the error matrix. The *error of commission* starts from the interpretation result and refers to the rows in the error matrix. The error of commission refers to incorrectly classified samples. Consider class d: only two of the 21 samples (10%) are correctly labelled. Errors of commission and omission are also referred to as type I and type II errors respectively.

User accuracy is the corollary of commission error, whilst omission error is the corollary of producer accuracy. The user accuracy is the probability that a certain reference class has also been labelled that class. The producer accuracy is the probability that a sampled point on the map is that particular class. Another widely used measure of map accuracy derived from the error matrix is the kappa or κ' statistic. Kappa statistics take into account the fact that even assigning labels at random results in a certain degree of accuracy. Based on Kappa statistics one can test if two data sets have a statistically different accuracy. This type of testing is used to evaluate different sources (image data) or methods for the generation of spatial data.

12.5 Problems in image classification

Image classification is a powerful technique to derive 'thematic classes' from multi-band image data. However, it has certain limitations that you should be aware of. The most important constraints of pixel based image classification are that it results in (i) spectral classes, and that (ii) each pixel is assigned to one class only.

Spectral classes are classes that are directly linked to the spectral bands used in the classification. In turn, these are linked to surface characteristics. In that respect one can say that spectral classes correspond with land cover classes (see also Section 11.3.3). In the classification process a 'spectral class' may be represented by several 'training classes'. Among others this is due to the variability within a spectral class. Consider a class such as 'grass'; there are different types of grass, which have different spectral characteristics. Furthermore, the same type of grass may have different spectral characteristics when considered over larger areas due to, for example, different soil and climate conditions. A related topic is that sometimes one is interested in land use classes rather than land cover classes. Sometimes, a land use class may be comprised of several land cover classes. Table 12.2 gives some examples of linking spectral land cover and land use classes. Note that between two columns there can be 1-to-1, 1-to-n, and n-to-1 relationships. The 1-to-n relationships are a serious problem and can only be solved by adding data and/or knowledge to the classification procedure. The data added can be other remote sensing image data (other bands, other moments) or existing spatial data, such as topographic maps, historical land inventories, road maps, *et cetera*. Usually this is done in combination with adding expertise knowledge to the process. An example is using historical land cover data and defining the probability of certain land cover changes. Another example is to use elevation, slope and aspect information. This will prove especially useful in mountainous regions where elevation differences play an important role in variations in surface cover types.

Spectral Class	Land Cover Class	Land Use Class
water	water	shrimp cultivation
grass1	grass	nature reserve
grass2	grass	nature reserve
grass3	grass	nature reserve
bare soil	bare soil	nature reserve
trees1	forest	nature reserve
trees2	forest	production forest
trees3	forest	city park

Table 12.2: Spectral classes distinguished during classification can be aggregated to land cover classes. 1-to-n and n-to-1 relationships can exist between land cover and land use classes.

The other main problem and limitation of image classification is that each pixel is only assigned to one class. When dealing with (relatively) small pixels, this is not a problem. However when dealing with (relatively) large pixels more land cover classes will occur within this pixel. As a result, the spectral value of the pixel is an average of the reflectance of the land cover present within the pixel. In a standard classification these contributions cannot be traced back and the pixel will be assigned to one of either classes or even to another class. This phenomenon is usually referred to as the *mixed pixel*, or *mixel* (Figure 12.12). This problem of mixed pixels is inherent to image classification: assigning the pixel to one thematic class. The solution to this is to use a different approach, for example, assigning the pixel to more than one class. This brief introduction into the problem of mixed pixels also highlights the importance of using data with the appropriate spatial resolution.

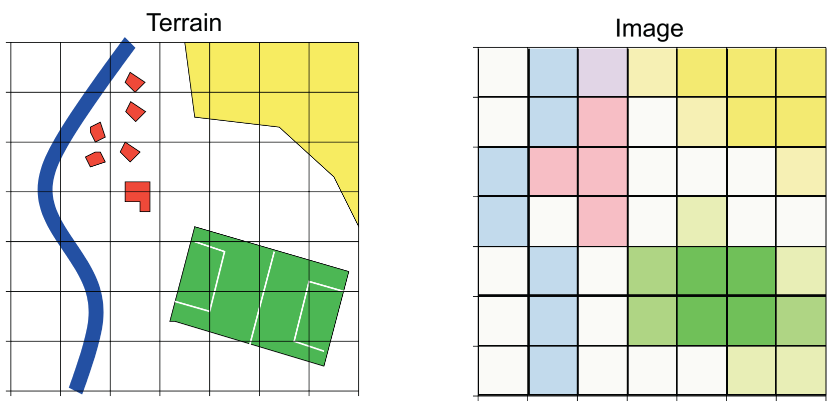


Figure 12.12: The origin of mixed pixels: different land cover types occur within one pixel. Note the relative abundance of mixed pixels.

Summary

Digital image classification is a technique to derive thematic classes from image data. Input are multi-band image data; output is a raster file containing thematic (nominal) classes. In the process of image classification the role of the operator and additional (field) data is significant. The operator needs to provide the computer with training data and select the appropriate classification algorithm. The training data are defined based on knowledge (derived by field work, or from secondary sources) of the area being processed. Based on the similarity between pixel values (feature vector) and the training classes a pixel is assigned to one of the classes defined by the training data.

An integral part of image classification is validation of the results. Again, independent data are required. The result of the validation process is an error matrix from which different measures of error can be calculated.

Questions

The following questions can help you to study Chapter 12.

1. Compare digital image classification with visual image interpretation in terms of input of the operator/photo-interpreter and in terms of output.

2. What would be typical situations in which to apply digital image classification?

3. Another wording for image classification is 'partitioning of the feature space'. Explain what is meant by this.

The following are typical exam questions:

1. Give the different steps of the process of image classification.

2. What is the principle of image classification?

3. What is a classification algorithm? Give two examples.

4. Image classification is sometimes referred to as automatic classification. Do you agree or disagree? (give argumentation)

5. Draw a simple error matrix, indicate what is on the axes and explain how the overall accuracy is calculated.

Bibliography

[1] Stan Aronoff. *Geographic Information Systems: A Management Perspective*. WDL Publications, Ottawa, 1989.

[2] Henk J. Buiten and Jan G. P. W. Clevers. *Land Observation by Remote Sensing: Theory and Applications*, volume 3 of *Current Topics in Remote Sensing*. Gorden & Breach, 1993. 30

[3] Peter A. Burrough and Andrew U. Frank. *Geographic Objects with Indeterminate Boundaries*. GISDATA Series. Taylor & Francis, London, 1996. 331

[4] Peter A. Burrough and R. McDonnell. *Principles of Geographical Information Systems*. Oxford University Press, Oxford, 1998.

[5] European Community. *CORINE Land Cover Technical Guide*. ECSC–EEC–EAEC, Brussels, Belgium, 1993. EUR 12585 EN. 325, 330

[6] Rolf A. de By, editor. *Principles of Geographic Information Systems*, volume 1 of *ITC Educational Textbook Series*. International Institute for Aerospace Survey and Earth Sciences, Enschede, second edition, 2001. 18, 32, 254

[7] J. A. Deckers, F. O. Nachtergaele, and O. C. Spaargaren, editors. *World reference base for soil resources: introduction*. ACCO, Leuven, Belgium, 1998. 318

[8] G. Edwards and K. E. Lowell. Modeling uncertainty in photo-interpretated boundaries. *Photogrammetric Engineering and Remote Sensing*, 60(4):337–391, 1996. 334

[9] P. Gunn. Airborne magnetic and radiometric surveys. *Journal of Australian Geology and Geophysics*, 17:1–216, 1997. Special Issue. 220

[10] John Horn. *Aerial Photography*. ITC Lecture Notes PHM.80. ITC, Enschede, The Netherlands, 2000. 119

[11] ITC. ITC Textbook of Photo-interpretation. Four volumes, 1963–1974. 18

[12] ITC. ITC Textbook of Photogrammetry. Five volumes, 1963–1974. 18

[13] Jane's. *Jane's Space Directory 1997–1998*. Alexandria, Jane's Information Group, 13th edition, 1997. 84

[14] J. H. Kramer. *Observation of the Earth and its Environment: Survey of Mission and Sensors*. Springer Verlag, Berlin, Germany, third edition, 1996. 84

[15] Robert Laurini and Derek Thompson. *Fundamentals of Spatial Information Systems*, volume 37 of *The APIC Series*. Academic Press, London, 1992.

[16] Thomas M. Lillesand and Ralph W. Kiefer. *Remote Sensing and Image Interpretation*. John Wiley & Sons, New York, NY, third edition, 1994. 30, 58, 119, 170

[17] Keith R. McCloy. *Resource Management Information Systems*. Taylor & Francis, London, U.K., 1995. 75, 76, 119

[18] Hans Middelkoop. Uncertainty in a GIS, a test for quantifying interpretation output. *ITC Journal*, 1990(3):225–232, 1990. 335

[19] Martien Molenaar. *An Introduction to the Theory of Spatial Object Modelling*. Research Monographs in GIS Series. Taylor & Francis, London, 1998. 331

[20] V. Perdigao and A. Annoni. *Technical and Methodological Guide for Updating CORINE Land Cover Data Base*. EC-JRC, EEA, Brussels, Belgium, 1997. EUR 17288 EN. 325

[21] Donna J. Peuquet and D. F. Marble, editors. *Introductory Readings in Geographic Information Systems*. Taylor & Francis, London, 1990.

[22] Colin V. Reeves. Continental scale and global scale geophysical anomaly mapping. *ITC Journal*, 1998(2):91–98, 1998. 220

[23] David G. Rossiter. *Lecture Notes: Methodology for Soil Resource Inventories*. ITC Lecture Notes SOL.27. ITC, Enschede, The Netherlands, 2nd revised edition, 2000. 318

[24] F. F. Sabins. *Remote Sensing: Principles and Interpretation*. Freeman & Co., New York, NY, third edition, 1996. 30

[25] Toni Schenk. *Digital Photogrammetry*, volume 1. TerraScience, Laurelville, 1999. 254

[26] W. Smith and D. Sandwell. Measured and estimated seafloor topography, version 4.2. Poster RP–1, 1997. World Data Center for Marine Geology and Geophysics. 224

[27] Michael F. Worboys. *GIS: A Computing Perspective*. Taylor & Francis, London, U.K., 1995.

[28] Alfred J. Zinck. *Physiography & Soils*. ITC Lecture Notes SOL.41. ITC, Enschede, The Netherlands, 1988. 318

Glossary

A

Absorption The process in which electromagnetic energy is converted in an object into other forms of energy (e.g., heat).

Active sensor Sensor with a built in source of energy. The sensor both emits and receives energy (e.g., radar and laser).

Additive colours The additive principle of colours is based on the three primary colours of light: red, green, blue. All three primary colours together produce white. Additive colour mixing is used, for example, on computer screens and television.

B

Backscatter The microwave signal reflected by elements of an illuminated surface in the direction of the radar antenna.

Band Usually related to wavelength band, which indicates a specific range of the electromagnetic spectrum to which the sensor is sensitive. In general, it can indicate any 'layer' of an n-dimensional image.

C

Charge coupled device (CCD) Semi-conductor elements usually aligned as a linear (scanner) or surface array (video, digital camera). CCDs produce image data.

Class Classes, usually defined in visual image interpretation and image classification, are a *variable* of the nominal type. Class schemes are applied to describe hierarchical relationships between them.

Cluster Used in the context of image classification to indicate a concentration of observations (points in the feature space) related to a training class.

Colour Colloquial term to indicate hue (IHS space).

Colour film Also known as *true colour film* used in (aerial) photography. The principle of colour film is to add sensitized dyes to the silver halide. Magenta, yellow and cyan dyes are sensitive to red, green and blue light respectively.

Colour infrared film Film with specific sensitivity for infrared wavelengths. Typically used in surveys of vegetation.

D

Di-electric constant Parameter that describes the electrical properties of a medium. Reflectivity of a surface and penetration of microwaves into the material are determined by this parameter.

Digital Elevation Model (DEM) Special case of a DTM. A DEM stores terrain elevation (surface height) by means of a raster. *Elevation* refers to a height expressed with respect to a specific reference.

Digital Terrain Model (DTM) Term indicating a digital description of the terrain relief. A DTM can be stored in different manners (contour lines, TIN, raster) and may also contain semantic, relief-related information (breaklines, saddlepoints).

E

Earth Observation (EO) Term indicating the collection of remote sensing techniques performed from space.

Electromagnetic energy Energy with both electric and magnetic components. Both the wave model and photon model are used to explain this phenomenon. The measurement of reflected and emitted electromagnetic energy is an essential aspect in remote sensing.

Electromagnetic spectrum The complete range of all wavelengths, from gamma rays (10^{-12} m) up to very long radio waves (10^{12} m).

Emission Radiation of electromagnetic energy. Each object with a temperature above $0\ K$ (-273 $°C$) emits electromagnetic energy.

Emissivity The radiant energy of an object compared to the energy of a blackbody of the same temperature, expressed as a ratio.

Error matrix Matrix that compares samples taken from the source to be evaluated with observations that are considered as correct (reference). The error matrix allows calculation of quality parameters such as overall accuracy, error of omission and error of commission.

F

False colour infrared film see Colour infrared film.

Feature space The mathematical space describing the combinations of observations (DN values in the different bands) of a multispectral or multiband image. A single observation is defined by a feature vector.

Feature space plot A two- or three-dimensional graph in which the observations made in different bands are plotted against each other.

Field of view (FOV) The total swath as observed by a sensor-platform system. Sometimes referred to as total field of view. It can be expressed as an angle or by the absolute value of the width of the observation.

Filter (1) Physical product made out of glass and used in remote sensing devices to block certain wavelenghts, e.g., ultraviolet-filter. (2) Mathematical operator used in image processing for modifying the signal, e.g., a smoothing filter.

G

Geo-spatial data Data that includes positions in the geographic space. In this book, usually abbreviated to 'spatial data'.

Geocoding Process of transforming and resampling image data in such way that these can be used simultaneously with data that are in a specific map projection. Input for a geocoding process are image data and control points, output is a geocoded image. A specific category of geocoded images are orthophotos and orthoimages.

Geographic information Information derived from spatial data, and in the context of this book, from image data. Information is what is relevant in a certain application context.

Geographic Information System (GIS) A software package that accommodates the capture, analysis, manipulation and presentation of georeferenced data. It is a generic tool applicable to many different types of use (GIS applications).

Georeferencing Process of relating an image to a specific map projection. As a result, vector data stored in this projection can for example be superimposed on the image. Input for a georeferencing process are image data and coordinates of ground control points, output is a georeferenced image.

Global Navigation Satellite System (GNSS) The Global Navigation Satellite System is a global infrastructure for the provision of positioning and timing information. It consists of the American GPS and Russian Glonass systems. There is also a proposed European Galileo system.

Ground control points (GCPs) Points which are used to define or validate a geometric transformation process. Sometimes also referred to as Ground Control Points stating these have been measured on the ground. Ground control points should be recognizable both in the image and in the real world.

Ground range Range direction of the side-looking radar image as projected onto the horizontal reference plane.

Ground truth A term that may include different types of observations and measurements performed in the field. The name is imprecise because it suggests that these are 100% accurate and reliable, and this may be difficult to achieve.

H

Histogram Tabular or graphical representation showing the (absolute and/or relative) frequency. In the context of image data it relates to the distribution of the (DN) values of a set of pixels.

Histogram equalization Process used in the visualization of image data to optimize the overall image contrast. Based on the histogram, all available grey levels or colours are distributed in such way that all occur with equal frequency in the result.

I

Image A digital file comprising pixels that represent measured local reflectance (emission or backscatter) values in some designated part of the electromagnetic spectrum. Typically, images are stored in a row-column system. An image may comprise any number of bands (or channels). After the reflectance values have been translated into some 'thematic' variable, the image becomes a raster. With an image, we talk of its constituent pixels; with a raster we talk of its cells.

Image classification Image classification is the process of assigning pixels to nominal, i.e., thematic, classes. Input is a multi-band image, output is a raster in which each cell has a (thematic) code. Image classification can be realized using a supervised or unsupervised approach.

Image enhancement The process of improving the visual representation of an image, for example by histogram equalization or using filters.

Image interpretation The process of information extraction from image data. What is information is defined by the application context. Visual interpretation and pattern recognition techniques can be used.

Image processing system A computer system that is specifically designed to process image data and to extract information from it by visualizing the data, application of models and pattern recognition techniques.

Image space The mathematical space describing the (relative) positions of the observations. Image positions are expressed by their row and column index.

Incidence angle Angle between the line of sight from the sensor to an element of an imaged scene and a vertical direction to the scene. One must distinguish between the nominal incidence angle determined by the geometry of the radar and the Earth's geoidal surface and the local incidence angle, which takes into account the mean slope of the pixel in the image.

Infrared waves Electromagnetic radiation in the infrared region of the electromagnetic spectrum. Near-infrared (700–1200 nm), middle infrared (1200–2500 nm) and thermal infrared (8–14 μm) are distinguished.

Instantaneous field of view (IFOV) The area observed on the ground by a sensor, which can be expressed by an angle or in ground surface units.

Interferometry Computational process that makes use of the interference of two coherent waves. In the case of imaging radar, two different paths for imaging cause phase differences from which an interferogram can be derived. In SAR applications, interferometry is used for constructing a DEM.

Interpretation elements The elements used by the human vision system to interprete a picture or image. Interpretation elements are: tone, texture, shape, size, pattern, site, association and resolution.

L

Latent image When exposed to light, the silver halide crystals within the photographic emulsion undergo a chemical reaction, which results in an invisible latent image. The latent is transformed into a visible image by the development process in which the exposed silver halide is converted into silver grains that appear black.

Look angle The angle of viewing relative to the vertical (nadir) as perceived from the sensor.

M

Microwaves Electromagnetic radiation in the microwave window, which ranges from 1–100 cm.

Mixel Acronym for *mixed pixel*. Mixel is used in the context of image classification where different spectral classes occur within the area covered by one pixel.

Monoplotting Process that enables extraction of accurate (x, y) coordinates from image data by correcting for terrain relief.

Multispectral scanning Remote sensing technique in which the surface is scanned at the same moment in different wavelength bands. This book distinguishes two types of scanners: pushbroom and wiskbroom scanners.

N

Nadir The point (or line) directly under the platform during acquisition of image data.

O

Objects Objects are real world features and have clearly identifiable geometric characteristics. In a computer environment, objects are modelled using an object-based approach in contrast to a field-based approach, which is more suited for continuous phenomena.

Orbit The path of a satellite through space. Types of orbits used for remote sensing satellites are, for example, (near) polar and geostationary.

Orientation Refers to relative or absolute position of a sensor and/or platform. Interior orientation relates to the components within a camera or sensor. Relative orientation refers to (successive) positions of the sensor-platform system. In photogrammetry, absolute orientation refers to the orientation of the image relative to a spatial reference system.

Orthoimage An orthoimage and orthophoto have been corrected for terrain relief. As a result, an orthoimage can be used directly in combination with geocoded data.

P

Panchromatic Indication of one (wide or narrow) spectral band in the visible and near-infrared part of the electromagnetic spectrum.

Passive sensor Sensor that records energy that is produced by external sources such as the Sun and the Earth.

Pattern recognition Term for the collection of techniques used to detect and identify patterns. Patterns can be found in the spatial, spectral and temporal domains. An example of spectral pattern recognition is image classification; an example of spatial pattern recognition is segmentation.

Photogrammetry The science and techniques of making measurements from photos or image data. Photogrammetric procedures are required for accurate measurements from stereo pairs of aerial photos, image data or radar data.

Photograph Image obtained by using a camera. The camera produces a negative film, which is can be printed into positive paper product.

Pixel Acronym for *picture element*, which is the elementary unit of image data. The ground pixel size of image data is related to the spatial resolution of a sensor system it was produced by.

Pixel value The representation of the energy measured at a point, usually expressed as a Digital Number (DN-) value.

Q

Quantization The number of discrete levels applied to store the energy as measured by a sensor, e.g., 8-bit quantization allows 256 levels of energy.

R

Radar Acronym for *Radio Detection And Ranging*. Radars are active sensors at wavelengths between 1–100 cm.

Radar equation Mathematical expression that describes the average received signal level compared to the additive noise level in terms of system parameters. Principal parameters include the transmitted power, antenna gain,radar cross section,wavelength and range.

Radiance The energy (flux) leaving an element of the surface.

Radiometric resolution The smallest observable difference in energy.

RAR Acronym for *Real Aperture Radar*.

Raster A regularly spaced set of cells with associated (field) values. In contrast to a *grid*, the associated values represent *cell* values, not point values. This means that the value for a cell is assumed to be valid for all locations within the cell.

Reflectance The ratio of the reflected radiation to the total irradiation. Reflectance depends on the wavelength.

Reflection The process of scattering of of electromagnetic waves by an object.

Relief displacement A shift of the position of an imaged object in the image due to the local elevation of the object. The shift (magnitude and direction) depends on sensor-platform characteristics and on the elevation of the object.

Remote sensing (RS) Remote sensing is the instrumentation, techniques and methods to observe the Earth's surface at a distance and to interpret the images or numerical values obtained in order to acquire meaningful information of particular objects on Earth.

Resampling Process to generate a raster with another orientation and/or different cell size and to assign DN-values using one of the following methods,nearest neighbour selection, bilinear interpolation and cubic convolution.

Resolution Indicates the smallest observable (measurable) difference at which objects can still be distinguished. In remote sensing context used in spatial, spectral and radiometric resolution.

RMS error Root Mean Squared error. A statistical measure of accuracy, similar to standard deviation, indicating the spread of the measured values around the 'true' value.

S

SAR Acronym for *Synthetic Aperture Radar*. The (high) azimuth resolution (direction of the flight line) is achieved through off-line processing. The SAR is able to function as if it has a large virtual antenna aperture, synthesized from many observations with the (relative) small real antenna of the SAR system.

Scale The ratio of the distance measured on a (printed) map or image to that measured on the ground surface between the two same points. It is expressed as a ratio, e.g.: 1 : 10,000. The '10,000' is referred to as scale factor.

Scanner (1) remote sensing sensor that is based on the scanning principle, e.g., a multispectral scanner (2) office device to convert analogue products (photo, map) into digital raster format.

Slant range Image direction as measured along the sequence of line of sight rays from the radar to each reflecting point in the scene.

Spatial data In the broad sense, spatial data is any data with which position is associated.

Spatial resolution See resolution.

Speckle Interference of backscattered waves stored in the cells of a radar image. It causes the return signals to be extinguished or amplified resulting in random dark and bright pixels in the image.

Spectral band The range of wavelengths to which a channel (single band) of a multispectral scanner is sensitive.

Spectral resolution See resolution.

Specular reflection Mirror-like reflection; 'bounced-off' radiation.

Stereo model A 3D relief model observed through stereoscopic vision of a stereo pair.

Stereo pair A pair of overlapping photos or images that (partially) cover the same area from a different position. When appropriately taken, stereo pairs form a stereo model that can be used for stereoscopic vision and stereoplotting.

Stereoplotting Process that allows to measure accurate (x, y, z) coordinates from stereo models.

Stereoscopic vision The ability to perceive distance or depth by observation with both eyes. In remote sensing, stereoscopic vision is used for the three-dimensional observation of two images (photos) that are made from different positions. Stereoscopy is used in visual image interpretation and stereoscopic measurements (stereoplotting). It can yield 3D coordinates.

Subtractive colours The subtractive principle of colours is based on the three printing colours: cyan, magenta and yellow. All printed colours can be produced by a combination of these three colours. The subtractive principle is also used in colour photography.

Sun-synchronous Used to indicate a satellite orbit that is designed in such way that the satellite always passes the same location on Earth at the same local time.

T

Training stage Part of the image classification process in which pixels representative for a certain class are identified. Training results in a training set that comprises the statistical characteristics (signatures) of the classes of interest.

Transmittance The ratio of the radiation transmitted to the total irradiation.

V

Variable, interval A variable that is measured on a continuous scale, but with no natural zero. It cannot be used to form ratios.

Variable, nominal A variable that is organized in classes, with no natural order, i.e., cannot be ranked.

Variable, ordinal A variable that is organized in classes with a natural order, and so it can be ranked.

Variable, ratio A variable that is measured on a continuous scale, and with a natural zero, so can be used to form ratios.

Viewing angle Angle of observation referring to the vertical (nadir) of the sensor.

W

Wavelength Minimum distance between two events of a recurring feature in a periodic sequence such as the crests of a wave. Wavelength is expressed as a distance (e.g., μm or nm).

Index

Appendix A

SI units & prefixes

Quantity	SI unit
Length	metre (m)
Time	second (s)
Temperature	kelvin (K)
Energy	joule (J)
Power	watt (W) (J/s)

Table A.1: Relevant SI units in the context of remote sensing

Prefix	Multiplier
tera (T)	10^{12}
giga (G)	10^{9}
mega (M)	10^{6}
kilo (k)	10^{3}
centi (c)	10^{-2}
milli (m)	10^{-3}
micro (μ)	10^{-6}
nano (n)	10^{-9}
pico (p)	10^{-12}

Table A.2: Unit prefix notation

Unit	SI Equivalent
centimetre	10^{-2} m
millimetre	10^{-3} m
micron	10^{-6} m
micrometre	10^{-6} m
nanometre	10^{-9} m

Table A.3: Common units of wavelength

Parameter	Value
speed of light	$2.9979 \cdot 10^{8}$ m/s
°C	(°C $+ 273.15$) K
inch	2.54 cm
foot	30.48 cm
mile	$1,609$ m

Table A.4: Constants and non-SI units